Paladyn, J. Behav. Robot. 2019; 10:140–159

DE GRUYTER

**Research Article**                                                              **Open Access**

Zachary Henkel*, Kenna Baugus, Cindy L. Bethel, and David C. May

# User expectations of privacy in robot assisted therapy

**Abstract:** This article describes ethical issues related to the design and use of social robots in sensitive contexts like psychological interventions and provides insights from one user design study and two controlled experiments with adults and children. User expectations regarding privacy with a therapeutic robotic dog, Therabot, gathered from a 16 participant design study are presented. Furthermore, results from 142 forensic interviews about bullying experiences conducted with children (ages 8 to 17) using three different social robots (Nao, Female RoboKind, Male RoboKind) and humans (female and male) as forensic interviewers are examined to provide insights into child beliefs about privacy and social judgment in sensitive interactions with social robots. The data collected indicates that adult participants felt a therapeutic robotic dog would be most useful for children in comparison to other age groups, and should include privacy safeguards. Data obtained from children after a forensic interview about their bullying experiences shows that they perceive social robots as providing significantly more socially protective factors than adult humans. These findings provide insight into how children perceive social robots and illustrate the need for careful consideration when designing social robots that will be used in sensitive contexts with vulnerable users like children.

**Keywords:** child-robot interaction; therapeutic robots; privacy

## 1 Introduction

As robotic systems become more reliable and autonomous, their use in therapeutic applications presents social and ethical design challenges. Our ongoing research includes the design of a therapeutic robotic dog, Therabot[TM], and the development of social robots for conducting forensic interviews with children. The development and use of these systems presents ethical questions, which must be incorporated into the research process [1]. In this article we examine results from two different projects that provide data towards answering the question: *How do user expectations and beliefs about autonomous social robots for psychological interventions affect ethical usage?*

## 2 Background

As the research community has grappled with ethical questions relating to robot-assisted therapy, many have arrived at mechanisms for limiting or augmenting full autonomy [2, 3]. In addition to developing standards and recommendations, which advise clinicians and patients on the usage of social robots, it is also helpful to consider the beliefs and expectations that users possess about the robots that will assist them. Social robots have been deployed across many domains to serve as tools for assisting humans in being entertained, coping with stress, and developing specific skills [4–6]. Robots which provide a therapeutic benefit [7–9] to their users typically engage with the user at regular intervals and obtain, process, and store information about their users. For example, social robots used in educational tutoring applications must track student progress in order to effectively assist its users [10–12].

The information collected by robots deployed for therapeutic uses, like Therabot[TM], or by robots for use in sensitive contexts, like forensic interviews or psychotherapy sessions, is likely to be considered by users of these systems to be private and deserving of protection. Researchers in the field of human-robot interaction working with children have raised concerns about how well users under-

---
**\*Corresponding Author: Zachary Henkel:** Mississippi State University, USA; E-mail: zmh68@msstate.edu
**Kenna Baugus:** Mississippi State University, USA;
E-mail: kbb269@msstate.edu
**Cindy L. Bethel:** Mississippi State University, USA;
E-mail: cbethel@cse.msstate.edu
**David C. May:** Mississippi State University, USA;
E-mail: dmay@soc.msstate.edu

stand a robot's collection and potential sharing of information in its environment [13, 14]. Similarly, our prior work has also outlined potential ethical concerns related to using social robots as forensic interviewers [1].

# 3 Approach

While pursuing a variety of independent projects within the domain of social robotics, our research group often encounters questions related to ethical design and usage of social robots in sensitive domains. In developing a therapeutic robotic dog for supplemental use during the course mental health interventions, concerns related to patient/user privacy, autonomy, and ethical usage are fundamental to understanding interactions between human and robot. Concurrently, our research inquiries in the area of using social robots as forensic interviewers for children in order to increase the amount of factual information obtained have led us to further investigate the beliefs children hold about the social robots they interact with. In both of these areas the expectations and beliefs users maintain about the systems they interact with prove to be rich sources of information for understanding the dynamics between the human and robot. Furthermore, when social robots are incorporated into domains of a sensitive social nature (e.g., mental health, police investigations, etc.) user expectations may also contribute significantly to ethical and legal considerations.

This article details results from three studies across two distinct areas of social robotics as they relate to user expectations of privacy when using robots in sensitive contexts. Though the two areas are independent research efforts each set of findings indicates that participants are not averse to the usage of social robots in sensitive domains but do have preconceived understandings of the types of privacy and social protections that will be provided by each system. The three studies presented investigate these topics through the use of semi-structured verbal interviews between a researcher and participant. The resulting qualitative data has been coded into discrete categories and distilled into higher-level insights that pave the way for further investigation in each area.

# 4 Therapeutic robot design

As part of the development of the therapeutic robotic dog, Therabot™ (see Figure 1), participants took part in a collaborative design process consisting of surveys and a semi-



**Figure 1:** The therapeutic robotic dog: Therabot™.

structured interview with a researcher concerning the potential uses of Therabot™. During the session participants were free to demonstrate their ideas using words, drawings, and by manipulating physical prototypes. The study included several questions to request feedback related to privacy while using a therapeutic robot.

## 4.1 Study protocol

Each session was conducted as a collaborative design process between a researcher and participant. Three short paper surveys were administered, but the primary focus of the session was a semi-structured interview that discussed the robot and its current and potential functions. Each session lasted no longer than 60 minutes and was organized into the following steps:

– **Explanation & consent** *(10 minutes)* - the purpose of the study was explained by a researcher and participants were given the opportunity to ask questions before they signed consent forms to participate.

– **Past experiences & attitudes towards comfort objects survey** *(10 minutes)* - a paper survey was administered that requested information about participants' past experiences with trauma and/or the use of comfort objects.

– **Guided interview** *(20 minutes)* - participants were provided with two different versions of the robot (a stuffed version without robotic hardware and a nonpowered robotic version) and the researcher conducted a semi-structured interview concerning the participants' thoughts about the robot and possible future enhancements that they felt may be beneficial.

- **Experience evaluation survey** *(5 minutes)* - a short paper survey that measured participants' perceptions of the robot, general beliefs about robots and technology, and overall enjoyment of the study was administered.

- **Demographics survey** *(5 minutes)* - a paper survey that inquired about participants' gender, age, formal education level, ethnic group(s), experience with technology, and experience with pets was administered.

All surveys were conducted using paper and pencil with open-ended response areas or Likert scales (anchored on each end, ranging 1 to 7 unless otherwise noted). After consent documents were signed, the *Past Experiences & Attitudes Towards Comfort Objects Survey* was conducted and included the following items:

- Have you ever received care for any type of trauma or traumatic event?

- Were you provided any type of comfort object (e.g. stuffed animal, pillow, etc.) as part of your care? If so, please describe.

- How familiar are you with the idea of using a comfort object, like a stuffed animal, as part of therapeutic care after trauma? *(Likert scale)*

- How harmful or beneficial would an interactive (robotic) comfort object be for the following age groups: Children (0 to 11), Adolescents (12 - 17), Adults (18 - 59), Older Adults (60+) *(Likert scale)*

- Describe any specific benefits you think an interactive comfort object, like a robotic dog, might have in comparison to a non-interactive comfort object.

- Describe any specific drawbacks an interactive comfort object, like a robotic dog, might have in comparison to a non-interactive comfort object.

The researcher then conducted the semi-structured interview with each participant, and presented the "stuffed animal" and robotic version of Therabot either simultaneously or sequentially depending on the study condition. While the interviewer followed the lead of the participant, the general structure the interviewer followed was:

- What do you feel this support companion would be most useful for and how would you use it if you were involved in a traumatic and/or stressful situation?

- How could this support companion work together with a therapist to support someone?

- How could this support companion work together with friends and family to support someone?

- Movements: How should each part of the robot move?

- Sounds: What types of sounds should the robot make? Should it use synthetic speech? Should it allow humans to speak through the robot?

- Touch: How would the participant touch the robot and what would be the expected responses?

- Should the robot have a companion app for users to communicate with it? If so what features should this app have?

- Would you be willing to let the robotic dog record your conversations with the therapist during therapy sessions?

- What should be taken into consideration to protect a user's privacy when using this companion robot?

Participants were specifically asked *How willing they would be to let the robot record their conversations with their therapist during therapy sessions?*, as well as *What should be taken into consideration with regards to protecting a user's privacy?*

Following the interview portion of the session, participants completed the *Experience Evaluation* survey, which included the following items:

- How large of a role do you think robots will play in the future? *(Likert scale)*

- How would you rate your enthusiasm for robots? *(Likert scale)*

- How interesting were the concepts presented in the study? *(Likert scale)*

- How understandable were the concepts presented in the study? *(Likert scale)*

- Was this study was worth the time it took to complete? *(Likert scale)*

Finally, participants completed the *Demographics* survey, which included the following items:

- What is your gender?

- How old are you?

- What is your highest level of education?

- Which ethnicities apply to you?

- How would you rate your level of experience with computers? *(5pt Likert scale)*

- How would you rate your level of experience with video games? *(5pt Likert scale)*

- How would you rate your level of experience with robots? *(5pt Likert scale)*

- In your life, how much time have you spent interacting with pets? *(5pt Likert scale)*

- Do you have any comments? If so, we would like to hear them.

Participants were then thanked for their contributions and given course-related credit if eligible. All sessions were audio- and video-recorded for later analysis.

## 4.2 Participants

Sixteen participants (9 female, 7 male) with a median age of 20 were recruited from the university's psychology program and completed the study. All participants were undergraduates at Mississippi State University with 5/16 pursuing science or engineering degrees, 8/16 pursuing psychology degrees, and 3/16 not reporting their area of study.

On average participants indicated they had high levels of experience with pets (M=4.63 of 5, SD=0.89) and low experience with robots (M=2.25 of 5, SD=0.78). Two participants reported receiving care in the past for a traumatic event, with one indicating that a cat had been used as part of their therapy.

## 4.3 Results

The interview portion of the sessions had a mean duration of 13 minutes, 2.25 seconds and were conducted by three different interviewers, each familiar with the robot.

Overall participants were not familiar with the use of a comfort object as part of therapeutic care ( M=3.44, SD=1.9). Reported potential benefits of a robotic dog included encouraging play, providing comfort, and increasing coping abilities. Participant's speculated that drawbacks might include the robot breaking, the user becoming overly attached, and not being as comforting as a real animal.

A majority of participants (14/16, 88%) reported that the audio-recording of therapy sessions or collecting user activity data would be acceptable under certain conditions. Participants provided the following restrictions: obtaining consent from users (6/16, 38%), recording only data useful for therapy (3/16, 19%), and limiting recording to therapy sessions (3/16, 19%). Some discussed features such as indicator lights informing users of data recording, while others focused on limiting who would have access to the recorded data.

The *Past Experiences & Attitudes Towards Comfort Objects Survey* included a series of questions that asked participants to rate *How harmful or beneficial would an interactive (robotic) comfort object be for each age group?* for the age groups: *Children (0-11), Adolescents (12-17), Adults (18-59), and Older Adults (60+).*

Responses to paper survey questions concerning the harm or benefit an interactive (robotic) comfort object would have on each age group indicated a significant difference between the ratings for each age group.

**Table 1:** Harmful/Helpful ratings for each age group, on a 1 (very harmful) to 7 (very beneficial) scale.

| Age group | Mean | SD |
|---|---|---|
| Children (0-11) | 6.27 | 1.03 |
| Adolescents (12-17) | 5.40 | 0.63 |
| Adults (18-59) | 4.73 | 0.88 |
| Older adults (60+) | 5.47 | 1.25 |

Table 1 shows the mean and standard deviations for each age group. A Maulchy's Sphericity Test indicated data did not violate assumptions of sphericity ($\chi^2(5) = 8.31, p = 0.141$). A significant difference was found between the helpfulness rating for each age group (one-way repeated measures ANOVA, $F(3, 42) = 6.17, p = .001, \omega^2 = 0.19$, *small effect*). Post-hoc pairwise comparisons indicated significant differences between Children and Adults ($p = .009$). Contrasts revealed that in comparison to adults, children were perceived benefit significantly more from social robots $F(1, 14) = 15.559, p = .001, r = 0.73$ with a large effect size.

Though from a small sample of 16 participants, these responses indicate that the therapeutic platform was thought to have differing levels of usefulness across age groups. This finding coupled with the general privacy concerns highlighted by participants indicates that further research into how different user groups (e.g., age groups) understand the privacy afforded by the system may provide important insights for ethical usage.

# 5 Interviewing children about sensitive information

As part of an effort that explores the use of social robots in forensic interview scenarios with children, we conducted two studies focused on bullying experiences at school with children (8-17 years old). In the first study (Study A), participants were ages 8 to 12 and engaged in a thirty minute forensic interview with either a human male, human female, male humanoid robot, or female humanoid robot acting as the interviewer. Humanoid RoboKind R25 robots (see Figure 4) were used for the robot conditions. In the second study (Study B), participants were ages 12 - 17 and they participated in a similar study with either a human male, male humanoid RoboKind robot, or a male humanoid Nao robot. Following the interview, a researcher verbally administered an open-ended survey to better understand the participants' perception of the interviewer with whom they interacted.

Though the studies were designed primarily to investigate differences in the likelihood of children to disclose sensitive information to robotic interviewers, participant responses to the follow-up survey about the forensic interviewer provided insight into expectations of privacy associated with social robots in this type of role. This section describes the study design and the findings associated with participants' social expectations of human and robot interviewers.

## 5.1 Study design

The primary focus for both studies was conducting an interview with participants concerning their personal experiences with bullying at school. The interview was designed to use the structure of a forensic interview (as described by Lamb et al.[15]) as a mechanism for obtaining data that is typically investigated in current research concerning bullying among children and adolescents. A dynamic script was developed for the interview, with the primary difference between the two study scripts being the inclusion of questions about cyberbullying in the second study with adolescents. Though the majority of the content was pre-scripted, participant responses determined the questions asked and several prompts incorporated information provided by the participant. Additionally, interviewers (human and robot) were able to clarify questions and respond to spontaneous requests if needed.

Both studies used a between-participants design, assigning participants to a specific type of interviewer using random assignment. In *Study A* children were assigned to one of the following interviewers: female human, male human, female humanoid RoboKind robot, or male humanoid RoboKind robot. In *Study B* participants were assigned to an interviewer from the following: male human, male humanoid RoboKind robot, or male humanoid Nao robot. This research was approved by and conducted with guidance from the Mississippi State University Institutional Review Board.

### 5.1.1 Study space, control methods, and robots

Study sessions took place in a dedicated lab space located in the Social Science Research Center on the campus of Mississippi State University. The study space consisted of a large segmented waiting area and three smaller rooms used for the study. Figure 2 illustrates the layout of the study space. One room was used for interactions between participants and a researcher, another for the actual forensic interview, and the third served as a control center for the study. Each portion of the study was conducted in a closed room and white noise was used between the study rooms and the waiting area to ensure privacy.

All sessions (including those with human interviewers) were conducted using a Wizard-of-Oz approach [16] in which two researchers collaboratively directed the flow of the interaction and entered participant provided information into a custom software system designed as part of this research effort for conducting interviews and storing that information into a database. Researchers responsible for production of the interview session monitored the interview room through three high-definition cameras, three microphones, and two Microsoft Kinect V2 sensors. Figure 3 shows the robot control room and the interview room.

One remote researcher served as the "director" of the interaction and was responsible for selecting the next statement the interviewer would deliver to the participant. The "director" would consider the participant's response and select the appropriate follow-up prompt. For example, if a participant reported that they had never experienced or witnessed verbal aggression at school, the "director" would bypass follow-up questions related to verbal aggression and move to the next area of the script. The second remote researcher served as the "producer" of the interaction and was responsible for inputting information into the system, which would be used in later portions of the script. For example, as a participant described the sequence of their activities during the previous day the "producer" added this information to the system so that the "director" could later choose prompts that inquired further
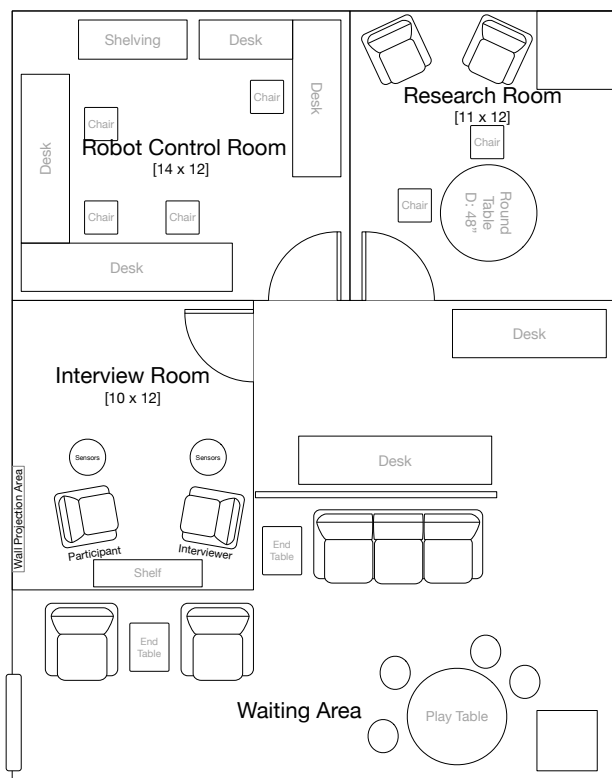
**Figure 2:** Physical layout of the study space used for conducting forensic interviews. (Room dimensions in feet.)

about a specific activity. The "producer" was also responsible for enabling back-channeling (i.e., head-nodding) and participant tracking during robot sessions.

When a robot interviewer was conducting the interview, the director's selection of a specific script statement would result in the robot's text-to-speech system delivering the statement to the participant. When a human interviewer was conducting the interview, the director's selection was displayed on a projected teleprompter concealed directly behind the participant. Human interviewers were also equipped with an electronic tablet that displayed the director's view and provided basic controls to fine tune the projected teleprompter (e.g., height adjustments allowing the interviewer maintain eye contact with the participant throughout the interview). Robot interviewers automatically maintained eye contact using an external Microsoft Kinect V2 sensor for tracking the participant's face.

Three robotic platforms were used between the two studies. In *Study A* male and female versions of the RoboKind R25 robots were used. In *Study A* the male R25 robot used the default Acapela Josh voice, the female R25 robot used the default Acapela Ella voice, and both used the automatic lip sync provided by the system's text-to-

speech system. In *Study B* the male R25 robot and a V5 humanoid Nao robot were used. During *Study B* both robots used the Cereproc Andy voice and custom lip-sync software was developed for the R25. The synthetic voice selection was modified to allow consistency between Nao and the R25, as well as to improve upon the performance of the default R25 voice.

As prescribed by forensic interviewing protocols, human interviewers worked to avoid providing excessive positive or negative feedback to participant responses through facial expressions, vocal intonation, or other mannerisms. In general the disposition of the interviewer was attentive and welcoming, listening to the participant's statements without providing judgment (similar to the sentiment of unconditional positive regard [17] in client-centered therapies). The RoboKind R25 robot faces were configured to a neutral position with the only movement being the robot's head, eyes, and jaw. The Nao robot moved its head and simulated eye blinking with LEDs surrounding its eyes. The R25 robot's eyes did not blink, as the actuation mechanism was found to be loud, unreliable, and ultimately distracting when prototyping the study.

### 5.1.2 Forensic interview

Both the interviewer and the script adhered to the forensic interview protocol to ensure a neutral interaction was conducted. After the introduction was performed by the researcher, the interviewer articulated guidelines and presented an opportunity for the participant to rehearse an example with each of the guidelines. The interviewer conducted the introduction and guidelines portions as follows:

INTERVIEWER: Hi <u>&lt;participant name&gt;</u>, it's nice to meet you. I'm the interviewer for the study today and would like to talk to you for a little while about your experiences at home and school. Does that sound okay to you?

PARTICIPANT: Yes [*nods*]

INTERVIEWER: Okay, great. If you can take a seat over there, we can get started. Once we are finished <u>&lt;researcher name&gt;</u> will come back and take you to the next part of the study.

INTERVIEWER: Like <u>&lt;researcher name&gt;</u> said before, my name is <u>&lt;interviewer name&gt;</u> and I spend a lot of time talking with kids about the different things they see and do at home and at school. Before we get started, do you have any questions for me?

PARTICIPANT: No.

**Figure 3:** Left: The interview room with a Nao robot sitting in the interviewer chair and sample teleprompter projection visible on the wall behind the participant's chair. Right: The study control room where two researchers collaboratively guide study sessions.
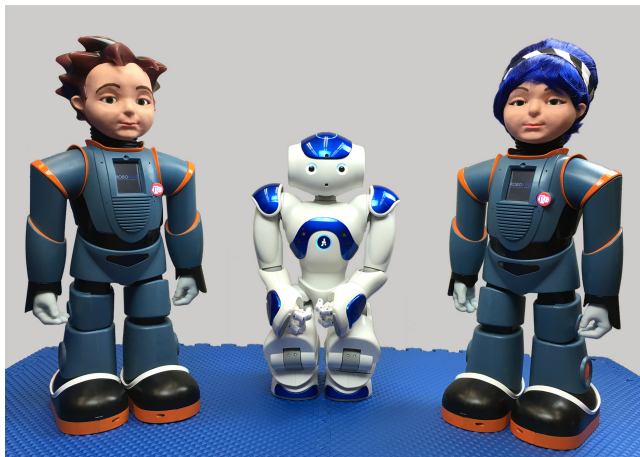


**Figure 4:** The three robotic platforms used in forensic interviews (RoboKind R25 robots left and right, Nao robot center).

INTERVIEWER:  Before we begin, there are a few important things you should know. The first is if at any-time during the interview you are not sure about the answer to a question, just tell me that you don't know. Let's practice this. So if I asked you, what's my dog's name, what would you say?

PARTICIPANT:  I don't know.

INTERVIEWER:  If I ask a question during the interview that is confusing or that you don't understand, just tell me and I will explain what I mean. Let's practice this. So if I asked you, what is your ocular color, what would you say?

PARTICIPANT:  I don't know what that means.

INTERVIEWER:  It's okay to tell me you don't know what that means. So I'll ask it in another way. What color are your eyes?

PARTICIPANT:  Blue.

INTERVIEWER:  It's also important that if I say something wrong during the interview, you should tell me. For example, if I said you were a 4 year old boy, what would you say?

PARTICIPANT:  I am not a 4 year old boy.

INTERVIEWER:  And the last important thing is that you tell the truth during the interview. Do you promise you will tell the truth today?

PARTICIPANT:  Yes, I promise.

Next, the interviewer began building rapport with the participant and asked open-ended questions about the participant's hobbies, family, and events that occurred the previous day. Sharing the sequence of events in his or her day or previous day allowed the participant to practice giving a narrative description and prepared the participant for the next section of the interview. The interviewer also asked three or four follow-up questions, using detailed information the participant provided in his or her own words. These questions were customized to best fit the participant's statement and assist in rapport building. An example exchange is as follows:

INTERVIEWER:  So first, I'd like to get to know you better. What's something you like to do for fun?

PARTICIPANT:  I like speech and debate. I'm in the speech and debate team at school.

INTERVIEWER:  How did you learn to do that?

PARTICIPANT: Well this is my first year doing it. My best friend is on the team and she recruited me to do it because she thought I'd be good at it because I'm into politics. She asked me if I would join the team and then I did. My coaches have taught me a lot.

INTERVIEWER: How often do you get to do that?

PARTICIPANT: We have tournaments almost every weekend, and we have practice almost every day after school.

INTERVIEWER: Are there other people that you do that with?

PARTICIPANT: Yeah, we have a pretty big team.

INTERVIEWER: Thank you for telling me about that. I'd like to know about things that happen to you. Can you tell me everything that happened yesterday in as much detail as possible, from the time you woke up until you went to bed?

PARTICIPANT: I went to school, ate breakfast, went through my classes, and went home. I took a nap, then played a video game for a little while. I ate supper with my mom then went back to bed.

INTERVIEWER: Tell me more about <u>playing a video game</u>.

PARTICIPANT: I wasn't really doing much.

INTERVIEWER: Earlier you mentioned <u>eating</u> <u>supper</u> <u>with</u> <u>your</u> <u>mom</u>, tell me more about that.

PARTICIPANT: She had bought something for us to eat.

INTERVIEWER: Is there anything else you remember about <u>yesterday</u>?

PARTICIPANT: No.

After building rapport, the interviewer began asking the participant questions about substantive issues. First, the interviewer asked each participant a number of questions about his or her friends and family. The interviewer then pivoted to discussing how kids get along with each other at school and progressed through prompts inquiring about verbal aggression, relational aggression, cyber aggression (*Study B* only), and physical aggression. If the participant reported aggression at any stage, a series of follow-up questions were asked. A sample interaction is as follows:

INTERVIEWER: Now I'd like to learn a little bit more about you and your family. [*pause*] There are a lot of different types of families today. Tell me about your family and who lives with you.

PARTICIPANT: My parents are still married. They have been married for a really long time, and I have three other siblings. I have an older brother, older sister, and younger brother.

INTERVIEWER: Who is the person that spends the most time with you when you are at home?

PARTICIPANT: Probably my sister. Me and sister are best friends.

INTERVIEWER: How do you feel about that person?

PARTICIPANT: I love her. We act the exact same. I got really lucky that she was my sister.

INTERVIEWER: Now let's talk about your friends. Can you tell me about your closest friends at school?

PARTICIPANT: One of my closest friends, the one who recruited me to the debate team, we spend a lot of time together. Especially now that I got my license. We hang out all the time.

INTERVIEWER: If you were going to tell someone a secret, who would it be and why?

PARTICIPANT: Probably her because I know I can trust her with all of it, and she won't tell anyone else.

INTERVIEWER: I'd like to talk some about how the kids at your school get along with each other. Let's start with the way kids talk to each other. Please tell me about the different ways kids talk to each other at school.

PARTICIPANT: Well, it's really really clique-y at my school so there is the cheerleading clique and then there is like the baseball clique and just like different things like that but most people, they're pretty nice to you. A lot of people are pretty nice to you, to your face, but a lot of people talk when they're not with you. There are people who are a little bit rude to people who are different than them, which isn't fun to see but for the most part people are pretty nice face-to-face. The real problem is whenever they're not together.

INTERVIEWER: Sometimes kids say mean or unkind things to other kids to make them feel bad. Can you tell me about this happening at your school?

PARTICIPANT: Yeah, I've witnessed that a few times, especially with people who are different, like people with disabilities and stuff. I've witnessed that a lot, which is terrible to see. I have a friend who has Asperger's Syndrome, so obviously he functions a little bit differently, but he's still really really nice, but people just treat him so differently and kinda talk down on him and make fun of him when he does stuff that they're just not really used to, which isn't fair. There are some just rude things that that–especially whenever people don't agree with each other about things, you shouldn't be mean about but a lot of people, if someone is different they automatically criticize.

INTERVIEWER: Do the kids who say mean things to other kids do this a lot?

PARTICIPANT: Yeah

INTERVIEWER: How often do they do this?

PARTICIPANT: I have witnessed it. Not so much about the disabilities thing but whenever people disagree about pretty much about anything; it happens almost everyday. Especially on social media, it happens so much, within the school. Then making fun of people that are different like people with disabilities and stuff, I witness that maybe like twice a week or three times a week.

INTERVIEWER: Which kids do they normally say mean things to?

PARTICIPANT: Recently, surprisingly, you wouldn't think this would be going on today in 2017, but recently it's been a lot of like political issues, which is crazy. Our volleyball team knelt during the national anthem so it made a big deal in the school, so there were people who agreed with it and people who disagreed with it, and all day people were just criticizing each other about it literally all day. High school students today have a really hard time accepting that someone else can believe something different than they do, and if someone thinks differently it is so quick and easy to criticize than it is to listen, so that happens a lot.

INTERVIEWER: How often do these kids say mean things to you?

PARTICIPANT: I am pretty confident in my opinions, if I think someone is doing something that I don't think is okay, I'll be like, "hey that's not okay," and I get criticized about that a good deal.

INTERVIEWER: What happens when they say mean things to other kids?

PARTICIPANT: A lot of times people don't stand up and say anything, and sometimes people will agree and say the same thing. But sometimes students will say, "hey that's not okay." But it's a lot easier to just go with the crowd and just laugh about it.

INTERVIEWER: How do you feel when they say mean things to others?

PARTICIPANT: It really sickens me to see it, because it's just so unfair I think.

INTERVIEWER: Sometimes people talk about other people and say mean or untrue things about them. For example, kids might spread rumors, gossip, or tell lies about someone to hurt them. Can you think of examples of this happening at your school?

PARTICIPANT: Oh yes. That happens all the time. I think that's the primary issue at my high school: gossiping and rumors being spread all the time.

INTERVIEWER: Let's think about the different groups of kids at your school. Sometimes kids will leave others out of the group or ignore them on purpose. Can you tell me about this happening at your school?

PARTICIPANT: Yeah, I think that does happen a lot.

INTERVIEWER: Are there kids who do this kind of thing a lot?

PARTICIPANT: Probably everyday. You see it everyday at lunch.

INTERVIEWER: Which kids do they normally leave out?

PARTICIPANT: The people who are not so popular, not dressed with the trends or whatever.

INTERVIEWER: Are there kids at your school who leave you out?

PARTICIPANT: Yeah, there has been friend groups that I was part of that I necessarily wouldn't always like get invited to things.

INTERVIEWER: What happens when they do this to you?

PARTICIPANT: It does kinda hurt cause it's like, "Why wasn't I good enough?" but most of the time I get over it pretty quickly and go talk to someone else.

INTERVIEWER: How do you feel when kids leave others out?

PARTICIPANT: It makes me feel bad, and sometimes I'll be like, "Do you want to come sit with us?" but sometimes I don't, because [*pause*] it's – it's a lot easier to not say anything. It is sad to admit, but sometimes it's just easier to say, "Well that really sucks," and kind of pity them but not do anything about it.

INTERVIEWER: Is there anything else you would like to share about that?

PARTICIPANT: I don't think so.

After examining each type of aggression, participants who reported aggressors were asked to characterize the aggressors using a series of questions designed to identify power imbalances between the aggressor and victim.

The final phase of the interview involved asking the participant to define bullying and discuss any additional experiences with bullying. The interviewer then entered the closure phase, thanking the participant for talking with them and asking them if they had any fun things they were planning to do soon.

### 5.1.3 Perceptions of the interviewer interview (PII)

Immediately following the forensic interview, a researcher verbally engaged the participant in a semi-structured interview about their experience with the interviewer. The *Perceptions of the Interviewer Interview (PII)* is a set of open-ended questions developed and refined over the course of six different studies involving robots and humans interviewing children about sensitive topics. These questions were initially designed as a way to assess general feelings towards to interviewer in a study. As a result of the responses provided by children in earlier studies (specifically responses after the introduction of a character guessing game [18]) questions were tailored to include items concerning the socially protective factors children attributed to the interviewer. The interview was structured as follows:

– **Q1:** What did you think about *interviewer name* during the study?
– **Q2 - Robot only:** Do you think *interviewer name* was aware of what was going on around *her/him*? Why or why not?
– **Q3:** How well do you think *interviewer name* understood what you said?
– **Q4:** How well do you think *interviewer name* understood how you felt?
– **Q5:** Do you think *interviewer name* could give you helpful advice if you had a problem? Why or why not?
– **Q6:** Sometimes people hide how they feel from others. Do you think you could hide how you feel from *interviewer name*? Why or why not?
– **Q7:** Are there things you could talk to *interviewer name* about that you could not talk to other people about? What kind of things could you talk about with *interviewer name*?
– **Q8 - Robot only:** How is *interviewer name* different from a human?
– **Q9:** Was *interviewer name* helpful to you?
– **Q10:** In what ways did you feel like *interviewer name* was helpful to you?
– **Q11:** How could *interviewer name* be more helpful to you?
– **Q12:** Did you like *interviewer name*? Why or why not?
– **Q13:** Do you think *interviewer name* liked you? Why or why not?
– **Q14:** What would you do if *interviewer name* did not listen to you while you were trying to talk to *her/him*?
– **Q15:** What would happen if you did not listen to *interviewer name* while he was trying to talk to you?
– **Q16:** Do you have any other thoughts about *interviewer name* from the study that you'd like to share with us?

### 5.1.4 Character guessing game

All participants (regardless of interviewer condition) were invited to play a popular character guessing game with a robot at the end of the study. The participant was asked to think of a character and the robot asked questions about the character, with the participant answering "Yes", "No", "Probably", "Probably not", or "I don't know.". After a number of questions the robot attempted to guess the character (see Henkel et al. [18] for comprehensive details).

## 5.2 Study protocol

The study was divided into four sections: 1) pre-interview tasks, 2) interview, 3) post-interview tasks, and 4) game. The study spanned approximately an hour, and the duration of the interview portion was approximately 30 minutes. During the pre-interview section, the researcher introduced herself to the participant and his or her parent in the lab waiting area. The researcher read a brief description of study, then asked the participant if he or she wished to participate. The researcher explicitly informed the participant and his or her parent that the participant's actions and speech would be recorded. Audio and visual consent forms were provided to and signed by the parent and child. The parent was then given a survey about his or her own experiences with bullying and remained in the waiting area, completing the survey throughout the duration of the study. The participant and researcher transitioned to the "Research Room," where the participant completed a demographics questionnaire.

Once the questionnaire was completed, the researcher led the participant to the "Interview Room." The researcher opened the door and introduced the participant and interviewer to one another. The interviewer was named either Hannah or Zach and was either human or robot. The researcher asked the participant to take a seat and informed the participant that she would return as soon as the interview was over. If the participant expressed concern, the researcher offered to leave the door ajar; if concern was still present, the researcher remained in the room with the interviewer and participant. Otherwise, the researcher closed the door, went to the "Robot Control Room," and observed the interview remotely.

Throughout the duration of the interview portion of the study, if the participant became visibly upset or said he or she was uncomfortable, the researcher interrupted the interview and offered the participant an opportunity to take a short break. The researcher reminded the participant that the study was voluntary. Regardless of whether or not the participant wished to continue, his or her parent was given a document summarizing psychological services provided by the university, and the researcher offered to assist in arranging an appointment for the participant if his or her parent believed it may be beneficial.

As the interview ended, the researcher returned to the 'Interview Room," and escorted the participant to the "Research Room." Then, in the post-interview portion of the study, the researcher discussed with the participant his or her perceptions of the interviewer by conducting the Perceptions of the Interviewer Interview.

After the Perceptions of the Interviewer Interview, the researcher and participant returned to the "Interview Room", and the participant played a character guessing game with a robot. Once the participant finished the game, the researcher led the participant to the waiting area and awarded them a small prize and $10. The researcher also gave the participant a debriefing form containing additional resources if he or she wished to discuss bullying experiences further.

## 5.3 Participants

Participants were recruited from a database of local children and parents that had expressed an interest in participating in research studies. The database is maintained by university researchers and is advertised through fliers, newspaper advertisements, and targeted advertisements on popular social media networks. Researchers used the database to contact parents with children that were eligible for each study. Participants who took part in the first study were ineligible for participation in the second study. Table 2 fully describes the number of participants in each study and condition. Between the two studies, 70 female participants and 71 male participants were interviewed; 75 interviews were conducted by robots, and 67 interviews were conducted by humans.

### 5.3.1 Study A - younger children

Between June and August of 2017, a total of 91 participants, ages 8 through 12, took part in the first study. Of the 91 participants, five cases were excluded due to a variety of causes: one participant did not wish to proceed with the interview, one participant took part in the study twice due to a clerical error; only his or her initial interaction was considered, during one session a significant robot malfunction occurred, and two participants declined to answer the Perceptions of the Interviewer survey. The conditions for the study were randomly assigned: 21 participants were interviewed by a female human, 26 by a male human, 20 by a female R25 robot, and 19 by a male R25 robot. There were a total of 47 interviews given by human interviewers and 39 given by robotic interviewers. Of the participants, 42 were female, and 43 were male.

### 5.3.2 Study B - older children

From October through December of 2017, a total of 56 participants, ages 12 through 17, took part in the second study. No cases were excluded. As in the prior study, the conditions were assigned randomly; however, in this study, the conditions were male human, male R25 robot, and Nao robot interviewer. The human interviewer completed 20 interviews, the R25 robot completed 17, and the Nao robot completed 19. The robot interviewers accounted for a total of 36 interviews; while the human interviewer accounted for 20. There were 28 female participants and 28 male participants.

## 5.4 Results

A total of 147 one-hour sessions were conducted during the summer and fall of 2017, yielding 142 usable cases. Participants in *Study A* (younger children) were distributed between conditions to balance the interviewer-participant gender pairings (see Table 2). In *Study B* (older children), participants were randomly assigned to one of three interviewers.

The same male and female human interviewers were used across all human condition sessions in *Study A*. Both the male and female interviewers in *Study A* had recently completed their undergraduate educations in social science fields and were in their early twenties. In *Study B* human interviews were conducted by a male undergraduate in the last year of his social science program and in his early twenties. In *Study A* one of three female research assistants guided participants through the study and conducted verbal surveys about the interviewer. In *Study B* one of three research assistants (one male, two female) guided participants through the session. Across both studies each

**Table 2:** Number of participants assigned to each condition between Study A and Study B. Sorted by Study, Interviewer Type, Interviewer Gender, and Participant Gender.

| Combined Interviews By Participant Age | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 142 | | | | | | | | | | | | | |
| Study A (Ages 8 - 12) | | | | | | | | Study B (Ages 12 - 17) | | | | | |
| 86 | | | | | | | | 56 | | | | | |
| Robot | | | | Human | | | | Robot | | | | Human | |
| 39 | | | | 47 | | | | 36 | | | | 20 | |
| Female R25 | | Male R25 | | Female Human | | Male Human | | Male Nao | | Male R25 | | Male Human | |
| 20 | | 19 | | 21 | | 26 | | 19 | | 17 | | 20 | |
| Female P. | Male P. | Female P. | Male P. | Female P. | Male P. | Female P. | Male P. | Female P. | Male P. | Female P. | Male P. | Female P. | Male P. |
| 10 | 10 | 12 | 14 | 11 | 10 | 12 | 14 | 6 | 13 | 8 | 9 | 14 | 6 |

session utilized two members of a four member interview production team to support the interaction.

This article examines results from the *Perceptions of the Interviewer Interview* conducted by the researcher after the participant completed their interaction concerning their experiences with bullying. Verbal and behavioral data captured during the main interview is currently undergoing analysis and will be reported in a future publication.

*Study A* was designed to explore the effects of interviewer type (human or robot) as well as any differences that might occur as a result of the genders of the interviewer and participant. *Study B* consistently used male human and robot interviewers and was designed to examine differences that might occur between the Nao humanoid robot, the male R25 robot, and a male human interviewer. Both studies used the same *Perceptions of the Interviewer Interview* and *Study B's* main interaction concerning bullying experiences expanded upon *Study A* by incorporating additional questions about cyberbullying.

The analysis in this section examines effects present when responses from *Study A* and *Study B* are pooled as well as when they are analyzed independently.

### 5.4.1 Data coding approach

Verbal answers provided by participants during the *Perceptions of the Interviewer Interview* were converted to text by two independent research assistants using audio recordings of each session and the ELAN software package [19]. A third research assistant examined and resolved any discrepancies between the transcriptions, yielding a final text transcript for each participant. If responses were unable to be determined from audio recordings alone, video of the session was consulted to clarify verbal responses and capture non-verbal responses.

Two researchers (the first and second authors) independently evaluated text transcriptions of participant responses to items Q2-Q13, coding responses first for agreement or appraisal (depending upon the question) and then for any social factors present in the response.

A five-point coding scheme for indicating agreement or disagreement was developed for items Q2, Q5, Q6, Q7, Q9, Q12, Q13 using the following coding guidelines:

– **No** - A verbal or non-verbal response indicating complete disagreement.

– **Indecisive negative** - A verbal response that primarily indicated disagreement but also included reservations, conditions, minor uncertainty, or hypothetical alternatives.

– **Indecisive** - A non-verbal or verbal response that ultimately indicated uncertainty.

– **Indecisive positive** - A verbal response that primarily indicated agreement but also included reservations, conditions, minor uncertainty, or hypothetical alternatives.

– **Yes** - A verbal or non-verbal response indicating complete agreement.

Similarly, a five point coding scheme for appraising performance was developed for items Q3 and Q4 with the following coding guidelines:

– **Very poor** - A verbal response that indicated exceptionally poor performance.

– **Poor** - A verbal response that indicated performance that was slightly problematic or did not fully meet expectations.

– **Indecisive** - A verbal or non-verbal response that ultimately indicated uncertainty.

- **Well** - A verbal response that primarily indicated performance was acceptable or met expectations.

- **Very well** - A verbal response that indicated superb performance or exceeding expectations.

In addition to the five established agreement and appraisal codes, a not applicable (NA) category was created for cases when a participant was not asked or did not provide an answer to a specific item. For the analysis presented in this article the agreement and appraisal scales were collapsed from five points to three points by combining the first two and last two categories on each scale.

Each response was also examined for supporting social factors that participants used to justify their answers. Items Q2, Q5-Q7 and Q9-Q13 included explicit follow-up prompts, which often elicited social factors, while responses to Q8 were primarily composed of social factors. Two researchers (the first and second authors) collaboratively generated a list of six base social factors from their observations of study sessions and by examining a small sample of transcribed responses to each item. During the coding process researchers discussed and created additional sub-categories within these six main factors when doing so assisted in more precisely characterized responses. Responses were coded for the following main social factors:

- **Appearance:** Responses that referenced the interviewer's physical appearance but did not incorporate elements of the interviewer's behavior. (Positive Example: *"yeah because I liked her blue hair"*, Negative Example: *"He doesn't have life in his, in his eyes so you don't feel like you're talking to somebody alive".*)

- **Demeanor:** Responses that referenced the interviewer's behavior or personality, often as it related to being understanding, interested, concerned, or helpful. (Positive Example: *"she was really nice and, uh she would listen a lot more than some other people would to me"*, Negative Example: *"I don't think she could see me uh really feel".*)

- **Interviewer behavior:** Responses that highlighted a specific action the interviewer took during the interaction with the participant. (Example: *"I could tell he was listening at least most of the time because he like when I would say something he would ask like questions like specifically about that and stuff like that".*)

- **Knowledge:** Responses that referenced the interviewer's knowledge. (Positive Example: *"he seemed pretty smart"*, Negative Example: *"I don t think she un-*

*derstood it at all because like I said before she was most likely programmed to do that".*)

- **Social confidence:** Responses that discussed social judgment, privacy, or trust as it related to the interviewer.(Positive Example: *"I feel probably better talking to him than a person because I mean like when you re saying it to the person it's basically just like a guilt trip right there"*, Negative Example: *"it felt good to talk about it and at the same time I had to stop myself because it's a robot and I don't know where it was going to, so i just kinda like stop myself".*)

- **Non-specific:** Responses that expressed only uncertainty, were not specific, referenced intuition, or emphasized the participant's own traits rather than evaluating the interviewer's traits or behaviors.

Responses to each item were coded for all social factors that applied to the response. A single response could be tagged with multiple social factors, though the most specific social factor available was always used. For example, if a participant indicated they trusted the interviewer because they felt the interviewer would maintain their privacy, the sub-category of privacy within the social confidence category was selected rather than simple social confidence. Responses to Q7 were coded into the main social factors when possible, but an additional set of codes was developed for responses to the follow-up question *What kind of things could you talk about with the interviewer?*. This additional list included the categories: *Anything, Bullying, Personal, School,* and *Superficial*.

The first and second authors coded each response for social factors independently and resolved any conflicts through discussion. A majority of conflicts in coding involved one coder selecting a less specific area of the same main category or one coder applying a single code when multiple were merited. After all coding conflicts were resolved, a final review of the responses associated with each social factor was conducted to ensure consistency.

### 5.4.2 Agreement and appraisal

After the initial data coding process, responses to items, which included an agreement or appraisal prompt were sorted into three categories (positive, negative, or indecisive) for analysis. For agreement prompts *Yes* and *Indecisive Positive* were grouped into the *Positive* category and *No* and *Indecisive Negative* were placed in the *Negative*. For appraisal prompts *Very Poor* and *Poor* were grouped as *Negative* while *Very Well* and *Well* formed the *Positive*

group. In a small number of cases a researcher inadvertently skipped an item, a participant offered no decipherable response, or a technical error prevented capturing the participant's response and the participant was excluded from the analysis for the affected items.

The responses from both studies were merged and the frequencies of coded responses for each item were compared when participants were grouped by study, interviewer type (human or robot), and participant gender. Further analyses were conducted within the context of each study to better understand the source of any significant differences.

No statistically significant differences on any agreement or appraisal items were identified when splitting responses into two groups based on the study they participated in (younger children versus older children).

### Robot vs. human interviewers

When responses were split into those with a human interviewer (male or female) and those with a robot interviewer (female R25, male R25, or male Nao), statistically significant differences were observed for items related to how well the interviewer understood the participant and the perceived ability of the interviewer to give helpful advice (Q3-Q5). Table 3 shows the response frequencies for each question when divided into responses for human and robot interviewers.

Participants with a human interviewer were more likely to respond that they were uncertain how well the interviewer understood what they said (Q3) (14.93%) in comparison to those with a robot interviewer (4.23%, Fisher's Exact Test $p = 0.01$). Furthermore, 7.04% participants with robot interviewers felt the interviewer did not understand what they said, while none of the participants in human interviewer conditions reported a lack of understanding. When asked how well the interviewer understood the way they felt (Q4), participants in the robot conditions were more likely to endorse *Negative* (14.29%) or *Indecisive* (22.86%) in comparison to the human interviewer conditions (1.54% and 13.85% respectively, Fisher's Exact Test $p < 0.01$). Participants in robot interviewer conditions were more likely (5.33%) than those with a human interviewer (1.54%) to indicate that they were uncertain if the interviewer could provide them helpful advice if they had a problem (Fisher's Exact Test $p = 0.01$). Furthermore, those in the robot conditions were more likely report that the interviewer would be unable to provide helpful advice (9.33%) in comparison to those in the human condition (0%).

An examination of *Study A* independently shows the only statistically significant difference (Fisher's Exact Test $p < 0.01$) that occurs between human and robot interviewers is in responses to Q3 (see Figure 5). In the human conditions (male and female interviewers) 78.72% participants appraised the interviewer's ability to understand what they said as *Positive*, while 21.28% were *Indecisive*. In comparison, 88.58% of participants in robot interviewer conditions (male and female) appraised the interviewer's ability to understand what they said as *Positive*, 2.86% were *Indecisive* and 8.57% reported the interviewer did not understand well.

As shown in Figure 5 when the results for *Study A* are sorted into categories for each interviewer (i.e., Human Female, Human Male, R25 Male, R25 Female) and compared, statistically significant differences are found for Q3 and Q4 (Fisher's Exact Test, $p = 0.04$ and $p = 0.02$ respectively). These results indicate that the female human interviewer was rated as understanding what participants said in 71.43% of the cases, while all other types of interviewer were rated as understanding in at least 84.62% of cases. Furthermore, this breakdown shows that 44.44% of participants interacting with the male R25 robot were uncertain if the robot understood how they felt.

Figure 6 illustrates that when analyzed independently, responses from *Study B* yield a statistically significant difference between participants in the different interviewer conditions (Human male, Nao male, R25 male) on Q4 (Fisher's Exact Test, $p = 0.05$). In the male Nao robot condition 66.67% of participants reported that the interviewer understood how they felt, while 22.22% indicated the robot did not have a good understanding of how they felt. Of the participants in the male R25 condition 62.5% indicated the robot understood how they felt, while 12.5% responded that the robot did not understand how they felt. In the human interviewer condition 95% of participants felt the interviewer understood how they felt.

### Participant gender

When responses were split into groups based on participant gender (male or female) statistically significant differences were present for items concerning interviewer understanding of feelings and ability to provide advice (Q4, Q5). Participants identifying as female were more likely (26.47%) to report uncertainty when asked how well the interviewer understood how they felt in comparison to male participants (10.45%, Fishers Exact Test $p = 0.01$), who were more likely to endorse the *Negative* option (13.43%) in comparison to females (2.94%). Furthermore, participants identifying as female were more likely (97.18%) to indicate the interviewer could provide them helpful advice in comparison to participants identifying as male (85.51%, Fisher's Exact Test $p = 0.04$).

**Table 3:** Participant responses to Q3-Q7, Q9, Q12, and Q13 in the Perceptions of the Interviewer Interview categorized by interviewer type (human or robot). Statistically significant test statistic cells shaded.

| Responses to Perceptions of the Interviewer Survey | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Question 3 | | | Question 4 | | | Question 5 | | | Question 6 | | |
| | Well | Not Well | Unsure | Well | Not Well | Unsure | Yes | No | Unsure | Yes | No | Unsure |
| Human | 55 | 0 | 10 | 55 | 1 | 9 | 64 | 0 | 1 | 24 | 32 | 11 |
| | (85%) | (0%) | (15%) | (84.62%) | (1.54%) | (13.85%) | (98.46%) | (0%) | (1.54%) | (35.85%) | (47.76%) | (16.42%) |
| Robot | 63 | 5 | 3 | 44 | 10 | 16 | 64 | 7 | 4 | 30 | 34 | 7 |
| | (88.73%) | (7.04%) | (4.23%) | (62.86%) | (14.29%) | (22.86%) | (85.33%) | (9.33%) | (5.33%) | (42.25%) | (47.89%) | (9.86%) |
| Test Statistic | Fisher's Exact Test $p = 0.01$ | | | $\chi^2(2) = 10.37$ $p = 0.01$ | | | Fisher's Exact Test $p = 0.01$ | | | $\chi^2(2) = 1.5$ $p = 0.47$ | | |

| Responses to Perceptions of the Interviewer Survey *(cont...)* | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Question 7 | | | Question 9 | | | Question 12 | | | Question 13 | | |
| | Yes | No | Unsure | Yes | No | Unsure | Yes | No | Unsure | Yes | No | Unsure |
| Human | 25 | 34 | 6 | 64 | 1 | 2 | 67 | 0 | 0 | 40 | 1 | 25 |
| | (38.46%) | (52.31%) | (9.23%) | (95.52%) | (1.49%) | (2.99%) | (100%) | (0%) | (0%) | (60.61%) | (1.52%) | (37.88%) |
| Robot | 39 | 24 | 8 | 69 | 3 | 3 | 68 | 2 | 3 | 44 | 21 | 8 |
| | (54.93%) | (33.8%) | (11.27%) | (92%) | (4%) | (4%) | (93.15%) | (2.74%) | (4.11%) | (60.27%) | (28.77%) | (10.96%) |
| Test Statistic | $\chi^2(2) = 4.8$ $p = 0.09$ | | | Fisher's Exact Test $p = 0.76$ | | | Fisher's Exact Test $p = 0.12$ | | | Fisher's Exact Test $p = 0.06$ | | |

Figure 7 illustrates significant differences found within each study. When examined separately, responses from *Study A* show significant differences between participant reported genders and appraisals of the interviewer for Q6 and Q13. Compared to 35.71% of male participants, 66.67% of female participants reported that they could not hide how they felt from the interviewer (Fisher's Exact Test $p = 0.02$).

Furthermore, within *Study A* female participants (64.29%) were more likely than male participants (50%) to perceive that the interviewer liked them (Fisher's Exact Test $p = 0.02$). While 16.67% of male participants responded that they felt the interviewer did not like them, none of the female participants reported the perception that the interviewer did not like them.

All female participants (100%) in *Study B* reported that the interviewer could provide helpful advice, while 78.57% of male participants indicated the interviewer could provide helpful advice (Fisher's Exact Test $p = 0.02$).

### 5.4.3 Social factor mentions

Following the analysis of agreement and appraisal responses, we examined the explanations that participants provided for their responses. Each response was tagged with all relevant social factors (described in Section 5.4.1). Data from *Study A* and *Study B* were combined and analyzed as a whole.

For each participant we computed whether or not the participant made mention of each social factor across their entire response to the *Perceptions of the Interviewer Interview*. This was done to limit the influence of participants who cited the same factors repeatedly for multiple items. As each factor could have a positive or negative valence, this resulted in a set of twelve binary variables for each participant indicated whether or not the participant discussed the factor. For the analysis in this article, the subcategories for each factor were not examined, rather they were counted as representing their higher level category.

Figure 8 compares the percentage of participants in human and robot interviewer conditions citing each social factor. With the exception of the *Knowledge+* social factor, all other positive social factors were referenced significantly more by participants in a robot interviewer condition.

Since participants in robot interviewer conditions had the opportunity to respond to two additional questions (Q2 and Q8), we also conducted an analysis in which we removed all responses from these questions. Removing these items results in a loss of statistical significance between human and robot interviewers for the area of *Social Con-*
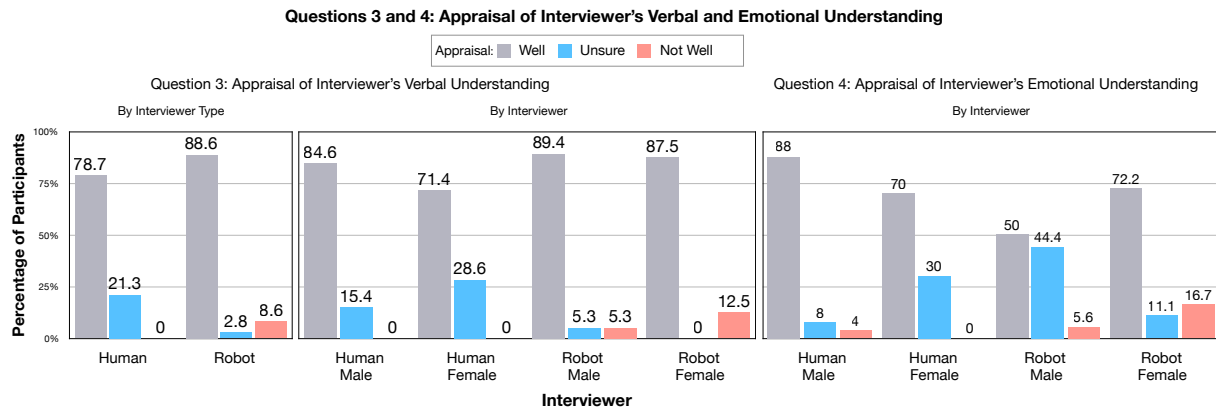
**Questions 3 and 4: Appraisal of Interviewer's Verbal and Emotional Understanding**

Appraisal: ▇ Well  ▇ Unsure  ▇ Not Well

Question 3: Appraisal of Interviewer's Verbal Understanding

Question 4: Appraisal of Interviewer's Emotional Understanding

By Interviewer Type | By Interviewer | By Interviewer

**Figure 5:** Participant responses to questions about the interviewer's ability to understand what they said and how they felt.
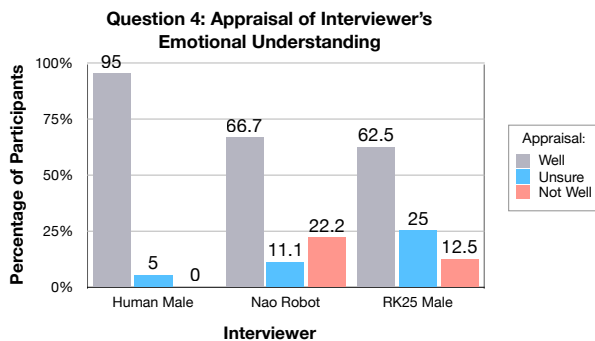
**Question 4: Appraisal of Interviewer's Emotional Understanding**

**Figure 6:** Older participant's assessment of how well the interviewer understood their feelings, divided by interviewer.

*fidence*, which includes components of general trust, social judgment, and maintaining privacy ($\chi^2(1) = 2.67, p = 0.10$, Cramer's $V = 0.19$, small effect). When incorporating all questions, 41.33% of participants in the robot conditions discussed factors related to *Social Confidence*, but when excluding responses to the question *How is the interviewer different from a human?* (Q8) this declined to 34.67% of participants in the robot interviewer conditions. Of the participants in the human interviewer conditions, 20.9% identified positive factors related *Social Confidence*.

To further understand this difference as it relates to privacy and social judgment, we examined the difference between human and robot conditions when removing generic references of trust and instead included only cases where participants described specific privacy or social judgment protections provided by the interviewer. This analysis found that 50% of those in the human interviewer conditions endorsing areas of *Social Confidence* referred only to general trust (e.g., "I felt like I could trust the interviewer") rather than specific elements dealing with pri-

vacy or social judgment (e.g., "I could trust that the interviewer would not share what I said with others."). The resulting analysis finds that regardless of the inclusion or exclusion of items only asked in robot interviewer conditions, participants in robot interviewer conditions were significantly more likely to report the interviewer would maintain their privacy and refrain from social judgment ($\chi^2(1) = 6.05, p = 0.01$, Cramer's $V = 0.21$, small effect size).

Significant differences in negative mentions related to areas of *Demeanor* and *Knowledge* were found between the robot and human interviewer conditions ($\chi^2(1) = 18.99, p < 0.001$, Cramer's $V = 0.92$, large effect, and Fisher's Exact Test $p < 0.01$ respectively). Participants in the robot interviewer conditions were more likely to identify deficits in the interviewer's demeanor (28% robot, 1.5% human) and knowledge (12% robot, 0% human).

# 6 Discussion

The results from *Study A* and *Study B* indicate several significant differences between participants depending on the groupings selected for analysis. This section discusses the implications for differences found between interviewer types and participant genders and provides further insight into the social factors participants cited in their responses. Finally a discussion of potential limiting and confounding factors is provided.

## 6.1 Interviewer differences

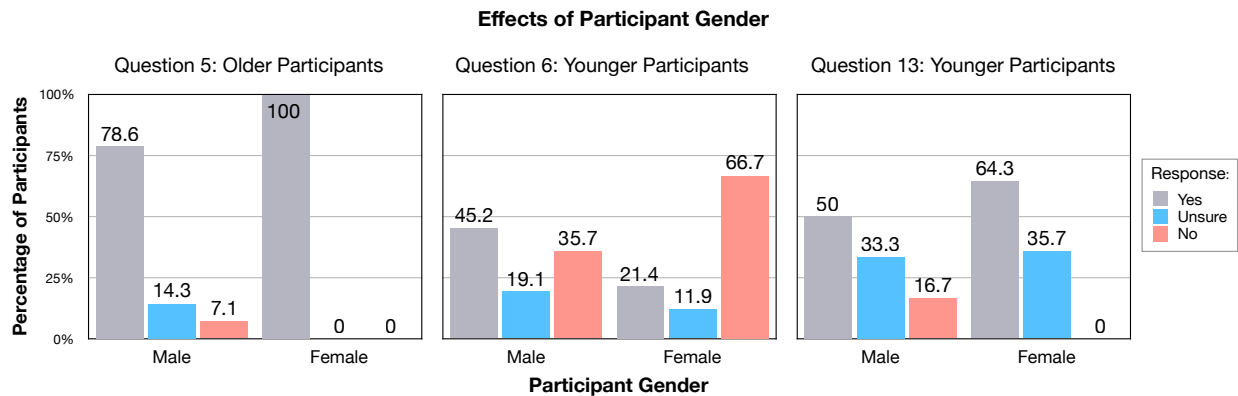Results from the merged data of *Study A* and *Study B* indicated significant differences in the way participants per-

**Effects of Participant Gender**



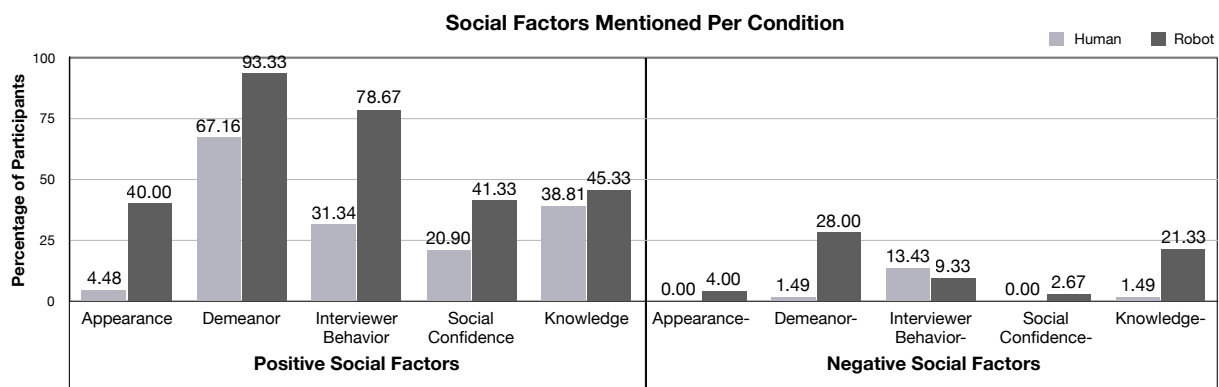**Figure 7:** Gender differences among participants in each study across all types of interviewers.



**Figure 8:** Positive and negative mentions of each social factor for human and robot interviewers.

ceived the ability of robot and human interviewers to understand what they were saying, how they felt, and to provide helpful advice.

Younger participants (in *Study A*) when analyzed separately only differed in their reporting of the robot and human interviewers ability to understand what they were saying. On average, *younger participants reported uncertainty more for human interviewers than robotic interviewers*. It is possible that the question itself was awkward or confusing when applied to a human interviewer, yielding more responses of uncertainty.

Older participants (in *Study B*) when analyzed separately only differed in their evaluations of human and robot interviewers' ability to understand how they felt. On average participants reported that *human interviewers understood how they felt 95% of the time compared to 64.71% of the time for robot interviewers*. The R25 robot received more negative ratings of understanding feelings than the human interviewer, while the Nao robot received more indecisive ratings than the human interviewer. This suggests

older participants may be more critical or skeptical of the robot's ability to understand human emotions.

While the effects for understanding are likely driven independently by each study, differences in the human and robot interviewer's ability to provide helpful advice were only significant in the merged data. While 98.46% of participants in the human condition felt the interviewer could provide helpful advice, only 85.33% in robot conditions reported the same. Further analysis of other data collected during the study may lend insights into this result. This may also be a result driven by participant demand characteristics, such as the good-subject effect, in which participants attempted to provide responses they believed the researcher desired [20].

## 6.2 Participant differences

In addition to differences between human and robot interviewers, differences were also identified when splitting participants by their reported gender.

Among younger participants (in *Study A*) 66.67% of female participants felt they could not hide how they felt from the interviewer, while only 35.71% of male participants felt this way. This effect was not present among participants in *Study B*. Furthermore, among those in *Study A*, male participants were more likely to report that the interviewer did not like them (16.67%) in comparison to female participants who only provide positive or uncertain responses.

In *Study B* all female participants reported that the interviewer could provide helpful advice, while only 78.5% of male participants reported the same.

In the merged data gender effects related to the interviewer providing advice remain with 97.18% of females and 85.51% of males indicating the interviewer could provide helpful advice. Additionally, a significant result between genders for how well the interviewer understood how the participant felt appears. Males were more likely to report the interviewer did not understand how they felt, while females were more likely to respond indecisively.

## 6.3 Social factors and social judgment

Examining the social factors presented by participants during their evaluation of the interviewer, significant differences are present between robot and human interviewers in areas of appearance, demeanor, behavior, and social confidence.

It is likely that cultural social norms, along with the novelty of the experience contributed to participants with robot interviewers being more likely to comment on the appearance of the interviewer. Interestingly, comments regarding the appearance of the robot were mostly positive and though several participants characterized the R25 robots as "creepy" (perhaps due to their location in the "uncanny valley" [21]) they noted this was a first impression that did not affect the remaining interview.

Similarly, increased positive comments about the demeanor of robot interviewers in comparison to human interviewers may be driven by the novelty factor associated with the robot or lower social expectations toward the robot. However, participants were more likely to report negative aspects of the robot's demeanor in comparison to human interviewers. These included statements about the robot's speech being unclear, being unable to understand emotions, having no emotions, and lacking the ability to fully understand the participant.

While describing their experiences, participants were more likely to cite an example of specific behaviors robot interviewers performed during the interview than human interviewers. This may be a result of lower expectations for the robot's performance or the desire to describe a novel experience.

Participants with robot interviewers were significantly more likely to describe factors related to social confidence in the interviewer than those with human interviewers. Of those with robot interviewers 41% expressed general trust, lack of social judgment, or a guarantee of privacy with the interviewer; only 21% of participants with human interviewers expressed similar sentiments.

The *Social Confidence* category is composed of three factors (privacy, social judgment, and general trust). When responses coded as "general trust" (i.e. statements like "I can trust her") are removed from the analysis, a significant difference between human and robot interviewers remains. Examining factors related to social judgment (e.g.,"He listened without judging you." ) alone shows that 3% of participants with human interviewers and 11% of participants with robot interviewers report feeling protected by a lack of social judgment from the interviewer. When factors related to privacy (e.g., "I know that she would not tell anyone.") are isolated in the merged data, 7% of participants with human interviewers and 19% of participants with robot interviewers reported that the interviewer would protect their privacy and that they could disclose sensitive information without worry.

Splitting social factor results by study (i.e. younger vs. older participants), it can be observed that differences related to privacy and social judgment between human and robot interviewers are more pronounced among older participants (*Study B*). For those in *Study B*, explanations related to social judgment and privacy were reported for 15% of interviews conducted by the human interviewer, while they were reported for 44% of the interviews conducted by robots. In comparison, of those in *Study A* (younger participants), 9% of human conditions and 18% of robot conditions reported the same factors.

One potential explanation for the disparity between age groups is that older participants had more privacy concerns or were better able to express their thoughts to the researcher. However, only a small increase is observed in reporting between age groups for human interviewers (19% for younger participants, 25% for older participants) while a larger increase is recorded for robot interviewers between age groups (33% for younger participants to 50% for older participants).

Overall these findings indicate that social robots, even when positioned in research contexts, evoke assurances of social protection from children interacting with them. Children interviewed in both studies expressed confidence that robots would not judge them in the same way a hu-

man would. Additionally, participants in both studies reported that the robot interviewer would not disclose their personal information, with several reasoning that doing so would not benefit the robot.

## 6.4 Limitations

While the data gathered in these studies has provided valuable insights into the perceptions of social robots used in sensitive contexts with children, there are several important factors to consider when interpreting these results.

First, studies were conducted in the Southern United States and participation in interview studies required either the participant or their parent to be able to travel to the study location to participate. This may affect the generalizability of these results.

Second, within child interview studies the small sample size for each possible combination of interviewer type, interviewer gender, participant gender, and participant age prevents rigorous analysis of each of these factors. For example, *Study B* used only male interviewers and the assignment of interviewer was disproportionate across participant genders (e.g., female participants conditions: 14 human interviewer, 6 Nao robot, 6 R25 robot). In addition between the two studies the male R25 robot conducted 48% of all robot interviews, creating an unequal sample size for comparisons of merged data.

Third, interview studies were conducted following a forensic interview protocol which places constraints on the social interaction that would not exist in other interaction paradigms. Specifically, interviewers work to avoid expressing significant emotions or feedback (verbal and nonverbal) that could influence the responses of the participant and adhere to a structured script with limited adaptations to incorporate the information disclosed by the participant. While ecologically the interview studies are most aligned with forensic interviews, all interviewers observed the same constraints fostering internal validity when comparing responses to the Perceptions of the Interviewer Interview across conditions.

Finally, approximately one month prior to *Study A* a local 12 year old student died by suicide linked to bullying [22]. Many of the participants in the study referenced this event and several had been classmates with the student. As a result the subject-matter of the interaction was particularly sensitive for these participants.

## 7 Conclusion and future work

Findings from the small interactive design study for a therapeutic robotic dog highlight that participants had basic privacy concerns and felt the platform would be most useful for interactions with children. The findings associated with the two forensic interview studies with children indicated that participants were more likely to ascribe socially protective factors including privacy to social robot interviewers than human interviewers. Specifically, children were more likely to cite the lack of social judgment and guarantee of privacy a social robot provides as a positive and protective social factor. Future work should investigate and better qualify these user beliefs and expectations across other domains and robotic platforms.

As we continue to investigate robot-assisted therapy and move towards the use of more autonomous systems, ethical practices merit further focused inquiry. Specifically, questions related to how user characteristics shape expectations and beliefs about therapeutic systems should be examined and incorporated into the design process.

## References

[1]  Z. Henkel, C. L. Bethel, A robot forensic interviewer: The bad, the good, and the undiscovered, In: Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, ACM, 2017, 10–20

[2]  C. J. Calo, N. Hunt-Bull, L. Lewis, T. Metzler, Ethical implications of using the Paro robot with a focus on dementia patient care, In: Proceedings of the 12th AAAI Conference on Human-Robot Interaction in Elder Care, AAAI Press, 2011, 20–24

[3]  S. Thill, C. Pop, T. Belpaeme, T. Ziemke, B. Vanderborght, Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook, Paladyn, Journal of Behavioral Robotics, 2012, 3(4), 209–217

[4]  R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, C. Pozzi, Child-robot interaction in the wild: advice to the aspiring experimenter, In: ICMI '11 Proceedings of the 13th International Conference on Multimodal Interfaces, ACM, 2011, 335–342

[5]   J. Saldien, K. Goris, S. Yilmazyildiz, W. Verhelst, D. Lefeber, On the design of the huggable robot Probo, Journal of Physical Agents, Red de Agentes Físicos, 2008, 2(2), 3–11

[6]   K. Goris, J. Saldien, I. Vanderniepen, D. Lefeber, The huggable robot Probo, a multi-disciplinary research platform, In: International Conference on Research and Education in Robotics, Springer, 2008, 29–41

[7]   B. Robins, K. Dautenhahn, R. Boekhorst, A. Billard, Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?, In: Universal Access in the Information Society, Springer, 2005, 4(2), 105–120

[8]   B. Scassellati, H. Admoni, M. Mataric, Robots for use in autism research, Annual Review of Biomedical Engineering, Annual Reviews, 2012, 14, 275–294

[9]   N. Giullian, D. Ricks, A. Atherton, M. Colton, M. Goodrich, B. Brinton, Detailed requirements for robots in autism therapy, In: 2010 IEEE International Conference on Systems Man and Cybernetics (SMC), IEEE, 2010, 295–2602

[10]  M. Saerbeck, T. Schut, C. Bartneck, M. Janse, Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor, In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2010, 1613–1622

[11]  T. Komatsubara, M. Shiomi, T. Kanda, H. Ishiguro, N. Hagita, Can a social robot help children's understanding of science in classrooms?, In: Proceedings of the 2nd International Conference on Human-Agent Interaction, ACM, 2014, 83–90

[12]  D. Leyzberg, S. Spaulding, B. Scassellati, Personalizing robot tutors to individuals' learning differences, In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, ACM, 2014, 423–430

[13]  J. Westlund, C. Breazeal, A. Story, Deception, secrets, children, and robots: What's acceptable, In: Workshop on the Emerging Policy and Ethics of Human-Robot Interaction, held in conjunction with the 10th ACM/IEEE International Conference on Human-Robot Interaction, ACM, 2015

[14]  J. M. Kory Westlund, C. Breazeal, Transparency, teleoperation, and children's understanding of social robots, In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, 2016, 625–626

[15]  M. E. Lamb, Y. Orbach, I. Hershkowitz, P. Esplin, D. Horowitz, A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD investigative interview protocol, In: Child Abuse & Neglect, Elsevier, 2007, 31(11-12), 1201–1231

[16]  J. F. Kelley, An iterative design methodology for user-friendly natural language office information applications, In: ACM Transactions on Information Systems (TOIS), ACM, 1984, 2(1), 26–41

[17]  C. R. Rogers, L. Carmichael, Client–Centered Therapy: Its Current Practice, Implications, and Theory, with Chapters Contributed by E. Dorfman, T. Gordon, and N. Hobbs, Houghton Mifflin Company, 1951

[18]  Z. Henkel, C. L. Bethel, J. Kelly, A. Jones, K. Stives, Z. Buchanan, et al., He can read your mind: Perceptions of a character-guessing robot, In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE, 2017, 242–247

[19]  P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, H. Sloetjes, Elan: a professional framework for multimodality research, In: 5th International Conference on Language Resources and Evaluation (LREC 2006), 2006, 1556–1559

[20]  A. L. Nichols, J. K. Maner, The good-subject effect: Investigating participant demand characteristics, Journal of General Psychology, 2008, 135(2), 151–166

[21]  M. Mori, K. F. MacDorman, N. Kageki, The uncanny valley [from the field], IEEE Robotics & Automation Magazine, IEEE, 2012, 19(2), 98–100

[22]  A. Holloway, Goodbye Mariah: More than 1,000 attend vigil in memory of Armstrong middle school student, In: The Commercial Dispatch, May 2, 2017