What? That's Not a Chair!: How Robot Informational Errors Affect Children's Trust Towards Robots

Denise Y. Geiskkovitch
Department of Computer Science
Department of Psychology
University of Manitoba
Winnipeg, Canada
d.geiskkovitch@cs.umanitoba.ca

Raquel Thiessen
Department of Computer Science
University of Manitoba
Winnipeg, Canada
thiess83@myumanitoba.ca

James E. Young
Department of Computer Science
University of Manitoba
Winnipeg, Canada
young@cs.umanitoba.ca

Melanie R. Glenwright
Department of Psychology
University of Manitoba
Winnipeg, Canada
melanie.glenwright@umanitoba.ca

***Abstract*— Robots that interact with children are becoming more common in places such as child care and hospital environments. While such robots may mistakenly provide nonsensical information, or have mechanical malfunctions, we know little of how these robot errors are perceived by children, and how they impact trust. This is particularly important when robots provide children with information or instructions, such as in education or health care. Drawing inspiration from established psychology literature investigating how children trust entities who teach or provide them with information (informants), we designed and conducted an experiment to examine how robot errors affect how young children (3-5 years old) trust robots. Our results suggest that children utilize their understanding of people to develop their perceptions of robots, and use this to determine how to interact with robots. Specifically, we found that children developed their trust model of a robot based on the robot's previous errors, similar to how they would for a person. We however failed to replicate other prior findings with robots. Our results provide insight into how children as young as 3 years old might perceive robot errors and develop trust.**

Keywords—child-robot interaction; trust; robot errors.

I. Introduction

As social robotics continues to grow and develop, robots are increasingly finding their way into more areas of society, including those that require interacting with children. Social robots can now be found in hospital rooms, where they help children feel better and heal [1], in homes, with some robots specifically designed to interact with children [2], and in daycare centers and schools, where they act as teaching assistants and companions [3]. The study of Child-Robot Interaction has emerged, with an emphasis on understanding how children may perceive and interact with robots. For example, research has looked at the types of characteristics that children find important in robots [4], and whether robots can serve as tutors, helping children learn classroom material (e.g., [5], [6]). We contribute to this emerging field by exploring how young children develop their trust of robots that provide them with information.

In many potential interaction scenarios robots will be providing children with information or instructions; in some cases, it may be important for children to trust the information that robots provide, such as in education or safety situations. We know from psychology that children as young as three years old develop a theory of mind of adults relating to trust, and use a range of factors (e.g., familiarity, age, previous errors) in deciding whether to trust information coming from an adult (e.g., [7]–[9]). Work has also demonstrated that children apply similar models of trust to puppets [10]. However, robots are different from puppets, as they create a strong sense of agency and anthropomorphism which shapes interaction [11]. Due to this anthropomorphism, robots (unlike puppets) tend to be perceived as more life-like, and children specifically attribute them with mental and social capabilities (e.g.,[12], [13]). Further, some research has suggested that people may expect robots to have perfect or elevated reasoning abilities, memory, and cognitive performance [14], potentially leading to different expectations than from puppets or even humans. However, we do not yet know how children perceive robots, and to what extent and in what form they will apply their trust models (of adults) to robots. Our work provides insight into this topic. Specifically, it provides insight into how young children perceive robots, as opposed to adults or puppets, helping us better understand how children will trust robots.

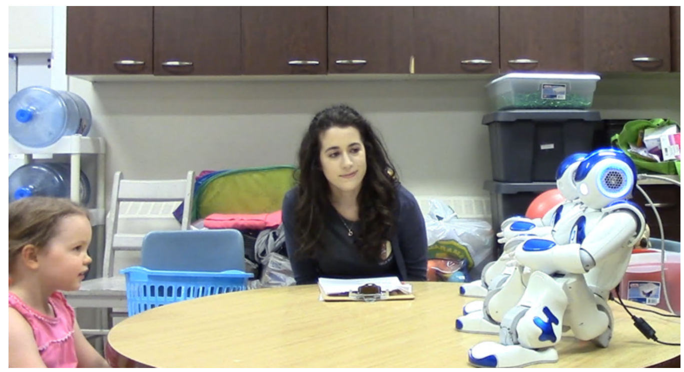


Fig. 1. Child (left) interacting with two robots (right), one of which makes errors. Child's picture used with permission.

Specifically, we investigate the impact of a robot making a factual mistake on how children trust the robot. This follows a body of investigation into how robot errors in general affect interaction with people, including work detailing how an erroneous or faulty robot can impact how people perceive or trust it. However, prior trust work with robots has primarily focused on the impact on adults. As children view robots differently than adults do [15], we investigate them specifically. We contribute to this body of work by exploring how young children (3-5 years old) may perceive robot errors, and how these errors may affect children's trust towards robots that provide them with information.

Funded by the Natural Sciences and Engineering Research Council of Canada.

We focused on young children (3-5 years old) as previous research has found that, from age 3, children begin to develop theory of mind – the ability to attribute mental states to themselves and others [16] – including epistemic mental states (i.e., knowledge and beliefs) of others. For that reason, this age group has been widely studied in both Psychology and Human-Robot Interaction (HRI; e.g., [10], [17]–[22]). As children develop theory of mind, it enables them to take into account another's mental state (human or perhaps non-human) and prior knowledge, to inform interaction, such as when deciding whether to trust an informant (i.e., individuals that provide information; e.g., [9]). However, we know little of how children's developing theory of mind affects how they perceive robots, and how this may shape how they trust them.

We conducted an experiment to examine how children's trust towards a robot is affected by whether the robot makes informational errors, providing insight into young children's perceptions of robots. Our experiment had three phases: a history phase, where a child observes a robot making errors and another one not making errors, and two testing phases, where children had to make decisions about which robot to trust in a range of scenarios. Our results indicate that children do attribute trust to robots similarly to how they do people, with children trusting a previously correct robot more than a previously incorrect one. However, this was only the case in some scenarios, while in others we found some possible discrepancies with prior work on how children trust people and puppets. We also found initial indications of a potential gender effect (girls and boys may trust robots differently) not found in prior work, that requires further exploration. In the remainder of this paper we detail our experiment and findings, and provide a discussion on why we believe we may have observed these findings.

II. Related Work

A number of studies in HRI have examined how robot errors are perceived by adults, and how they affect the interaction between a person and a robot. Ragni et al. [14] suggested that because of expectations that we have of robots, if a robot makes a mistake, it might severely impact how it is perceived and interacted with. Through an experiment, they found that participants that interacted with an erroneous robot perceived it as less intelligent, reliable, and competent than those that interacted with a non-erroneous robot. This suggests that robotic errors may lead to more negative perceptions of a robot.

Other research has looked at how robot errors can be perceived as intentionally cheating. Short et al. [23] conducted an experiment in which participants played a game of Rock Paper Scissors with a robot that sometimes either reported the results incorrectly (i.e., saying that it had won when it had not), or changed its answer after seeing the participant's. They found that participants perceived the incorrect robot as making an error, while the robot that changed its answer was perceived as cheating. This suggests that factors such as the timing of an error, and context in which it occurs, can influence the perception of a robot and its errors. While research has been conducted to examine how adults perceive robot errors, there is a lack of similar research that focuses on children. This is especially problematic because children tend to perceive robots differently than adults do, often believing them to be alive, and interacting with them as social beings [15]. Therefore, robot errors may have a different effect on children's perceptions and trust towards robots than they do on adults.

In Child-Robot Interaction, research has looked at various aspects of how children interact with robots. Researchers, for example, have studied what features robots should have to best interact with children, finding that some robot characteristics (e.g., head and eyes, gestures) are more important than others [4], [24]. Another area of research within Child-Robot Interaction looks at different methods for conducting such research. For example, Charisi et al., [25] found that studies investigate a variety of outcomes such as children's views of robots, and their emotional involvement with robots, while other researchers have explored data collection methods to observe children's interactions with robots, such as longitudinal studies [26], [27], and field studies [28], [29]. Although research in Child-Robot Interaction is quite varied, as witnessed above, robot errors with children have yet to be examined. In our research we provide a new experimental methodology within our field for conducting Child-Robot Interaction experiments, specifically investigating trust towards robots that make errors.

The inspiration for our work comes from psychological research that investigated children's trust towards informants (i.e., individuals that provide some sort of information). Several experiments have looked at children's trust towards informants when one of the informants provides incorrect information (i.e., makes errors; e.g., [9], [10], [17]–[20], [30]). In addition, researchers have also examined how different characteristics of an informant, such as how familiar they are to the child or whether they were previously dishonest, can have an effect on how much children trust them (e.g., [7], [31]). The inspiration for our research stems from an experiment that investigated whether young children trusted two puppets differently based on the quality of information they had previously provided [10]. Researchers conducted an experiment in which 3- and 4-year old children observed two puppets name objects. During a history phase, one of the puppets provided correct object labels, while the other one did not. Two testing phases then assessed children's trust towards the robots. The researchers found that even when the children were not familiar with the labeled objects, they were more likely to trust the puppet that had previously labeled common objects correctly than the one that had not. This experimental methodology has become a standard for investigating children's trust towards informants; we conduct a follow-up variant focusing on trust towards robots.

III. Background: Trust

Trust is a complex construct and has been explored in a variety of fields and scenarios, including nursing, medicine, sociology, psychology, and even HRI, with each discipline utilizing a different definition of trust [32]. Within psychology, one core focus is interpersonal trust, which is defined as "an expectancy held by an individual or a group that the word, promise, verbal, or written statement of another individual or group can be relied on" [33]. This type of trust has been explored with children through experiments on trust towards informants, such as those mentioned above. For example, Corriveau and Harris [7] investigated whether children trust information provided by a familiar informant more than information provided by an unfamiliar one. They recruited 3- and 4-year old children from two different schools, and one teacher from each school to be the familiar and unfamiliar informants. During the experiment, children watched videos in which the two informants labeled, and demonstrated the use of, familiar and novel objects. Children were then asked to indicate which of the provided labels and uses were correct, and which informant had been better at answering the questions. The researchers found that children trusted the answers provided by the familiar informant (the teacher from their school) more than those provided by the unfamiliar informant. This research provides insight into young children's developing theory of mind affects how children see informants, and how children's perceptions of informants are affected by factors such as familiarity.

Other research in Psychology has looked at whether children trust informants differently depending on whether the informant is a child or an adult. VanderBorght and Jaswal [8], for example, found that children aged 3-5 years old trusted adults with information about food and nutrition, but trusted other children with information about toys. Another experiment comparing adult and child informants found that children trusted an adult informant, over a child, when either both informants (child and adult), or just the adult informant, were reliable (had previously answered a question correctly). However, children trusted the child informant when it was the only one that had been reliable [19]. These findings suggest that children apply selective trust towards informants, which may affect how they trust robots. For example, trust towards robots may also differ depending on the type of information that they provide, or different attributes they are given (e.g., child versus adult robot).

Trust has also been explored in a variety of contexts within HRI, where researchers have examined whether people trust robots, and the types of attributes that may affect such trust. Salem et al., [34], for example, looked at how adult participants perceived a robot that moved irrationally, remembered things incorrectly, and made unusual requests, in comparison to one that behaved normally. They found that while participants reported trusting the faulty robot less than the non-faulty one, they still completed the unusual requests that the faulty robot made, such as pouring orange juice into a plant. This suggests that even if a robot appears to be malfunctioning, people may still follow its unusual instructions. Other researchers have found additional factors that affect trust towards robots, and their related outcomes. For example: *i*) people express more trust towards a vulnerable robot than a non-vulnerable one [35], *ii*) a robot is more trustworthy when it attempts to fix a mistake it has made, and when it is able to communicate through speech [36], and *iii*) people will over-trust a robot and allow it to enter a secured facility, even when it could be dangerous [37]. Trust has become an invaluable area of research within HRI, providing insight into the types of behaviors that robots should and should not possess, and the potential outcomes that are tied to such behaviors. While we have gained much knowledge in this area, further exploration is required, especially as it relates to robot errors, and how such errors may affect children's trust.

Children's trust towards technology has also been explored in a variety of contexts. For example, Danovitch and Alzahabi [21] explored how children trust two computer informants (not robots) differently depending on whether they have provided correct or incorrect information about familiar objects, similarly to the Psychology research previously discussed. The researchers found that when young children (3-5 years old) needed information about novel objects, they relied on the computer agent that had previously provided correct information, rather than one that had not. This implies that young children also develop mental models of computer informants' states, and not just humans or puppets. While these findings may hint that this would also apply to robots, robots are inherently different from computers, and possess a sense of agency that computers do not have [11]. Therefore, it is important to further explore whether children trust robot informants differently due to errors.

Additional research has also suggested that children not only view puppets and computers as informants, but also robots, and endorse (i.e., trust) a robot that provides information in a timely manner more than one that does not [22]. However, research in Child-Robot Interaction has yet to explore how children perceive robot informants that make errors. In our work, we therefore further explore children's trust towards robot informants, and how it is influenced by errors that a robot makes.

With previous literature in mind, the goal of our research is to examine how children's trust towards robot informants is affected by informational errors made by robots. To that end, we designed and conducted an experiment closely related to the literature discussed above that examined whether children trust two robots differently depending on whether the robots had previously made informational errors. This research is a first step towards exploring how robot errors affect children's perceptions of robots, and in turn how this affects their trust.

IV. Experiment: Robot Informational Errors and Children's Trust

We designed and conducted an experiment, with components and phases taken from prior similar work [10], to investigate how children's trust towards robots is influenced by robot informational errors. Our experiment had three phases: *history*, *same label*, and *contrast label* phases, as well as a *clean-up* task (Figure 2). The *history*, *same label*, and *contrast label* phases were obtained from Birch et al. [10], and modified to be conducted with robots (e.g., using made-up words that robots could pronounce) and for additional statistical power (i.e., including 4 trials in each phase instead of 3). In addition, we added the *clean-up* task in order to test whether robot informational errors also affect children's behavior, in addition to trust. In our experiment, a child first observed two robots label objects familiar to children (e.g., a ball, a car), such that one of the robots consistently labeled objects incorrectly (i.e., made errors) while the other robot consistently labeled them correctly (Figure 4a). The goal of the experiment was to explore how children trust these two robots differently, based on one making errors and one not. Thus, our independent variable was whether the robot made errors or not, presented during the *history* phase. Following, we investigated the impact of this manipulation on children's future trust of information provided by the robots from two angles: in cognitively simple and complex situations. We further included a task where the robots gave instructions to the child, and then the child had to decide which robot's directions to comply with. Below we describe our methodology in detail, including the different phases and how they took place.

A. Robots

We used two Aldebaran Nao H25 V5 robots for this experiment (Figure 1). They are 23 inches high, and widely used in HRI research with children. Both robots had the same white and blue design, and behaved and sounded the same except for the information provided. We gave the robots gender neutral names (Casey and Taylor) to reduce the likelihood of children perceiving the robots as having the same or different gender as them. The robots were controlled by a research assistant, who was present in the study room, through a Wizard of Oz interface.

B. Materials

During the experiment, the robots named objects placed in front of the participant (i.e., child). In total, the robots named 20 objects across the phases, 4 of which were familiar to the children and 16 which were unfamiliar to the children (see Table 1 for the full list of objects). We accept that some children may recognize some of the unfamiliar items, but we intended for the names of the objects to be unknown. Most objects were taken from prior similar experiments (e.g., [10]).

C. Experiment Phases and Tasks

Our experiment was comprised of three phases, presented in order: *history*, *same label*, and *contrast label*, with the addition of a *clean-up* task (Figure 2, Figure 4a).

1) History Phase: The experiment started with this phase, which served as our independent variable presentation, demonstrating to children which robot was correct and which one made errors. During this phase children observed the two robots label common objects, such as a ball. One of the robots (the previously correct robot) provided correct labels for the objects, while the other robot (previously incorrect robot) provided incorrect labels (e.g., labelling a car as a book; see Figure 4a).

2) Same Label Phase: The goal of this phase was to test children's trust towards the two robots when information is simple and clear. During this phase, the robots labeled uncommon objects, providing made-up names for them (e.g., ferber; see in Table 2 for full list of labels). Some of the made-up words were obtained from previous related literature from psychology, and others were created for this experiment. All words were comprised of two syllables and could be easily understood by children.

The two robots provided the same made-up name (e.g., ferber) for two different unfamiliar objects. The experimenter then asked the child to grab the object that matched the name (e.g., ferber; Figure 4b). This tested which robot the child trusted to know what the label (e.g., ferber) represented. We repeated this process four times, each time with different unfamiliar objects and made-up names (Figure 2, Figure 4b, 8 objects total, 4 labels). As the child handed the objects to the researcher they were placed in a blue basket on the side of the table. The first hypothesis was as follows:

H1: Children will trust the previously correct robot more than the previously incorrect one, and will therefore choose the object that the previously correct robot had named, when asked to hand an object name that they had previously heard (i.e., during the *same label* phase).

3) Contrast Label Phase: The contrast phase was meant to test children's trust towards robots in a more complex situation than the same label phase, with some ambiguity introduced.

The robots labeled two uncommon objects just as in the same label phase. However, the experimenter asked the child to pass an object with a new made-up label, which had not been provided by the robots. For example, when both robots labelled the objects gilly, the experimenter asked for cheena (see Figure 4c). The logic is that, if the child trusts the previously correct robot regarding what a gilly is, then they would pass the object labelled by the previously incorrect robot. We repeated this process four times, each time with different unfamiliar objects and a made-up name, and a different request label (Figure 2, Figure 4c, 8 objects total, 4 object labels, 4 request labels). As the child handed the objects to the researcher they were placed in a blue basket on the side of the table. The second hypothesis was therefore the following:

H2: Children will trust the previously correct robot more than the previously incorrect one, and will therefore choose the object that the previously incorrect robot had named, when asked to hand an object name that they had not previously heard (i.e., during the *contrast label* phase).

4) Clean-up Task: We introduced a new component, a clean-up task, which was not explored in previous trust research. This was motivated by HRI work showing how adults can sometimes follow unusual instructions from faulty robots [34]. We aimed to explore the relationship between children's trust toward robots and compliance to robots' instructions.

Between the *same label* and *contrast label* phases, after all the objects had been labelled, the researcher placed two coloured pieces of paper (red, white) on the table, one in front of each robot (Figure 4d), and informed the child that it was time to clean up. The robots then provided conflicting instructions: one of them said to put the blue basket of toys on the red paper, while the other said to put it on the white paper. The experimenter then asked the child where to place the basket, and then placed it there for them. We administered this task twice, once after the *same label* phase, and after the *contrast label* one (Figure 2). The third hypothesis was as follows:

H3: Children will trust the previously correct robot more than the previously incorrect one, and will therefore place the

TABLE I. OBJECTS SHOWN TO CHILDREN AND LABELED BY THE ROBOTS DURING THE EXPERIMENT. WE EXPECT THAT CHILDREN DO NOT KNOW THE NAMES OF THE UNFAMILIAR OBJECTS.

Familiar Objects	Unfamiliar Objects	
Ball	Turkey baster	Garlic press
Car	Peach slicer	Pineapple slicer
Dog	GPS/phone car holder	Tape dispenser
Plate	Honey dipper	Lobster cracker
	Faucet aerator	Towel holder
	Dog toy	Measuring spoon
	Door stop	Door spring

TABLE II. MADE-UP WORDS USED BY THE ROBOTS AND EXPERIMENTER DURING EXPERIMENT (COMPLETE LIST).

Labels used by robots		Labels used by experimenter in Contrast Phase
Ferber	Koba	Modi
Turly	Gilly	Cheena
Jeebus	Mizule	Claster
Plakil	Gleblu	Maloo

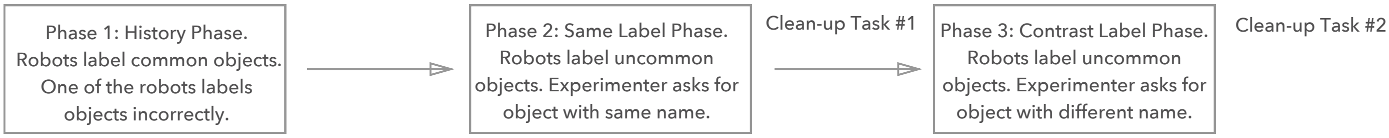


Fig. 2. Flow chart of experimental procedure.

blue basket with the objects where the previously correct robot indicated during the *clean-up* tasks.

D. Participants

We performed an a-priori sample size analysis using G*Power, with the following parameters taken from previous relevant literature (e.g., [7]): d = .60, α = .05, and power = .80, for a single-sample t-test comparing our findings against expected chance. The analysis yielded a preferred sample size of 19. We recruited 21 children aged 3-5 from a public daycare in our city, and their parents provided consent before the study took place. Children were able to pick a small toy to thank them for their participation, and parents received a $15 honorarium.

E. Procedure

Participants were escorted into a room at the daycare by a daycare worker, and were introduced to the experimenter. The daycare worker remained in the room for the entire session, but was not in the child's immediate field of view. The child sat at a table, and the experimenter introduced the child to the two robots, who were sitting on the table (see Figure 3). The experimenter informed the participant that they would be observing the two robots name objects, and asked the child if they would like to do that (verbal consent from child). The robots then introduced themselves to the child, and the experimenter informed the participant that they could stop the study at any time.

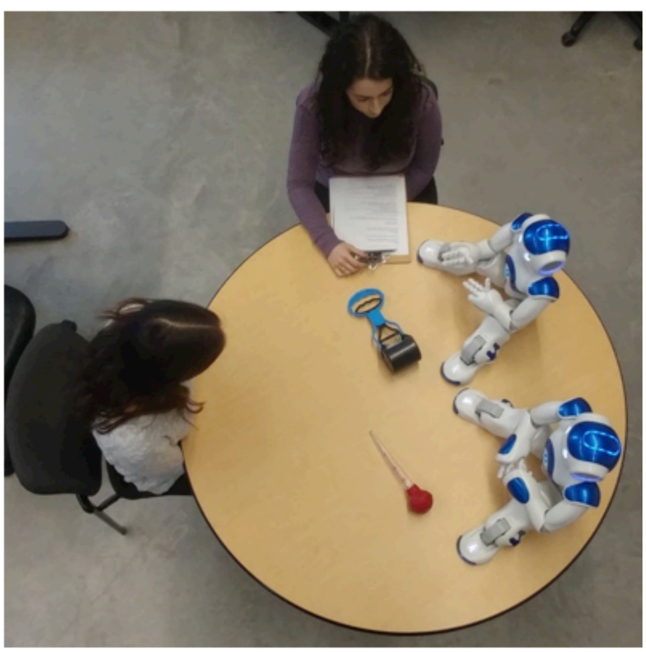


Fig. 3. Top-view of experiment set-up. Child, robots, and experimenter sat at a table. Two robots sat on the table (right), across from the child (left). Experimenter sat perpendicularly between the child and the robots (top).

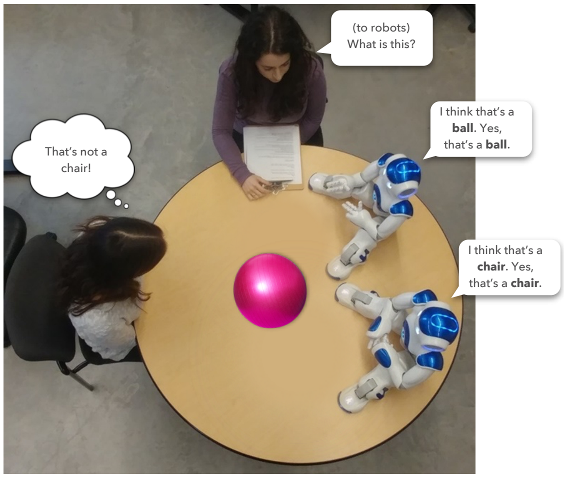
The experimenter administered the three experiment phases, and the clean-up tasks (as in Figure 2): first the *history* phase, then the *same label* phase, then the first *clean-up* task, followed by the *contrast label* phase, then the second *clean-up* task. Children were also asked to label the familiar objects from the *history* phase to ensure that they had been properly exposed to the independent variable. This was a within-subjects experiment, with the placement of the robots (right versus left), and the order in which they named objects being counterbalanced between participants. Therefore, there were four counterbalancing conditions: speaking order (Casey vs. Taylor), and seating order (Right vs. Left).

V. Results

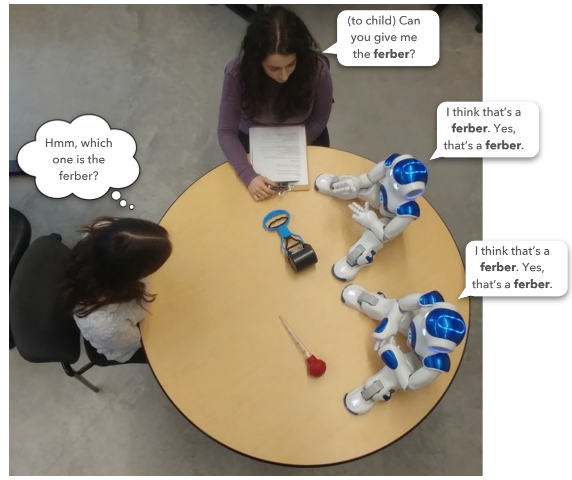
Out of the 21 participants that were recruited, 2 children were too shy to interact with the robots, 1 child was afraid of the robot, and 1 child did not follow task directions (in every trial they gave both objects to the experimenter instead of picking one). Therefore, 17 children (10 male, 7 female), aged 3-5 (M = 3.65, SD = .61, 7 3-year olds, 9 4-year olds, 1 5-year old) participated in this experiment. For analysis we combined children's actions for each phase into a percentage of how often they sided with the previously-correct robot, following established analysis techniques for this study design (e.g., [10]). Thus, scores closer to 100% mean that the child sided with the previously correct robot more, and scores closer to 0% indicate they sided with the previously incorrect robot most often. For the *same-label* phase, a one-sample t-test comparing children's actions against expected chance (of 50%, similar to previous work [10]), revealed a statistically significant difference between which robot's object children handed to the experimenter (t_{16} = 2.640, p = .018, d = .640). Children trusted the previously correct robot more than the previously incorrect one (M = .706, SD = .322, Figure 5).

For the *contrast-label* phase (when the experimenter asked for an object with a different name than what the robots had said) we did not find a statistically significant difference between which robot was trusted more (t_{16} = -.753, p = .463, M = .427, SD = .403; see Figure 6). Post-hoc, we conducted an ANOVA with gender as a between-subjects factor. Results suggest a potential gender effect for the *contrast label* phase ($F_{1,15}$ = 13.840, p = .002), with girls more often siding with the previously correct robot (M = .750, SD = .289), and boys picking the previously incorrect robot's answer (M = .200, SD = .307; see Figure 7). However, we note that this splits the data into smaller sample sizes, and so further inquiry is required.

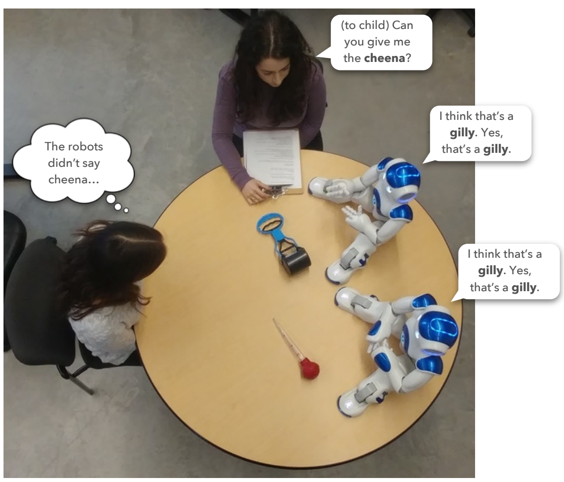
For the *clean-up* tasks, we did not find a statistically significant difference of which robot children complied with (clean-up 1: χ^2 = .529, p = .467, clean-up 2: χ^2 = 2.288, p = .090; see Table 3). No other analyses, such as age, robot placement, or robot speaker order, were significant.



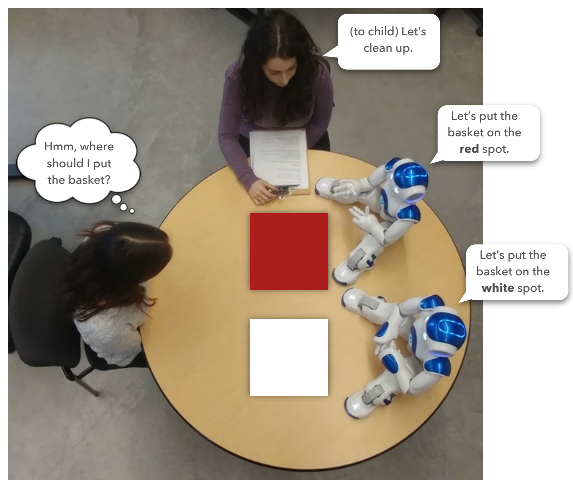
b) History Phase



a) Same Label Phase



c) Contrast Label Phase



d) Clean-up Task

Fig. 4. Experimental phases, including the experimenter's and robot's scripted speech.

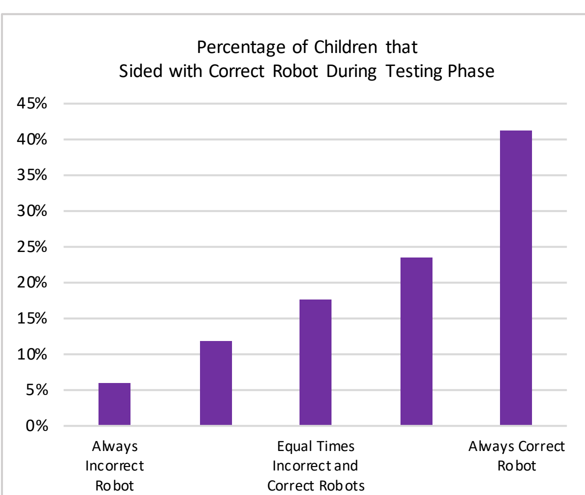


Fig. 5. Same Label Phase: percentage of children that sided with the previously correct robot during *same label* phase. We expected children to trust, and therefore side with, the previously correct robot (i.e., the one that did not make errors).

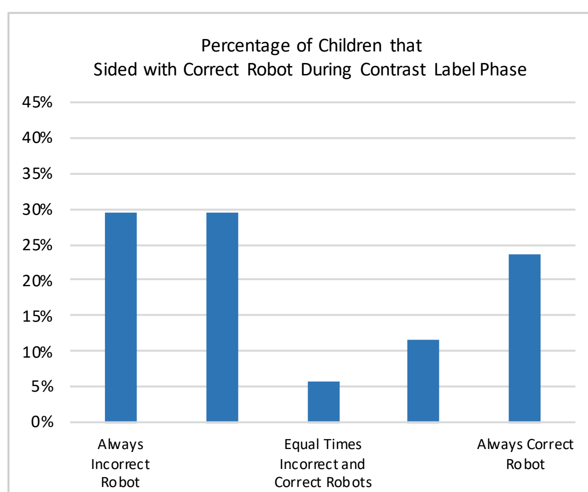


Fig. 6. Contrast Label Phase: Percentage of children that sided with the previously correct robot during *contrast label* phase. We expected children to trust the previously correct robot, and therefore side with the previously incorrect robot (i.e., the one that had made errors).

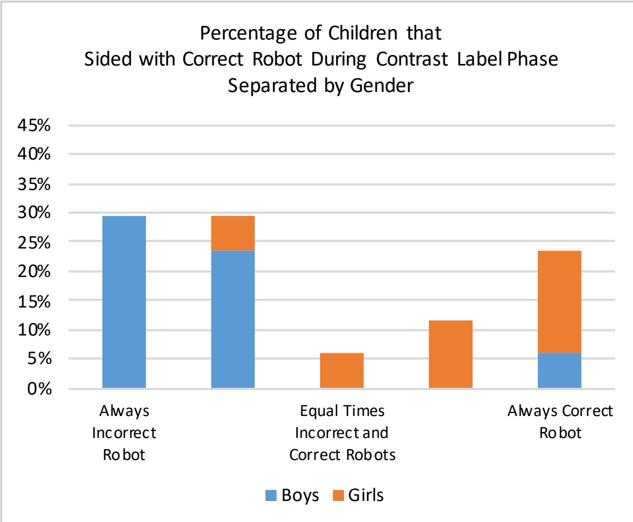


Fig. 7. Percentage of children that sided with the previously correct robot during *contrast label* phase, divided by gender.

TABLE III. PERCENTAGE OF CHILDREN WHO COMPLIED WITH THE PREVIOUSLY CORRECT AND INCORRECT ROBOTS DURING CLEAN-UP TASKS.

	Previously Correct	Previously Incorrect
Clean-up 1	47.5%	52.9%
Clean-up 2	23.5%	76.5%

VI. DISCUSSION

Our results demonstrate that children do indeed build their trust models of a robot based on their previous experience with that robot, in a pattern consistent with how children trust people. This provides insight into how children perceive robots, and how they attribute mental states to robots. However, we were not able replicate prior findings in all cases, and in addition, found a potential gender effect. This suggests that there might be outcomes specific to robots, requiring ongoing study. In the remainder of this section we discuss the results in greater detail.

A. Same Label Phase

In our experiment, we found that errors that a robot had made during the *history* phase had an effect on how much children trusted it during the *same label* phase. Children sided with the robot that had not made errors more often than the one that had. This supports our first hypothesis, and aligns with previous work that found that children expressed more trust towards a previously correct puppet, than an incorrect one [10]. This also implies that children were properly exposed to our manipulation. Our findings suggest that prior errors that a robot makes could have an effect on how much it is trusted afterwards, at least in the short term, but further investigation is necessary in more externally valid situations.

These findings highlight the impact that robots' informational errors can have and call for further inquiry into how such errors may affect children's perceptions of robots. Scenarios in which robots are in positions of authority, and children are required to listen or follow their directions (e.g., daycares, schools, hospitals), are of particular importance. For example, a daycare robot that continuously calls a child by an incorrect name for a period of time, may lose that child's trust. The child might therefore not follow its directions, even when they are important. Alternatively, a hospital robot might make a mistake while playing with a child, causing the child to not follow its safety instructions thereafter (e.g., to not jump out of bed). If this is the case, we need to be aware that errors that robots make, such as providing incorrect or nonsensical information, could greatly impact children's future trust.

B. Contrast Label Phase

We failed to replicate the previous finding by Birch et al., [10] that children will continue to trust a previously correct informant, in a more complex situation. We did not find that children continued to trust the previously correct robot during the *contrast label* phase, when children were asked for an object with a name the robots had not mentioned. We expected children to trust the robot that had not made errors, and therefore pick the object labeled by the one that had. Our second hypothesis, that children would trust the robot that had not made errors, and thus side with the one that had, was therefore not supported.

These findings may have occurred for a number of reasons. For example, it may be that interactions with robots are more complex than interactions with puppets, with multiple mechanics taking place. Prior work has found that children view robots differently than humans or even computers, attributing them with feelings and liveliness [15]. Children's cognitive load while interacting with robots (which are arguably also more novel and uncommon than puppets) may therefore be higher than when interacting with puppets, especially during a more cognitively demanding task. It is also possible that throughout the experiment, the novelty and agency of the robots was distracting or too attention-grabbing for the children, and they forgot which robot had made errors. Further work is necessary to explore whether this was the case, and additional variables that may have played a role.

Our experimental procedure contained some deviations from previous research that may have also led to different results, mainly the addition of the *clean-up* task. One of the iterations of the *clean-up* task took place between the *same label* and *contrast label* phases, and may have led to children forgetting which robot had made errors. Further, due to the break in the phases, children may have interpreted the *contrast label* phase as a blank canvas for robot information.

In our experiment, the two robots were identical in terms of how they looked and talked, with the only differentiators being their placement (right or left) and whether they had provided correct information during the *history* phase. In previous research, puppet informants used for similar experiments have not been as similar, with the puppets' genders [10] or animal species [30] serving as visual differentiators. While we intentionally had the robot informants be as similar as possible, to reduce bias towards a particular informant, it may be that children in our experiment had trouble differentiating between the two robots, leading to our second hypothesis not being supported. However, we did find that children trusted the robots differently during the *same label* phase of the experiment, suggesting that children were able to differentiate between the two robot informants, at least during that particular phase.

Alas, we are not the only researchers who failed to find a difference in more complex scenarios (see, e.g., [10], similar to our *contrast label* phase). The researchers stated that because they had asked children about the uses of objects, as opposed to names, the task was more cognitively advanced and difficult for children. It therefore may be that although, in our experiment, robots provided object labels (and not uses), having robot informants instead of puppets was more cognitively demanding for children, especially when being asked to complete a cognitively challenging task (i.e., *contrast label* phase).

We also found indications of a potential gender effect during the *contrast label* phase of the experiment, with boys performing as expected (choosing the previously incorrect robot), and girls performing the opposite of what we expected (choosing the previously correct robot). It is possible that girls and boys perceive and anthropomorphize robots differently than they do puppets and each other, which may be why we obtained these results. However, it is unclear whether this is why we found these results in our experiment, and further exploration with a larger sample size is needed to assess whether there exist gender differences on children's trust towards erroneous robots.

C. Clean-up Task

We were unable to reject our third hypothesis that children would comply with the previously correct robot's directions during the *clean-up* tasks, and would therefore place the basket with all the objects where that robot indicated. It is possible that whether a robot previously made mistakes does not impact children's compliance with its instructions, or that other unknown factors were at play. A future replication of this task, with a larger sample size and minor modifications, may provide further insight into whether robot informational errors do (or do not) affect children's compliance with robots' instructions.

D. Limitations and Future Work

The findings outlined in this paper are a first step towards understanding how robot informational errors may affect children's trust towards them. In the future, we plan to continue this path of inquiry by further examining additional variables. Additional data may shed some light on the results that we found, and why some of them align with previous findings while others do not, as well as the gender effect that we found. Our research also highlights the potential for robot errors to intentionally decrease trust towards a robot, in the case that it might be unwanted or unsafe. Future research could explore the potential of programing certain types of robots to make errors once in a while to decrease children's trust towards them.

One of the main limitations of our research is the controlled nature of it. Our experiment was conducted at a daycare center, and the experimental set-up was not organic or a natural interaction that children may experience. Future research should explore how children respond to robot errors in a more ecologically valid setting, such as having unscripted interactions, or allowing the robot to actually make errors (not controlled through Wizard of Oz).

Future research should also explore other types of errors or characteristics that may affect children's trust towards robots, such as mechanical and technical malfunctions including jittery movements, or sound problems. In addition, other research could investigate how different robot attributes, such as whether a robot is introduced as a child or an adult, or its embodiment (i.e., anthropomorphic, zoomorphic, mechanical-looking) may affect how its errors are perceived. Further research can provide insight into how different types of robot errors may affect children's trust towards robots, and how designers could use this information to design child-friendly robots.

VII. Conclusion

In this paper we explored how robot informational errors affect children's trust towards robots. Through an experiment conducted with 3-5–year old children and two humanoid robots, we found that children trusted a robot that had not made errors (i.e., previously correct robot) more than one that had (i.e., previously incorrect robot), but only when the information provided by the robots matched what the child was asked for. This finding aligns with previous research in psychology that observed that children trusted a previously correct puppet more than a previously incorrect one [10]. However, we did not find that children trusted the robots differently when they were asked for an object with a name that the robots had not mentioned. Nor did we find that whether a robot had made errors had an effect on the children following its instructions during a *clean-up* task. These results suggest that robot informational errors do have an effect on children's trust towards them, but further research is needed to explore what parameters may affect such trust.

Acknowledgments

We would like to thank On The Move Inc. daycare in Winnipeg, Manitoba, the staff, and the parents for their help and support in conducting our research. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] T. Belpaeme, P. E. Baxter, R. Read, R. Wood, H. Cuayáhuitl, B. Kiefer, S. Racioppa, I. Kruijff-Korbayová, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerincx, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Cañamero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert, "Multimodal child-robot interaction: Building social bonds," J. Human-Robot Interact., vol. 1, no. 2, pp. 33–53, 2012.

[2] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, and M. Blow, "KASPAR - a minimally expressive humanoid robot for human-robot interaction research," *Appl. Bionics Biomech.*, vol. 6, no. 3–4, pp. 369–397, 2009.

[3] J. H. Han, D. H. Kim, and J. W. Kim, "Physical learning activities with a teaching assistant robot in elementary school music class," in *Proceedings of the 5th International Joint Conference on INC, IMS, and IDC*, 2009, pp. 1406–1410.

[4] A. Sciutti, F. Rea, and G. Sandini, "When you are young, (robot's) looks matter. Developmental changes in the desired properties of a robot friend," in *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 567–573.

[5] J. De Wit, T. Schodde, B. Willemsen, K. Bergmann, S. Kopp, and P. Vogt, "The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies," in *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction*, 2018, pp. 50–58.

[6] J. Kennedy, P. Baxter, and T. Belpaeme, "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 67–74.

[7] K. Corriveau and P. L. Harris, "Choosing your informant: Weighing familiarity and recent accuracy," *Dev. Sci.*, vol. 12, no. 3, pp. 426–437,

2009.

[8] M. VanderBorght and V. K. Jaswal, “Who knows best? Preschoolers sometimes prefer child informants over adult informants,” *Infant Child Dev.*, vol. 18, no. 1, pp. 61–71, 2009.

[9] E. Nurmsoo and E. J. Robinson, “Children’s trust in previously inaccurate informants who were well or poorly informed: When past errors can be excused,” *Child Dev.*, vol. 80, no. 1, pp. 23–27, 2009.

[10] S. A. J. Birch, S. A. Vauthier, and P. Bloom, “Three- and four-year-olds spontaneously use others’ past performance to guide their learning,” *Cognition*, vol. 107, no. 3, pp. 1018–1034, 2008.

[11] J. E. Young, J. Sung, A. Voida, E. Sharlin, T. Igarashi, H. I. Christensen, and R. E. Grinter, “Evaluating Human-Robot Interaction,” *Int. J. Soc. Robot.*, vol. 3, no. 1, pp. 53–67, Oct. 2010.

[12] P. H. Kahn, T. Kanda, H. Ishiguro, N. G. Freier, R. L. Severson, B. T. Gill, J. H. Ruckert, and S. Shen, “‘Robovie, you’ll have to go into the closet now’: Children’s social and moral relationships with a humanoid robot,” *Dev. Psychol.*, vol. 48, no. 2, pp. 303–314, 2012.

[13] S. Turkle, C. L. Breazeal, O. Dasté, and B. Scassellati, “Encounters with kismet and cog: Children respond to relational artifacts,” *Digit. media Transform. Hum. Commun.*, no. September, pp. 1–20, 2006.

[14] M. Ragni, A. Rudenko, B. Kuhnert, and K. O. Arras, “Errare humanum est: Erroneous robots in human-robot interaction,” in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 2016, pp. 501–506.

[15] A. S. Rao and M. P. Georgeff, “BDI agents: From theory to practice,” in *Proceedings of the 1st International Conference on Multiagent Systems*, 1995, vol. 95, pp. 312–319.

[16] D. Premack and G. Woodruff, “Does the chimpanzee have a theory of mind?,” *Behav. Brain Sci.*, vol. 1, no. 4, pp. 515–526, 1978.

[17] K. Corriveau and P. L. Harris, “Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information,” *Dev. Sci.*, vol. 12, no. 1, pp. 188–193, 2009.

[18] M. Fusaro, K. H. Corriveau, and P. L. Harris, “The good, the strong, and the accurate: Preschoolers’ evaluations of informant attributes,” *J. Exp. Child Psychol.*, vol. 110, no. 4, pp. 561–574, 2011.

[19] V. K. Jaswal and L. A. Neely, “Adults don’t always know best: Preschoolers use past reliability over age when learning new words,” *Psychol. Sci.*, vol. 17, no. 9, pp. 757–758, 2006.

[20] M. A. Koenig and P. L. Harris, “Preschoolers mistrust ignorant and inaccurate speakers,” *Child Dev.*, vol. 76, no. 6, pp. 1261–1277, 2005.

[21] J. H. Danovitch and R. Alzahabi, “Children show selective trust in technological informants,” *J. Cogn. Dev.*, vol. 14, no. 3, pp. 499–513, 2013.

[22] C. Breazeal, P. L. Harris, D. Desteno, J. M. Kory Westlund, L. Dickens, and S. Jeong, “Young Children Treat Robots as Informants,” *Top. Cogn. Sci.*, vol. 8, no. 2, pp. 481–491, 2016.

[23] E. Short, J. Hart, M. Vu, and B. Scassellati, “No fair!! An interaction with a cheating robot,” in *Proceedings of the 5th International Conference on Human Robot Interaction*, 2010, pp. 219–226.

[24] H. Kose-Bagci, E. Ferrari, K. Dautenhahn, D. S. Syrdal, and C. L. Nehaniv, “Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot,” *Adv. Robot.*, vol. 23, no. 14, pp. 1951–1996, 2009.

[25] V. Charisi, D. Davison, D. Reidsma, and V. Evers, “Evaluation methods for user-centered child-robot interaction,” in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 2016, pp. 545–550.

[26] A. Coninx, P. Baxter, E. Oleari, S. Bellini, B. Bierman, O. Blanson Henkemans, L. Cañamero, P. Cosi, V. Enescu, R. Ros Espinoza, A. Hiolle, R. Humbert, B. Kiefer, I. Kruijff-Korbayovà, R. Looije, M. Mosconi, M. Neerincx, G. Paci, G. Patsis, C. Pozzi, F. Sacchitelli, H. Sahli, A. Sanna, G. Sommavilla, F. Tesser, Y. Demiris, and T. Belpaeme, “Towards long-term social child-robot interaction: Using multi-activity switching to engage young users,” *J. Human-Robot Interact.*, vol. 5, no. 1, p. 32, 2015.

[27] J. M. K. Westlund, H. W. Park, R. Williams, and C. L. Breazeal, “Measuring children’s long-term relationships with social robots,” in *Workshop on Perception and Interaction Dynamics in Child-Robot Interaction*, 2017.

[28] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, “Child-robot interaction in the wild: advice to the aspiring experimenter,” in *Proceedings of the 13th International Conference on Multimodal Interfaces*, 2011, pp. 335–342.

[29] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, “Interactive robots as social partners and peer tutors for children: A field trial,” *Human-Computer Interact.*, vol. 19, no. 1, pp. 61–84, 2004.

[30] T. Kushnir, C. Vredenburgh, and L. A. Schneider, “‘Who can help me fix this toy?’ The distinction between causal knowledge and word knowledge guides preschoolers’ selective requests for information.,” *Dev. Psychol.*, vol. 49, no. 3, pp. 446–453, 2013.

[31] Q. G. Li, G. D. Heyman, F. Xu, and K. Lee, “Young children’s use of honesty as a basis for selective trust,” *J. Exp. Child Psychol.*, vol. 117, no. 1, pp. 59–72, 2014.

[32] J. E. Hupcey, J. Penrod, J. M. Morse, and C. Mitcham, “An exploration and advancement of the concept of trust,” *J. Adv. Nurs.*, vol. 36, no. 2, pp. 282–293, 2001.

[33] J. B. Rotter, “Personality and generalized expectancies for interpersonal trust,” *Am. Psychol.*, vol. 26, no. 5, pp. 443–452, 1971.

[34] M. Salem, G. Lakatos, F. Amirabdollahian, and K. Dautenhahn, “Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust,” in *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 141–148.

[35] N. Martelaro, V. C. Nneji, W. Ju, and P. Hinds, “Tell me more: Designing HRI to encourage more trust, disclosure, and companionship,” in *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction*, 2016, pp. 181–188.

[36] A. Hamacher, N. Bianchi-Berthouze, A. G. Pipe, and K. Eder, “Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-robot interaction,” in *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*, 2016, pp. 493–500.

[37] S. Booth, J. Tompkin, H. Pfister, J. Waldo, K. Gajos, and R. Nagpal, “Piggybacking robots: Human-robot overtrust in university dormitory security,” in *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 426–434.

.