The Robot Who Knew Too Much: Toward Understanding the Privacy/Personalization Trade-Off in Child-Robot Conversation

Iolanda Leite
Disney Research
Pittsburgh, USA
iolanda.leite@disneyresearch.com

Jill Fain Lehman
Disney Research
Pittsburgh, USA
jill.lehman@disneyresearch.com

ABSTRACT

In human-human conversation we elicit, share and use information as a way of defining and building relationships – how information is revealed, and by whom, matters. A similar goal of using conversation as a relationship-building mechanism in human-robot interaction might or might not require the same degree of nuance. We explore what happens in the increasingly likely situation that a robot has sensed information about a child of which the child is unaware, then discloses that information in conversation in an effort to personalize the child's experience. In a pilot study, 28 children conversed with a social robot that either told a story with characters already introduced into the conversation by the child (control) or characters hidden by the child in a treasure chest that the child was holding (experimental). Cumulative evidence showed that all participants in the experimental condition noticed the robot's violation of expectations, but younger children (4 to 6 years) exhibited more contained emotional reactions than older children (7 to 10 years), and girls expressed more negative affect than boys. Despite the immediate response, post-conversation measures suggest that the single event did not have an impact on children's ratings of robot likeability or their willingness to interact with the robot again.

ACM Classification Keywords

H.1.2 User/Machine Systems: Human factors

Author Keywords

Conversational robotic companions, child-robot interaction, private information, personalization.

INTRODUCTION

Children are growing up in a world increasingly populated by sensors that communicate with each other to personalize interactions and create unique user experiences [2]. In many cases, users are not even aware that personal information is being collected and used later, or by other applications. Extensive research has investigated the privacy-utility trade-offs in adult users, particularly how adults reason about the disclosure of information in exchange for personalization [10, 11, 20]. However, little is known about how these questions affect children, despite the trend in commercial products that seek to offer them personalized experiences as well [25, 15].

We are particularly interested in the case of conversational robotic companions for children. Robots have great potential to leverage information from data collected in previous activities and use it to make a child feel special. Suppose, for example, that a robot knows that the child won a soccer tournament at school. How should it plan its dialog, given the goal of creating a moment of positive regard? If it asks a general question ("What happened today?"), the desired topic might not arise and the opportunity will be lost. But will simply raising the topic of sports be enough? Will introducing the win explicitly be too much? In human-human conversation we make choices about if and when to bring information into the common ground [6] as a way of defining and building relationships – how information is revealed, and by whom, matters. Although conversational interaction is extremely promising as a rapport-building mechanism, it could also be discordant if there is a mismatch between what the robot discloses and what the child believes the robot knows. There are two separate questions of interest in such a scenario: If the robot reveals seemingly inaccessible information, will the child notice? If the child notices, will the event have a positive or negative effect on perception of the interaction and the robot?

To begin to address these questions we consider a case in which the robot demonstrates that it has knowledge that should be inaccessible by conversing about information only the child should be able to reveal. The ability to understand that two people can have different knowledge states about the same event is often referred to as Theory of Mind (ToM) [3]. Developmental research indicates that ToM is established in children by four years [7], the age of the youngest children in our study. Prior research also suggests that children perceive and interact with robots as social entities [21, 4, 8, 5]. In theory, then, our participants should posit the same processes that they do in another human or familiar human-like object (such as a doll), and notice that the robot's knowledge is different from expected, but it is possible that they will not make the same ToM assumptions about our robot (Figure 1), or that they will do so only at a different point in development. Developmental effects might also apply to the second question of interest. For children who do posit separate and different knowledge states at the outset of the conversation, the revelation that the robot actually has the unanticipated knowledge might be construed as magical or invasive, and which reaction occurs might depend on the child's age.

Because of these potential age differences and the general difficulty of accessing the subjective states of young children in an unfamiliar environment, we look for cumulative evidence from multiple measures. Within-conversation metrics include both social-referencing through eye gaze and judgments of emotional reaction by familiar adults at the moment the robot reveals its knowledge of private information. Post conversation, we look to self-report, using a survey format that the child has practiced in earlier robot-centric activities. Overall, we find that children do extend typical ToM expectations to our robot and notice the violation when it is revealed in conversation, but whether they experience it as a negative event is a function of age and gender.

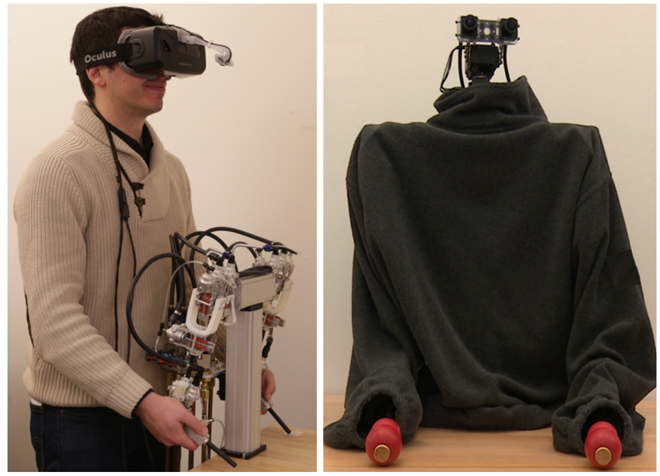


Figure 1: The immersive setup for controlling Jimmy (left) and the robot as it appeared to the child (right).

RELATED WORK

Over the past decade, researchers in Human-Robot Interaction (HRI) have been investigating the impact of a robot's unpredictable behavior, such as cheating [19, 22] or deception [23], on users' perceptions of robot intelligence, intentionality, and trust, as well as with the purpose of making interactions more engaging [14]. While most of this research has been conducted with adults, a few of the studies have included children. Lemaignan et al. [12] evaluated the responses of 4- and 5-year-olds in a playful scenario in which a social robot occasionally exhibited unexpected behaviors such as getting lost or disobeying voice commands. They found that the robot was more engaging when it misbehaved than when it was behaving appropriately. Although this study was inconclusive as to whether children of different age groups interpreted the robot's misbehaviors as mechanical malfunctions or as intentional, the authors suggest that there might be age effects in the way children perceive different misbehavior categories. This study is the most similar to the one we report here, but the context was not conversational and the authors' main goal was to investigate the effect of unexpected movement rather than unexpected knowledge.

Prior work has established that the way children socialize with robots is more similar to the way they do so with other people than with toys or other non-living objects [21, 8]. As robots are expected to interact with children over longer periods of time, one increasingly relevant question becomes whether children attribute social and moral standing to them, and which social and moral conventions are assumed. Kahn and colleagues [9] investigated this question in a study with 9- to 15-year-olds. In their experiment, a brief social conversation with a humanoid robot was followed by an incident of "robot moral harm" – the child observed the robot being put in a closet by one of the experimenters, despite its objections. The main findings from a semi-structured interview were that most children believed the robot had moral standing and considered it as a social being who deserved fair treatment. Interestingly, 15-year-olds rated the robot lower in many of the study measures than children with ages between 9 and 12 years. More recently, Chernyak and Gary [5] found that children's moral judgment regarding robots is influenced by their perception of the robot's autonomy. This finding implies that the way a robot is presented to the child can largely impact the way s/he perceives and behaves toward the robot.

Another source of evidence that children view robots as social entities is the work of Bethel and colleagues [4]. They found that preschoolers interacted with a social robot using the same type of social conventions they use while interacting with each other. More importantly, their study showed that children were as likely to share a secret (that they were asked to keep) with a robot as they would with an adult. The ethical issues of sharing private information, particularly when a child shares something with the robot, are now starting to be raised and discussed in HRI [26, 24].

Although the focus of each of the studies we presented so far is different from the one we address, the results are relevant to our work for two main reasons. First, because we start from the assumption that children will perceive and interact with the robot in our study as a social entity, and second, because they point to several developmental age differences that might also effect the outcomes in a conversational context.

EXPERIMENTAL METHOD

The questions we want to explore require establishing a conversational moment in which the robot (hereafter, Jimmy) reveals knowledge of information that is presumed private. We create this moment during a longer sequence of activities in an hour-long session of play we call "The Winter Games." Children were recruited in close-aged pairs and participated in a total of six games, some together, some alone, some with robots and some without. This winter's activities included:

- Robot Rating: sitting together, children were shown and asked to rate five robot faces for friendliness and likeability using the *Smileyometer* [17] rating scale. All faces were displayed on a Furhat robot head [1], accompanied by experimenter-controlled expressions, speech and movement.

- *Mole Madness*: children played a cooperative, voice-controlled videogame that moves an animated character through its environment in response to the words "go" and "jump."
- *Mole Madness*/Robot Bear Hug: one child played *Mole Madness* with a Furhat-headed robot as co-player while the other was asked to hug a sensor-laden, soft-technology bear. Each child filled out two Smiley scales about the experience, one task-dependent question and one likeability question, then switched to the other task.
- *Storybook Adventure/Storytelling and Conversation with Jimmy*: one child chose the appearance of different characters in a storybook, while the other child participated in the robot interaction activity that is the focus of this paper, then they switched. We explain the Jimmy activity and the control and experimental conditions in detail, below.

Each activity in the Winter Games has its own research goals and measures. We enumerate the larger context here to point out three relevant features. First, all children had experiences with three different robots prior to the conversation with Jimmy in an effort to control for any novelty effects that might arise from first-time interaction with a robot. Similarly, all children had repeated exposure to the scale used for self-report in the Jimmy task, and had practiced using the scale for both task-specific and likeability judgments. Finally, because children participated in pairs and conversation with Jimmy was always the last robot-related task, by randomly assigning one child from each pair to the control and experimental conditions of the Jimmy task we made a best effort to ensure that at least their most recent history with robots was comparable.

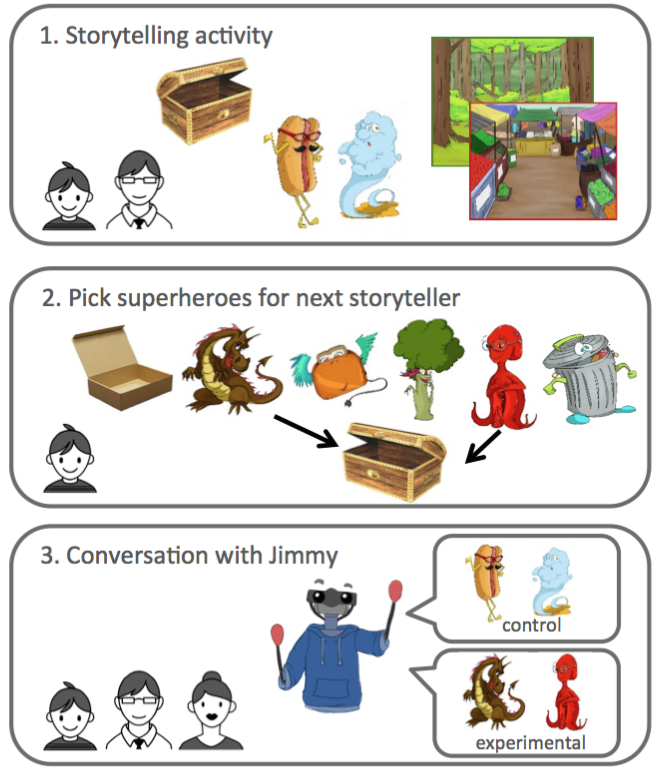


Figure 2: The three phases of the Jimmy task: children tell a superhero story about two characters, make a private decision about which characters some other child should use, then converse with Jimmy about the story they told. Jimmy follows the child's retelling with a story of its own, using either the characters revealed by the child in the conversation (control) or the characters that have not been revealed (experimental).

Participants

We recruited 28 children (17 females and 11 males) through postings in physical and online community bulletin boards. Children's ages ranged from 4 to 10 years ($M = 6.7, SD = 1.82$). Participants were scheduled in closely-aged pairs with one child assigned to each condition in order to balance with respect to age and gender. The control group consisted of 9 females and 5 males ($M = 6.6$ years, SD = 1.95), and the experimental group contained 8 females and 6 males ($M = 6.9$ years, $SD = 1.75$). The average age difference between participants within a pair was eight months. Children were compensated for their participation. The research was approved by the Human Subjects Institutional Review Board of Carnegie Mellon University.

Accompanying adults – in most cases a parent – were offered the opportunity to participate in the Jimmy task as uncompensated volunteers, as described below. Those who agreed (25/28) were predominantly female (18/25).

Procedure

The purpose of the Jimmy task was to explore two questions: would children across our age range notice if the robot revealed knowledge which should have been inaccessible? And, if noticed, would Jimmy's revelation have a positive or negative impact on the child? Examining these questions requires that we establish and reveal the same piece of "private" knowledge for each child in the experimental condition.

For the child's comfort, we deliberately chose **not** to request and reveal personal information. Instead we created a three-phase process in which every child established two facts out of Jimmy's view and had one elicited by Jimmy in natural social conversation. The distinction between conditions hinges on the remaining piece of information. In the control condition, Jimmy elaborated on the fact already introduced by the child into the common ground. In the experimental condition, Jimmy introduced the comparable fact that the child believed was known to him/her alone. Figure 2 summarizes the procedure, described more completely in the following sections.

Phase 1: Storytelling

In the first phase (top panel of Figure 2), the experimenter walked the child to a table in one area of the lab, announcing that the next activity involved storytelling. Once there, the

Control / Experimental

E	Hi, Jimmy!
J	Hi there, nice to see you! (to child) Are you having fun?
C	Yes.
J	You kids get to do all the fun stuff. I get so bored stuck here all day... Which game did you just play?
C	A storytelling game.
J	Oh, I like that one a lot! Where was your story? The forest or the market?
C	The forest.
J	The forest is my favorite!! Who was in your story?
C	Whirlwind and the Hot Dog Guy.
J	That is a very interesting super hero pair. What happened in your story?
C	They went into the forest to rescue a lost kitten.

Control	*Experimental*
J It sounds like a great story! I wish they'd let me tell one...**I'd start with Whirlwind and Hot Dog Guy, just like you did**, but in the market. And then Whirlwind would blow everything all over the place until Hot Dog Guy shows up and saves all the people and kittens and stuff by spraying him with ketchup and mustard and chasing him away. What do you think?	J It sounds like a great story! I wish they'd let me tell one...**I'd pick Red Octopus and Dragon Lady, just like you did**, but in the market. And Red Octopus would slither into the market and stink everything up with his fishy smell until Dragon Lady comes and saves all the people and kittens and stuff by swooping in and carrying Red Octopus away. What do you think?

Control / Experimental

C	It's a good story.
E	All right Jimmy, we have to go now.
J	Oh, I guess our time is over... I need to recharge! It was great meeting you. See you later!

Table 1: An example of one Child's (C) responses during the scripted interaction with Jimmy (J) and the Experimenter (E), with the Critical Moment (CM) that distinguishes control and experimental conditions shown in bold.

experimenter invited the child to open a treasure box containing two superhero characters printed on cardboard. The same two characters, Whirlwind and Hot Dog Guy, were used by all participants. The child then chose a scenario placemat (forest or market) where the story action would occur, and was told s/he could use any or all of a small set of story tokens (e.g., magic potion, kitten, etc.) to help tell the story. Children who had trouble getting started were given a story prompt ("One day Whirlwind and Hot Dog Guy were in the forest when all of a sudden..."). This task lasted approximately 5 minutes and was recorded by a camera that was in plain view.

Phase 2: Establishing Private Knowledge

When the story was finished to the child's satisfaction, the experimenter asked him/her to move to a different table that was clearly out of camera view and on which there was a closed, empty treasure chest and a closed, plain box containing five new story characters. A commercially available near-field RFID antenna was completely hidden on the underside of the table beneath the box, and each of the story characters was augmented with a non-visible passive RFID tag. Through this arrangement the identity of the characters in the box was continually broadcast to the robot.

The child was instructed to pick two of the five new story characters in the plain box, place them in the treasure chest for a future storyteller, then close both boxes (middle panel of Figure 2). To reinforce the idea that the child's choice was private, the experimenter turned her back after the instruction, saying she wanted to be surprised by the new superhero characters when the box was opened in the future.

When the child signaled that the choices were made and the boxes were closed, s/he was told to take the treasure chest and was guided by the experimenter to the area of the laboratory where Jimmy was located, picking up the accompanying adult along the way. As the RFID tags of the two characters inside the treasure chest moved out of range of the reader, their absence in the broadcast set identified them to Jimmy's software.

Phase 3: Conversation with Jimmy

As shown in Figure 1, Jimmy was an upper body torso that allowed human teleoperation via an identical master through a hybrid air-water configuration [27]. The robot has 4 degrees of freedom (DOF) in each arm and a 2-DOF neck. Stereo cameras mounted on Jimmy's neck stream real-time video to the operator's head-mounted display, which in turn maps the head orientation of the operator to the neck servos of the robot. The operator is visually immersed in the robot's physical space and can "puppeteer" the robot on the other side of the wall in real-time, making this setup extremely suitable for conducting Wizard-of-Oz human-robot interaction experiments. In this study, Jimmy was controlled by two human operators: one operator was responsible for controlling the body movements

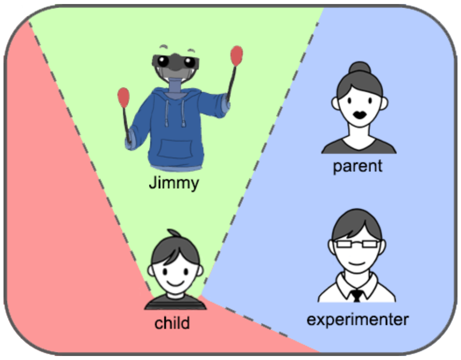


Figure 3: Spatial locations of participants during the conversation with Jimmy.

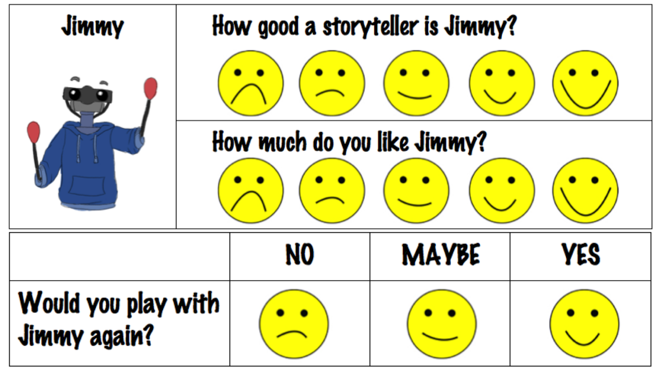


Figure 4: Survey questions used for collecting the storytelling judgment and likeability measures.

and head orientation, and the second operator selected Jimmy's text-to-speech utterances using a touch-based interface.

The experimenter initiated the conversation with Jimmy after positioning the participants as depicted in Figure 3. The conversation followed the same script for each child; an example is shown in Table 1. After a simple greeting, Jimmy asked the child a series of questions about the storytelling activity, eliciting the names of the characters and location used in a natural way. If the child could not remember the answer and looked to the experimenter or parent for help, the experimenter provided a brief response.

The conversational turn following the child's story retell was the point that distinguished control and experimental conditions (middle section of Table 1). The robot made a positive comment about the child's story then proceeded to tell a story of its own. In the control condition, Jimmy told a story about Whirlwind and Hot Dog Guy, the two characters that the child had already introduced into the conversation. In the experimental condition, the robot told a story about the two characters inside the treasure box that the child was holding.[1] The boldface text in Table 1 contrasts the Critical Moment (CM). If the child makes the typical Theory of Mind assumption for Jimmy, then Jimmy should not know what is in the box and the child should show some reaction that acknowledges the inconsistency.

Following the critical moment in both conditions, Jimmy picked a location for its story that was the alternative to where the child's took place (in the example of Table 1, the market). This was a deliberate design choice because we wanted both conditions to have a piece of information that was non-congruent to the child's retell. At the end of Jimmy's story, the robot asked the child to evaluate its efforts. After the child did so, the experimenter ended the interaction (bottom section of Table 1) and the child filled in the Smileyometer questions for the Jimmy task.

[1] Because we could not know in advance which pair of characters the child would choose, there was a separate version of the experimental story for each character pair, but each version had a main message, duration and word complexity that was similar to the others and to the control story.

Measures

We collected measures both during the conversation and at the end of the interaction. Within-conversation measures included the accompanying adult's judgment of the child's *affective reaction* at the critical moment. Adults who volunteered to participate received a paper questionnaire that instructed them to pay special attention to the moment where Jimmy tells his story in order to answer the two questions. The first question asked them to characterize the child's emotional reaction on a 5-point scale ranging from 1 (very negative) to 5 (very positive), while the second asked them to list one or more words that described the reaction. It is important to note that parents knew nothing about the purpose of the experiment and were not present during the storytelling and character selection tasks.

Following prior research on analyzing child behavior [13], the second within-conversation measure was an analysis of *gaze orientation* from the CM to the end of Jimmy's story. Using video recorded during the interaction, one female coder with experience in behavioral analysis annotated children's gaze orientation according to three categories: "looking at the robot," "looking at the experimenter or the accompanying adult," or "looking elsewhere," which included all other gaze orientations (the green, blue and red areas of Figure 3, respectively). Combining the annotation results with the interaction logs from Jimmy's software, we calculated the number of different annotations during the period in which Jimmy told his story and coded it as the number of *gaze shifts*. A disproportionate number of gaze shifts can be a signal of discomfort or confusion. The story lasts 19 seconds in the control condition, and an average of 18 seconds across the experimental versions.

The post-conversation measures included children's self-reports on their impression of the robot. In particular, children were asked to rate the robot's *storytelling* (How good a storyteller is Jimmy?) and *likeability* (How much do you like Jimmy?) with the same five-point Smileyometer [17] scale they had used in the previous robot activities. They were also asked to fill out an Again Table [17], which uses the same cues but a nominal (yes, no, maybe) scale. The Again Table queried whether the child would *play again* with each of the robots

Condition	Control		Experimental	
Age group	*4 to 6*	*7 to 10*	*4 to 6*	*7 to 10*
Affective Reaction	3.88 (0.35)	4.60 (0.89)	3.8 (0.45)	3.29 (0.76)
Gaze Shifts	3.00 (2.27)	1.25 (0.50)	4.0 (1.22)	3.57 (1.72)
Storytelling	4.88 (0.35)	4.83 (0.41)	4.40 (0.55)	4.22 (0.55)
Likeability	4.50 (0.76)	5.00 (0.00)	3.80 (1.64)	4.56 (0.73)
Gender	*Female*	*Male*	*Female*	*Male*
Affective Reaction	4.38 (0.52)	3.80 (0.84)	3.14 (0.69)	4.00 (0.00)
Gaze Shifts	2.00 (1.83)	3.00 (2.35)	3.86 (1.57)	3.60 (1.52)
Storytelling	4.89 (0.33)	4.80 (0.45)	4.50 (0.53)	4.00 (0.89)
Likeability	4.67 (0.71)	4.80 (0.45)	4.25 (1.39)	4.33 (0.82)

Table 2: Means(standard deviations) of the within-conversation (affective response, gaze shifts) and post-conversation (storytelling judgment, likeability) measures for participants in the control and experimental condition, divided by age group and gender.

they had interacted with during The Winter Games. Figure 4 shows the visuals used for the self-report measures. The experimenter read and explained the text associated with each scale to non-readers. [2]

RESULTS

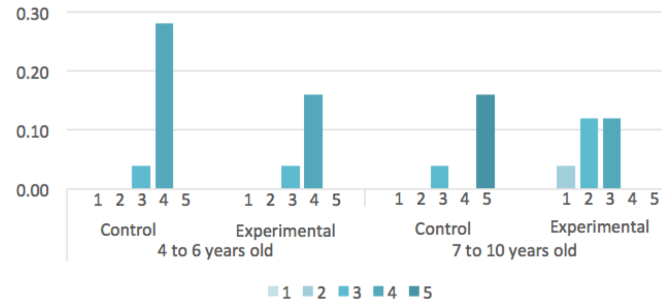
We conducted between-subjects analyses of variance (ANOVA) with study condition (control or experimental), participant age, and gender as factors. Due to sample size, participants were divided into two age groups (4 to 6 and 7 to 10 years) based on the child developmental literature [16]. Also due to the number of participants, we considered interactions of condition with age and gender individually rather than grouping them into a three-way ANOVA. The means and standard deviations of the measures reported below are presented in Table 2.

Within-conversation Measures

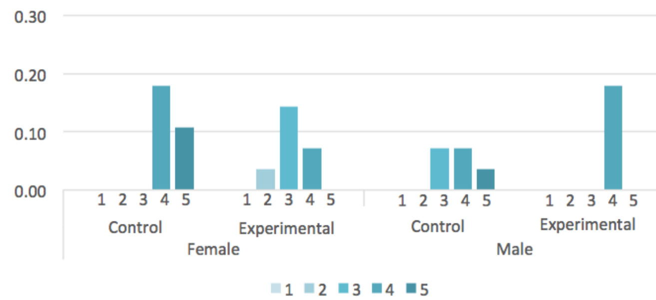
Affective Response

An ANOVA was conducted to investigate the impact of study condition and age group on parents' report of their child's affective response at the critical moment (see Figure 7 for examples of facial expressions). There was a significant main effect of study condition, $F(1,21) = 7.31, p < .05, \eta^2 = .22$, with parents considering their child's affective state as more positive in the control condition than in the experimental condition. No significant main effect of age was found for this measure, $F(1,21) = .17, p = .69, \eta^2 = .00$. However, a significant interaction effect was found between affective response and age group, $F(1,21) = 5.82, p < .05, \eta^2 = .17$. Parents of children in the younger age group provided similar ratings in the control condition and in the experimental condition, whereas parents of the older children considered their child's affect more positive in the control condition than in the experimental condition. The distribution of ratings by condition and age, normalized by the total number of parent judgments, is shown in the left graph of Figure 5a.

[2]We also asked children a few open-ended debrief questions. Most children, however, were not able to remember or articulate the critical moment of Jimmy's reveal, so these answers were not included in our analysis.



(a) Age group



(b) Gender

Figure 5: Affective reaction to the CM as rated by accompanying adults. Each graph shows the percentage of adults' responses at each value on the 5-point scale ranging from 1 (very negative) to 5 (very positive), normalized by the total number of judgments.

When including gender instead of age in the analysis, the main effect of study condition remained significant $F(1,21) = 4.45, p < .05, \eta^2 = .12$, and there was no significant main effect of participant's gender $F(1,21) =, p = 0.57, \eta^2 = .01$. A significant interaction effect was found between study condition and gender $F(1,21) =, p < .01, \eta^2 = .23$. The distribution of ratings by condition and gender is shown in the right graph of Figure 5b. While girls experienced more positive affective reactions in the control than in the experimental condition, the affective response of boys was not significantly different across the two conditions.

With only one older boy and one younger girl in the experimental condition whose parent volunteered to participate, it is impossible to tell from our data whether age and gender interact or one factor is the better predictor of the more negative affect observed.

Gaze Orientation

While investigating the effects of study condition and age group on the number of gaze shifts of the child during the period of the interaction in which Jimmy tells his story, we

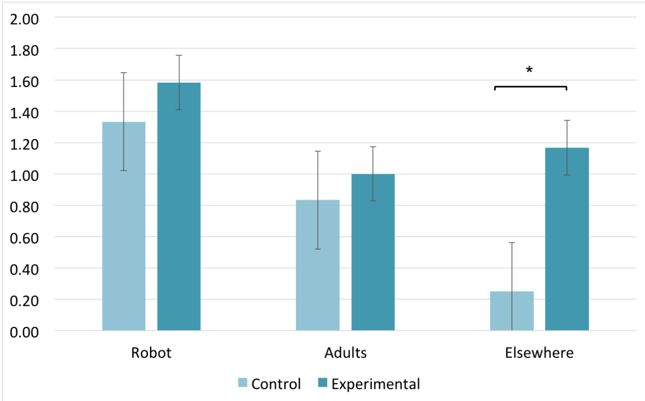


Figure 6: Average number of gaze shifts toward the robot, the adults or elsewhere for children in the control and experimental conditions during Jimmy's story.

found a significant main effect of study condition $F(1,20) = 5.08, p < .05, \eta^2 = .19$, with children in the experimental condition exhibiting a higher number of average gaze shifts than children in the control condition. There was no main effect of age group $F(1,20) = 2.19, p = .15, \eta^2 = .08$, or in the interaction between participant condition and age group $F(1,20) = .8, p = .38, \eta^2 = .03$. In the ANOVA conducted to study the effects of condition and gender on the number of gaze shifts, there were no significant main effects of condition $F(1,20) = 2.67, p = .11, \eta^2 = .11$, gender $F(1,20) = .24, p = .63, \eta^2 = .01$ nor in the interaction between both $F(1,20) = .70, p = .41, \eta^2 = .03$.

To further investigate these results, we performed pair-wise comparisons between each coded gaze region (robot, adults and elsewhere) across the study conditions with Bonferroni adjusted alpha levels of .017 per test (.05/3). As shown in Figure 6, the average number of gaze shifts "elsewhere" (i.e., avoiding the robot and the adults) was significantly higher in the experimental condition ($M = 1.17, SD = 0.68$) than in the control condition ($M = 0.25, SD = 0.43$). No significant differences were found in the average number of gaze shifts to the robot or to the adults between the two study groups.

Post-conversation Measures

When an adult asks a young child to make a value judgment there is always the risk that the child will answer with what s/he thinks the adult wants to hear. Our participants were handed "official" clipboards with their survey scales at the beginning of the Winter Games and told that we were counting on them as junior scientists to always tell us what they really thought. Looking across the use of the Smileyometer scale in all Winter Game activities, we find that all but one child used values in the lower, middle and upper part of the range, giving us confidence that the results reported in this section are a meaningful reflection of the children's differences with respect to condition.

Storytelling Judgment

When Jimmy asked for the child's opinion of its story during the conversation (see Table 1), all participants politely replied that it was a good story, independent of the study condition they were in. The comparable Smileyometer scale (top of Figure 4) was filled in after the interaction was over, in a non-social context, and an ANOVA was conducted to analyze the effects of study condition and age group on the children's judgment of Jimmy as a storyteller. We found a significant main effect of study condition, $F(1,24) = 5.55, p < .05, \eta^2 = .18$, with children in the control group rating Jimmy as a better storyteller than in the experimental condition, in contrast to children's judgment about the story itself, given in the presence of the robot. There was no significant main effect of age group, $F(1,24) = .23, p = .63, \eta^2 = .01$, nor in the interaction between condition and age group, $F(1,24) = .09, p = .78, \eta^2 = .01$.

Similarly, a significant main effect for study condition was found in the ANOVA that investigated the effects between condition and gender, $F(1,24) = 7.32, p < .05, \eta^2 = .22$. No significant main effect was found for gender, $F(1,24) = 1.8, p = .19, \eta^2 = .05$, nor for the interaction between study condition and gender, $F(1,24) = .88, p = .35, \eta^2 = .03$. The distribution of ratings by condition and age or gender, normalized by the total number of judgments, is shown in of Figure 8.

Robot Likeability

No significant main effect was found in the robot likeability measure with study condition as a factor, $F(1,24) = 2.74, p = .11, \eta^2 = .09$. Similarly, no significant main effect was found for age group, $F(1,24) = 3.3, p.08, \eta^2 = .11$, or in the interaction between condition and age group, $F(1,24) = .71, p = .71, \eta^2 = .0$. When considering gender as a factor, we also did not find any significant main effect of study condition, $F(1,24) = 1.44, p = .24, \eta^2 = .06$, gender, $F(1,24) = .09, p = .78, \eta^2 = .0$, or in the interaction between the study condition and gender, $F(1,24) = 0, p = .94, \eta^2 = .0$.

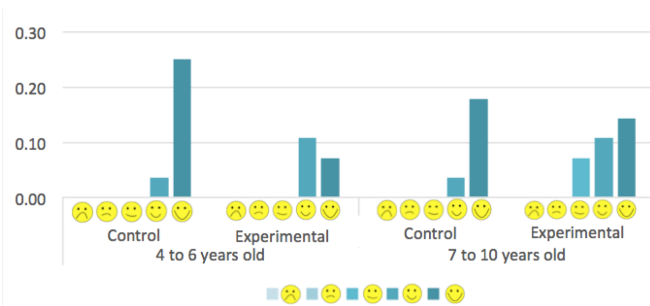
Willingness to Interact Again

The final self-report scale was answered as part of the nominally-valued Again Table. A chi-square test analyzing the effects of willingness to interact again by study condition yielded no significant result, $\chi^2(2) = 1.76, p = .41$. Similarly, no significant relationship was found between willingness to interact again and age group, $\chi^2(1) = 1.29, p = .23$, nor between willingness to interact again and gender, $\chi^2(1) = .15, p = .5$.

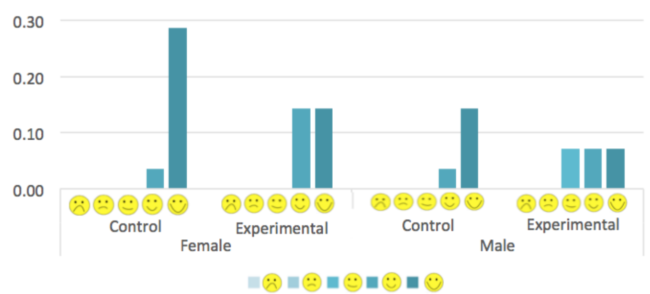
DISCUSSION

Our main goal in this pilot experiment was to explore two questions. First, whether children notice when a robot demonstrates in conversation that it has access to information they presume is private. Second, how they feel about that fact, both in the moment and at the end of the interaction. It is important to note that a meaningful analysis of the second question (Does it matter?) presupposes an affirmative answer to the first question (Did they notice?) – we want to be sure that if there is no difference in affective reaction between conditions it stems from not caring about the revelation rather than not noticing it.

Figure 7: Sample snapshots of children's facial expressions immediately after the critical moment in the control (top) and experimental conditions (bottom).



(a) Age group



(b) Gender

Figure 8: Children's responses to the question "How good a storyteller is Jimmy?" Each graph shows the percentage of children's responses for each Smiley face, normalized by the total number of judgments.

The cumulative evidence across all measures strongly argues that children in the experimental group did perceive the critical moment as a violation of expectations, but that their feelings about Jimmy's use of that private information depended on age and/or gender.

The experimental manipulation occurred when Jimmy named the two characters it would use in its story: characters already introduced into the conversation by the child (control) or hidden in the treasure chest the child was holding (experimental). At that moment, the accompanying adults of children in the control group reported significantly more positive ratings on the affective reaction of their children when compared to the parents in the experimental group. The type of words that parents used to describe their children's reactions echoes the divergent valence between conditions (see Figure 9).

We also found age and gender differences in the parents' report. While older children were rated as being in a more positive affective state in the control condition than in the experimental condition, we cannot make such strong claims about the difference in affective reactions of the younger children. These results are in line with previous research suggesting that younger children tend to inhibit expressions of affect more often in the presence of others, for fear of not conforming with social norms [18]. When considering gender, girls exhibited a wider range of affect (more positive in the control condition) while boys remained at the higher end of the scale in both conditions. This finding is also supported by prior work showing that girls, in general, are more likely to express negative emotions than boys [28].

Behavioral evidence also showed that, while the robot was telling his story, children in the experimental group shifted their attention away from both the robot and the adults significantly more than children in the control condition. In fact, the immediate reaction of most children in the experimental group

(a) Control condition (b) Experimental condition

Figure 9: Word clouds of the terms used by the accompanying adults to describe their children's reactions to Jimmy's story.

upon hearing the robot name the characters they picked was to look at the treasure box to make sure it was closed. One reason for avoiding subsequent eye contact might be that children felt they had done something "wrong" (such as inadvertently opening the treasure box) and were avoiding social contact as a sign of embarrassment.

The post-conversation measures indicate that children in the control group considered the robot a better storyteller than children in the experimental group, regardless of their age or gender. It is worth noting that despite the less positive reactions in the moment and a more critical judgment of the robot's storytelling abilities, there were no significant differences in how much children liked the robot, or whether they wanted to play with Jimmy again. It might have been the case that participants based their likeability and Again ratings on the overall interaction with Jimmy, and a single uncomfortable moment was not enough to affect their overall opinion.

CONCLUSION

While using sensor data collected outside a conversation can help robots to personalize interactions with children, that feature can also make them more vulnerable to violating conversational norms. In this paper, we conducted a pilot experiment to evaluate children's reaction when a robot exposes that it has knowledge about the child's actions that were not volunteered as part of the conversation. Upon noticing the violation of expectations, older children (7 to 10 years) visibly (to their parents) responded with more negative affect. Although younger children also noticed the conversational violation, they did not have the same degree of emotional reaction (or else expressed it in a less visible manner). Despite the immediate response, the post-conversation measures suggest that the reaction did not generalize much beyond the event, as children in both conditions provided similar ratings for robot likeability and willingness to interact with the robot again.

Our work has several limitations that we plan to address in the future. The conversation with the robot was short and contained only a single revelation of private information. It is unclear how the trends seen here might change if the robot violated this norm multiple times in a single conversation or across repeated interactions. The type of information disclosed might also be critical to the way children react. In this study, Jimmy disclosed a fact that had minimal personal importance to the child, but what if the robot revealed it knew something the child cared about?

Finally, given the modest sample size and children's inability to articulate their internal state, it was beyond the scope of this work to speculate about which social construct was violated from the child's point of view. Those who reacted negatively may have conceptualized the issue as one of fairness, trustworthiness or misbehavior, or as more than one of these, depending on age and/or gender. Future work is needed to explore this question, which is becoming more relevant as persistent information gathered by passive technologies becomes ubiquitous.

SELECTION AND PARTICIPATION OF CHILDREN

We recruited 28 children (17 females and 11 males) from Allegheny County, Pennsylvania, USA, where the population is approximately 1.2 million. The study was advertised through postings in physical and online community bulletin boards. Children were compensated for their participation. The research was approved by the Human Subjects Institutional Review Board of Carnegie Mellon University.

ACKNOWLEDGMENTS

This work was funded by Disney Research. We thank our participants and their families, and also André Pereira, Albert Li, Nikolas Wolfe, Jo Ana Vaz, Kyna McIntosh, Mo Mahler and Eric Brockmeyer.

REFERENCES

1. Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems*. Springer, 114–130.

2. Luigi Atzori, Antonio Iera, and Giacomo Morabito. 2010. The internet of things: A survey. *Computer networks* 54, 15 (2010), 2787–2805.

3. Simon Baron-Cohen. 1997. *Mindblindness: An essay on autism and theory of mind*. MIT press.

4. Cindy L Bethel, Matthew R Stevenson, and Brian Scassellati. 2011. Secret-sharing: Interactions between a child, robot, and adult. In *Proceedings of the IEEE International Conference on Systems, man, and cybernetics (SMC)*. IEEE, 2489–2494.

5. Nadia Chernyak and Heather E Gary. 2014. Autonomous Movement Predicts Children's Moral Regard and Prosocial Behavior Towards a Social Robot Dog. In *Proceedings of 36th annual meeting of the Cognitive Science Society*. 2032–2037.

6. H.H. Clark. 2006. Context and Common Ground. In *Encyclopedia of Language & Linguistics (Second Edition)* (second edition ed.), Keith Brown (Ed.). Elsevier, Oxford, 105 – 108. http://www.sciencedirect.com/science/article/pii/B0080448542010889

7. Alison Gopnik and Janet W Astington. 1988. Children's understanding of representational change and its relation

to the understanding of false belief and the appearance-reality distinction. *Child development* (1988), 26–37.

8. Peter H Kahn, Heather E Gary, and Solace Shen. 2013. Children's Social Relationships With Current and Near-Future Robots. *Child Development Perspectives* 7, 1 (2013), 32–37.

9. Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G Freier, Rachel L Severson, Brian T Gill, Jolina H Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology* 48, 2 (2012), 303.

10. Vera Khovanskaya, Eric P.S. Baumer, Dan Cosley, Stephen Voida, and Geri Gay. 2013. "Everybody Knows What You'Re Doing": A Critical Design Approach to Personal Informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3403–3412. DOI: http://dx.doi.org/10.1145/2470654.2466467

11. Alfred Kobsa, Hichang Cho, and Bart P Knijnenburg. 2016. The effect of personalization provider characteristics on privacy attitudes and behaviors: An Elaboration Likelihood Model approach. *Journal of the Association for Information Science and Technology* (2016).

12. Séverin Lemaignan, Julia Fink, Francesco Mondada, and Piérre Dillenbourg. 2015. You're Doing It Wrong! Studying Unexpected Behaviors in Child-Robot Interaction. In *Social Robotics*, Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Lecture Notes in Computer Science, Vol. 9388. Springer International Publishing, 390–400. http://dx.doi.org/10.1007/978-3-319-25554-5_39

13. Vanessa LoBue, Tracy Nishida, Cynthia Chiong, Judy S DeLoache, and Jonathan Haidt. 2011. When getting something good is bad: Even three-year-olds react to inequality. *Social Development* 20, 1 (2011), 154–170.

14. Omar Mubin and Christoph Bartneck. 2015. Do as I say: Exploring human response to a predictable and unpredictable robot. In *Proceedings of the 2015 British HCI Conference*. ACM, 110–116.

15. Marina Nazario. 2015. This talking dinosaur for kids was named one of the best inventions of the year. (8 December 2015). Retrieved January 21, 2016 from http://www.businessinsider.com/cognitoys-talking-dinosaur-2015-12.

16. Jean Piaget. 1997. *The moral judgement of the child*. Simon and Schuster.

17. Janet C. Read and Stuart MacFarlane. 2006. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children (IDC '06)*. ACM, New York, NY, USA, 81–88. http://doi.acm.org/10.1145/1139073.1139096

18. Carolyn Saarni. 1988. Children's understanding of the interpersonal consequences of dissemblance of nonverbal emotional-expressive behavior. *Journal of Nonverbal Behavior* 12, 4 (1988), 275–294.

19. E. Short, J. Hart, M. Vu, and B. Scassellati. 2010. No fair!! An interaction with a cheating robot. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 219–226.

20. Juliana Sutanto, Elia Palme, Chuan-Hoo Tan, and Chee Wei Phang. 2013. Addressing the personalization-privacy paradox: an empirical assessment from a field experiment on smartphone users. *Mis Quarterly* 37, 4 (2013), 1141–1164.

21. Fumihide Tanaka, Aaron Cicourel, and Javier R Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences* 104, 46 (2007), 17954–17958.

22. Daniel Ullman, Iolanda Leite, Jonathan Phillips, Julia Kim-Cohen, and Brian Scassellati. 2014. Smart human, smarter robot: How cheating affects perceptions of social agency. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci2014)*.

23. Marynel Vazquez, Alexander May, Aaron Steinfeld, and Wei-Hsuan Chen. 2011. A deceptive robot referee in a multiplayer gaming environment. In *Proceedings of the 2011 International Conference on Collaboration Technologies and Systems (CTS)*. 204–211.

24. E Fosch Villaronga and J Albo-Canalsb. 2015. Boundaries in Play-based Cloud-companion-mediated Robotic Therapies: From Deception to Privacy Concerns. In *Proceedings of the 1st International Conference on Social Robots in Therapy and Education (New Friends)*.

25. James Vlahos. 2015. Barbie Wants to Get to Know Your Child. (15 September 2015). Retrieved January 21, 2016 from http://www.nytimes.com/2015/09/20/magazine/barbie-wants-to-get-to-know-your-child.html?smtyp=cur&_r=1.

26. Jacqueline Kory Westlund and Cynthia Breazeal. 2015. Deception, Secrets, Children, and Robots: What's Acceptable? *Workshop on The Emerging Policy and Ethics of Human-Robot Interaction* (2015).

27. J.P. Whitney, T. Chen, J. Mars, and J.K. Hodgins. 2016. A hybrid hydrostatic transmission and human-safe haptic telepresence robot. In *2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden (to appear)*. IEEE.

28. Janice Zeman and Judy Garber. 1996. Display rules for anger, sadness, and pain: It depends on who is watching. *Child development* 67, 3 (1996), 957–973.