



Age- and Gender-Based Differences in Children’s Interactions with a Gender-Matching Robot

Anara Sandygulova¹ · Gregory M. P. O’Hare²

Accepted: 6 March 2018 / Published online: 29 March 2018
© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract

Social robots are increasingly being used to encourage social, emotional and cognitive growth in children. However, in order to establish social and bonding interactions, social robots need to be able to exhibit adaptive strategies to keep children engaged and interested. Adaptive strategies of a social robot based on children’s age and gender are motivated by the comprehensive theory on gender development. Given the strong influence of gender in children’s cognitive development, the experiment first examined the responses of 107 children, ages 5–12, whether synthesized voice evokes gender associations in children. The results suggest that young children (ages 5–8) are not able to successfully attribute gender to the robot in correspondence with the synthesized voice. In addition, we explicitly investigated children’s preferences for the robot’s gender, and the results were contrary to our expectations: young children indicated their preference for a robot with a matching gender while there was no difference in preferences for a robot’s gender by older children (ages 9–12).

Keywords Child–robot interaction · Human–robot interaction · Social robotics · Adaptive strategies · Gender · Age

1 Introduction

Human–Robot Interaction (HRI) is no longer the exclusive preserve of adults. Research and commercial robots have infiltrated homes, hospitals and schools, becoming attractive and proving impactful for children’s healthcare, therapy, education, entertainment and other applications. As a result, within the field of HRI, a particular sector has emerged, which addresses research and practice of interaction between robots and children—Child–Robot Interaction (cHRI). According to Belpaeme et al. [5], cHRI creates entirely different conditions for human–robot interaction to operate within since children’s neurophysical, physical and mental developments are ongoing. Accordingly, a number of recent large-scale interdisciplinary projects, such as CoWriter [20], ALIZ-E

[6] and L2TOR [7] have explored child-centered research with the mission to enable the design, implementation, and evaluation of robots that encourage social, emotional, and cognitive growth in children, including those with social or cognitive deficits.

Since the value provided by social robots can only be obtained if the child actually spends time interacting with it, motivating a child to stay engaged is an important interaction design challenge. One possible means is a hypothesis of *gender segregation* [57]—the separation of boys and girls into same-gender groups in their friendships and casual encounters [33]—shows potential for attracting motivation and engagement with social robots. Indeed, *gender segregation* has been referred to as one of the most persistent and reliable of developmental phenomena [42]. And indeed, children tend to pretend play [25] and anthropomorphize [18], and as a result robots are readily treated as being alive and having “beliefs, desires and intentions” [5,39].

Gender effects have been found throughout HRI studies whether researchers are explicitly looking for them [14,55] or not [28,34,54]. Prior cHRI initiatives have identified that male and female children tend to engage with robotic technologies differently [16,22,23,27,46]. At the same time, children demonstrate similarity preferences for computer-based animated agents and human-like robots, their age and

This research was conducted when the author was at University College Dublin, Belfield, Dublin 4, Ireland.

✉ Anara Sandygulova
anara.sandygulova@nu.edu.kz

Gregory M. P. O’Hare
gregory.ohare@ucd.ie

¹ School of Science and Technology, Nazarbayev University, 53 Kabanbay Batyr Avenue, Astana, Kazakhstan

² University College Dublin, Belfield, Dublin 4, Ireland

gender [36,52]. Literature in cHRI exhibits a gap on how children perceive and interact with a gendered robot across age and gender groups. This paper addresses this void with an explicit research on age-related gender effects in the context of child–robot interaction.

Although gender discrimination is commonly studied and discussed in the adult psychological literature [15], research on children’s same-gender peer preferences, evaluations, and interactions are rarely framed in terms of discrimination [57]. While gender differences along a variety of dimensions have been successfully elicited by both on-screen agents [29,36] and robots for adult users [14,37,49], the impact of age-related gender effects [48] has not been addressed in cHRI. As this research aims to undertake an analysis of whether boys and girls retain *gender segregation* with a gendered robot, this work also investigates whether synthesized voice evokes gender associations in children.

Findings from previous work [45] and a preliminary study [43] strengthen the hypothesis that children would prefer to interact with a robot of a matching gender. However, with the experiment presented in this paper we aim to explore age- and gender-based differences in children’s opinion and engagement when playing with a robot of a matching gender.

2 Related Work

2.1 Gender Similarity Attraction in HCI and HRI

According to similarity attraction theory [29], people tend to exhibit similar gender favoritism in their daily life social interactions [17]. Nass et al. report the desirability of same-gender interactions within Human–Computer Interaction (HCI) [35,40].

When examining responses of 80 fifth-grade elementary school children to computer-synthesized speech in educational media in an experiment by Lee et al. [29], results show that children apply gender-based social rules to synthesized speech. More specifically, children evaluate more positively, trust the speech more, and learn more effectively when a voice’s gender matches either content gender and/or children’s own gender.

Another study by Ozogul et al. [36] investigated whether middle-school learners (11–13 years old), when given a choice of animated pedagogical agent, will select a young agent that matches their gender. The findings supports the similarity attraction hypothesis with significant preference ($p < 0.001$) in children’s choice for the computer-based animated agent that matched their age and gender.

Past HRI findings highlight the importance to explore the relation between adult’s gender and perceived robot gender in human–robot interaction [11,14,47]. An investigation of the perceptions of a gendered humanoid robot [37] found

that male and female participants treat and interact with the robot differently according to its gender, similar to a human–human communication. Persuasiveness of the robot’s gender was explored by Siegel et al. [49] concluding that adults were more often persuaded by a robot of the opposite gender to make a donation. The above findings highlight the importance of exploring the relation between subject gender and perceived robot gender in child–robot interaction. Woods [55] indicated the importance of gender in structuring even small details of human–robot interactions such as specific physical characteristics of robotic models, such as color and shape.

2.2 Age- and Gender-Based Differences in cHRI

In 2004, Kanda et al. [23] performed a field trial evaluation for two weeks (9 school days) with two age groups of elementary school Japanese students, 119 children aged 6–7 years old and 109 children aged 11–12 years old, and two English-speaking interactive humanoid robots (Robovie) behaving as peer English tutors. The study revealed that children of the younger age group spent more time interacting with the robot than older children did, and the robot sustained their interest longer. Unfortunately, authors do not comment on whether there were any gender differences in interactions with the robot.

In a study by Kennedy et al. [26], who investigated whether the social strategy of the robot tutor has an impact on learning with 45 children aged 7–8 years old. Findings were unexpected since children’s learning improved after interacting with a non-social robot in contrast to a more social robot. Authors suggest that it might be due to the distractions caused by the social behaviors of the robot on child’s concentration. There were also interesting gender differences found: girls improved significantly more with the physical robot present while boys barely improved in this condition.

An observational HRI study was conducted at the science museum in 2002 with a small wheeled teleoperated robot Sparky that had an expressive face [46]. The authors found interesting age and gender groups differences. Young children (4–7) tended to be very energetic and generally very kind to it regardless of gender. Older children (7 to early teens) had different interaction patterns depending on gender. Older boys were usually aggressive: they pushed it backwards and engaged in verbal abuse. Older girls were generally gentle with the robot: they often touched it, and were, on occasion, protective of the robot.

In 2014, Fink et al. [16] performed a field trial evaluation of the mobile robot “Ranger”—a robotic toy box that aims to motivate young children to tidy up their room. The authors found that younger children (3–6 years) gained the most benefit from using Ranger. They were more fascinated by the audio and light cues than older children (7–10 years).

There were also interesting qualitative gender differences. Boys mistreated and gestured toward Ranger more often than girls, who petted the robot slightly more often than boys.

Sandygulova and O’Hare [45] conducted an observational study with the NAO robot at the children’s museum with 74 children aged 3–9 year old aiming to explicitly investigate whether children would spend more time playing with the same-sex robot. The results support this hypothesis confirming that children of this age group demonstrate preferences for a gender-matching robot in their free play.

This section presented related work that addresses the role of gender in cHRI. This topic is often controversially discussed, with opinions ranging from accommodating cHRI by using same-gendered robots to using robots to change gender stereotypes and gender-based social roles in early education. According to the review of the child psychological literature on gender development by Zosuls et al. [57], research on children’s same-gender peer preferences, evaluations, and interactions are rarely framed in terms of discrimination, and is valuable to prevent acts of discrimination among children. Likewise, research on children’s same-gender robot preferences, interactions, and evaluations is required for better understanding the dynamics of child–robot interactions and for designing strategies to prevent discrimination (for example, using only male robot) and address gender-based stereotypes in child’s environments. This paper aims to address this issue by explicitly investigating children’s preferences, interactions, and evaluations of a gendered robot.

3 Robotic Platform

This research makes use of the NAO humanoid robot created by Aldebaran Robotics as a common development and evaluation platform. This robot platform has been used in a number of recent European projects such as CoWriter [20] and L2TOR [7]. Using such a shared platform facilitates the exchange of code and the transfer of results. The NAO is a small humanoid robot, measuring 58 cm in height, weighing 4.3 kg and having 25 degrees of freedom. The NAO has a generally friendly and non-threatening appearance, which is therefore particularly well suited for studies involving children [6].

NAO’s default voice (Operating System version 1.x) is an artificial child male voice, namely Kenny. However, artificial voices are often difficult to understand [51], to express meaning and intent and to relate to. In contrast, synthesized speech is a simulated speech created by concatenating pieces of prerecorded human narrator’s speech that are stored in a database. In fact, synthesized speech has already achieved an intelligibility level comparable to real human speech [29].

Aldebaran Robotics states that NAO and other company’s robots are gender-neutral i.e. have no gender [50]. How-

ever, the company refers to NAO as “him” on its website. In contrary, Belgian company Zorabots [58] designed the Zora software running on a NAO robot to have a personality and behaviours of an 8-year-old girl. Zora has been adopted by nearly 200 robots and deployed to nursing homes and children’s rehabilitation centers in Europe and USA. In Aldebaran Robotics’ press release [50], it states that people tend to assign a varying gender to NAO, depending on individual perceptions and cultural differences.

The colour of our NAO robot is orange, which is categorized as gender-neutral in the research of colour and toys [3]. Therefore, the assignment of the robot’s gender is decided to be manipulated with a gendered voice rather than modifying its appearance.

4 Research Questions and Hypotheses

Developmental research on gender [32] suggests that as early as preschool, children report feeling more positively about their own gender [56], and differential liking is also seen among older children [19,53]. However, findings are mixed regarding age trends: for example, Egan and Perry [12] and Powlishta et al. [38] suggest that intergroup biases decline in primary school while Yee and Brown [56] do not show the decline in liking the in-group better, at least not until early adolescence [53].

These findings in the research on gender development fuel the following hypotheses:

- **H1** Synthesized gendered voice evokes gender and age associations in social robots.
- **H2** A match between perceived robot gender and participant’s gender will have a positive effect on social interaction with the robot.
- **H3** Young children (approximately 5–8 years) will tend to like the interaction with robots with voices of the same gender better.
- **H4** There will be no difference in preferences for robots with male and female voices by older children (approximately 9–12 years).
- **H5** There will be Age \times Gender interaction effect (e.g. it might turn out that younger/older girls and younger/older boys both significantly prefer a robot with a particular gendered voice).

5 Method

The experiment was conducted over a three-week period and involved three meetings with a robot (i.e. one meeting once a week) for each child participant. All participants were assigned to a condition in a 2 (age group) \times 2 (child

Table 1 Breakdown of participants by age and gender

Grade level	Age (y.o.)	Gender		Group total
		F	M	
Senior infants	5	1	1	27
	6	12	13	
Year 2	7	2	2	22
	8	8	10	
Year 4	9	3	3	22
	10	7	9	
Year 5	11	12	13	36
Year 6	12	6	5	
Total				107

gender) \times 2 (robot voice gender) mixed-subject design, with age and child gender as between-subject variables and robot voice gender as within-subject variable.

Each child interacted with a robot for approximately ten minutes on three separate occasions one week apart. One week, the robot had a female voice, another week the robot had a male voice, and the final week the child got to choose what voice the robot would have. A three-phase counterbalancing method was implemented to reduce the chance of the order influencing the results. Half of the children interacted with the female voice first, the other half of the children interacted with the male voice first. Half of the children interacted with one version of the game first (Shapes), another half of the children interacted with the other version of the game first (Animals). Counterbalancing was also applied in terms of gender and year group, so that each condition has a balanced number of boys and girls in every age group. Assignment to each of the two gender conditions was otherwise random for any particular child.

5.1 Participants

A sample of 107 children was recruited from local primary school in the south Dublin area of the Republic of Ireland. The children came from diverse racial and socio-economic backgrounds and all were native or fluent English speakers. Children belonged to four classes: 27 children were senior infants, 22 students in year 2, 22 children were fourth graders, 25 students in year 5 and 11 six-graders. Table 1 presents the breakdown of participants by age and gender.

5.2 Ethics Statement

This research was approved by the University College Dublin's ethics committee for studies involving child participants. Informed consent was obtained in writing from all children and their parents, participating in the study. Sup-

porting information included Informed Consent Form for Children, Informed Consent Form for Parents or Guardians, Information Leaflet, and Questionnaires used before and after the experiment.

5.3 Scenario

The experiment was structured as an interactive game with a stationary sitting NAO robot. A child was instructed to sit in a chair at the table facing the robot (see Fig. 1). The second experimenter controlled the launch of each session through an iPhone application. This made the robot seem fully autonomous. As the child sat at the table, the NAO woke up and greeted the child.

5.3.1 First Two Meetings: Shapes/Animals Game

There were large cards lying on the table with printed pictures in between the robot and the child. As the robot asked for a particular card, the child had to find that picture among available cards and show it to the robot. The robot would check if the card was correct, and then would ask for another one. The game would be stopped either by a child or after five cards were tried. The robot would then summarize the number of correctly shown cards and congratulate the child.

There were two versions of the card game: Shapes and Animals. Only one version was played at one meeting with the robot. The Shapes version had pictures of a moon, a tree, a house, a star, a hand, a face, a heart, a puzzle, a butterfly and a triangle. Furthermore, pictures of a dog, a cat, a cow, a goat, a duck, a horse, a chicken, a sheep and a pig were in the Animals version. The game logic, robot's verbal and non-verbal behaviours were otherwise exactly the same between meetings.

The game was developed using robot's text-to-speech, face and image recognition engines. Throughout the game, the robot performed a series of relevant emotions of happiness, approval or sadness with the use of arm gestures and head movements. In addition, the robot expressed non-verbal social cues such as acknowledging child presence with eye contact and deictic gestures.

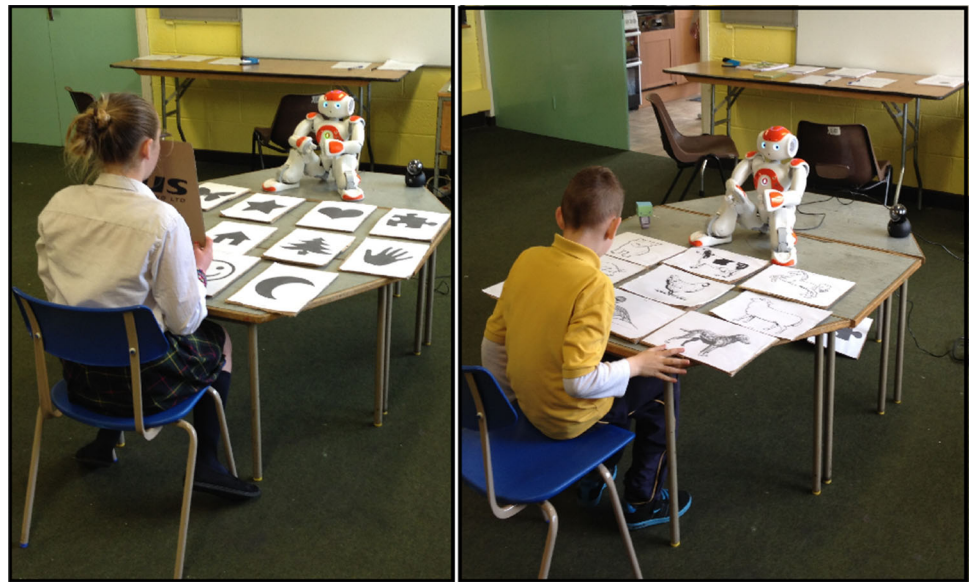
Each child was asked for a different set of five pictures as they were selected randomly by the game logic. An example of the robot's speech:

R: Hi! I am so glad to see you. I am a robot and my name is NAO. We are going to play a game with you. That's so exciting. Let's begin? (Pause). Now, please show me a butterfly [random shape or animal].

C: (child picks up and shows a particular card to the robot)

R: Yes. That is a butterfly! What's about showing me a puzzle [random shape or animal]?

Fig. 1 Children are playing Shapes (left) and Animals (right) card interactive games with the NAO robot



C: (child picks up and shows a particular card to the robot)
 R: Of course, that is a puzzle. Now, please show me a triangle [random shape or animal]!
 C: (child picks up and shows a particular card to the robot)
 R: Yes, that is a triangle! What about showing me a moon [random shape or animal]?
 C: (child picks up and shows a particular card to the robot)
 R: Good! That is a moon! Great! That was the last one! You made no mistakes. I could not do it better myself. It's been great playing with you. Thank you. See you next time!

5.3.2 Final Interaction Session: Stories

For the final meeting, NAO performed a story of a child's choice: "Three Musketeers" or "Monkey King". In addition to choosing a story, children were asked to choose the robot from the first or second meeting. They were also asked whether they prefer the robot to be a boy or girl.

5.4 Conditions

Voice was the only quality of the robot that was varied in the assignment of the gender. The robot's appearance was not modified, nor was any aspect of the robot's behavior. According to the findings from the previous study [44] on children's perceptions of synthesized voices, our robot's speech utterances were produced utilizing Acapella Group's children voices of English UK: Harry (male voice) and Rosie (female voice). Acapella Group [1] provides text-to-speech solutions to vocalize speech with authentic and original voices that express meaning and intent. In addition, this text-to-speech also provides the prosody of the human speech: a grammat-

ical and syntactic analysis enables the system to define how to pronounce each word in order to reconstruct the sense. As a result, these voices sound natural, express a particular accent and resemble the narrator's personality [1]. The Acapella Group toolkit was used to produce two versions of the speech utterances for each scenario.

5.5 Measures

Data were collected from both self-reported questionnaires and a camera that recorded the interactions.

Questionnaire

In order to choose the pictorial representations of scales, several versions of pictorial Likert scales were explored. In the end, our questionnaire utilized 5-Likert Smileyometer scale (Fig. 2 top) and the 5-Likert Self-Assessment Manikin (SAM) (Fig. 2 bottom) since both of these scales use 5 points and one of them is reversed to account for the primacy effect [4]. What is more, both of these scales have been previously used in the literature of cHRI to assess social robots specifically with children [6,30].

Similar to Belpaeme et al. [6], we included two questions to compare the robot to force-choice descriptors and to compare the perception of the robot as a social actor (see Social Pie Chart [6], Fig. 3).

The questions were approved by the child psychologist. Older children were self-reporting their perception of the robot, whereas younger children were assisted by the first researcher. In the end, the pictorial questionnaires were kept short and simple in order not to overwhelm the children.

Automatic Emotion Analysis

In order to increase the robustness of the results in evaluating children's attitude toward the interaction experience,

Fig. 2 Smileyometer (top) and Self-Assessment Manikin (SAM) (bottom)

Which picture best describes your mood now:



And which one of these pictures:

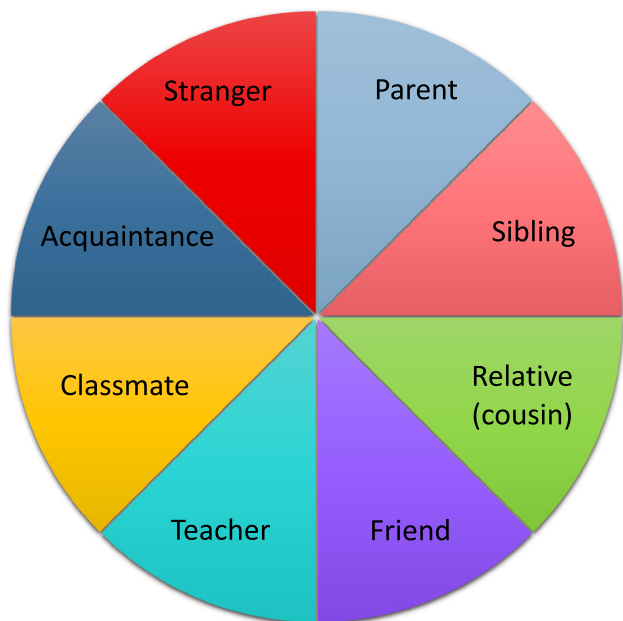


Fig. 3 Social pie chart

there was a camera placed in front of the child capturing facial expressions for real-time emotion analysis.

When compared to other face-based emotion recognition solutions, Sophisticated High-speed Object Recognition Engine (SHORE) [13] is reported to provide the best performance [2,24]. This package gives the intensity values of the following emotional states: happiness, sadness, surprise and anger. According to Alonso et al. [2], SHORE has 100% success rate for recognizing happiness.

For the purposes of this experiment, we only accounted for the expressed happiness of the participants for the first ten seconds of interaction i.e. greeting, introduction and invitation to play the game. This was to ensure that this measurement is not affected by the game flow. The intensity score of the expressed *Happiness* [0–100] is recorded every second over the first ten seconds to calculate the average score of the expressed happiness for each participant. We also counted the number of smiles during these ten seconds, which is reported

as *Percentage of Smiling*. However, in addition to SHORE's data, first ten seconds of all interactions were also manually video coded by the researcher.

In summary, as a result of careful consideration of questionnaire scales and emotion analysis tools, the following measures were utilized in this study:

- *Perceived robot gender*
- *Perceived robot age group*
- *Gender of child's most good friends*
- *Pre-and post-mood*: change in mood measured by the SAM (Fig. 2 bottom) and the Smileyometer (Fig. 2 top).
- *Valence*: mean of ratings of Enjoyment and child's stressfulness (Feeling) when playing with the robot (5-Likert pictorial scale)
- *Likeability*: mean of ratings of robot's perceived Friendliness and Kindness (5-Likert pictorial scale)
- *Robot type*: the robot is compared to forced-choice descriptors [6]
- *Social actor*: perception of the robot as a social actor (see Social Pie Chart [6], Fig. 3)
- *Happiness*: average score of the expressed happiness [0–100] during first ten seconds of interaction
- *Percentage of smiling*: number of smiles multiplied by 10 during first ten seconds of interaction
- *Preference for the robot's gender*: child's preference for the robot to be a boy or a girl
- *Preference for the robot*: child's choice for the robot from the first or second meeting

5.6 Experimental Setup

The experiment took place in a large empty room of the participating primary school, which is usually used for children's play. The interactions with the robots were in the main central area of the room at the large table. A web camera was behind the robot to the left recording the front of the children to enable the capture of facial expressions. The robot's movements involved moving its arms and head, otherwise it was always stationary sitting. Cards were placed on a table

in between the robot and a child necessary for the game. There was a table with questionnaires across the room from the entry door with two chairs, for the first researcher and the child participant. As a consequence of school policy it was necessary that a second adult be present in the room. Thus, the second researcher was sitting in the left corner of the room and was responsible for the timely launch of the robot's behaviour via a smartphone application, but was as non-reactive as possible in order to avoid interferences with the experiment. Experimental setup and room layout are illustrated in Figs. 1 and 4.

5.7 Procedure

Children were given a brief group introduction to the NAO robot at the school before the experiment commenced. In this introductory session children learned about the robot and the purpose of the study (investigate child–robot interaction). It was emphasized that they wouldn't be assessed or graded on what they did or said and that there were no right or wrong answers. That was explained because children should not be worried or distressed about the experiment [9]. It was to ensure that all the children had the background of the robot prior to the experiment, as having varying beliefs could affect the results [9]. In that introduction, it was explained that they would be expected to spend about ten minutes interacting with the robot and to answer a few questions of what they think of the robot on three separate occasions over three weeks.

At each of the three experimental meetings, each child participant was called out of the class and walked with the first researcher for approximately two-three minutes to the room with experimental setup (Fig. 4). While walking with the child, the researcher started with an icebreaker warm-up talk necessary to relax and engage the child. Cindy Dell Clark (2011) [9] in her book on doing child-centered qualitative research, "In a Younger Voice", suggests that it is a good idea to start by getting to know each other's name,

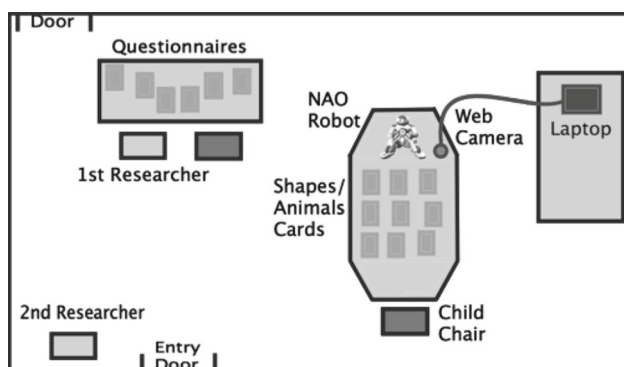


Fig. 4 Room setup

child's age, whether they have any siblings and the young person's birth order. "Tell me what it's like to be (the oldest/the youngest/the middle/the only one)." "Do you like being the oldest/the youngest/the middle/the only one?" Children often have a lot to say about the privileges and problems of their sibling position [9]. A lot of children had their siblings studying in the same school, who were also involved in this experiment so they were happy to share their stories with the researcher.

Upon entering the room, children were invited to take a seat at the table with questionnaires and answer a few questions about their age, gender and mood prior to the interaction with the robot. After the first questionnaire was filled in, children were invited to sit at the table with cards facing the robot. After the interaction, children were asked to fill in the second questionnaire. In the end, the first researcher brought the child back to the class and called out the next participant.

After the experiment had taken place, the children were given a debrief as a class to explain how the NAO robot works, presented with some demonstrations and given opportunity to ask questions.

5.8 Results

The first part of this section reports the results of the conducted manipulation checks. It is followed by the report of the results from the first robot encounter i.e. the data is analyzed with between-subject statistical tests. The second encounter results are not detailed here since the majority of younger children did not perceive the Rosie voice as female. These findings are discussed in detail in the next subsection. The last part of this section reports on children's choices for the robot gender.

5.8.1 Manipulation Check

Perceived Robot Age Group

First it was assessed, whether the perceived robot age group corresponded with the intended robot age group. Most participants ($N=63$) believed the robot to belong to the primary school age group (58.9%). Of these 63 participants, 20 were in age group 1 (5–8 years old children), while 43 were in age group 2 (9–12 years old children). 11.2% ($N=12$) thought the robot was a secondary school child, 12.1% ($N=13$) thought the robot was an adult and 9.3% ($N=10$) were unsure. It should be noted that of those who believed the robot to be an adult, all but one belonged to a younger age group ($N=12$), suggesting that children at that young age might be less likely to attribute robot age to voice characteristics.

Perceived Robot Gender

Then it was assessed, whether the perceived robot gender corresponded with the intended robot gender. The data from

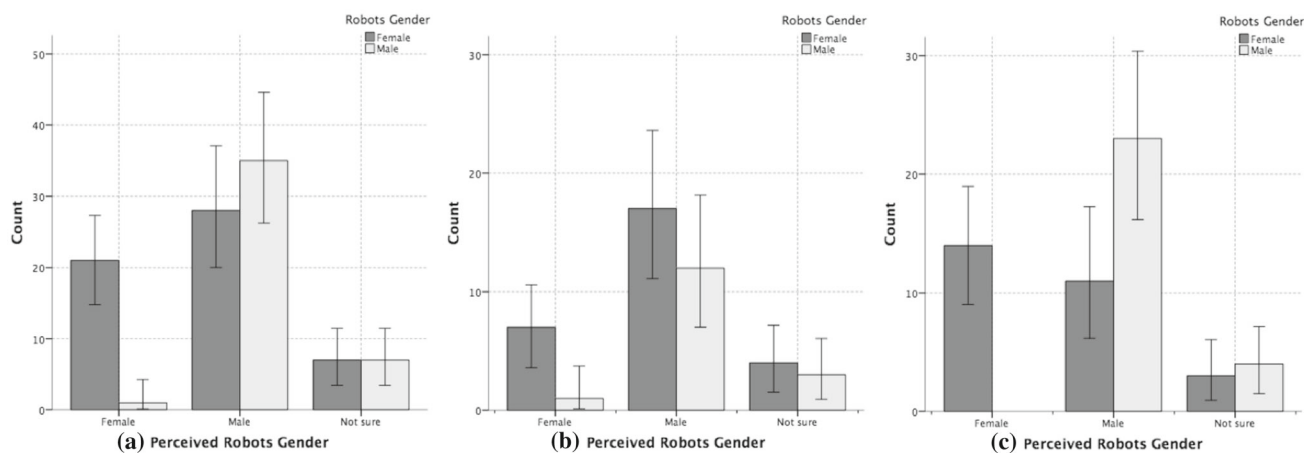


Fig. 5 **a** Overall robot gender classification based on voice. **b** Robot gender classification by younger children. Rosie was classified as male by the majority of young children. **c** Robot gender classification by older children

Fig. 5a show that when the robot used Rosie's voice, 21 participants classified it correctly as female, while 28 participants believed it to be male. In contrast, when the robot used Harry's voice, only one participant believed it to be female, whereas 35 identified it correctly as male. In both cases, 7 participants were unsure. Hence, approximately two thirds of participants believed that they interacted with a male robot (63 out of 99, excluding 8 incomplete data sets). Male and female children were approximately equally likely to identify the robot as male ($\chi^2(1, N = 107) = 3.31, p = 0.069$). However, there was a distinct difference as determined by Pearson's Chi-square test in how younger and older children were able to identify the gender of the robot: $\chi^2(1, N = 107) = 8.89, p = 0.003$. Robot gender classification by two age groups, younger children and older children, is presented in Fig. 5b, c respectively. Rosie was perceived to be male by the majority of young children, while fewer than half of the older children classified it incorrectly.

Also, the responses of children's perception of gender were analyzed after the second encounter with the robot. Again, the robot's gender was perceived in accordance with robot's gendered voice by 56 participants. Similarly, boys and girls were equally likely to pass the manipulation check ($\chi^2(1, N = 91) = 0.054, p = 0.82$). Once again, there was a statistical significant difference between age groups ($\chi^2(1, N = 91) = 7.08, p = 0.008$). 24 younger children and 32 older children perceived the robot's gender in accordance with a gendered voice.

Lastly, the final manipulation check included combining the data of those participants who correctly perceived the robot gender after *both* robot encounters i.e. the robot was perceived to be male when the robot spoke with Harry's voice and it was perceived to be female when the robot spoke with Rosie's voice. There were 23 participants (10 female and 13 male) that satisfied this final manipulation check. Again, girls

and boys did not have a significant difference in their ability to evoke gender association ($\chi^2(1, N = 96) = 0.625, p = 0.63$). However, there was a statistically significant difference in this ability between younger and older age groups: ($\chi^2(1, N = 96) = 14.59, p < 0.001$). In the end, there were 4 participants in the younger age group and 19 participants in the older age group who passed both manipulation checks.

Gender of Participants' Friends

We performed Chi-square test for independence in order to assess children's responses about the gender of their best friends. We wanted to see if there is an association between the child's gender and the reported friend's gender. Pearson Chi-square value indicates a significant association between the child's gender and their friend's gender, $\chi^2(2, N = 100) = 67.77, p < 0.001, \phi = 0.823$.

5.8.2 Independent samples analysis from the first robot encounter

First it should be noted that we excluded the data of the participants that failed manipulation checks (i.e. robot gender was not perceived as expected) from further statistical analysis. Thus, this section reports the results of 55 children: 22 girls and 33 boys. Secondly, all remaining participants were classified as "same-gender" when their own gender matched with the perceived robot gender, and "opposite-gender", when the children believed they interacted with a robot of the opposite sex. Overall, 34 participants were classified as "same-gender", with 21 classified as "opposite gender" for the between-subjects statistical tests on the data from the first robot encounter. Similarly, all remaining participants were classified into two age groups based on their grade year: younger children (senior infants and 1st year students) and older children (2nd year, 3rd year and 4th year

students). Manipulation check was passed by 18 younger and 37 older children.

Tests of Normality Within Groups

A series of Kolmogorov–Smirnov (K–S) and Shapiro–Wilk tests was conducted on all dependent variables overall and within groups (i.e. gender, age group, robot gender, and robot condition) to check the assumption of normality. All scores were significantly non-normal at 0.001 level. Consequently, non-parametric tests were used for the statistical data analysis presented in the next sections. It should be noted that overall *Average Expressed Happiness* scores were significantly non-normal. However, when normality was analyzed on each category separately, the scores did not deviate significantly from normal ($p > 0.05$).

Pre-and Post-mood

First it was tested, whether the two mood measures (Smileyometer and reversed SAM) correlated significantly. A bivariate Spearman's rank-order correlation for ordinal data showed a small, non-significant correlation between the two pre-intervention measures ($r=0.17$, $p = 0.21$) and between the two post-intervention measures ($r=0.03$, $p = 0.81$) for the first encounter. For the second encounter, the items correlated more strongly, with medium-sized significant correlations for the pre-intervention measures ($r=0.31$, $p = 0.03$) and the post-intervention measures ($r=0.45$, $p = 0.002$). Overall, the results indicate that the two mood measures assessed different aspects of mood and thus constituted two independent measures.

Then, a series of independent-samples Mann–Whitney U tests was conducted to separately compare *Mood Change* and *Mood Change 2* rates between groups. Firstly, Mann–Whitney U comparisons showed no significant effect on participants' ratings between same-gender and opposite-gender groups (Hypothesis 2). Secondly, the difference in ratings between age groups was not significant. Thirdly, there was no statistically significant difference between boys' and girls' ratings of mood changes. Finally, the difference in ratings between perceived robot gender was not significant.

In order to test Hypotheses 3 and 4, *Age × Gender Matching* effect was explored with independent-samples Mann–Whitney U tests for younger and older children. Change in mood measured by SAM did not differ significantly between gender matching conditions, $U = 19.00$, $z = -1.10$, $p = 0.38$, $r = -0.26$, for younger age group. However, change in older children's ratings of mood before and after the interaction was significantly lower in the opposite-gender condition ($M = -0.24$ (decreased by 0.24) ± 0.66) compared to children who interacted with the robot that matched their gender ($M = 0.55$ (increased by 0.55) ± 1.43), $U = 236.00$, $z = 2.15$, $p = 0.045$, $r = -0.35$.

In order to test Hypothesis 5, *Age × Gender* and *Age × Robot Gender* effects were explored with separate independent-samples Mann–Whitney U tests for younger and older children. However, we did not find significant differences in mood ratings before and after the interaction between younger boys and girls and older boys and girls as well as between female and male robots for younger and older age groups.

Valence

Firstly, *Enjoyment* variable was reversed. There was one outlier case that was not included in the analysis. Secondly, Cronbach's alpha showed poor (0.18) internal consistency of children's ratings of their *Enjoyment* and *Feeling*. Consequently, these measures were analyzed as separate dependent variables.

Then, the same procedure of statistical analysis was conducted for the ratings of *Enjoyment* and *Feeling* to test Hypotheses 2–5. However, these ratings did not show significant interaction between independent variables retaining null hypotheses.

Likeability

The questionnaire results of children's ratings of the robot's *Friendliness* and *Kindness* showed poor internal consistency with Cronbach's alpha score 0.58 for all participants. Therefore, these results were considered within separate evaluations. A series of Mann–Whitney tests was conducted to examine the effect of gender matching, gender and age group on children's ratings of *Friendliness* and *Kindness*. We did not find any statistical significance in interactions between groups for these measurements retaining null hypotheses.

Robot Description

A series of Pearson Chi-square tests for independence was conducted in order to assess children's choices for the *Robot Type* and *Social Actor* questions. However, we did not find any statistical significant differences in children's responses for these questions between robot conditions, gender or age groups.

There was a significant association between age groups and what children replied for the *Social Actor* multiple choice question: $\chi^2(1, N = 54) = 16.97$, $p = 0.004$. Only 35.3% of younger children compared the robot to a friend while 75.7% of older children made this comparison. The rest of the answers for younger children were as follows (sorted in descending order): brother or sister (17.6%), stranger (17.6%), classmate (11.8%), cousin or other relative (11.8%), teacher (5.9%). Older children compared the robot to brother or sister (10.8%), parent (5.4%), teacher (5.4%), and relative (2.7%). Interestingly, younger children did not compare the

robot to a parent while older children never compared the robot to a stranger or classmate.

There were non-significant differences in responses between boys and girls as well as between perceived male and female robots. *Robot Type* measure did not evoke significant differences between groups.

Percentage of Smiling

A series of independent-samples Mann–Whitney U tests was conducted to compare *Percentage of Smiling* between groups. We found a statistically significant difference between perceived robot gender: $U = 219.00$, $z = -2.00$, $p = 0.045$, $r = -0.28$. The mean values indicate a tendency for children to smile more during the interaction with a robot that they perceived to be female ($M = 81.5 \pm 35.73$) in comparison to male robot ($M = 74.68 \pm 26.15$). However, the difference in smiling was not significant between age groups as well as between boys and girls. Finally, Hypothesis 2 was not supported either i.e. child's age had no effect on smiling at the robot.

In order to test Hypotheses 3–5, *Age × Gender Matching*, *Age × Gender* and *Age × Robot Gender* effects were analyzed with independent-samples Mann–Whitney U tests for individual groups. However, the differences in smiling were not significant retaining null hypotheses.

Average Expressed Happiness

A series of one-way ANOVA tests was conducted to determine if there was any statistical significant difference in *Expressed Happiness* behavioral data between groups. The difference between same-gender and opposite-gender was not significant. Significance between age and gender groups was not found. Levene's tests were non-significant, thus indicating that the assumption of variance equality was not violated.

To test our Hypothesis 3, a two-way ANOVA test was conducted which examined the effect of *Age × Gender-matching Interaction*. The interaction between the effects of age and gender matching was statistically significant, $F(1, 45) = 4.64$, $p = 0.037$. There was a statistically significant interaction found for younger children: ($F(1, 14) = 6.06$), $p = 0.027$) between opposite-gender (73.00 ± 19.82) and same-gender (40.08 ± 23.98) groups. In contrast, older children had a non-significant differences in Happiness between same-gender (40.89 ± 26.97) and opposite-gender (34.07 ± 31.32) conditions.

5.8.3 Preference for the Robot's Gender

Finally, a series of Pearson Chi-square test for independence was performed in order to analyze children's responses when asked "Would you prefer the robot to be a boy or a girl?".

Children could reply with a particular gender or respond that it does not matter. Firstly, we included all 107 children in the analysis. There was a statistical significant difference in children's answers between boys and girls: $\chi^2(2, N = 91) = 27.97$, $p < 0.001$. 42.9% ($N = 18$) of female participants said that they would prefer the robot to be a girl while 50.0% ($N = 21$) of female respondents indicated that the gender of the robot was not important. At the same time, 55.1% ($N = 27$) of male participants said that they would prefer a boy robot and 36.7% ($N = 18$) said that they do not have a particular gender preference for the robot. The remaining 3 boys and 4 girls indicated that they would prefer the robot of the opposite gender to them.

There was also a significant difference in children's preferences depending on age: $\chi^2(2, N = 91) = 30.65$, $p < 0.001$. A girl robot and a boy robot were preferred by 36.7 and 46.9% of younger children respectively. A remaining 16.3% replied that they did not mind either gender. In contrast, 73.8% of older children did not have a particular preference for the robot's gender. A boy robot and a girl robot were selected by 16.7 and 9.5% of older children respectively.

In order to investigate *Age and Gender* effect, we conducted separate Pearson Chi-square tests for younger and older groups in order to explore an association between the child's gender and their choice for the robot's gender. Pearson Chi-square value indicates a significant association between child's gender for younger children: $\chi^2(2, N = 49) = 20.01$, $p < 0.001$. 60.9% of younger girls and 76.9% of younger boys indicated their preference for the same-gendered robot i.e. a girl robot and a boy robot respectively. 13.0% of younger girls and 15.4% of younger boys replied that they would prefer a boy or a girl robot respectively. No particular gender preference was reported by 26.1% of younger girls and 7.7% of younger boys.

When analyzing older children's responses there was a statistically significant difference between boys and girls as indicated by Pearson Chi-square test: $\chi^2(2, N = 42) = 10.75$, $p = 0.005$. 78.9% of older girls and 69.6% of older boys reported that it did not matter to them. 21.1% of older girls and 30.4% of older boys preferred the robot of the same gender as them. Interestingly, none of the girls and none of the boys indicated the preference for the robot of the opposite gender.

5.9 Discussion

In this study, we addressed the role of age and gender on children's interactions with a social robot. Robot's gender was manipulated with the help of two synthesized child gendered voices namely Rosie and Harry. This design of manipulations was not successful in evoking gender and age associations in younger children. In particular, the majority of younger children (5–8 years old) perceived the robot as male. In con-

trast, older children were able to successfully attribute age and gender to the robot in correspondence with the synthesized voice. These findings align themselves with the results from the previous study [44]. Thus, these findings reject our Hypothesis 1 for younger children and support Hypothesis 1 for older children i.e. synthesized voice evokes gender and age associations in social robots for older children but not for younger children.

Unfortunately, the design of robot gender manipulation failed for young children since most of children in the younger age group believed that they interacted with the male robot. Consequently, we cannot generalize the results from questionnaires and video coding analysis, thus we can't accept or reject Hypothesis 2 for younger children. As a result, we only discuss the results for older children.

Hardly any significant effects were found with questionnaire data which suggests that gender matching is not really a factor in the children's rating of the human–robot interaction. However, the results from the self-reported mood change suggest that our Hypothesis 2 is supported for older children. Older children reported their mood to be significantly lower on average in the opposite-gender condition compared to older children in the same-gender condition.

An important finding from behavioural data is the fact that children reacted significantly more positively towards the female robot—regardless of their own gender. This finding supports our Hypothesis 2 for girls but not for boys suggesting that Hypothesis 5 is true for older boys toward the female robot. As stated earlier, more than 69% of boys reported that the gender of the robot did not matter, however they smiled significantly more towards the female than towards the male robot.

The main measurement of this experiment—children's choice for the robot's gender—concludes an important finding that Hypothesis 3 is supported for younger children. When asked for their preference for the gender of the robot younger children preferred the robot of the same gender as them ($p < 0.001$). It should be noted that the responses of all 49 children of this age group were included in the analysis. This finding supports our Hypothesis 3: young girls and boys (5–7 years old) prefer to interact with a social robot that is of the same gender (60.9% of girls and 76.9% of boys). Interestingly, younger girls were more flexible in their choice: 26.1% of girls would not mind the gender of the robot. In fact, 13.0% of girls and 15.4% of boys had indicated their preference towards the opposite gender of the robot.

Our Hypothesis 4 was also supported for older children: a significant majority of older children, 78.9% of girls and 69.6% of boys, reported that the gender of the robot did not matter to them. The remaining 21.1% of older girls and 30.4% of older boys preferred the robot of the same gender as them.

The robot of the matching gender was compared to a friend by the majority of children in both age groups when

asked to compare the robot to either a parent, a teacher, a cousin, a stranger, and other social roles (Fig. 3). Similarly, the majority of older children compared playing with the robot to playing with a friend even in cases when the robot's and child's gender did not match. According to Cook [10], acceptance of cross-gender behavior and appearance increases during middle primary school (around the age of 9) which supports the theory that older children were significantly more likely to compare the robot of the opposite gender to a friend. On the contrary, only 35.3% of younger children compared the robot to a friend. They often compared the robot to their sibling, stranger, classmate, relative, or teacher. These results suggest that younger children were significantly less likely to consider a robot of the opposite gender as a friend since younger children are less likely to have friends of the opposite gender, which is in line with the trends in gender development known as gender segregation [10,21,31], which rigidity starts to gradually decrease after 7–8 years of age.

5.10 Limitations

There are a few limitations of this experiment concerning child–robot evaluation and its measurements that should be discussed.

Firstly, it would help to measure pre-concept gender and age attribution of the robot. Since the robot used in the study has had an extensive media coverage in the last years, children might have predetermined attribution based on previous experiences. Thus, it would be useful to check whether this attribution changes as they hear the voice of the robot. It would also be useful to ask why children think the robot is of a particular gender and age and whether it is due to the robot's appearance or voice. It would help to make clear conclusions about child–robot interaction similar to the theory of doubly disembodied language (i.e. tendency to imagine social characteristics of a speaker (e.g. age and gender) when communicating with software agents or robots) [29] but for young children.

Another limitation of our evaluation was the perceived age of the robot. It would be useful if children stated the exact age of the robot as a continuous variable instead of their perception of the age group. It would allow us to measure whether they think the robot is younger or older than them. This would help to further investigate the peer segregation hypothesis.

In addition, it would be useful to know whether a particular child has siblings of the opposite gender. This would help to investigate it further whether their preference for the particular gendered robot is due to having siblings of the same gender. Finally, there might be differences among cultures, which will be important to investigate in future work.

6 Conclusion and Future Work

To shed light on the impact of children's developmental tendency of *gender segregation* within child–robot interaction, a large-scale study was conducted, inviting 107 children aged between 5 and 12 years old to interact with the robot over three sessions one week apart. The robot's gendered voice was manipulated in two experimental conditions to either child male or female synthesized voice. The children's choice of the robot's gender for the third robot encounter was then assessed. The results of this question's responses suggest that girls and boys have varying preferences for the gender of the robot according to their age. Younger children (5–8 years old) reported that they would prefer the robot have the same gender as them while older children (9–12 years old) responded that the gender of the robot did not matter to them. In contrast to what older children said, they reacted more positively (smiled more) to the female robot regardless of their own gender.

In contrast to these findings, the results from subjective and behavioural measures (expressed happiness) adopted during the study did not support our gender-matching hypothesis (Hypothesis 2). Since children report to prefer same-gendered robots, the reason for the insignificant results could be explained by the fact that children's overall positive responses hid the effect. The children were obviously excited to interact with a robot so their baseline is already near the top of the questionnaire with no room for them to show a dramatic change in the responses. A similar explanation was reported by Chiasson [8] in their investigation of Media Equation effect [40] with children.

Another explanation could be the challenging nature of child-centered research. Other researchers in the field of cHRI have reported similar challenges with Likert scales' extreme responses [5]. Additionally, behavioural measures such as facial expressions have their advantages but are also subject to individual differences between children (expressive vs. non-expressive interaction styles) [41].

This experiment has raised a number of questions, and further work is needed in several directions to adequately determine how gender affects children. First, the challenges of effectively evaluating child–robot interaction need to be addressed and alternative methods are required, in particular for young children. Second, we will look at the issue of children's overall positive responses by finding tasks where the children's baseline is lower similar to Chiasson's suggestion [8]. Our future studies need to include a scenario that may be not particularly fun to do and children might not otherwise choose to do it to try and lower the baseline response. Third, the synthesized voice did not evoke gender association in younger children (5–8 years old). It is important to address this limitation in the future studies by adding an addi-

tional gender association, for example for a robot to have a gendered name.

In summary, the findings from the experiment can be summarized in the following recommendations:

- Children aged between 5 and 8 years old do not attribute age and gender characteristics to robot's synthesized voice.
- Children aged between 9 and 12 years old recognize gender and age identity of robot's synthesized voice.
- Children aged between 5 and 8 years old prefer to interact with a same-gender robot.
- Children aged between 3 and 8 years old conform to gender segregation in their interactions with a gendered robot. These findings are based on their self-reported preferences for the robot and on the observation of their natural engagement with the same-gender robot reported in [45].
- Children aged between 9 and 12 years old are flexible about robot gender. These findings are based on their self-reported preferences for the robot and on their expressed happiness during interactions with same-gender and opposite-gender robot.

These recommendations provide a strong motivation to design a social robot that evokes gender associations in young children in order to improve and maintain rapport and engagement, which are essential for educational and therapeutic benefits to take effect. In scenarios when a robot takes a role of a friend or a peer with young children, interaction designers should consider accommodating children's preferences for same-gender robotic peers. It is also sensible to provide children with a choice or an option to interact with a gender-matching robot. This could simply be achieved by having a robot introduce itself with a gendered name (e.g. Rosie vs. Robie). In relation to robot's appearance, social robots should be designed to look as gender-neutral as possible (e.g. a humanoid robot Pepper) to support gender manipulation per needs and preferences of users from various age and gender groups. This type of adaptation would be relevant for one-to-one interactions rather than for group contexts as it would be strange to have a robot dynamically change its gender. Future work is needed to inform whether robot's gender has an effect when a social robot takes a role of a tutor, learner, game opponent, and other social roles on child's learning or therapy. Long-term studies investigating long-term gender effects need to be carried out too. This paper contributes to the literature of HCI and, in particular, HRI, by reporting upon important considerations when designing social robots and robotic applications for children.

Acknowledgements The authors would also like to express great appreciation to the members of staff of the Queen of Angels primary school for their time and help with the experiment. Special thanks to Dr. Maria Nikolayev (Sumaroka), who provided very valuable advice and suggestions in designing the experiment. In addition, great gratitude should be given to Dr. Verena Nitsch for her help with data analysis. The authors would like to thank the reviewers for their valuable feedback and suggestions to improve the article.

Funding This study was funded by Irish Research Council and Science Foundation Ireland (07/CE/11147).

Compliance with ethical standards

Conflicts of interest Anara Sandygulova has received research grants from Irish Research Council.

References

1. Acapela group. <http://www.acapela-group.com/>. Accessed 30 Aug 2015
2. Alonso-Martín F, Malfaz M, Sequeira J, Gorostiza JF, Salichs MA (2013) A multimodal emotion detection system during human–robot interaction. *Sensors* 13(11):15549–15581
3. Auster CJ, Mansbach CS (2012) The gender marketing of toys: an analysis of color and type of toy on the disney store website. *Sex Roles* 67(7–8):375–388
4. Bell A (2007) Designing and testing questionnaires for children. *J Res Nurs* 12(5):461–469
5. Belpaeme T, Baxter P, de Greeff J, Kennedy J, Read R, Looije R, Neerinx M, Baroni I, Zelati M (2013) Child-robot interaction: perspectives and challenges. In: Herrmann G, Pearson M, Lenz A, Bremner P, Spiers A, Leonards U (eds) *Social robotics, volume 8239 of Lecture Notes in Computer Science*. Springer, Berlin, pp 452–459
6. Belpaeme T, Baxter P, Read R, Wood R, Cuayahuitl H, Kiefer B, Racioppa S, Kruijff-Korbayová I, Athanasopoulos G, Enescu V, Looije R, Neerinx M, Demiris Y, Ros-Espinoza R, Beck A, Cañamero L, Hiolle A, Lewis M, Baroni I, Nalin M, Cosi P, Paci G, Tesser F, Somavilla G, Humbert R (2012) Multimodal child-robot interaction: building social bonds. *J Hum-Robot Interact* 1(2):33–53
7. Belpaeme T, Kennedy J, Baxter P, Vogt P, Krahmer EE, Kopp S, Bergmann K, Leseman P, Küntay AC, Göksun T et al (2015) L2TOR-second language tutoring using social robots. In: *Proceedings of the ICSR 2015 WONDER workshop*
8. Chiasson S, Gutwin C (2005) Testing the media equation with children. In: *Proceedings of the SIGCHI conference on human factors in computing systems, CHI '05*, New York, NY, USA. ACM, pp 829–838
9. Clark C (2010) In a younger voice: doing child-centered qualitative research. *Child development in cultural context series*. Oxford University Press, Oxford
10. Cook J, Cook G (2003) *Child development: principles and perspectives*. Allyn & Bacon, Boston
11. Crowelly C, Villanoy M, Scheutzz M, Schermerhornz P (2009) Gendered voice and robot entities: perceptions and reactions of male and female subjects. In: *IEEE/RSJ international conference on intelligent robots and systems, 2009. IROS 2009*. IEEE, pp 3735–3741
12. Egan SK, Perry DG (2001) Gender identity: a multidimensional analysis with implications for psychosocial adjustment. *Dev Psychol* 37(4):451–463
13. Ernst A, Ruf T, Kueblbeck C (2009) A modular framework to detect and analyze faces for audience measurement systems. In: *2nd workshop on pervasive advertising at informatik*, pp 75–87
14. Eyssele F, Kuchenbrandt D, Bobinger S, de Ruyter L, Hegel F (2012) ‘If you sound like me, you must be more human’: on the interplay of robot and user features on human–robot acceptance and anthropomorphism. In: *Proceedings of the seventh annual ACM/IEEE international conference on human–robot interaction*. ACM, pp 125–126
15. Fine C (2010) *Delusions of gender: how our minds, society, and neurosexism create difference*. WW Norton & Company, New York
16. Fink J, Lemaignan S, Dillenbourg P, Rétornaz P, Vaussard F, Berthoud A, Mondada F, Wille F, Franinović K (2014) Which robot behavior can motivate children to tidy up their toys? In: *Proceedings of the 2014 ACM/IEEE international conference on human–robot interaction—HRI '14*, New York, NY, USA. ACM Press, pp 439–446
17. Foley S, Linnehan F, Greenhaus JH, Weer CH (2006) The impact of gender similarity, racial similarity, and work culture on family-supportive supervision. *Group Org Manag* 31(4):420–441
18. Fussell SR, Kiesler S, Setlock LD, Yew V (2008) How people anthropomorphize robots. In: *Proceedings of the 3rd ACM/IEEE international conference on human robot interaction*. ACM, pp 145–152
19. Heyman GD (2001) Children’s interpretation of ambiguous behavior: evidence for a ‘boys are bad’ bias. *Soc Dev* 10(2):230–247
20. Hood D, Lemaignan S, Dillenbourg P (2015) The CoWriter project: teaching a robot how to write. In: *Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction extended abstracts*. ACM, pp 269–269
21. Huston AC (1983) Sex-typing. In: Mussen PH, Hetherington EM (eds) *Handbook of child psychology, vol 4. Socialization, personality and social behavior*, 4th edn. Wiley, New York, pp 387–467
22. Johnson J (2003) *Children, robotics, and education*. *Artif Life Robot* 7(1–2):16–21
23. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. *Hum Comput Interact* 19(1):61–84
24. Katibeh Z (2013) *Techniques for real-time multi-person face tracking for human–robot dialogue*. Ph.D. thesis, Blekinge Institute of Technology
25. Kaufman SB, Singer JL, Singer DG (2012) The need for pretend play in child development. *Psychol Today*
26. Kennedy J, Baxter P, Belpaeme T (2015) The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning. In: *Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction*. ACM, pp 67–74
27. Kose-Bagci H, Ferrari E, Dautenhahn K, Syrdal DS, Nehaniv CL (2009) Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Adv Robot* 23(14):1951–1996
28. Kriz S, Anderson G, Trafton J, Bugajska M (2009) Robot-directed speech as a means of exploring conceptualizations of robots. In: *2009 4th ACM/IEEE international conference on human–robot interaction (HRI)*, pp 271–272
29. Lee KM, Liao K, Ryu S (2007) Children’s responses to computer-synthesized speech in educational media: gender consistency and gender similarity effects. *Hum Commun Res* 33(3):310–329
30. Leite I, Castellano G, Pereira A, Martinho C, Paiva A (2012) Long-term interactions with empathic robots: evaluating perceived support in children. In: Ge SS, Khatib O, Cabibihan JJ, Simmons R, Williams M-A (eds) *Social robotics*. Springer, Berlin, pp 298–307
31. Maccoby EE (1988) Gender as a social category. *Dev Psychol* 24(6):755
32. Martin CL, Ruble DN (2010) Patterns of gender development. *Annu Rev Psychol* 61:353–381

33. Mehta CM, Strough J (2009) Sex segregation in friendships and normative contexts across the life span. *Dev Rev* 29(3):201–220
34. Meiribekov S, Balkibekov K, Jalankuzov Z, Sandygulova A (2016) You win, I lose: towards adapting robot's teaching strategy. In: The Eleventh ACM/IEEE international conference on human robot interaction. IEEE Press, pp 475–476
35. Nass CI, Brave S (2005) *Wired for speech: how voice activates and advances the human–computer relationship*. MIT Press, Cambridge
36. Ozogul G, Johnson AM, Atkinson RK, Reisslein M (2013) Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Comput Educ* 67:36–50
37. Powers A, Kramer AD, Lim S, Kuo J, Lee S-I, Kiesler S (2005) Eliciting information from people with a gendered humanoid robot. In: IEEE international workshop on robot and human interactive communication, 2005. ROMAN 2005. IEEE, pp 158–163
38. Powlishta KK, Serbin LA, Doyle A-B, White DR (1994) Gender, ethnic, and body type biases: the generality of prejudice in childhood. *Dev Psychol* 30(4):526–536
39. Rao AS, Georgeff MP et al (1995) BDI agents: from theory to practice. *ICMAS 95*:312–319
40. Reeves B, Nass C (1996) *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, New York
41. Ros R, Nalin M, Wood R, Baxter P, Loojje R, Demiris Y, Belpaeme T, Giusti A, Pozzi C (2011) Child–robot interaction in the wild: advice to the aspiring experimenter. In: Proceedings of the 13th international conference on multimodal interfaces, ICMI '11, New York, NY, USA. ACM, pp 335–342
42. Ruble DN, Martin CL, Berenbaum SA (2007) *Gender development*. Wiley, Hoboken
43. Sandygulova A, Dragone M, O'Hare GM (2014) Investigating the impact of gender development in child–robot interaction. In: Proceedings of the 3rd international symposium on new frontiers in human–robot interaction at AISB 2014, Goldsmiths College, London, UK
44. Sandygulova A, O'Hare GM (2015) Children's perception of synthesized voice: robot's gender, age and accent. In: Tapus A, André E, Martin J-C, Ferland F, Ammi M (eds) *Social robotics*. Springer, Cham, pp 594–602
45. Sandygulova A, O'Hare GM (2016) Investigating the impact of gender segregation within observational pretend play interaction. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 399–406
46. Scheeff M, Pinto J, Rahardja K, Snibbe S, Tow R (2002) Experiences with Sparky, a social robot. In: *Socially intelligent agents*, pp 173–180
47. Schermerhorn P, Scheutz M, Crowell CR (2008) Robot social presence and gender: Do females view robots differently than males? In: Proceedings of the 3rd ACM/IEEE international conference on human robot interaction. ACM, pp 263–270
48. Serbin LA, Moller LC, Gulko J, Powlishta KK, Colburne KA (1994) The emergence of gender segregation in toddler playgroups. *New Dir Child Adolesc Dev* 65:7–17
49. Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In: IEEE/RSJ international conference on intelligent robots and systems, 2009. IROS 2009. IEEE, pp 2563–2568
50. Softbank robotics. <https://www.ald.softbankrobotics.com/en/press/faq-about-pepper>. Accessed 30 Aug 2015
51. Tamagawa R, Watson C, Kuo I, MacDonald B, Broadbent E (2011) The effects of synthesized voice accents on user perceptions of robots. *Int J Soc Robot* 3(3):253–262
52. Tung F-W (2011) Influence of gender and age on the attitudes of children towards humanoid robots. In: Jacko J (ed) *Human–computer interaction. Users and applications*, volume 6764 of Lecture Notes in Computer Science. Springer, Berlin, pp 637–646
53. Verkuyten M, Thijs J (2001) Ethnic and gender bias among Dutch and Turkish children in late childhood: the role of social context. *Infant Child Dev* 10(4):203–217
54. Wang Y, Young JE (2014) Beyond “pink” and “blue”: gendered attitudes towards robots in society. In: Proceedings of gender and IT appropriation. Science and practice on dialogue—forum for interdisciplinary exchange, gender IT '14, Siegen, Germany, European Society for Socially Embedded Technologies, pp 49:49–49:59
55. Woods S (2006) Exploring the design space of robots: children's perspectives. *Interact Comput* 18(6):1390–1418
56. Yee M, Brown R (1994) The development of gender differentiation in young children. *Brit J Soc Psychol* 33(2):183–196
57. Zosuls KM, Miller CF, Ruble DN, Martin CL, Fabes RA (2011) Gender development research in sex roles: historical trends and future directions. *Sex Roles* 64(11–12):826–842
58. Zora. <http://zorarobotics.be/?lg=en/> Accessed 30 Aug 2015

Anara Sandygulova is an Assistant Professor in the Department of Robotics at the School of Science and Technology. Dr. Sandygulova received her Ph.D. from the School of Computer Science at University College Dublin. For her research and outreach activities, Dr. Sandygulova was awarded the prestigious Google Anita Borg Scholarship in July 2014. As an undergraduate student, she studied Computer Science at University College Dublin and University of California Irvine. In addition to graduating with First Class Honors, she was the 2011 class recipient of the John Kelly Memorial Medal for her exceptional studies. Dr. Sandygulova's research lies at the intersection of human–robot interaction, machine learning, and social robotics to make robots more socially appropriate and acceptable by the end-users.

Gregory M. P. O'Hare (M'08) received the B.Sc., M.Sc., and Ph.D. degrees from the University of Ulster, Coleraine, Northern Ireland. He was the Head of the Department of Computer Science, University College Dublin (UCD), from 2001 to 2004. Prior to joining UCD, he was on the faculty of the University of Central Lancashire (1984–1986) and the University of Manchester (1986–1996). He is currently a Professor with the School of Computer Science and Informatics, UCD. His research interests are distributed artificial intelligence and multi-agent systems (MAS), and mobile and ubiquitous computing, autonomic systems, and wireless sensor networks. Dr. O'Hare is a Fellow of the British Computer Society. He is a member of the ACM and AAAI. He is a Chartered Engineer.