Children's Trust Towards Erroneous Robot Informants

By

Denise Y. Geiskkovitch

A Thesis submitted to

the Faculty of Graduate Studies of the University of Manitoba

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Individual Interdisciplinary Studies

Department of Computer Science, Department of Psychology

University of Manitoba

Winnipeg

Abstract

As social robotics continues to grow and develop, robots are increasingly finding their way into more areas of society, including hospitals, homes, daycare centres, and schools. It is essential for these robots to behave in ways that are appropriate for interacting with children, especially when they may need to elicit trust. As part of this thesis, we conducted two experiments with an overall total 115 participants investigating preschool-aged children's trust towards robots with human-like informational (experiment 1, Chapter 5) and robot-typical speech-recognition (experiment 2, Chapter 6) errors. Our findings suggest that children trust a robot that makes informational errors less than one that does not, but may trust a robot that exhibits speech-recognition errors more than one that does not. This suggests that children may perceive robot errors, and therefore trust robots differently, from other entities such as humans or puppets. We contribute the findings from these two experiments, as well as an initial framework of child-robot trust. This thesis provides a starting point for robot designers to consider trust when designing robots for children, and for researchers to further investigate young children's trust towards robots.

Dedication

Para mi familia,

lo más importante en la vida.

Acknowledgements

This Ph.D. would not have been possible without the support from many individuals.

I would like to thank my committee, for providing me with guidance, insights, and support throughout the years. This thesis would not be what it is without them.

I would like to especially thank my supervisor. It has now been close to ten years since you first introduced me to Human-Robot Interaction, and that changed my life. Who knows where I would be otherwise! Thank you for all your support over the years. As we both know, I had many struggles during my degree, and you were always there to listen and understand, and provide advice if needed. I can honestly say that my life (and certainly my Ph.D.) would not be the same without you in it, and that's only partly because I wouldn't get to play with AJ and Luca! You prepared me to be a researcher, a professor, and a valuable member of the academic community. For all of your support over the years, I cannot thank you enough.

Thank you to everybody at the lab, especially my bros, for making the last few years fun.

To my family, who have always supported me and done everything in their power help me be the best person I can be. Thank you to my parents, brother, and extended family, for always being present and willing to help with whatever I need. Thank you Vol, for being my second half and ensuring that I always have some grumpiness around me ☺. Thank you Charlie, Maguie, Luna, and Yankee for always making my life richer and more meaningful than it would be otherwise.

Table of Contents

List of Figures

Chapter 1 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 (Belpaeme et al., 2012), in homes, with some robots specifically designed to interact with children (Dautenhahn et al., 2009), and in daycare centers and schools, where they act as teaching assistants and companions (Han, Kim, & Kim, 2009). Whether as a robot teaching assistant, or a robotic nurse, it is essential for robots to behave in ways that are appropriate for interacting with children, to obtain desired outcomes, especially when children may need to trust a robot, or follow its directions or orders (e.g., a nurse robot). The study of Child-Robot Interaction (CRI) – within Human-Robot Interaction (HRI) and Human-Computer Interaction (HCI) – 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 (Sciutti, Rea, & Sandini, 2014), and whether robots can serve as tutors, helping children learn classroom material (Kennedy, Baxter, & Belpaeme, 2015; Wit et al., 2018). However, it is unclear whether young children trust robots; that is, do young children trust information provided by robots in ways similar to how they do from people or puppets? And if so, how is such trust affected by errors that a robot makes, or even the type of error? This dissertation contributes to this emerging field by exploring how young children alter and manage their trust of robots that provide them with information.

1.1 Trust and Robots

Trust is highly intertwined into various aspects of social life, and it is no different when it comes to robots. Robots can bring to question matters of obedience and persuasion, leading researchers to investigate how robots are able to elicit such constructs and outcomes that we might not expect from them (e.g., Bartneck, Bleeker, Bun, Fens, & Riet, 2010; Geiskkovitch, Cormier, Seo, & Young, 2016; Ham & Midden, 2014). Similarly, privacy and security are two aspects of great concern in HRI, as robots become more prevalent in society and are able to see, hear, and obtain information that people may not realize (e.g., Booth et al., 2017; Liao, Vitak, Kumar, Zimmer, & Kritikos, 2019; Rueben, Bernieri, Grimm, & Smart, 2017). Likewise, researchers are also interested in how the embodiment (Looije, Van Der Zalm, Neerincx, & Beun, 2012; Martelaro, Nneji, Ju, & Hinds, 2016; Rae, Takayama, & Mutlu, 2013) of robots and even the culture (Shahid, Krahmer, & Swerts, 2014; L. Wang, Rau, Evers, Robinson, & Hinds, 2010) that they are in can affect interactions and perceptions. However, most of this work focuses on how adults interact and perceive robots when robots function properly, and has yet to explore how these aspects apply to children or what occurs when robots make errors, making it difficult to design successful interactions with children.

The question of how people generally react to robot errors is an emerging research area, and investigations have covered various topics. For example, researchers have found that adults tend to react to different robot errors similarly, although somewhat unexpectedly. Unusual robot behaviour, such as illogical movements and unusual requests, tends to lead to lower reported trust (Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015). Similarly, a robot that provides incorrect responses is trusted less (Ragni, Rudenko, Kuhnert, & Arras, 2016), as well as one that takes a long time to complete a time-sensitive task (Robinette, Howard, & Wagner, 2017). However, while people report lowered trust, they continue to perform tasks as instructed by an erroneous robot (Robinette, Li, Allen, Howard, & Wagner, 2016; Salem et al., 2015), suggesting that additional trust-affecting factors might exist. Likewise, research cited above investigated how adults perceive robot errors, which may not generalize to children who are often the target user group of social robots, especially those of preschool age due to their still-developing verbal and social skills.

1.2 Children and Robots

Generally, children tend to perceive and respond to robots differently than adults do; while adults may interact with robots as though they are alive, due to existing social heuristics, children are more likely to think of them as alive (Rao & Georgeff, 1995) and attribute them with social and moral rights (Kahn et al., 2012). In addition, children and adults tend to interact with robots under different paradigms, leading to research with adults and children focusing on different aspects of interaction with robots. For example, unlike the research briefly described above with adults, investigations with children tend to focus on areas such as tutoring (e.g., Brown & Howard, 2014; Kanda, Hirano, Eaton, & Ishiguro, 2004; Kennedy, Baxter, Senft, & Belpaeme, 2016; Ramachandran, Huang, & Scassellati, 2017; Saerbeck, Schut, Bartneck, & Janse, 2010) and autism (e.g., Albo-Canals et al., 2018; Moorthy & Pugazhenthi, 2017; Shamsuddin et al., 2012) investigating how robots might be able to assist children with varying tasks. However, formal investigations into how preschool-aged children perceive robots, or even how they trust them as informants or as social beings, remain sparse. The investigation into this specific age group is extremely relevant, as robots are currently being designed for daycare or home use, where they might encounter young children. Further, preschool-aged children are still developing their language and social skills, and especially relevant to our research, their theory of mind (Wellman, 2002).

1.2.1 Children, robots, and trust

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. For example, robots are currently acting as teaching assistants in schools – providing children with information and behavioural guidance (e.g., "Robot teachers invade Chinese kindergartens," 2018), as well as security patrols in public spaces (e.g., Latella, 2021) – delivering safety information to children (and adults). However, whether these robots are able to elicit trust from young children to be of use is not yet known. We know from Psychology that children as young as three years old develop a theory of mind of adults relating to trust. Children can then use factors such as how familiar an individual is, what their age is, and whether they have made mistakes in the past when deciding whether to trust information that they provide. Specifically, children tend to trust familiar (Corriveau & Harris, 2009a) and previously accurate adults (Birch, Vauthier, & Bloom, 2008; Corriveau & Harris, 2009b; Fusaro, Corriveau, & Harris, 2011) more than unfamiliar or inaccurate individuals, as well as trusting adults with some types of information and children with others (VanderBorght & Jaswal, 2009). However, we do not yet know how young children perceive information provided by robots, specifically erroneous ones, and to what extent and in what form they may apply trust models of adults to robots.

Problem statement: it is unknown how young children perceive information and alter their trust towards erroneous and normally functioning robots. In this thesis, we explore how two different types of errors affect preschool-aged children's trust in robots.

In this work, we explore whether young children trust information provided by robots, and how this trust may be impacted by a robot making errors. To accomplish this, we conducted a series of experiments to investigate whether children trust erroneous and non-erroneous robots differently. This follows a body of investigation into how robot errors generally affect interaction with adults, and how young children trust erroneous (human or puppet) informants. We therefore investigated related work on how informant attributes and trust towards them translates to robots that make human-like errors (Chapter 5) and those that make robot-typical errors (Chapter 6). We then performed an extensive survey on trust towards humans, machines, and robots (Chapter 7) and developed an initial framework of child-robot trust, grounded on previous research and the findings from our experiments.

To explore this area, we designed and conducted a set of experiments, through the implementation of robot behaviours, to investigate how young children selectively trust erroneous and non-erroneous robot informants. This research provides an initial exploration into how children may trust robots that make mistakes, and will enable robot designers to account for robot errors when designing robots for use with children. It also provides researchers with a starting point to study robot errors with children, and information on how different errors may be interpreted by young children. Lastly, this work provides general information about how young children can perceive robots and how they may apply existing mental models, or create new ones, to understand the robots' behaviour and knowledge.

1.3 Research Questions

In this research we investigate the following broad questions:

1. Do preschool aged children exhibit trust towards robots in similar patterns to how they do towards people or puppets?
2. How do robot errors affect preschool aged children's trust towards robots?
3. Does the nature of a robot error impact preschool-aged children's trust toward the robot?
4. What do we know about trust in people (including children) and how might this inform how children may trust robots?

Thus, in this thesis we have four main prongs of investigation: similarities between young children's trust towards adults or puppet and trust towards robots, effects of robot errors on children's trust, impact of the nature of the robot error on children's trust, an exploration of the psychological and theoretical foundations of trust to shape our understanding of how children may trust robots. Below, we elaborate on how we explore these themes.

1.3.1 Do preschool aged children trust information from robots in similar ways as they do that from adults or puppets?

To investigate these objectives we first followed in the footsteps of previous research in Developmental Psychology that investigated trust towards human and puppet informants. This work has found that various attributes about informants (e.g., how familiar they are to the child, their age, the type of knowledge that they possess, as well as their prior accuracy) can affect how much children trust a particular informant. This is because starting at around age three, children begin to be able to reflect on their own and others' mental states and knowledge about their environment, and thus are able to adjust their trust accordingly, based on information obtained (Corriveau & Harris, 2009a; Nurmsoo & Robinson, 2009a; VanderBorght & Jaswal, 2009). A well validated experimental methodology exists in which two humans or puppets act as informants, who have different attributes but provide similar information (e.g., Birch et al., 2008; Corriveau & Harris, 2009a; Nurmsoo & Robinson, 2009; VanderBorght & Jaswal, 2009). In this methodology, children then have to make a decision that reflects their trust of the informants, based on the informants' differing attributes. In our work, we adapted and modified this well-established methodology to apply it with robot informants instead of humans or puppets.

1.3.2 How do robot errors affect preschool aged children's trust towards robots?

To investigate how robot errors affect children's trust, we conducted two experiments in which two robots (which were controlled through a Wizard-of-Oz implementation) provided information to a young child; one of the robots made errors while the other one did not. The child then indicated which robot's information they trusted. This allowed us to compare trust towards a robot that makes errors and one that does not, and to investigate how such errors affect children's trust.

1.3.3 Does the nature of the robot error impact preschool-aged children's trust in robots differently?

In the two experiments that we conducted (as mentioned above), we focused on different types of errors. The first experiment (Chapter 5) explored how human-like informational errors that a robot makes affect children's trust towards that robot (versus a robot that does not make any mistakes). Our second experiment (Chapter 6) examined how robot-typical speech-recognition errors affect children's trust towards robots. The combination of the two experiments in this thesis therefore provides us with knowledge on how different errors in robots can influence children's trust.

1.3.4 What do we know about trust in people (including children) and how might this inform how children may trust robots?

We explored the foundations of trust, drawing from related literature from Psychology, Computer Science (HCI and HRI), Sociology, and Management, to develop a framework of child-robot trust (Chapter 7). In our framework, we synthesized the trust research from these fields, along with the findings from our experiments, to provide a basis for the types of factors and dimensions that may affect the various types of trust that people (including children) can experience. We start broadly with human-human trust, and then narrow our insights to human-system trust, human-robot trust, and finally arrive at child-robot trust.

1.4 Significance

In this thesis, we investigate how a robot's errors influence preschool aged children's trust in that robot. To examine this topic, we conducted two experiments with over 100 children aged 3-5 years old, where each child interacted with and observed two robots provide information. However, one of the robots made errors: in the first experiment these were human-like informational errors, while in the second experiment we explored robot-typical speech-recognition errors. The results from our work demonstrate that human-like informational errors are likely to decrease young children's trust in informational robots (Chapter 5), while robot-typical speech-recognition errors might actually increase trust towards robots (Chapter 6). Our work suggests that children may be likely to trust robots in similar ways to how they do humans, with some marked differences. For example, similarly to how children trust a previously accurate human (or puppet) more than an inaccurate one, we found similar themes in our research.

As the presence of social robots continues to increase in places such as schools, hospitals, shopping malls, and even homes, it is crucial that we understand how children will perceive these robots. For example, will children pay attention and follow the instructions of a teaching assistant robot? Will children perceive a home robot to be like a parent, a friend, a sibling, or neither? Research suggests that children view robots as live beings (Rao & Georgeff, 1995), however, several gaps in knowledge remain, such as how young children may trust robots, how robot errors may affect such trust, and even general knowledge about young children's perceptions of robots. Our work provides insight on these topics. By investigating preschool-aged children's trust towards erroneous robots, we gain insight into how young children might react to robot errors, how young children may perceive different types of errors, and how such errors in turn influence children's trust. The rise in robots being targeted for young children (e.g., pet robots, companion robots, tutoring robots), compounded by the necessity for many of these robots to elicit trust, means that our work is highly relevant to current societal trends, helping robot designers and researchers to make informed decisions.

1.5 Contributions

This work makes several contributions to the study of young children's trust in informational robots:

1) We present an original and fully implemented set of robotic behaviours for facilitating interaction with preschool-aged children and investigation of trust-related concepts with them.

2) We provide data from two experiments conducted with over 100 young children (combined) that highlight how different types of robot errors are perceived differently by young children, and therefore affect their trust in different ways. Specifically, our findings demonstrate that robot informational errors decrease preschool-aged children's trust in robots, while more robot-typical speech-recognition errors increase trust.
3) We provide analyses and results from our data, which contribute novel and pioneering knowledge on how robot errors (human-like informational and robot-typical speech-recognition) influence young children's trust in robots and how we can expect children to generally trust robots.
4) We provide a framework of child-robot trust, based on research on human-human trust, human-system trust, human-robot trust, and child-robot trust, to aid robot designers and researchers interested in applications of social robots with young children, to elicit or mitigate trust. Our framework provides dimensions and factors affecting various types of trust towards humans, systems, and robots.

In the remainder of this document, we explore related work in the areas of Human Computer Interaction, Human-Robot Interaction, Child-Robot Interaction, and Psychology (Chapter 2 and Chapter 3). Following, Chapter 4 details our experimental methodology, while Chapter 5 and Chapter 6 present the two experiments we conducted. Chapter 7 presents our framework of child-robot trust, and we finish with a discussion of the thesis as a whole, limitations, future work (Chapter 8), and concluding statements (Chapter 9).

Chapter 2 Interacting With Robots

2.1 Overview of Human-Computer Interaction

The field of Human-Computer Interaction (HCI) is multidisciplinary, and encompasses many topic areas and subfields, with its main goal being to study the ways in which individuals use technology, and provide suggestions for improvement of such technology. This objective has broadened from a focus on making work more productive, to emotion and user experience design as technology became part of all areas of life and not just work (Dix, 2017). Nowadays, the field's focus has shifted towards personal and social applications providing solutions for people's health, education, well-being, sociality, and the use of cloud-based, multi-device, and ubiquitous computing systems (Dix, 2017). The focus is specifically on the usefulness and usability of systems for such application areas, ensuring that systems are helpful and easy to use (Helander, Landauer, & Prabhu, 1997).

A particularly important focus in HCI, given the topic of this thesis, is people's tendency to treat computers as social beings, even if they do not believe that computers are social beings (Nass, Steuer, Tauber, & Reeder, 1993). Individuals tend to apply social rules to computers, for example, using politeness rules and applying gender stereotypes to them (Nass, Steuer, & Tauber, 1994). Researchers have found that, similarly to how they would with other people, individuals are more polite to and find praise more meaningful from a computer they have built rapport with than one they have not (Nass et al., 1994). However, the computer itself is not the only factor affecting interaction, as people tend to perceive different voices coming from the same computer as different social actors, and even evaluate them differently depending whether the voice sounds stereotypically male or female (Nass et al., 1994).

2.1.1 Child-Computer Interaction

In the subfield of Child-Computer Interaction, researchers study the "activities, behaviours, concerns and abilities of children as they interact with computer technologies" (Read & Bekker, 2011). Major themes in this area include coding, education, interaction design, general design, and child-specific concerns such as privacy and content control (Giannakos, Papamitsiou, Markopoulos, Read, & Hourcade, 2020). However, research has predominantly focused on how technology and experimental methodologies designed for adults need to be modified and adapted for use with children (Read & Markopoulos, 2013). Researchers suggest a number of future opportunities or challenges to explore with children and computers, including children's participation in child-computer interaction and the role of mobile, pervasive, tangible, and embodied interaction (Read & Markopoulos, 2013), highlighting the importance of specifically studying children and their interactions with technology as an important aspect of HCI. Thus, in this thesis, we follow this trend by investigating how young children interact with, perceive, and trust robots (a type of embodied technology).

2.1.2 Research on Trust Towards Computers

Many computer system designs rely on users' trust in one way or another. For example, online banking systems rely on clients trusting the system enough to use it and be comfortable conducting financial transactions. Thus, trust is a key element of inquiry in HCI. Within the field, trust is divided into two broad categories: trust towards machines, such as automated and recommender systems, and trust in technology-mediated applications, such as e-commerce websites and online reviews. Due to the variety of application areas that trust plays a role in, many definitions of trust exist in HCI. In the context of automation, for example, trust is most popularly defined as an "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (J. D. Lee & See, 2004). Research suggests that there are many factors that affect trust in automated systems, including the relationship between an individual's self-confidence and trust in the system – that is, the less confident the user is in their abilities, the more likely they are to trust the system (J. D. Lee & Moray, 1994). Similarly, the accuracy or performance of an automated system can affect how much it is trusted (Khasawneh, Bowling, Jiang, & Gramopadhye, 2003; Merritt & Ilgen, 2008). Trust is therefore key in many systems, as it mediates the use and adoption of such systems (J. D. Lee & See, 2002; Muir, 1987). For recommender systems, however, trust is defined as the belief of a user that the information provided by another user will be helpful (Meyffret, Médini, & Laforest, 2012). This definition emphasizes the role that other individuals play in feeding information into the recommender's algorithm, as opposed to self- or system-confidence.

In technology-mediated applications – those that are used to complete a task, but not as an end in themselves – trust definitions differ depending on the type of system. Trust in online systems is described as a "confident expectation in an online situation of risk that one's vulnerabilities will not be exploited" (Corritore, Kracher, & Wiedenbeck, 2003). Research on trust in online systems mainly focuses on markers of trust, and many dimensions and models have been proposed. Antecedents and determinants of trust have been suggested by many researchers, the first including integrity, ability, and benevolence (Gefen, 2002), while the latter can be divided into six dimensions: information content, product, transaction, technology, institutional, consumer-behaviours (D. J. Kim, Song, Braynov, & Rao, 2001). However, concern exists over research of online systems and its reliance on the researchers thoughts and perspectives, as opposed to experimental data (Y. D. Wang & Emurian, 2005). On the other hand, there is ample research on trust in Computer-Supported Cooperative Work (CSCW), which has been defined as a "positive expectation that one's vulnerabilities will not be exploited" (Riegelsberger, Sasse, & McCarthy, 2005). Nonetheless, research in this area focuses on how human collaborators trust one-another (e.g., Al-Ani et al., 2013; Knowles et al., 2015; Merrill & Cheshire, 2017), as opposed to investigating trust towards the mediating system.

Through the above research it is evident that trust is a complex and highly relevant construct in HCI, due to its effects on usage and interaction outcomes. However, save for some research on how children perceive computer errors (Danovitch & Alzahabi, 2013), this area has largely focused on adults. In Chapter 3, we unpack what trust is and provide insight into different aspects of it.

2.2 Overview of Human-Robot Interaction

Having reviewed trust in HCI, we now switch our focus to robots, which are different to other types of machines due to their physical embodiment, the agency they project, and how people tend to anthropomorphize them. The field of Human-Robot Interaction (HRI), a subfield of HCI, specifically investigates interaction between people and physically embodied agents. HRI can be subdivided into four major themes: human supervisory control of robots in routine tasks, remote control of robots in hazardous or inaccessible environments, supervision or interaction with autonomous vehicles, and human-robot social interaction (Sheridan, 2016). For example, researchers have investigated which controls are best to use for operating robots, finding that keyboards might lead to better performance than gesture-based interfaces (Nagy, Young, &

Anderson, 2015; Radmard, Moon, & Croft, 2015), or how feedback can be improved to help the operator have a better sense of the robot and their surroundings (e.g., Nielsen, Goodrich, & Ricks, 2007; Rea, Seo, Bruce, & Young, 2017). This research emphasizes the importance of taking the user, their needs, and their state of mind into account when designing robots and interfaces to control robots, in addition to how the social aspect of the interaction between the operator and the robot(s) may play an important role during the operation of the robot.

There exist certain robots that elicit social interaction akin to people, by attempting to apply similar behavioural norms to people (Bartneck & Forlizzi, 2004). Social HRI, highlights the uniqueness of robots in comparison to other electronics such as computers, appliances, etc. Robots tend to encourage social interaction due to their physical manifestations and how they are able to move and interact with the environment. In addition, robots tend to be anthropomorphized more than other types of technologies (e.g., Bartneck, van der Hoek, Mubin, & Al Mahmud, 2007; Bartneck, Verbunt, Mubin, & Al Mahmud, 2007; Forlizzi & DiSalvo, 2006; Fussell, Kiesler, Setlock, & Yew, 2008), such as computers, which provides robots with agency (the ability to act

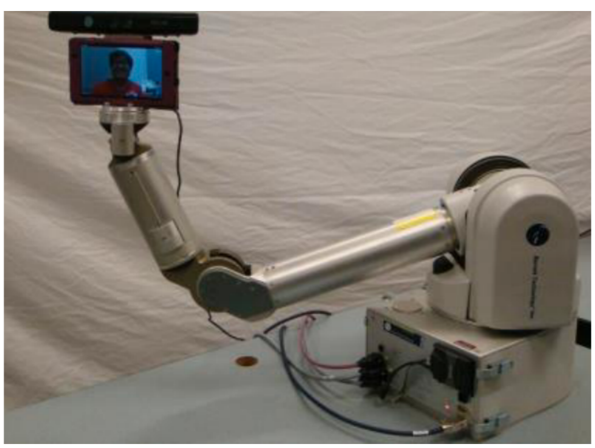


Figure 2. Controls for operating robots (Radmard, Moon, & Croft, 2015).

Figure 1. Displaying attention cues to the operator (Rea, Seo, Bruce, & Young, 2017).

with intention; Dewey, 1980). This motivates the expectation that children will also treat these robots as social beings.

Social HRI, specifically, studies how people behave and perceive robots in a social way, often using social heuristics to interact with them (Young, Hawkins, Sharlin, & Igarashi, 2009). One area of research in social HRI is related to investigating how robots should interact with people, with the goal of designing them in a way that benefits interaction. For example, the Kaspar robot was designed with autistic children in mind: the robot's features are purposely not very pronounced, and it moves and speaks in a way that makes it easier for autistic children to interact with it and learn social skills from the interaction (Dautenhahn et al., 2009). Social robots are also able to affect human behaviour in a variety of ways, such as persuading people to do things they do not want to do (Geiskkovitch et al., 2016), complete embarrassing tasks (Bartneck et al., 2010),

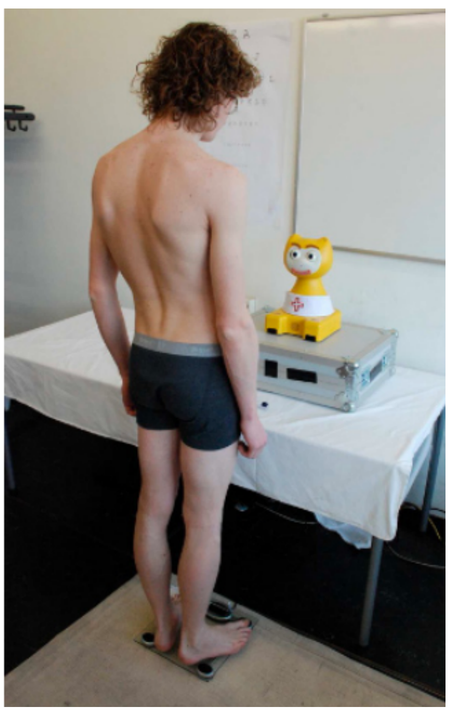


Figure 3. Zoomorphic robot convinces participants to undress on camera (Bartneck et al., 2010).

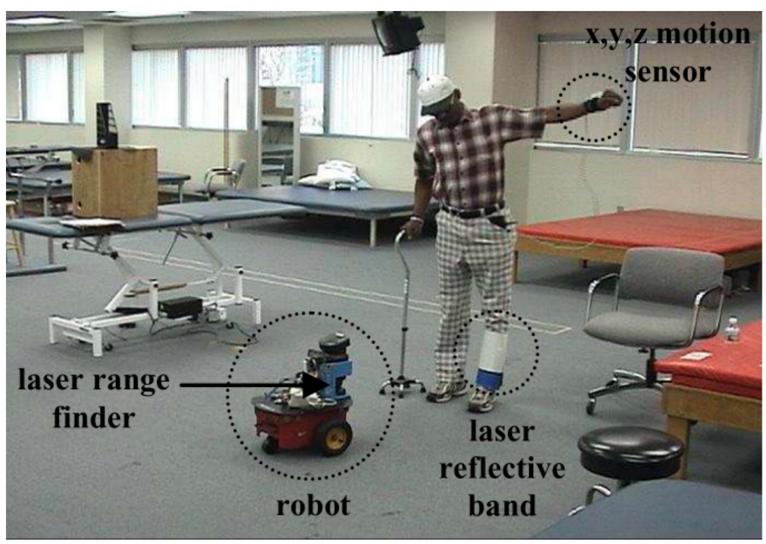


Figure 4. Robot engaging person in rehabilitation exercises (Matarić et al., 2009).

engage in rehabilitation therapy exercises (Matarić, Tapus, Winstein, & Eriksson, 2009), reduce energy consumption (Ham & Midden, 2014), and even maintain fitness goals (Kidd, 2008). Thus, the field is highly varied, and requires the collaboration of many disciplines (e.g., Computer Science, Psychology, Medicine) to carry out the necessary research.

The study of social robots and people is of particular importance because individuals can and do form emotional connections with robots (Sung, Guo, Grinter, & Christensen, 2007) and feel genuinely bad at the prospect of harming one (Bartneck, Verbunt, et al., 2007), in ways that we may not fully understand yet. Many factors affect these attachments. For example, researchers have found that physically embodied robots tend to elicit more empathy than virtual robotic agents (Seo, Geiskkovitch, Nakane, King, & Young, 2015).

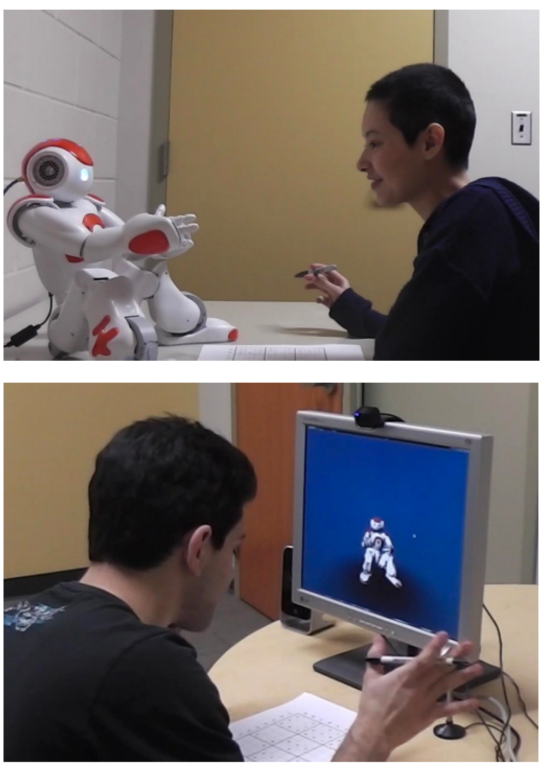


Figure 5. Physical robots are able to elicit more empathy than virtual ones (Seo, Geiskkovitch, Nakane, King, & Young, 2015).

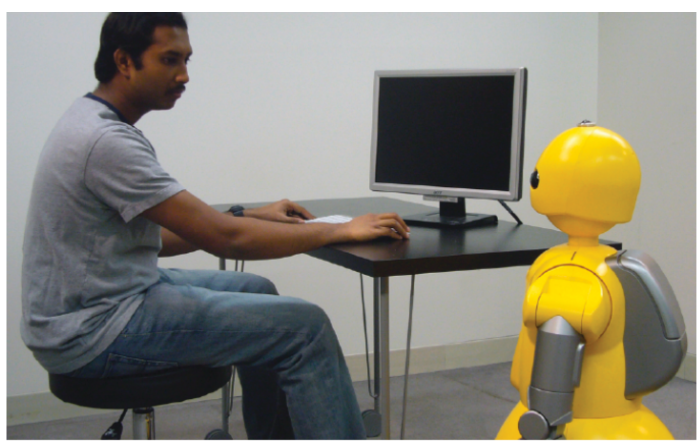


Figure 6. Robot is able to persuade people by using non-verbal cues (Beer et al., 2012; Chidambaram, Chiang, & Mutlu, 2012).

The research in this area has allowed researchers to propose robot design characteristics to elicit various human responses or for different populations (e.g., Beer et al., 2012; Chidambaram, Chiang, & Mutlu, 2012; Wu, Fassert, & Rigaud, 2012). Investigating people's social, emotional, and physical interactions with robots continues to be a large focus in HRI, examining various user groups, robot types, and application areas from a number of different perspectives and fields (e.g., Charisi, Davison, Reidsma, & Evers, 2016; Young et al., 2011). Although research with children and robots is still in its infancy, such implications are extremely important in this sub-population.

As witnessed by the large diversity in research topics within HRI as a whole, and social HRI more specifically, robots have the potential to influence people in many ways (whether intentionally or not), with many possible outcomes. One way is by eliciting trust from people. The next section provides an overview of what we know thus far about how people trust robots.

2.2.1 Research on Trust Towards Robots

Research has shown that people tend to be very trusting of robots. For example, people will trust robots to lead them in the correct direction (and will follow them) during emergency situations, even when they know the robot's instructions to be incorrect (Robinette et al., 2016). Individuals also allow robots into secured facilities (Booth et al., 2017), and complete strange tasks that a robot asks them to do (even if potentially harmful; Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015). These findings point to the imperative need to study trust towards robots in various scenarios and with diverse individuals.

Trust can also affect a number of other outcomes in HRI, including whether people collaborate with a robot (Malle, Fischer, Young, & Moon, 2020), accept its advice (Malle et al., 2020), or even use or interact with a robot to begin with (Freedy, DeVisser, Weltman, & Coeyman, 2007). Trust can also influence the ability for a human-robot team to accomplish goals, create effective relationships, and have effective interactions with the robot (Hancock, Billings, Schaefer, et al., 2011). Inadequate amounts of trust towards a robot, either too much or too little, can result in serious consequences (Hancock, Billings, Schaefer, et al., 2011). Therefore, it is not sufficient to simply consider trust, but we must also consider the context in which trust exists, as well as the population of interest, due to the very nuanced nature of trust towards robots.

Trust research in HRI traditionally focused on how adults trust robots, with an emphasis on task-oriented scenarios. Intuitively, one may assume that trust towards robots is analogous to that towards computers and automation, technologies that preceded the robots that we use today (e.g., Hancock, Billings, Oleson, et al., 2011). Thus, trust in HRI initially focused on functionality. Researchers explored how a robot's functionality, and its ability to carry out its intended functions directly influenced trust towards it, finding that functionality or lack thereof directly influences certain trust measures (Martelaro et al., 2016). However, the sole focus on functionality does not allow for factors such as sociality to be taken into account, and therefore may not capture different types of trust towards robots. Furthermore, more recent research has found that people's mental models of robots are not the same as those of computers, and they interact with them in different ways (Young et al., 2011). Thus, trust in now being explored in a variety of contexts within HRI,

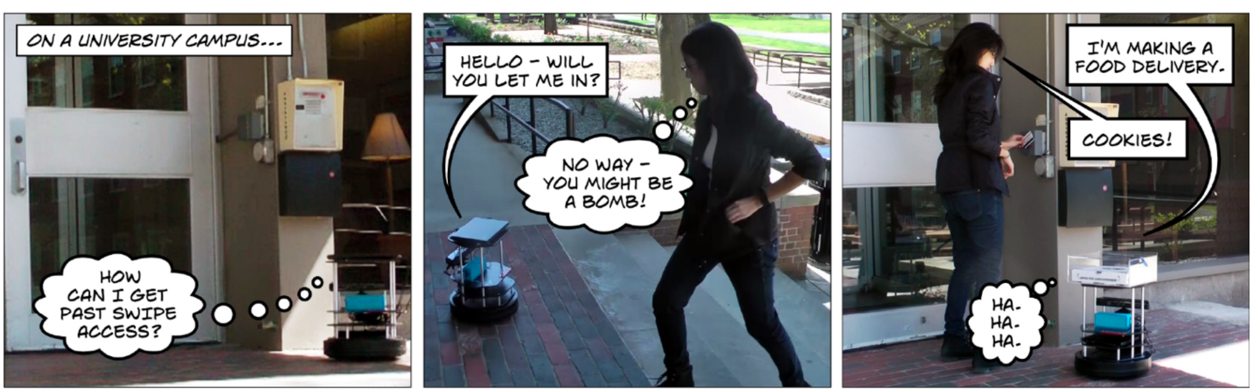


Figure 7. People let a robot into a secured facility, demonstrating trust (Booth et al., 2017).

where researchers examine whether adults trust robots and the types of attributes that may affect such trust, such as whether a robot is faulty or not (Salem et al., 2015), a robot's vulnerability level (Martelaro et al., 2016; Strohkorb Sebo, Traeger, Jung, & Scassellati, 2018), and whether it is able to fix a mistake it has made (Hamacher, Bianchi-Berthouze, Pipe, & Eder, 2016). In addition, new models of trust are being developed, which include dimensions such as moral and capacity trust, incorporating different angles of trust towards robots, although further research is still needed (Ullman & Malle, 2018, 2019). These models often apply knowledge and techniques from other fields, reinforcing the multidisciplinary nature of trust in HRI. Varying models suggest different components of trust in HRI, including ability, integrity, and benevolence (Martelaro et al., 2016), characteristics of the human, the robot, and the environment (T. Sanders, Oleson, Billings, Chen, & Hancock, 2011), and the reliability, competence, and moral integrity of the robot (Ullman & Malle, 2018, 2019). Trust has therefore become an invaluable area of research within HRI, providing insight into the types of behaviours that robots should and should not possess, and the potential outcomes that are tied to such behaviours.

2.2.1.1 Errors

Errors are commonplace with robots, as their components are unable to always function perfectly. Trust in the face of errors is an emerging area of HRI, and researchers have found that robot errors may lead to decreased reported trust, but may not affect behavioural outcomes (Hamacher et al., 2016; Ragni et al., 2016; Robinette et al., 2016; Salem et al., 2015). When trust is broken, however, providing an apology and a sense of regret can repair it (Hamacher et al., 2016). A detailed discussion of robot errors can be found in Chapter 6.

2.3 Overview of Child-Robot Interaction

The area of Child-Robot Interaction (CRI) has been growing for the past several years, with interest stemming from the growth in popularity of robots in society, and the desire to explore how they should be designed for better and safer interactions with children. Research in this area is quite varied, ranging from the design of robots to interact with children with Autism Spectrum Disorder (ASD) and help them learn social interactions (e.g., Robins et al., 2005; Robins, Dautenhahn, & Dubowski, 2006; Shamsuddin et al., 2012) and tutoring robots that help children learn (e.g., Gordon et al., 2016; Kennedy et al., 2015; Kennedy et al., 2016; Vogt et al., 2017), to exploring how to design speech-recognition mechanisms for robots to understand children's speech (e.g., Kruijff-Korbayová et al., 2012; Kennedy, 2017).

A large area of research investigates how robots can be designed to interact with children with ASD, to help them improve their social skills. Robins et al. (2005), for example, conducted a longitudinal case study in which they had four children with ASD interact with the humanoid robotic doll Robota. They found that through continued social interactions with the robot, the children displayed increased social behaviours with the robot, such as imitation and turn-taking. Learning these behaviours could be highly beneficial for children with ASD. Other researchers have explored the importance of a robot's appearance when interacting with children with ASD. Through trials with different types of robots, including humans pretending to be robots, researchers were able to narrow down the features that are most important for robots when interacting with children with ASD, and were able to design and develop the robot Kaspar, which has minimalistic expressions and features, for better interactions with children with ASD (Robins, Dautenhahn, & Dubowski, 2006).

Another area within CRI, which has grown in the past few years, looks at how robots can help typically developing children improve their learning. Research in this area explores how a robot's characteristics can affect its tutoring success. For example, Gordon et al. (2015) found that when a robot adjusted its feedback strategy based on the child's affective state, it was able to teach children a second language, and children perceived it more positively than when it did not adjust its feedback. Other research also found that robots are able to teach children a second language, but a robot's social behaviour, or lack thereof, does not have an effect on children's language learning (Kennedy et al., 2016; Kennedy et al., 2015). The research cited above highlights the breadth and multidisciplinary nature of the CRI sub-field, as well as how relevant it is in today's technologically-loaded society.

2.3.1 Research on Children's Trust Towards Robots

The exploration of trust towards robots has mainly been from the adult perspective. However, more scientific literature on the subject has been published recently, highlighting the focus of trust in CRI, the breadth of interests, and the multidisciplinarity of the field. For example, research suggests that children trust a robot that provides timely information, more than one whose

Figure 8. Children with Autism Spectrum Disorder interacting with robotic doll Robota (Robins et al., 2005)

Figure 9. Robot tutoring a child (Kennedy et al., 2016)

responses are delayed (Breazeal et al., 2016), and that young children (e.g., 3 years old) trust adults more than a robot, but older children (e.g., 7 years old) trust a robot more than adults (Di Dio et al., 2020). Additional research has focused on developing types of trust that children may experience towards robots, such as technological and interpersonal (the intent to act on a system's recommendation and vulnerability, respectively; van Straten, Peter, Kühne, De Jong, & Barco, 2018) or social affiliation and perceived competence/reliability (respectively, likeability and integrity and capabilities; Stower & Kappas, 2020). Thus, while the study of child-robot trust is still in its infancy, the importance of studying children's trust towards robots from a variety of perspectives is evident. This thesis adds to the existing knowledge base, by investigating how children trust robots that make mistakes. This literature review is also expanded in Chapter 7.

2.4 Chapter Summary

In this chapter, we provided an overview of existing research in the fields of Human-Computer Interaction and Human-Robot Interaction (including Child-Robot Interaction). These fields are largely multidisciplinary, and the research within them is just as varied as the fields themselves. However, the common thread between the work reviewed above is the social nature of people and trust. Individuals relate socially to one another, but they also do towards machines and systems, and even robots. In turn, treating machines and robots as social beings means that they can elicit a number of emotions and reactions from people, including trust. This is especially true of young children, who are not always able to differentiate between live beings and life-like things such as computer agents and robots.

Trust, however, is a complex multidimensional construct affecting almost all daily life interactions. We unpack this construct in Chapter 3, providing insight into what it is and how it is affected. We then relate our research on children's trust towards erroneous robots in Chapter 4, Chapter 5, and Chapter 6, and develop a framework of child-robot trust in Chapter 7. We conclude with a discussion of our work and implications in Chapter 8, and concluding remarks in Chapter 9.

Chapter 3 Children: Development and Trust

The topic of this thesis being children's trust towards robots, it is essential to examine what trust is, the different types of trust that exist, and specifically how children trust others. However, trust is multifaceted and investigating it within our context (i.e., of young children) requires the combination of knowledge from many areas of research such as Psychology, Sociology, and Child Development. In this chapter, we provide background on trust from these fields, with the goal of informing our research as well as how children may generally trust robots.

3.1 Child Development

Many theories of child development exist, each providing a different perspective and knowledge base for the study of children, some of which can be applied to child-robot interaction and child-robot trust. Although discredited in modern Psychology, initially, proponents of psychoanalytic theory proposed how the unconscious mind might be shaped during early childhood, impacting development. Freud, for example, developed the (now largely disproven) psychosexual stages of development, fixating on how the focus of pleasure changes from birth through puberty (Santrock, 2008). Erikson, on the other hand, although also a psychoanalyst, believed that human behaviour is motivated by social aspects, as opposed to sexual ones. He developed the psychosocial theory of development, detailing human development from the time we are born, through late adulthood. According to Erikson, individuals go through 8 stages of development, the first five taking place in childhood: learning to trust (birth to age 1), autonomy (1-3 years), initiative (3-5 years), industry (6-10 years), and identity (10-20 years). During any of these stages, development can progress in a positive or negative way, for example, by learning to trust or mistrust people and the world around us (Erikson, 1950).

Psychoanalytic theories may help provide insight into how different aspects of development and life may shape children's trust towards others, including robots. For example, a young child might be likely to trust freely according to these theories, but as they develop that might change if they have bad experiences.

Cognitive theories of development, unlike psychoanalytic theories, pay attention to how the cognitive and thinking systems of children develop. Piaget, one of the most recognized researchers within this area, believed that children develop in four stages (Piaget, 1954). According to him, from birth to age 2 children begin to understand the world by coordinating sensory experiences with physical actions (sensorimotor stage). From ages 2 to 7, children start to form stable concepts and begin to reason, and are able to represent the world with images, words, and drawings (preoperational stage). During ages 7 to 11, children develop the ability to perform concrete operations and reason logically about concrete examples (concrete operational stage). Lastly, from 11 years of age children are able to reason about more abstract concepts (formal operational stage). As such, Piaget begins to provide us with a concrete idea of how children's cognitive abilities develop. However, additional cognitive theories argue that other aspects are also responsible for development. Vygotsky's sociocultural cognitive theory emphasizes the roles that culture and society play on development, stating that cognitive development involves learning to use the inventions of society (e.g., language), and can therefore differ depending on culture (Vygotsky, 1962). Information processing theory instead, it poses human cognition as similar to that of a computer processor, where individuals manipulate and monitor information, and then strategize about it, to different extents depending on age (Santrock, 2008).

Cognitive theories of development can be extremely valuable to studying child-robot trust, as they suggest the changes that take place in children's brain development. For example, a child in the preoperational stage (Piaget) may view a robot as being alive, and having thoughts and emotions, however, a child in more advanced stages may better understand the underpinnings of the robot, and therefore that it is no more than a machine. Information processing theory may provide a similar point of view. In addition, depending on the culture that a child grows up in, their perspective on robots can greatly differ, simply due to how their culture talks about and represents technology (Vygotsky). Cognitive theories of development have the potential to be of great assistance in further understanding how children's trust develops, and how such trust may specifically develop towards robots.

Although not core to our current research questions, and in contrast to cognitive theories, behavioural theories of development completely ignore cognition, instead placing their focus solely on behaviour. Classical conditioning, pioneered by Pavlov (Pavlov, 1960) and expanded by Watson (Watson & Rayner, 1920), highlights how the pairing of stimuli can lead to behavioural responses, such as fear. For example, by pairing a stimulus (e.g., rat) with a fearsome stimulus, young children will learn to be afraid of the first stimulus (i.e., the rat). Operant conditioning, developed by Skinner, instead states that the consequences of behaviour can affect the likelihood of its future occurrence (Skinner, 1938). In this view, rewards and punishments, not thoughts and feelings, shape child development. Social cognitive theory (Bandura, 1986) pieces together some of the theories discussed above, stating that not only behaviour, but also cognition and the environment, are key factors in development. By observing their environment, children (as well as adults) acquire a large variety of behaviours, thoughts, and feelings, which they then cognitively represent, and sometimes adopt themselves. In this way, the environment, cognition, and behaviour are all linked to each other.

When it comes to robot trust, behavioural theories can provide insight into special instances or extremes, where children may have been primed to hold certain views about robots and technology, or how previous experiences may affect future ones. Social cognitive theory in particular, has great potential to be applied to child-robot trust, as it suggests that children can easily pick-up behaviours from their environment, something that needs to be seriously considered in relation to the types of behaviour and information that robots produce.

Other theories of child development also exist, which shift the focus to biology and ecology. Ethological theory, for example, states that behaviour is highly influenced by biology, is related to evolution, and is characterized by several sensitive periods (Bowlby, 1969; Lorenz, 1965). This theory suggests that the attachment that an infant possesses with their caregiver highly impacts the rest of their life. Ecological theory, however, views development as the reflection of influences from environmental systems, such as the child's home, school, and neighborhood, as well as the culture in which they live (Bronfenbrenner, 1986). These theories present evidence as to how biological development can impact trust development, and how a child's environment including the people they interact with on a regular basis, can greatly impact how they perceive things, including robots.

While all of the above theories of development differ in their focus and claims, there exist aspects of each of them that we can learn from to inform how children may develop trust towards robots. Cognitive theories (i.e., Piaget, Vygotsky, Information Processing), for example, are likely to be the most applicable to how children trust robots, as they could relate how children's cognition may affect their understanding of robots and their actions. Behavioural, social cognitive, and ecological theories though, have the potential to provide us with clues as to how the interaction environment in which a robot and a child engage, or how the robot and others react may play a role in the child's trust development. Psychoanalytic theories of development can provide some insight into what individual differences may be present between children.

3.1.1 How trust develops in children

Exploring how trust develops in childhood can provide insight into how the age and mental and emotional development of a child affect the type and amount of trust they might develop (towards people or robots). For our work, it is integral because debate exists as to whether trust is an innate part of the human experience, something that we are born with as children (Erikson, 1950), or is a learned skill, that we perfect with experience (Rotter, 1967). By around 16 months of age, infants are already able to distinguish between true and false statements and reject false claims. However, during any of the stages of development (which there are many of), growth can progress in a positive or negative way, for example, by learning to trust or mistrust people and the world around us (Erikson, 1950). As time passes, children learn to rely and seek information from those who have proven to be reliable (i.e., have provided accurate information in the past; Koenig & Harris, 2005b). This is thanks to a development in the understanding of epistemic states, also referred to as *theory of mind*, where an individual is able to perceive others' mental states (Premack & Woodruff, 1978). While most adults possess this ability, children must still develop it, leading trust in children to develop and take place differently depending on what developmental stage they are at, and of course distinctly from most adults. This early development can provide some insight on how children may trust robots in different situations.

While not all theories of development refer to trust, many of them can still provide insight into how child development relates to trust development (and therefore how it might to child-robot trust development). In this paragraph we provide our interpretation of how the theories mentioned above may apply to trust. For example, we believe that Piaget's theory of cognitive development (Piaget, 1954) may point to how children in the sensorimotor stage (birth to 2 years) may trust freely or blindly, while a child in the more advanced stages of preoperational (2 to 7 years), concrete operational (7 to 11 years), or formal operational (11 years on) may learn to deliberately trust (or not). Similarly, Vygotsky's sociocultural cognitive theory emphasizes the roles that culture and society play on development, stating that cognitive development involves learning to use the inventions of society (e.g., language), and can therefore differ depending on culture (Vygotsky, 1962). According to this theory, children may have different perceptions of robots (and we think therefore trust them in different ways or to different extents) depending on how the culture they are raised in views robots and technology. Social cognitive theory expands on this, and posits that children (as well as adults) acquire behaviours, thoughts, and feelings by observing their environment, which they then cognitively represent, and sometimes adopt themselves (Bandura, 1986). This theory would suggest that trust develops depending on what children witness and experience in their environment, and what they choose or are taught to adopt themselves. Theories based on biological or ecological processes (e.g., Bowlby, 1969; Bronfenbrenner, 1986; Lorenz, 1965) similarly focus on the child's environment, suggesting that the attachment that a child has with their caregiver or the type of environment where they live and go to school affects their development, and therefore is likely to influence their development of trust. The variety of ways in which child development can impact the likelihood of building trust is an important aspect to consider when studying child-robot trust, as it serves as a starting point of knowledge.

3.1.2 Theory of mind

One of the biggest factors affecting children's ability to trust and its progression is the development of theory of mind – the ability to perceive other's mental states (Premack & Woodruff, 1978). While children as young as two years old are able to reason and perceive somebody else's desires and actions, they are not yet able to grasp the concept of beliefs (Wellman & Woolley, 1990). However, many three-year-olds, and most four-year-olds, start being able to tell the difference between what an individual desires versus what they believe (Wellman & Woolley, 1990). For example, if a character observes an object being put into a box, and then while they are away the object gets moved, a child younger than four years old will likely state that the character will look for the object in the new spot, while a five-year-old will almost always comprehend that the character does not know that the object was moved, and will therefore look for it in the original box. This type of task is common in theory of mind research because, although many argue that disadvantageous to some, it shows the presence of false belief – the difference between contents of the mind and how the world is (Wellman, 2002).

The development of theory of mind is extremely important in relation to trust in young children, as whether a child has developed this capacity will inform the data they are basing their trust on. In addition, multiple experiments have found a relationship between theory of mind development and who or what children tend to trust (DiYanni & Kelemen, 2008; Moore, Pure, & Furrow, 1990). Theory of mind can also provide some clues as to how children may perceive and trust robots. Children with more developed theory of mind might be able to choose whether to trust a robot based on previously obtained information, while a child with less theory of mind may choose randomly or not at all (relying on others' trust decisions). This suggests that when studying child-robot trust, we must pay careful attention to the age and developmental progress of children, as any findings that are obtained can be greatly implicated by the presence, or lack thereof, of a theory of mind. We specifically focus our research on this age group to investigate how these developmental changes might affect trust towards robots.

3.1.2.1 Theories of theory of mind

Several theories of theory of mind exist, suggesting various ways in which theory of mind develops and takes place in children. These theories can shed light into how children of different ages and with different experiences may develop and exhibit trust. *Theory theory* (Gopnik & Wellman, 1994), for example, suggests that the way we think about the mind is not scientific, but instead an informal everyday framework. According to *theory theory*, children begin with a desire psychology where they perceive others' mental states mainly as desires. They later progress onto desire-belief psychology – where children may be able to reason about beliefs, but speak about actions in terms of desire (Bartsch & Wellman, 1995) – and then belief-desire psychology – when they are aware that what individuals believe and their desires can greatly impact their actions. Instead of beliefs and desires, *modularity theory* focuses on the acquisition of theory of mind through neurological pathways that are created through experiences (Leslie, 1994). Modularity theory therefore places a greater emphasis on hard-wired components of theory of mind (Flavell, 2004). *Simulation theory*, however, states that children learn to process the mental states other people through a role-playing simulation process that goes on in their heads (Harris, 1992). As children continue to practice this skill, their simulation abilities and therefore theory of mind improve, according to simulation theory. These three theories highlight different approaches to theory of mind research and understanding: *theory theory* provides a perspective from observable behaviour, *modulation theory* focuses on neurological changes in the brain, and *simulation theory* proposes an explanation of the inner workings on the mind. Overall, theories of theory of mind provide different standpoints on how children may experience this concept, which is so highly related to the ability to trust.

3.1.3 Children's trust towards informants

The topic of how children trust has been explored under a few contexts, with a large focus on trust towards informants – entities that provide some sort of information. There have been several experiments examining children's trust towards human informants in Developmental Psychology. Informants in this case are individuals or beings who know or provide information about a specific area of knowledge (e.g., which toys are the best, what a screwdriver is used for, etc.). Corriveau and Harris (2009), for example, conducted an experiment to examine whether children (aged 3-4 years old) trust information provided by a familiar informant more so than an unfamiliar one. The researchers found that children trusted the answers provided by the familiar informant (a teacher from their school) more so than those provided by the unfamiliar informant. In a similar experiment (Koenig & Harris, 2005a), children aged 3-4 years old watched videos in which two female actors labelled objects, one correctly, and the other one incorrectly (i.e., was ignorant). The researchers found that children trusted the correct informant more than the ignorant one when they were tasked with labelling novel objects. Other experiments suggest that this finding continues to apply after a week. That is, children continue to trust a previously accurate informant more than an inaccurate one, a week after being exposed to the accuracy information (Corriveau & Harris, 2009b). These studies show that the familiarity and accuracy of an informant can play a large role in whether children will trust the informant or not, providing one factor that can affect children's trust towards in formants, and therefore contribute to our research. Specifically, this research provides a methodological basis for our investigation and reinforces the importance of the children having the same level of familiarity with the two robots.

Other research has investigated whether children trust adult or child informants differently based on the types of information they need to provide. VanderBorght and Jaswal (2009) asked children aged 3-5 years old for help answering questions about food and toys. Being able to ask for advice from an adult or a child, children trusted adults with information about food, but other children with information about toys. Other experiments have also found that children trust information from adults and children differently. Jaswal and Neely (2006), for example, conducted an experiment in which they compared children's trust towards a child and an adult informant in three different cases: when only the adult informant was reliable, when only the child informant was reliable, and when the adult and child informants were both reliable. They found that children trusted the adult informant more in some situations (i.e., when the child and adult informant were reliable, and when only the adult informant was reliable), but the child one in others (i.e., when only the child informant was reliable). This research provides insight into how children perceive informants of various ages, and can therefore inform how children may view robots. For example, children may perceive and trust robots that are seen to be of different ages or at different life stages to varying extents, depending on how their standing relates to the children's.

Researchers have also investigated how children react to non-living informants, such as puppets, which can provide additional insight into how children may respond to robots (who may not be as “alive” as humans, but more so than puppets). Birch, Vauthier, and Bloom (2008), for example, conducted an experiment in which 3- and 4-year old children observed two puppets name objects. Throughout the experimental session one of the puppets always provided correct labels for objects, while the other puppet always provided incorrect labels. The researchers found that even when children did not know what the new objects were called, they were more likely to trust the puppet that had previously labelled familiar objects correctly than the one that had not. The experimental methodologies used in this experiment have become a standard for investigating children’s trust of informants, and other studies such as those mentioned above (Corriveau & Harris, 2009a; VanderBorght & Jaswal, 2009) as well as others (e.g., Fusaro, Corriveau, & Harris, 2011; Kushnir, Vredenburgh, & Schneider, 2013; Li, Heyman, Xu, & Lee, 2014) have built off of it. We also build on this research, modifying and extending it with robots in our experiments.

In a similar experiment, for example, Kushnir, Vredenburgh, and Schneider (2013) tested whether previous knowledge about whether a puppet was good at fixing things or labelling things would affect which task children trusted the puppets to do. They had two puppets, one that correctly labelled items, but incorrectly stated how to fix them, and one that labelled items incorrectly but was able to fix the objects. They found that children trusted the labeller puppet when they needed to label objects, and the fixer puppet when they had to fix objects. This research suggests that young children are able to discern between the types of information that informants are knowledgeable about (through the development of theory of mind), and can adjust their trust of the informants accordingly.

Previous research has also examined whether children trust informants differently depending on the informants' access (or lack thereof) to information. Nurmsoo and Robinson (2009) conducted an experiment in which puppets provided children with information when trying to indicate the colour of a toy (that they had limited access to), and whether it was hard or soft. Researchers found that 3 and 4 year-olds trusted a puppet that had been previously incorrect due to lack of information more than the one that was previously incorrect but well informed. These findings suggest that previous mistakes may be excusable if the information was not accessible by the informant. It therefore may be that if an informant makes an error, and children are aware of the reason for the error, that their trust towards the informant might not be affected by such error. This is highly relevant to our work on trust towards erroneous robots, as robots may have access to different types of information and abilities than humans, similar to the puppets in this experiment.

This body of work sheds light into how children may trust human informants differently based on previous social experience with them. Overall, this work provides invaluable insight into how children may trust robots. For example, the familiarity of a robot, its apparent age, and its availability to knowledge can affect children's trust towards the robot in the presence of errors. Thus, we can base CRI trust research on Developmental Psychology research, and evaluate how interactions with robots differ from those with people. However, children do not conceptualize robots as they do people, and sometimes treat robots in ways in which they may not humans (e.g.,

Table 1

Summary of Children's Trust Towards Informants Literature

Citation	Informant type	Independent variable (informant)	Age (years)	Medium (in-person, video, etc.)	Trust results
Birch, Vauthier, & Bloom (2008)	Humanoid puppet	Accurate/ inaccurate	3-4	In-person	Accurate > inaccurate
Corriveau & Harris (2009)	Human	Familiar/ unfamiliar	3-4	Video	Familiar > unfamiliar
Corriveau & Harris (2009b)	Human	Accurate/ inaccurate	3-4	Video	Accurate > inaccurate (after 1 week)
Fusaro, Corriveau, & Harris (2011)	Humanoid puppet	Accurate/ inaccurate, strong/weak	3-5	Video	Accurate > inaccurate for accuracy tasks, strong > weak for strength tasks, accuracy and strength not related
Jaswal & Neely (2006)	Human	Adult/child, accurate/ inaccurate	3-4	Video	Adult > child when (1) adult accurate and (2) when adult and child accurate child > adult when child accurate
Koenig & Harris (2005)	Human	Accurate/ inaccurate	3-4	Video	Accurate > inaccurate
Kushnir, Vredenburgh, & Schneider (2013)	Animal puppet	Accurate/ inaccurate, labeler/fixer	3-4	In-person	Labeler > fixer for labeling task, fixer > labeler for fixing task
Li et al., (2014)	Animated character	Honest/ dishonest	3-4	Video	Honest > dishonest in older children
Nurmsoo & Robinson (2009)	Animal puppet	Accurate/ Inaccurate, informed/ uninformed	3-5	In-person	Uninformed & inaccurate > informed & inaccurate
VanderBorght & Jaswal (2009)	Human	Adult/child	3-5	Hypothetical	Child > adult for toys, Adult > child for nutrition information

Brščić, Kidokoro, Suehiro, & Kand, 2015). Table 1 summarizes previous research on children's trust towards informants.

3.2 Unpacking Trust

While a plethora of types and definitions of trust exist, there remain disagreements within academic fields as to what trust truly is, and it when it may or be not be present. Some researchers believe that trust is a belief about the average character of another person (Rotter, 1971), referring to their general beliefs and behaviours. On the other hand, some maintain that trust is a larger concept, greater than any one individual, and pertaining to the social system and society as a whole (Lewis & Weigert, 1985). Furthermore, others view trust as the result from positive expectations, such as the confidence that one will obtain what they desire from others, as opposed to what they fear (Deutsch, 1958), one's willingness to rely on others because of the expectation that others will provide gratification (Scanzoni, 1979), or that trust happens when there are multiple satisfying interactions with another individual (Driscoll, Davis, & Lipetz, 1972). Yet other researchers do not believe that trust is innate or relies on personal experiences, but instead argue that it mainly depends on situational variables (Lewis & Weigert, 1985). From the literature, it is clear that many definitions of trust exist but they do not always coalesce. This makes trust research difficult, since depending on the context and individuals involved, some researchers would qualify certain actions as trust while others may not. Nonetheless, we can summarize the existing research (some presented above and below) to agree on the necessity of risk as a requirement for trust, and a calculated decision as the demonstration of trust. Thus, while there seem to exist many variants of trust, they at least have these two elements in common.

3.2.1 Types of trust

Research suggests that various types of trust exist in varying contexts and with different sets of people. Some of these types of trust might be present in some situations but not others, calling for the exploration of all varieties to assess when they should be considered. It is especially important when conducting research about children, because they may experience trust differently (e.g., relying on others to make trust decisions for them, being unable to conceptualize certain types of trust, etc.) Below, we summarize the different descriptions of trust that we found in the literature. We group them into three categories depending on what area of life the trust is towards. However, some overlap exists as these are aspects that researchers utilize to consider trust. In this thesis, we focus on interpersonal trust, as described below and in Chapter 7. We present this synthesizes here to illustrate the large scope of the topic of trust. We discuss and interpret these types of trust further in Chapter 7.

3.2.1.1 Personal types of trust

Interpersonal trust. This type of trust is characterized by a general expectancy that others will keep their word (Rotter, 1971). It has two components, cognitive-based trust and affective-based trust (McAllister, 1995), which control the extent and situations in which people experience interpersonal trust. This type of trust influences how much we generally trust others, and can affect our everyday lives.

Dyadic trust. In psychological research, dyadic trust refers to that which is present in intimate, sometimes romantic, relationships between two individuals (Larzelere & Huston, 1980). This type of trust is a subset of interpersonal trust, and is therefore affected by it. Thus, it relies on the belief that the other individual is honest and benevolent, and consequently encourages us to be vulnerable with them.

Relational trust. This type of trust develops due to repeated and positive interactions with other people (Lewicki, Tomlinson, & Gillespie, 2006; Rousseau, Sitkin, Burt, & Camerer, 1998). It does not rely on having a specific relationship with individuals, and can therefore be applied to people we know well, as well as those we have interacted with multiple times but are not close to. While positive interactions lead to relational trust, negatives ones can certainly hinder it.

3.2.1.2 Systematic types of trust

Deterrence-based trust. This type of trust exists when trust is placed due to the possible negative consequences of the trustee breaking such trust (Shapiro, Sheppard, & Cheraskin, 1992). Deterrence-based trust does not rely on qualities of a specific individual or body, but instead is present due to outer rules and expectations. Thus, we may trust a person or institution because of practical reasons, rather than prior information or connections.

Institution-based trust. This type of trust is based on group membership. An individual, item, service, etc. might be trusted because it is linked to a specific institution (e.g., company, school, group; Lewicki et al., 2006). Institution-based trust is used often, as it removes the need to make a trust decision and instead relinquishes the power to the institution.

System trust. Extremely important in everyday life, system trust is the sense that everything appears to be in good order (Luhmann, 1979). This type of trust is essential for society, as life as we know it would not be possible otherwise (Lewis & Weigert, 1985). System trust can certainly differ based on geographical location, culture, customs, etc. System trust is different from institution-based trust in that "system" does not refer to a specific institution of sorts, but is more general and can be applied to various areas of society.

3.2.1.3 Societal types of trust

Cognitive trust. Highly thought-through, cognitive trust relies on a person's cognition rather than emotion (Lewis & Weigert, 1985). Careful analysis and deliberation of potential risks and rewards of trust are the cornerstones of cognitive trust, and little else (e.g., emotion) is considered.

Emotional trust. This type of trust is based on the feelings and emotions of the trustor (Lewis & Weigert, 1985). Thus, trust is less of a decision, since thoughts are not taken into account or evaluated. Instead, an individual trusts because their emotions suggest they should.

Ideological trust. Ideological trust is based on the combination of emotional and cognitive components in a way that they complement each other, such that emotions can provide support for cognitive trust or cognitions justify emotional trust (Lewis & Weigert, 1985). This is often witnessed in organized religion, where there is an emotionally charged context and cognitive justifications (and vice versa).

Mundane trust. Everyday trust, which we place on people and society without the need to think or feel anything about it, is referred to as mundane trust (Lewis & Weigert, 1985). This type of trust is low on cognition and emotion, and relies on routine aspects of daily life that people have become accustomed to.

3.3 Chapter Summary

To investigate children's trust towards robots, it is essential to first understand child development and the specific qualities that make children of different ages and adults distinct. In this chapter, we presented research on how children develop psychologically and how development affects trust. In addition, we provided a summary of relevant literature on trust, to guide our research. The research presented suggests that various types of trust exist, and the different ways in which children may respond to informants. While robots differ from such informants, in physicality as well as the types of trust they may elicit, our research is inspired by that presented in this chapter, and follows similar paths of investigation. In the following chapters, we present our experimental design (Chapter 4), and two experiments (Chapter 5, Chapter 6) investigating children's trust towards erroneous robots. We then present a framework of child-robot trust (Chapter 7), and conclude with a general discussion on our research and ways to move forward (Chapter 8 and Chapter 9).

Chapter 4 Design of an Experiment to Evaluate Children's Trust Towards Erroneous Robot Informants

The broad goal of this research program is to investigate how young children react to robots making errors, and how this affects their trust towards the robots. As discussed in Chapter 2, robots are generally prone to making errors and researchers do not yet know how children react to these robot errors, and specifically how such errors affect trust. Thus, to be able to conduct this research we require an experimental design that: (*i*) is suitable for preschool-aged children, (*ii*) is feasible to conduct with robots, and (*iii*) enables us to measure children's trust. These requirements stem from our goal of investigating how preschool-aged children (requirement *i*) trust (requirement *iii*) erroneous robots (requirement *ii*). In this chapter we present an adapted experimental design based on research from Developmental Psychology (Birch et al., 2008), that enables us to investigate how robot errors impact preschool-aged children's trust towards the robots.

4.1 Validated Experimental Design

We borrowed from an experimental design from Developmental Psychology which follows the idea that two informants (typically either humans or puppets) with differing attributes (e.g., accuracy levels) label objects in front of a young child. The child is then asked to label the objects, with their indications or selections serving as an indication of which informant they trust (Birch et al., 2008). The independent variable is presented through the labeling task, where an informant makes mistakes or possesses differing characteristics that may affect children's trust (e.g., familiarity, age, motive, etc.). Trust is measured through the actions of the child, that is, which informant they choose to side with. The experiment has several phases through which the child observes the informants' abilities (and demonstrated differing behaviour or characteristics) and makes judgements about which one to trust, providing researchers with a measure of trust in varying informational situations. A detailed explanation of the experimental design, as well as the modifications that we made to it, can be found in the remainder of this chapter.

4.2 Modifications to Original Experimental Design

To investigate how robot errors influence young children's trust in robots, we adapted and modified a validated experimental methodology from Developmental Psychology (i.e., Birch et al., 2008). This experimental methodology was originally devised to examine young children's trust towards human or puppet informants, some of which make mistakes. Thus, it already meets two of our requirements above: it is suitable for young children (requirement *i*) and is intended to measure trust (requirement *iii*). However, the experimental design needs to be modified to be conducted with robots, as robots possess specific qualities that require special consideration. For example, robots have characteristic motion patterns which can affect possible behaviours, and a different type of embodied presence and perceived liveness than do humans or puppets. Certain modifications to the original experimental methodology are therefore necessary for use with robots.

4.2.1 Robot script

One of the required adaptations is to the informants' (in our case, robots') scripts to take into account the sounds that robots can and cannot make. Some robots have a limited range of sounds, and are not always able to properly annunciate some of the common sounds in any given language. Through trial and error with the robots we use in our experiments, we found that certain words needed to be modified (e.g., words with multiple syllables because the intonation is often incorrect). Thus, certain words in the robots' scripts were modified to ensure their understandability by children. A full version of the script can be found in Appendix A (experiment 1) and Appendix B (experiment 2).

4.2.2 Remote robot control

We implemented the robots' speech and movements through a Wizard-of-Oz (Dahlbäck, Jönsson, & Ahrenberg, 1993) approach, where the robot is remotely controlled by a research assistant. While this might appear to be a similar approach to the puppets used in psychological research (e.g., Birch et al., 2008), in the Wizard-of-Oz implementation children are not aware that the robots are controlled by a human, and they therefore appear to be autonomous. This is different from puppets, who are often visibly puppeteered by humans. In addition, remotely controlling robots requires prescribed motions and actions based on the robot's capabilities and morphology, to ensure that it is able to move properly when it is wizarded. This adds an additional layer of complexity when conducting research with robots as opposed to puppets or humans.

4.2.3 Robot introduction and goodbye

To help make children feel more comfortable during the experiment (since they have likely never interacted with similar robots), we introduced them to the robots about a week before the experiment took place, by hosting an introductory session with a similar robot and the group of the preschool children at each daycare centre (a more thorough description is provided later in the chapter). This addition to the experimental methodology is necessary with robots due to their novelty, as we often find that children can sometimes be scared, intimidated, or even overly excited during a first meeting.

In addition, we provided children with a chance to say goodbye to the robots at the end of the experimental session, where the robots and the child made some final remarks before the robots went "back to sleep." This allows children to have closure with the robots, whose novelty may make it difficult to part with.

4.2.4 Identical Informants

The original experimental methodology in Psychology utilizes two informants that have different appearances and voices to make it easy to distinguish between them. However, small details such as the gender of a robot, its voice, or even its colouring can have an impact on how children perceive it (Sandygulova & O'Hare, 2018; Walker, Girotto, Kim, & Muldner, 2016). Therefore, we intentionally made the robots look and sound the same, with the only defining characteristics being their gender-neutral names and that one robot exhibited errors, to reduce the likelihood of children exhibiting a bias.

4.2.5 Number of Trials

The series of experiments we base our work (Birch et al., 2008) on typically utilize 3 trials per experimental phase, to enable one informant to be trusted more than the other (i.e., 2 out of 3 trials). However, the 3-trial approach does not allow for a distinction between amounts of trust towards the informants. For example, a child may in reality trust both robots equally, and simply take turns siding with each. Thus, our experimental design incorporated 4 trials in each experimental phase, enabling children to side with each robot an even amount of times, thus demonstrating a similar level of trust in the information received from the two robots. Furthermore, a 4-trial approach enables us to gauge the amount of trust that a child exhibits towards the robots, as there is one more opportunity to demonstrate it. This provides us with a clearer view of children's trust towards the robots in our experiments.

4.2.6 Robot Errors

The original experiment design features two puppet or human informants, one of which in some cases makes informational mistakes (i.e., provides incorrect information). These are what we refer to as human-typical errors, since they are mistakes that people can commonly make but may not be as common in robots. For example, robots make a lot of speech-recognition errors (Mubin, Henderson, & Bartneck, 2014; Novoa et al., 2018), especially with children due to differences in their speech and non-representative robot training (Kennedy et al., 2017). These are errors that are common of robots but not necessarily of humans; we therefore refer to them as robot-typical errors. We built on previous research by investigating both of these types of errors (i.e., human-like and robot-typical) with robots, expanding the original experimental design to make this possible.

4.3 Methodology

4.3.1 Target Participants

Our target participants were children (3-5 years old), who are fluent in English. We selected this age group due to developmental changes that take place around this age, specifically, the development of theory of mind (i.e., the ability to attribute mental states to themselves and others; Premack & Woodruff, 1978). As children develop theory of mind, it enables them to take into account others' mental state (human or perhaps non-human) and prior knowledge, to inform interaction, such as when deciding whether to trust an informant (e.g., Nurmsoo & Robinson, 2009). We specifically targeted children in this age group to investigate how young children assign trust to robot informants.

To allow children to become better acquainted (and less scared of) the robots and experimenter, the children were able to meet a similar robot and the experimenter beforehand, during a group introductory session. The introductory session took place a few days before the experimental sessions, and it was a chance for the children at the daycare to acclimate themselves to the experimenter and robot in a fun and psychologically safe environment with their peers. During these introductory sessions, children were able to interact with a robot (controlled by a research assistant through a Wizard-of-Oz implementation) and ask it questions. The goal of the introductory session was to acquaint the children with a similar robot and the experimenter to make them more comfortable during the actual experimental session, and was highly recommended by child experts in the research team.

4.3.2 Materials

4.3.2.1 Robots

For this experimental platform we used two SoftBank Nao H25 V5 robots (Figure 10), which are 23 inches high, and widely used in HRI research with children (e.g., Bethel, Stevenson, & Scassellati, 2011; Chevalier et al., 2017; Johal, Jacq, Paiva, & Dillenbourg, 2016; Pulido et al., 2017; Shamsuddin et al., 2012; Tielman, Neerincx, Meyer, & Looije, 2014; Wit et al., 2018). We utilized this type of robot for a variety of reasons. First, the size of the robot, being smaller than the average 3-5 year old child, and its friendly appearance, enable experiments with children to be conducted with little to no fear or hesitation from the children. Other robots, such as the SoftBank Pepper robot (4 feet tall) which was created by the same company, may be intimidating for young

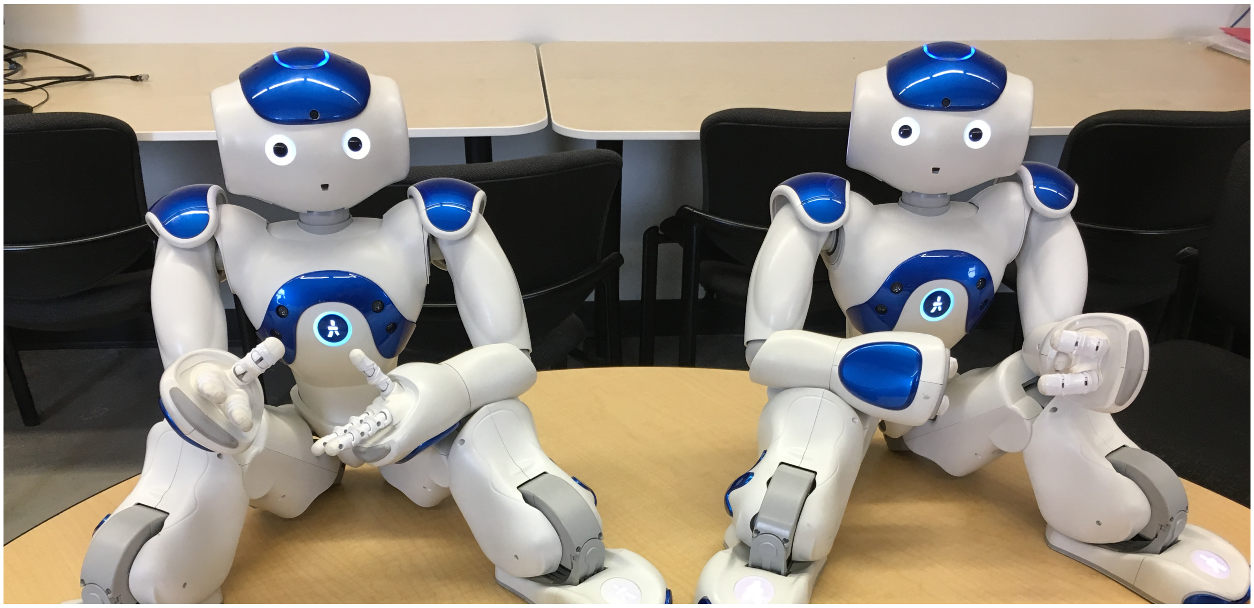


Figure 10. SoftBank robots used in our experiments. The robots look and sound the same, to avoid preference towards one of the robots. These robots are commonly used in research with children.

children. We could have instead used robots with other types of embodiments, such as animal-like robots (e.g., Sony's Aibo, or the PARO robot). However, an animal-like robot that provides information may not be perceived similarly enough to how human-like puppets, or even human informants are perceived, making it a large deviation from previous research. Further, humanoid robots are often touted as the solution for the care and companionship of young children, and research with adults suggests that they elicit more trust than other form factors (Vattheuer et al., 2020). Lastly, as explained earlier, Aldebaran Nao robots are heavily used in research with children, and we can therefore expect young children to interact with them naturally and not be afraid of them, seeing that other researchers have not encountered these problems. Thus, humanoid robots, and Aldebaran Naos in particular, were an acceptable choice for this research.

The two robots we used had the same white and blue design, and behaved and sounded the same. This was a purposeful choice we made to ensure that children were not biased towards one of the robots because of its appearance, gender, etc., which is common with robots (Sandygulova

& O’Hare, 2018; Walker et al., 2016). The robots were therefore only differentiable by their names (Casey and Taylor, purposefully gender neutral) and any errors that they exhibited.

The robots were controlled through a Wizard-of-Oz implementation (Dahlbäck et al., 1993), where a robot is controlled by a research assistant who is present in the experimental room, but does not appear to be affiliated with the robot. We utilized this approach, instead of autonomous robots, for a number of reasons. First, our primary goal was to be able to control the robots and their behaviour at all times, to make our experiments more methodologically sound and controlled. Autonomous robots, on the other hand, were unlikely to stick precisely to the interactions scripted in our experiments, and furthermore may have exhibited errors that were not intentionally related to the experiments. Moreover, implementing the necessary skills and behaviours for robots to be autonomous for our experiments would have taken a considerable amount of time (perhaps years). While observing how an autonomous robot authentically makes errors while interacting with a child would be an interesting topic of exploration, as would the implementation of autonomous behaviours, neither was the focus of our research. Thus, the robots in our experiments were remote-controlled by a trained research assistant.

4.3.2.2 Objects

During the experiment, the experimenter showed various objects to the two robots, and asked the robots to label them. During certain parts of the experiment, the child observed this take place, and then had to choose which robot’s information to trust. The experimenter placed the objects, one or two at a time, on a table between the child and the robots; the child and experimenter sat at the table, and the two robots on it. The robots labeled a total of 20 objects in front of the child; some of these objects were familiar to the children, while we intended for others to be unfamiliar.

Object selection was guided by a number of principles: 1) familiar objects needed to be easily labelled by young children, 2) unfamiliar objects should not be easily labelled by young children because the experimental methodology required children to not be aware of these objects' names, and 3) familiar and unfamiliar objects should be similar to those validated in previous related research (e.g., Birch et al., 2008).We took inspiration for these objects in terms of size and what they are used for from previous research. The search process for the objects involved browsing through local dollar stores in search for objects that even the experimenter may not know what they are called or used for, but that would be safe for children to handle. In this way we arrived at

Figure 11. Presumed unfamiliar objects labelled by the robots during the experiment. From top left to bottom right: dog waste scooper, turkey baster, peach slicer, GPS/phone car holder, honey dipper, faucet aerator, dog toy, door stop, pineapple slicer, tape dispenser, display plate holder, garlic press, lobster cracker, towel holder, measuring spoon, and door spring. Children may have seen these objects before, but we would expect them to not know the names of them.

the following objects. The 4 familiar objects were: a ball, car, dog, and plate. The 16 unfamiliar objects were: a dog waste scooper, turkey baster, peach slicer, GPS/phone car holder, honey dipper, faucet aerator, dog toy, door stop, display plate holder, garlic press, pineapple slicer, tape dispenser, lobster cracker, towel holder, measuring spoon, and door spring (see Figure 11).

4.3.2.3 Implementation

As described above, our experiments used two humanoid robots. The robots did not have any behaviours preloaded, but were controlled by a research assistant (i.e., the Wizard). The robots had local IPs and were connected to a dedicated laptop and a router through a closed-loop wired connection to a local area network. This enabled us to conduct our experiments in daycares, which did not have the internet connection or infrastructure that would otherwise be required. The robots' scripts and movements were implemented through an in-house platform (previously developed in our lab), in Visual Studio using C#. The platform for this experiment encompassed new behaviours to suit our experiments. During the experimental sessions, the Wizard (i.e., robot controller) used a graphical user interface to control both robots, allowing the Wizard to simply press on the appropriate buttons to make the robots move and speak as scripted (see Figure 12). Through the interface, the Wizard was able to see what the robots "see," and make adjustments to the robots' head positions if necessary. The robots continuously waited to receive commands from the Wizard and then executed them in the order received, providing similar response times for all responses.

Our implementation made it possible to conduct our experiments in daycares, which do not often have internet or wireless connection, and to experimentally control the robots' behaviours and responses.

4.3.3 Experimental Phases and Task

Our experiment was comprised of three phases, presented in order: *history*, *same label*, and *contrast label* (Figure 15). Throughout the experiment, one robot made errors (which we designate as the *erroneous robot*) and the other robot did not (which we designate as the *non-erroneous robot*). This was the independent variable of our research. Our experimental methodology was adapted and extended from that of Birch et al. (Birch et al., 2008), with the main differences being in the type of informants (we used robots), the addition of one trial in each phase (we had four trials in each phase, previous research used 3), and some additional object labels to account for the

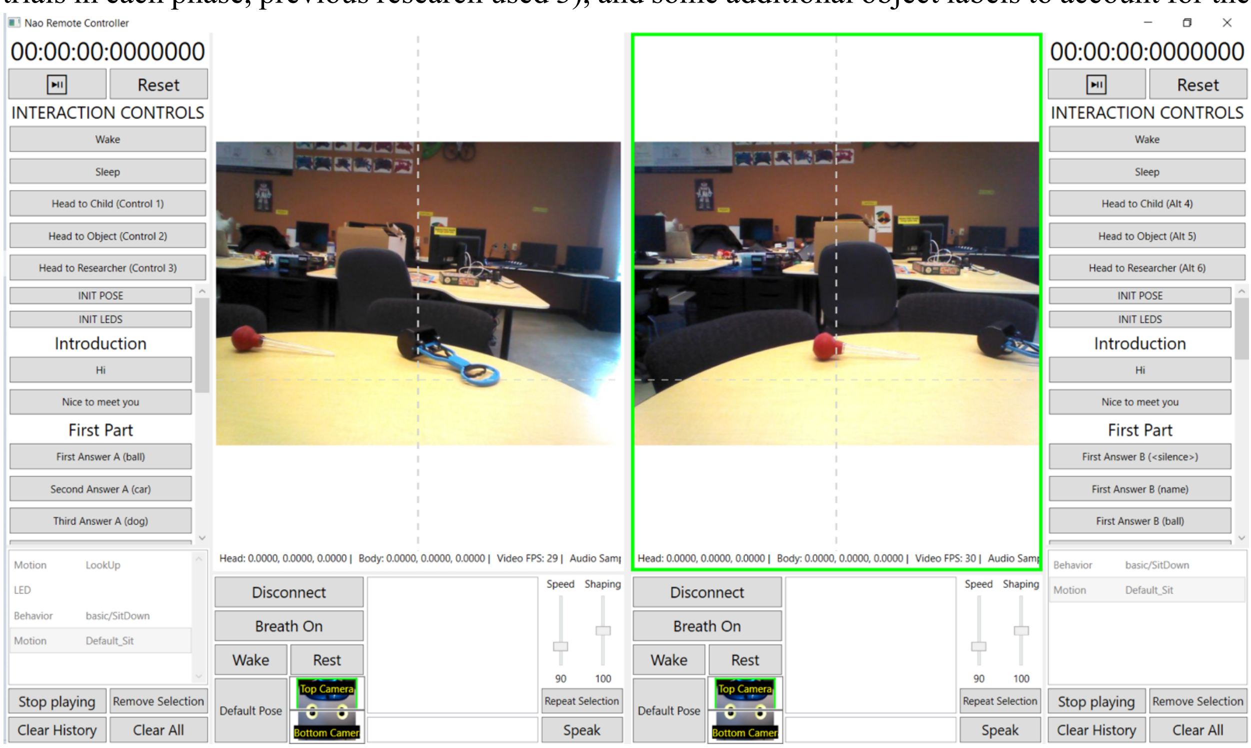


Figure 12. Graphical user interface through which the Wizard controls the two robots. Each half of the window comprises controls for one robot. The wizard clicks on a button and the robot executes that behaviour. This platform was created by a former lab student and has been used for experiments in the Human-Robot Interaction lab for the past 5 years.

added trials (see full list of modifications at the beginning of chapter and phase descriptions below). The order in which the robots spoke, in addition to their positioning in relation to the experimenter was counterbalanced.

4.3.3.1 History phase

The experiment started with this phase, which served as the initial independent variable presentation, demonstrating to children which robot was erroneous and which was not. During this phase children observed the two robots label four common objects, such as a ball. One of the robots (the *non-erroneous robot*) did not exhibit errors, while the other one (the *erroneous robot*) did. This phase was an extension of prior work with puppets (Birch et al., 2008), with the main difference being that we instead used robots. For example, when the experimenter showed the robots a ball, the *erroneous robot* labeled it as a cup (experiment 1). In addition, while prior work

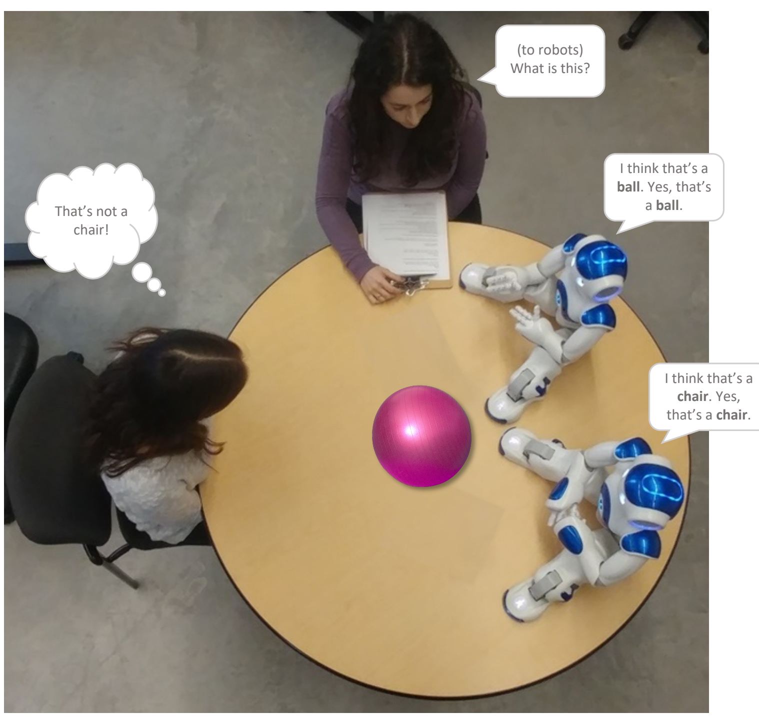


Figure 13. Example script for robots and experimenter during history phase. One robot makes errors while the other one does not.

used a 3-trial approach, we utilized 4 trials, to obtain an additional datapoint and allow for the possibility that both robots are chosen/trusted an equal number of times. Our *history* phase was otherwise methodologically similar to that of related work (Birch et al., 2008).

4.3.3.2 Same label phase

The goal of this phase was to test children's trust towards the two robots when information was simple and clear. This phase was also an extension of prior research (Birch et al., 2008), with the differences being the addition of one trial (for a total of 4), the objects shown and labelled, and some of the labels that the robots provided. During this phase, the robots labeled uncommon objects that children were not expected to know the names of, providing made-up names for them (e.g., ferber; see Table 1 for full list of labels). Some of the made-up words were obtained from previous related literature from Developmental Psychology, while others were created for this experiment. All words were comprised of two syllables, and could be easily heard and understood by children, so they were able to differentiate between different made-up words that the robots provided.

Table 1. A full list of the made-up words used by the robots and experimenter during the same label and contrast label phases.

Labels used by robots in *same label* and *contrast label* phases		Labels used by experimenter in *contrast label* phase
Ferber	Koba	Modi
Turly	Gilly	Cheena
Jeebus	Mizule	Claster
Plakil	Gleblu	Maloo

In this phase, 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 16 gives a visual representation of this procedure). This means that the child had to figure out which robot they trusted knew the correct answer, pick up that object, and hand it to the experimenter. This tested which robot the child trusted to know which object the label (e.g., ferber) represented. We repeated this process four times, each time with different unfamiliar objects and made-up names (8 objects total, 4 labels).

4.3.3.3 Contrast label phase

The contrast phase was designed to test children's trust towards robots in a more complex situation than the *same label* phase, with some ambiguity introduced. This phase was also an extension of prior research (Birch et al., 2008), with the addition of a fourth trial, a different but similar set of objects as the original experiment, and some additional object labels. Here, the robots labeled two uncommon objects, providing made-up names for them, just as in the *same label* phase. However, when the experimenter asked the child to hand them one of the objects, the experimenter provided a new label, which was not mentioned by either robot. This means that the child had to figure out which robot they trusted to know the correct response, and then select, pick up, and hand the other robot's object to the experimenter. For example, when both robots labeled the objects gilly, the experimenter asked for cheena (see Figure 16). The logic was that, if the child trusts the non-erroneous robot regarding what a gilly is, then they would choose the object labelled by the erroneous robot to be the cheena. We repeated this process four times, each time with different unfamiliar objects, made-up names, and a different request label (8 objects total, 4 object labels, 4 request labels).

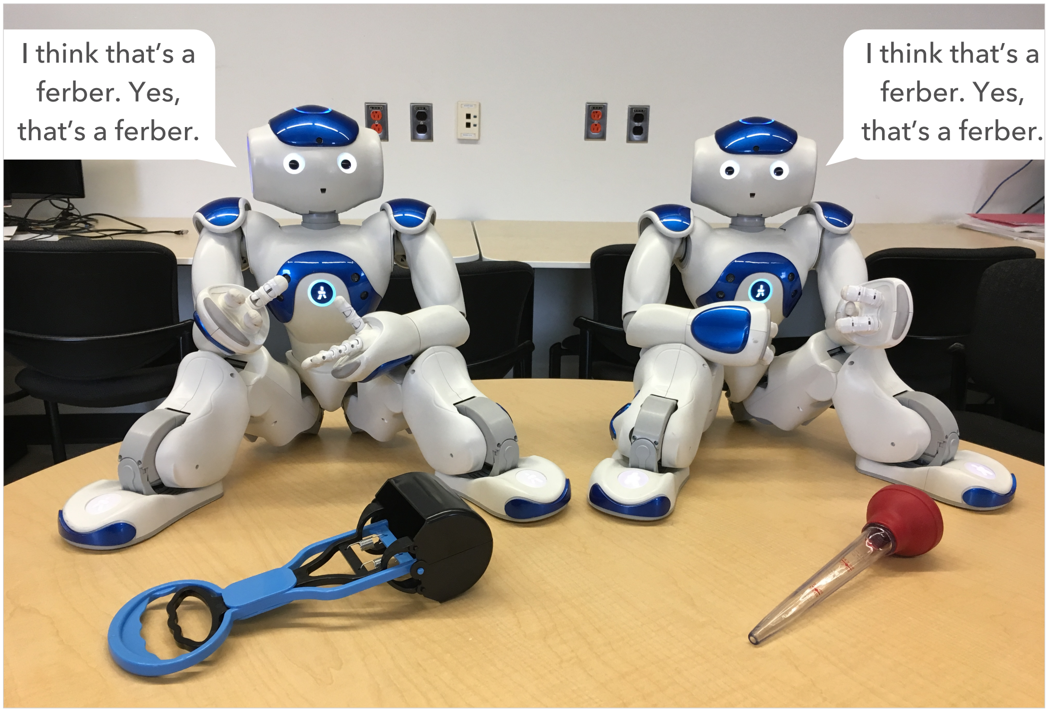


Figure 14. Visual representation of the robots labelling objects from the perspective of the child participant.

4.3.4 Procedure

For the experiment, participants were ideally escorted into a room at the daycare by a daycare worker, and introduced to the experimenter. A daycare worker was always present during the experimental sessions, 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 top of the table (see Figure 14). The experimenter informed the participant that they would be observing the two

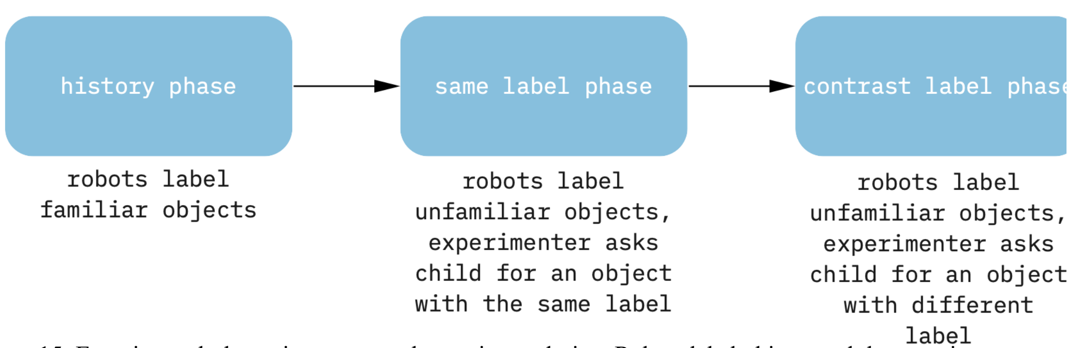


Figure 15. Experimental phases in our general experiment design. Robots label objects and the experimenter sometimes asks the child to side with a robot, therefore demonstrating which robot they trust.

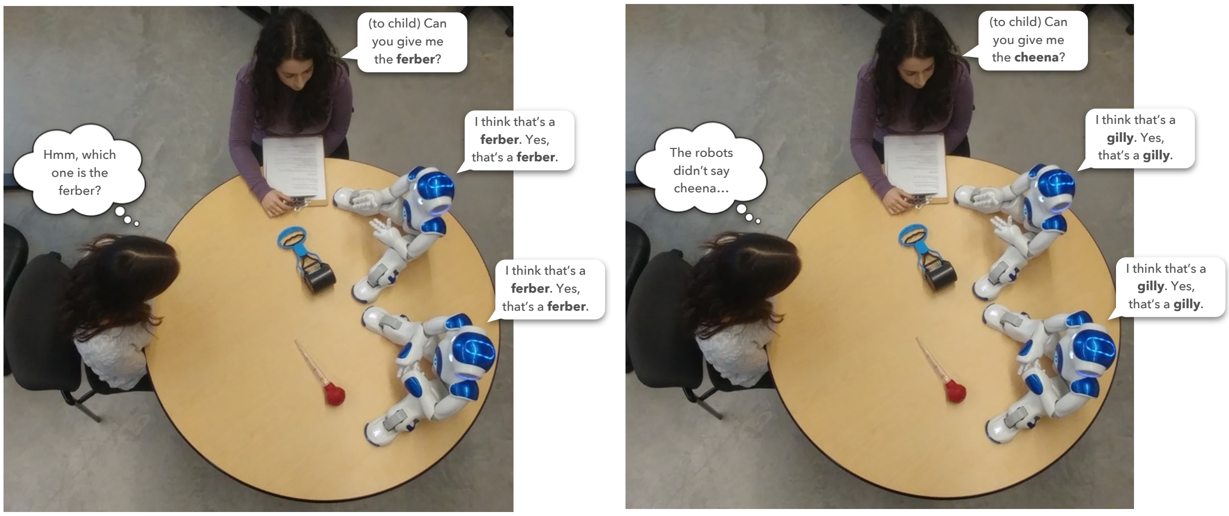


Figure 16. Example script for robots and experimenter during *same label* (left) and *contrast label* (right) phases. Both robots provide the same label for different objects. The child has to decide which robot's information to trust.

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.

The experimenter administered the three experiment phases (as in Figure 15): first the *history* phase, then the *same label* phase, followed by the *contrast label* phase. The experimental phases were always administered in this order and were not counterbalanced, with the complexity of the task increasing in each phase. This was a within-subjects experiment, with the placement of the robots (right versus left), and the order in which they labeled objects being counterbalanced between participants. Therefore, there were four counterbalancing conditions: speaking order (Casey vs. Taylor), and seating order (Right vs. Left).

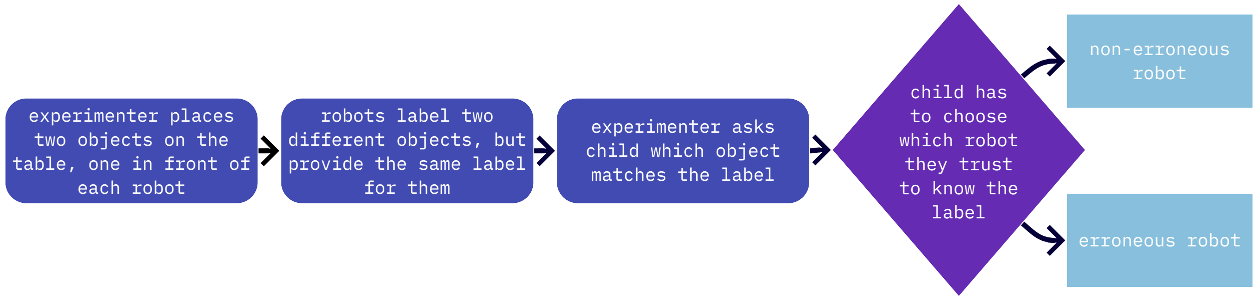


Figure 17. Procedural flow diagram of a trial during *same label* and *contrast label* phases of the experiment. During the four trials in each experimental phase, the robots provide the same (conflicting) label for two different objects and the child has to choose which robot to trust.

4.4 Data Analysis Strategy

We conducted an a-priori sample size analysis using G*Power (Erdfelder, Faul, & Buchner, 1996), with the following parameters taken from previous literature on trust towards informants: d = .60, α = .05, and power = .80 (Corriveau & Harris, 2009a). The analysis for a one-sample t-test comparing our findings against expected chance (of 50%) yielded a preferred sample size of 19 participants. However, we recruited approximately 60 participants per experiment due to the novel nature of our research with robots (as compared to puppets or adults) and established practices in similar areas of Psychology (e.g., Birch et al., 2008; Corriveau & Harris, 2009b; Holmes, 1983). This was meant to increase the statistical power of our experiments and decrease the likelihood of a Type II error. Appendix C shows the results of the sample size analysis.

The experimental sessions were videotaped and the objects that the children chose (i.e., which robot they sided with) were extracted for data analysis, as a measure of trust (our dependent variable). We aggregated this data for each experimental phase (averaging the 4 trials in each phase), and conducted one-sample t-tests against expected chance of 50% to examine whether children trusted one of the robots more than the other one. We chose this method of data analysis for a number of reasons. For one, aggregating the trials from each phase into a ratio provided us with continuous data from one group of participants. The data being continuous suggested the use of tests such as t-tests or ANOVAs. Due to the lack of comparison group in our experiments, a one-sample t-test was more appropriate to examine our data. Another approach could be to compare how each trial differs from the next, as opposed to aggregating them, thus observing the trajectory of children's trust. However, this approach would not have met our goal of examining the general patterns of trust in the experiment. Thus, our method of aggregating trial data and analyzing it using one-sample t-tests against chance (50%) was not only similar to that used in previous research on children's trust towards informants (Birch et al., 2008), but also aligned with our data analysis and experiment goals.

4.5 Chapter Summary

In this chapter, we presented an experimental design adapted and modified from Developmental Psychology to investigate young children's trust towards erroneous robots. We provided an outline of the experimental phases, the robots used, the sampling population, and our data analysis approach, as well as any modifications from the original methodology. In the forthcoming chapters, we detail an experiment investigating the effect of robot human-typical (i.e., informational) errors on how children trust them (experiment 1, Chapter 5), and an experiment exploring how robot-typical (i.e., speech-recognition) errors affect children's trust towards robots (experiment 2, Chapter 6). These experiments utilized the methodology outlined in this chapter to investigate young children's trust towards erroneous robots. Later on, we develop a framework of child-robot trust (Chapter 7), and provide a discussion of our approach, findings, and what we have learned in Chapter 8 and Chapter 9.

Chapter 5 Investigating the Impact of Informational Errors on Children's Trust Towards Robots

5.1 Publications

Ideas and content presented in this chapter have been published in the following papers:

Geiskkovitch, D. Y., Thiessen, R., Young, J. E., Glenwright, M. R. (2019). What? That's Not a Chair!: How Robot Informational Errors Affect Children's Trust Towards Robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction*, 48-56. ACM/IEEE, HRI '19. (24% acceptance rate)

Geiskkovitch, D. Y., & Young, J. E. (2020). Children's overtrust: Intentional use of robot errors to decrease trust. In *Proceedings of the 19th IEEE International Conference on Robot and Human Interactive Communication (SCRITA Workshop)*. IEEE, RO-MAN '20. 2 pages.

5.2 Introduction

The increased presence of robots in society calls for the study of how children perceive and interact with them. We specifically investigated young children's trust towards robots that provide information, as they are commonly targeted for use with children. Leveraging existing research methodologies from Developmental Psychology, which examine young children's trust towards informants, we employed the experimental design highlighted in Chapter 4 to investigate how informational errors affect children's trust in robots. The results from this experiment inform how children may trust erroneous robots in general, but more specifically, whether robot errors affect children's trust towards them (similar to how they do in humans).

5.2.1 Deviations from Chapter 4

We made some additions to the original methodology to explore additional constructs. Our goal was to examine how children trust erroneous robots, following the experimental methodology outlined in Chapter 4. To do so, we adapted and extended previous research conducted with young children and puppets, in Psychology, examining trust towards puppet informants. One of the modifications we made for this experiment, was the addition of the *clean-up* task. The purpose of this task was to assess whether robot informational errors affect children's actions and compliance relating to subsequent robot instruction. The following paragraph explains the inspiration behind the *clean-up* task and the methodology.

Due to the novelty of this research when examining children's trust towards erroneous robots, we looked to research with adults for guidance. Research suggests that while adults report trusting an erroneous robots less than a non-erroneous one, their behaviour does not reflect this, and they follow the erroneous robot's instructions nonetheless (Salem et al., 2015). Thus, we included the *clean-up* task to explore whether robot informational errors affected children's compliance to the robot's instructions. During the *clean-up* task, the experimenter informed the child that it was "time to clean up" (a phrase commonly used with young children at home and in daycare settings). The experimenter placed two coloured pieces of paper (one red, one white) on the table, one in front of each robot (Figure 18), and the robots provided conflicting instructions: one robot asked the child to put the already-labeled objects (which were collected in a blue basket) on the red piece of paper, while the other robot referred to the white piece of paper. The experimenter then asked the child where the basket with the objects should go, and the basket was placed there. We administered this task twice, once after the *same label* phase, and after the *contrast label* one

(Figure 19) to gain insight into how children's compliance might shift throughout the experimental session.

5.2.1.1 Clean-up task data analysis strategy

The *clean-up* task resulted in a set of nominal data indicating which of the two robots each child complied with. This data was obtained for each of the two *clean-up* tasks that took place, and analyzed separately. This data could not be analyzed using the same approach as the experimental phases, as here we only had one nominal data-point to analyze, and the one-sample t-tests we used to analyze the phases require continuous numerical data. Thus, we used a χ^2 test to analyze the data from the *clean-up* task. The χ^2 test assesses whether there is a relationship between two categorical variables (in our case the two variables are the robot that provided the instruction:

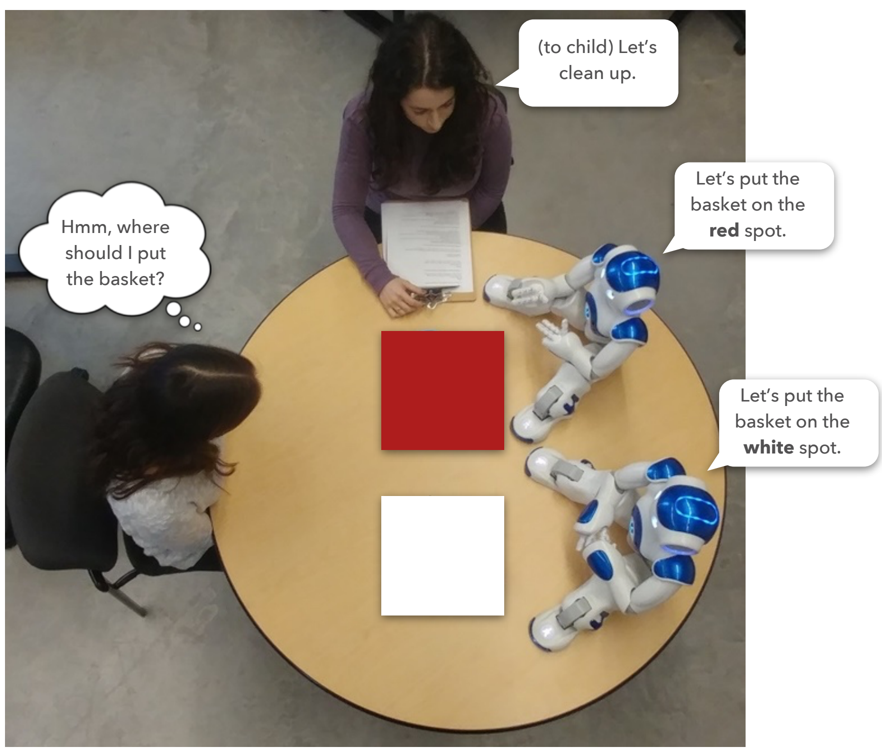


Figure 18. Script for robots and experimenter during the *clean-up* task. The child has to choose which robot's instructions to follow.

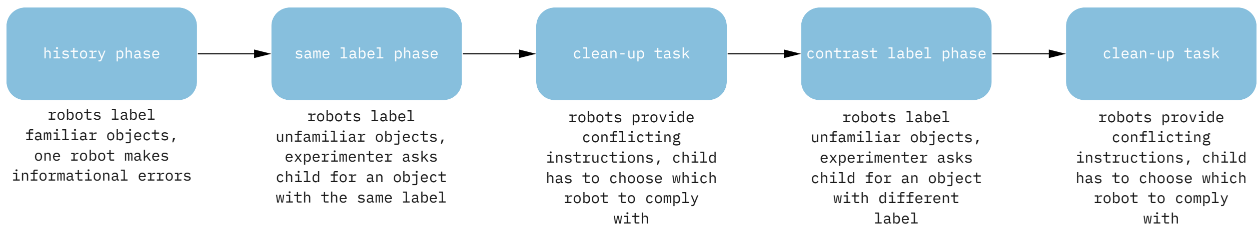


Figure 19. Procedure flow chart for experiment 1, children's trust towards robots with informational errors.

erroneous vs. non-erroneous, and the instruction that the child complied with: erroneous robot vs. non-erroneous robot). Our data also met the assumptions for this test as we had two nominal variables (which robot was erroneous and which robot the child complied with) and the groups were independent from each other. Thus, the χ^2 test was appropriate to use for this particular set of data.

5.3 Hypotheses

According to previous research, children tend to trust informants based on a number of factors, including accuracy information (see Chapter 3), preferring to trust humans and puppets that have not made errors (Birch et al., 2008; Corriveau & Harris, 2009b; Fusaro et al., 2011; Koenig & Harris, 2005a; Kushnir et al., 2013). The goal of our research was to see how these findings relate to robots, who children tend to perceive be to have human characteristics such as liveness, moral rights, values, and agency (Kahn et al., 2012; Rao & Georgeff, 1995). Thus, it was reasonable to hypothesize that we might see similar trust responses to informational robots than to informational humans or puppets, guiding the hypotheses we present below.

5.3.1 Same Label Phase

The intention of the *same label* phase was to test trust towards the robots in a simple situation (i.e., when the experimenter provided the same object label as the robots). Previous research in

Psychology demonstrated that preschool-aged children trust accurate (i.e., non-erroneous) human and puppet informants more than inaccurate (i.e., erroneous) ones (Birch et al., 2008; Corriveau & Harris, 2009b; Fusaro et al., 2011; Koenig & Harris, 2005a; Kushnir et al., 2013). Specifically, Birch et al. (2008) found that preschool-aged children trusted a puppet that provided non-erroneous information more than one that made errors. We believed that a similar finding would exist with robots, given their agency and anthropomorphism. Therefore, we hypothesized that:

H1: Children will choose the objects labeled by the erroneous robot more often than chance (50%), when asked to hand an object with a name that they had previously heard (i.e., during the same label phase).

5.3.2 Contrast Label Phase

The goal of the *contrast label* phase was to test young children's trust towards the robots in a complex situation (i.e., when the experimenter provided a different object label than the robots). While some experiments found that young children tend to trust non-erroneous human and puppet informants in simple situations (i.e., *same label* phase; e.g., Birch et al., 2008, experiment 1), others have not found any differences when the labeling task is more complex (e.g., informants labeling the use of objects as opposed to the name; i.e., Birch et al., 2008, experiment 2). While it was unclear how previous findings for the *contrast label* phase may translate to robots, the task itself (labeling objects) did not change. Therefore, we hypothesized that:

H2: Children will choose the object labeled by the erroneous robot more often than chance (50%), when asked to hand an object with a name that they had not previously heard (i.e., during the *contrast label* phase).

5.3.3 Clean-up Task

Through the *clean-up* task we aimed to examine the effects of robot informational errors on children's compliance. We accomplished this by having the robots provide conflicting instructions to the child (i.e., where to place the basket with the objects), and observing which robot they complied with. Researchers found that while adults report lowered trust towards an erroneous robot, they nevertheless continue to comply with the robot's instructions even though they know it is malfunctioning (Salem et al., 2015). As similar research is yet to be conducted with children (i.e., the goal of this research), we can predict a pattern similar to those from the other experimental phases. Thus, we hypothesized that:

H3: Children will comply with the instructions of the non-erroneous robot more often than the erroneous robot (i.e., during the *clean-up* task).

5.4 Experiment

5.4.1 Participants

We recruited 53 children (31 male, 22 female), aged 3-5 (M = 3.62, SD = 0.596, 23 3-year olds, 27 4-year olds, 3 5-year olds) to participate in this experiment. We provide additional information in the next sentences to support full understanding of our research methodology. Children were recruited from 4 daycare centres in Winnipeg, Manitoba, Canada and their parents provided written consent (see Appendix D) before the experiment took place. Children were able to pick a small toy to thank them for their participation, and parents received a $15 honorarium. This experiment was approved by the University of Manitoba Joint Faculty Research Ethics Board (see Appendix E).

5.4.2 Procedure

We utilized the experimental procedure from Chapter 4, with the modifications presented at the beginning of this chapter. The independent variable for this experiment was whether the robot made informational errors or not, which was demonstrated by the robot providing correct or incorrect responses during the *history* phase of the experiment. A detailed description of the procedure is found below (also see Figure 19).

The experiment took place at 4 daycare centres starting in the Summer of 2018 and ending in Spring 2019. One to two weeks prior to the experiment taking place, we conducted introductory sessions with the appropriately-aged children at each daycare (as explained in Chapter 4). Following, we conducted the experimental sessions. Daycares in which the experiment spanned multiple days (due to a large number of participants or time limitations) had all sessions completed within one week.

At the beginning of the experimental session, participants were escorted into a room at the daycare by a daycare worker, and were introduced to the experimenter (most children had previously met the experimenter, research assistant, and a slightly different robot). The daycare worker remained in the room for the entirety of the experimental session, but was not in the child's immediate field of vision. The child sat down at a table and was introduced to the two robots (Casey and Taylor), who were sitting on the table (see Figure 20). The experimenter informed the participant that they would be observing the two robots label objects, and asked the child if they would like to do that (verbal consent from child). The robots (which were remotely controlled by a researcher in the experimental room; implementation explained in Chapter 4) then introduced

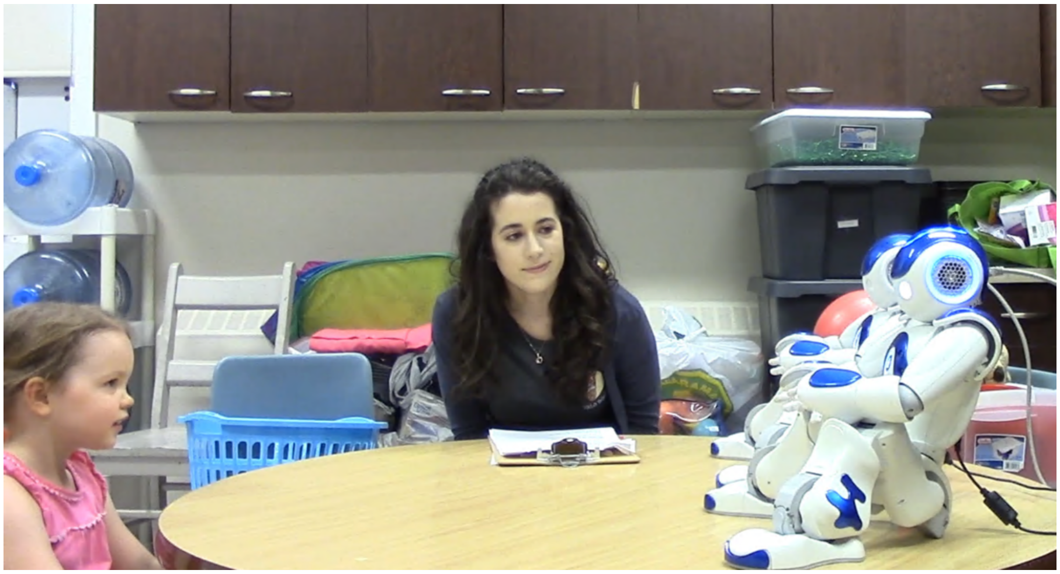


Figure 20. Example experimental set-up in a daycare centre, where child sits at table (left) with two robots (right), and experimenter (center). Picture used with permission.

themselves to the child, and the experimenter informed the participant that they could stop the study at any time.

For the *history* phase of the experiment, the robots provided labels for four common objects, following our established protocol (see Chapter 4). While naming the common objects, one of the robots provided the correct labels for the objects (the non-erroneous robot), while the other provided incorrect labels (e.g., labelling a car as a book; the erroneous robot). This was the independent variable in this experiment, focusing specifically on which robot provided accurate information and which did not.

During the *same label* phase of the experiment, the two robots provided labels for less common objects, each robot providing the same name for a different object (i.e., conflicting information) (see Figure 21). The child was then asked to hand the object with the matching name to the experimenter. During this phase of the experiment, we expected children to give the experimenter the objects labeled by the non-erroneous robot (H1). The object that the child handed to the experimenter was used as a measure of trust.

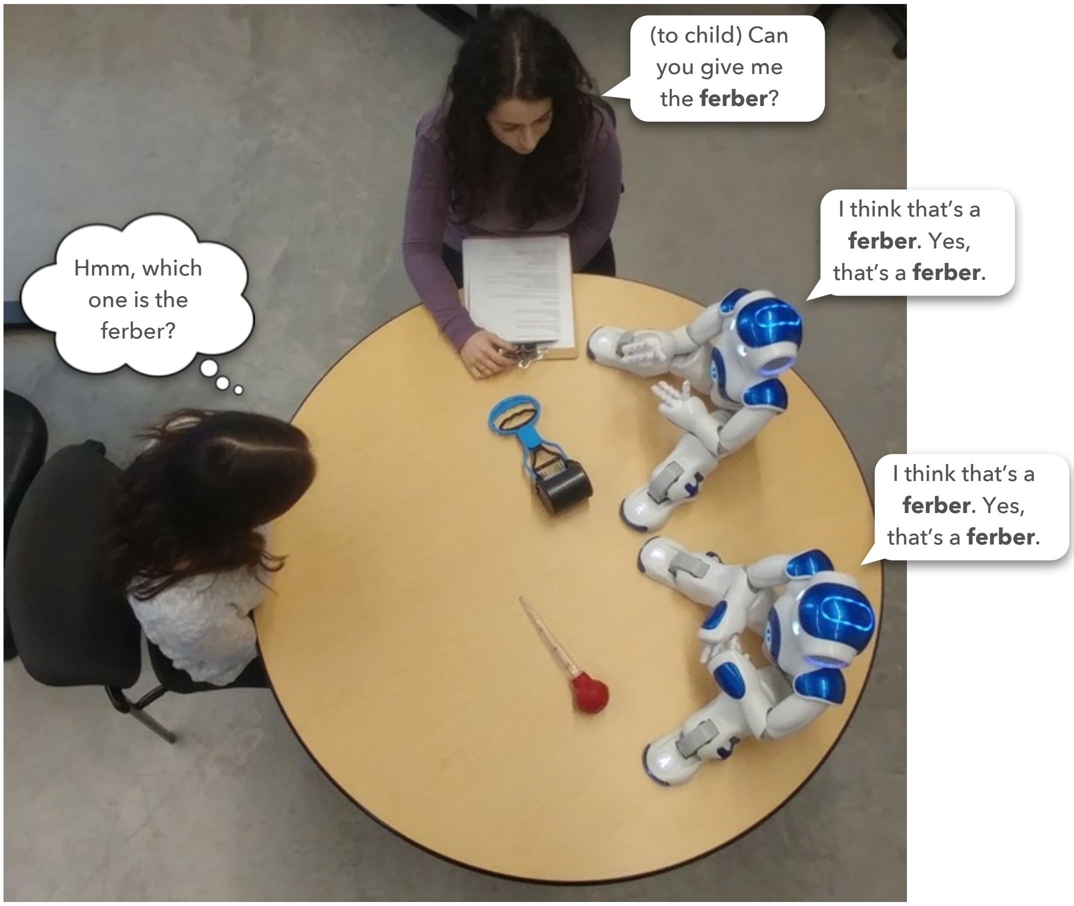


Figure 21. During the *same label* phase, both robots provide the same label for two different objects. The child has to choose which robot they trust to provide the correct information.

After the *same label* phase, we included the *clean-up* task (as described at the beginning of this chapter). The child was told that it was time to clean up, and the robots told the child where to place the objects that had already been labeled. The robots, however, provided conflicting information, and the child was asked to choose which robot to side with. We expected children to do what the non-erroneous robot told them to do (H3).

In the *contrast label* phase of the experiment, the robots once again provided labels for less common objects, each providing the same name for different objects, but in this case the experimenter asked the child for an object that did not match the labels provided by the robots.

During this phase of the experiment, we expected children to give the experimenter the objects labeled by the erroneous robot (H2; see Figure 22).

Before the experiment ended, the child was asked to label the four common objects from the first phase (i.e., ball, car, dog, plate) as a manipulation check, to ensure that they had been properly exposed to the independent variable.

Following, a second iteration of the *clean-up* task was conducted, and the child once again had to choose which robot to side with.

This was a within-subjects experiment, with the placement of the robots (right vs. left), and the order in which they named objects being counterbalanced between participants. Therefore,

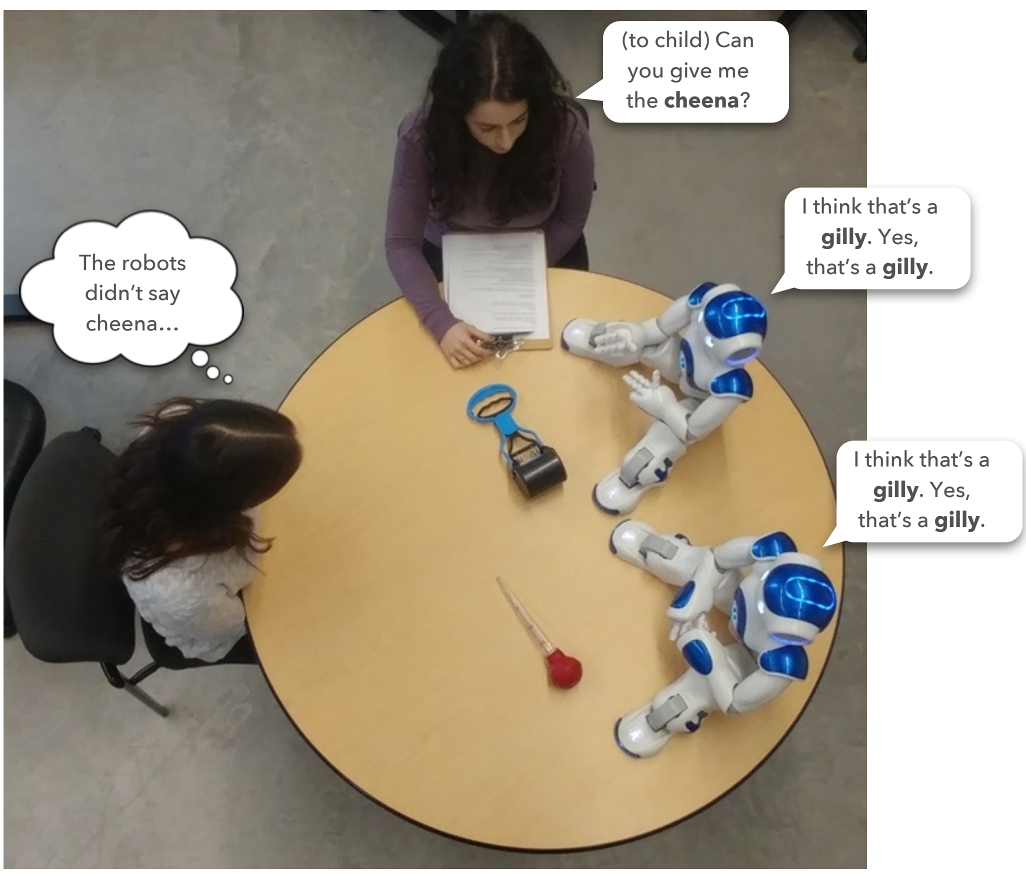


Figure 22. During the *contrast label* phase, both robots provide the same label for two different objects. However, the experimenter asks the child for an object with a different name. The child has to choose which robot they trust to provide the correct information.

there were four counterbalancing conditions: speaking order (Casey vs. Taylor), and seating order (Right vs. Left).

5.5 Results

Out of the 53 participants that were recruited; 9 participants were removed due to experimental confounds (the sessions took place in the same space as where other children played, so participants were very distracted), 2 children were too shy to interact with the robots, and 1 child did not follow task directions (in every trial they gave both objects to the experimenter instead of picking one). Therefore, the data from 41 (24 males, 17 females; 14 3-year olds, 24 4-year olds, 3 5-year olds; $M = 3.73$, $SD = .593$) valid participants was analyzed. A one-sample t-test on the aggregate robot choice ratio (choosing a particular robot from 0% to 100% of the four trials) against chance of 50% indicated a statistically significant difference between which robot children trusted during the *same label* phase, $t(40) = 2.667$, $p = .006$, $d = 2.667$ (see Figure 23). Children trusted the non-erroneous robot more than the erroneous one ($M = .652$, where children sided with the non-erroneous robot 65.2% of the time compared to 50% chance, $SD = .057$), thus the first hypothesis is supported. However, we did not find a statistically significant difference between which robot children trusted in the *contrast label* phase, $t(40) = .206$, $p = .419$, $d = -.032$, $M = .488$ (children sided with non-erroneous robot 48.8% of the time compared with 50% chance, $SD = .379$) and therefore we did not find support for the second hypothesis (see Figure 23).

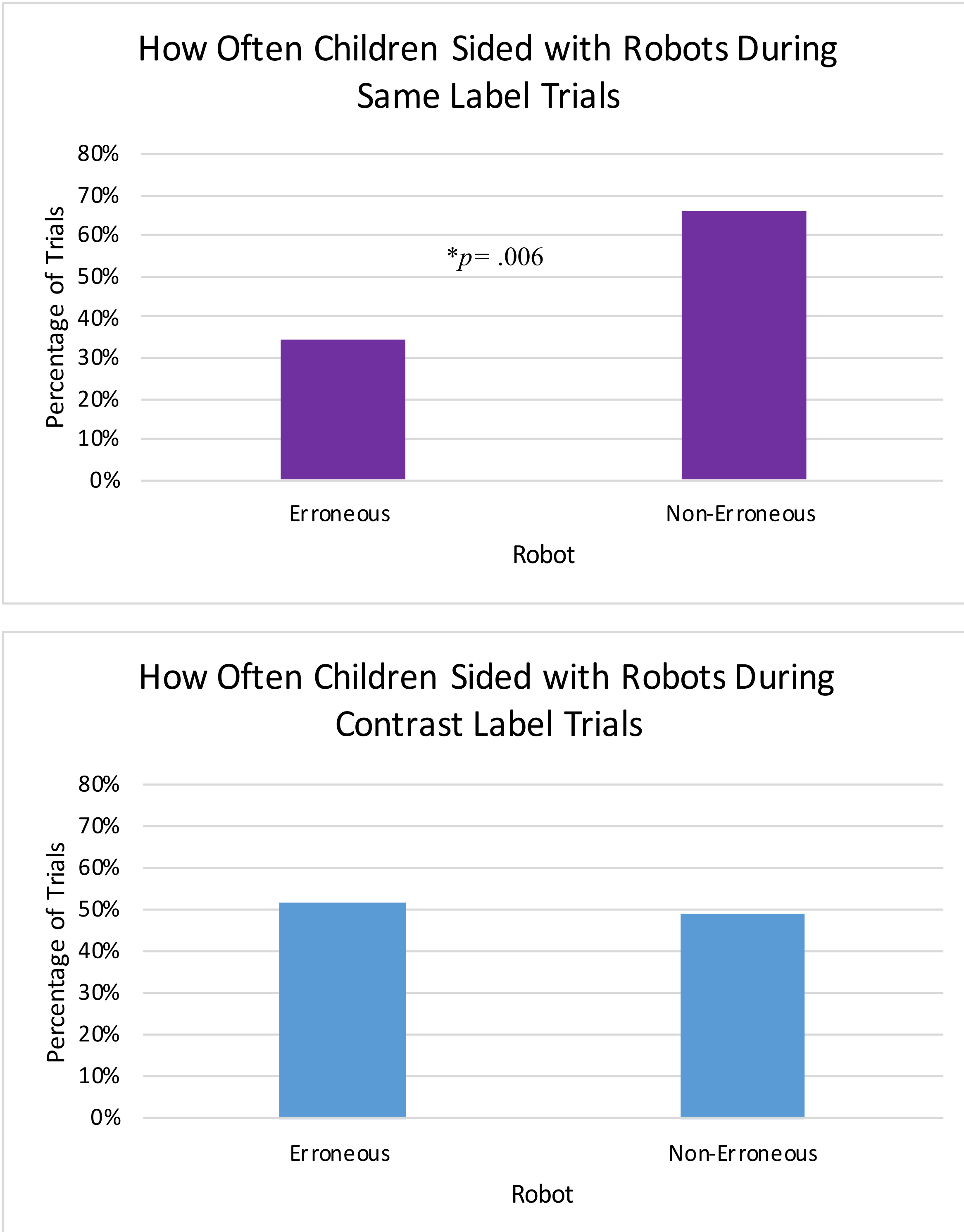


Figure 23. How often children sided with the robots during the *same label* and *contrast label* phases of experiment 1, where one robot made human-like informational errors. There was a statistically significant difference during the *same label* phase, with children trusting the non-erroneous robot more than the erroneous one ($t(40) = 2.667$, $p = .006$, $d = 2.667$). However, there was no such difference in the *contrast label* phase ($t(40) = .206$, $p = .419$, $d = -.032$).

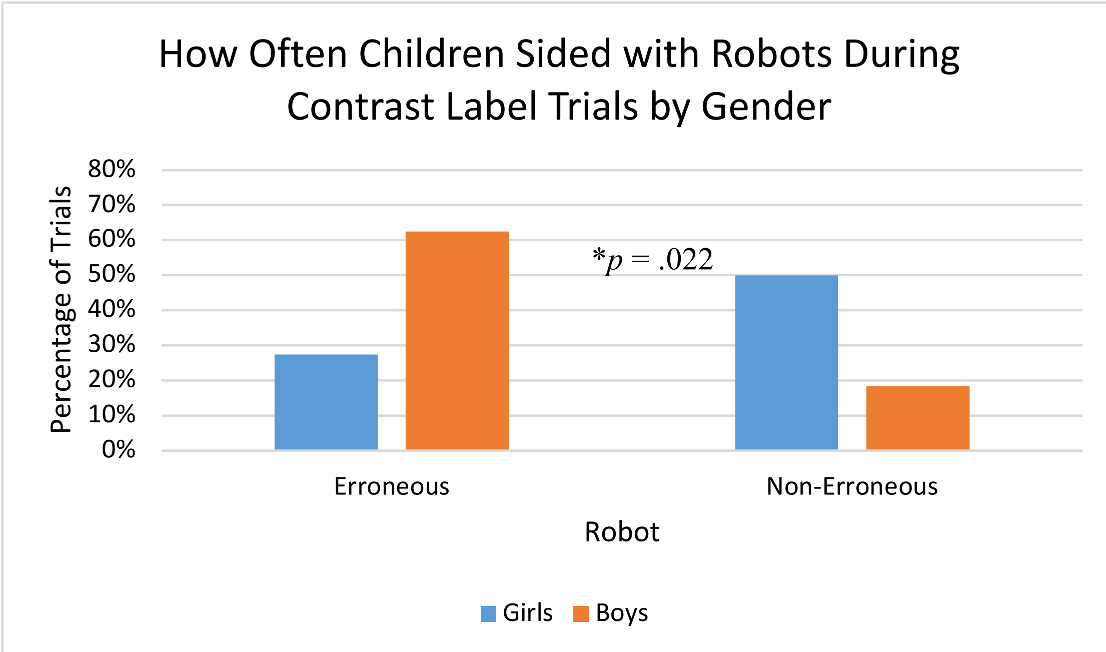


Figure 24. Post-hoc analysis of gender in the contrast label phase of experiment 1. Boys sided with the erroneous robot while girls sided with the non-erroneous robot. We expected both genders to trust the non-erroneous robot, and therefore side with the erroneous robot (as boys did). Independent samples t-test: $t(39) = 2.395$, $p = .022$, $d = .75$.

Informally, we noted that the responses seemed to differ between boys and girls during the *contrast label* phase. We therefore conducted a post-hoc independent samples t-test to investigate whether gender played a role in our lack of findings. The post-hoc results suggest a potential gender effect for the *contrast label* phase, $t(39) = 2.395$, $p = .022$, $d = .75$, with girls trusting the non-erroneous robot ($M_G = .647$, $SD_G = .386$) and boys the erroneous one ($M_B = .375$, $SD_B = .338$; see Figure 24).

For the *clean-up* tasks (Figure 25), we did not find a statistically significant difference between which robot children complied with (clean-up 1: $\chi^2 = 1.195$, $p = .274$, 17 erroneous, 24 non-erroneous; clean-up 2: $\chi^2 = 1.976$, $p = .160$, 25 erroneous, 16 non-erroneous) nor between the two *clean-up* tasks (exact McNemar's test $p = .216$). No other analyses, such as age, robot placement, or robot speaker order were statistically significant.

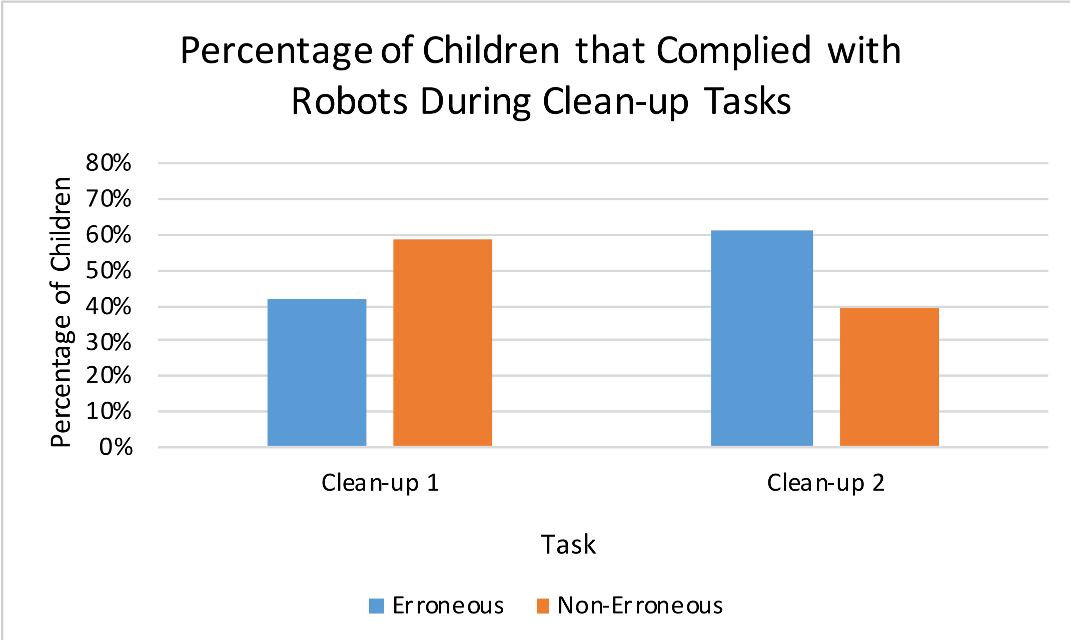


Figure 25. How often children complied with the robots during the *clean-up* task. We expected children to follow the instructions of the non-erroneous robot. However, we did not find a statistically significant difference between which robot's instructions children followed. Clean-up 1 $\chi^2 = 1.195$, $p = .274$, clean-up 2: $\chi^2 = 1.976$, $p = .160$.

5.6 Discussion

The results from this experiment partially support our hypothesis that young 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. Specifically, they choose objects labelled by a non-erroneous robot as the "correct" objects, signifying their trust in that robot's information, more than in the erroneous robot, during a simple situation. This provides insight into how children perceive robots, and how they attribute mental states to robots, that is, children appear to rely on information from a non-erroneous robot more than from one that makers errors. This is similar to patterns that are seen with human and puppet informants, where young children rely on information from accurate informants more than from informants that make mistakes (e.g., Birch et al., 2008; Koenig & Harris, 2005a). This suggests that young children may attribute mental states to robots similar to how they do people and even puppets, and build similar mental models of them, but we only found evidence for this in a simple situation (i.e., *same label* phase). 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.

5.6.1 Same Label Phase

Our data suggests that errors that a robot made during the *history* phase had an effect on how much children trusted it during the *same label* phase (i.e., when the labels provided by the robots and experimenter matched). Children sided with the robot that had not made errors more often than the one that had. Specifically, 65.2% of the time children chose to side with the robot that did not make errors, making this their choice almost two thirds of the time. This supports our first hypothesis that, in a simple situation, young children are more likely to trust a non-erroneous robot (i.e., choosing to trust its information) more than one that makes errors, and aligns with previous work that found that children expressed more trust towards a previously correct puppet, than an incorrect one (Birch et al., 2008). This also implies that children properly responded to our manipulation, as the robot errors did have an effect on children's trust. Our findings suggest that prior informational errors that a robot makes could have an effect on how much it is trusted afterwards, at least in the short term. However, further investigation is necessary as we found this to be the case only two thirds of the time.

These findings highlight that young children may, most of the time, apply how they trust adults (and similarly puppets), at least in simple cases, to inform how they should trust robots in similar situations. Cases similar to the *same label* are applicable to robots in positions of authority over children, and can be found in places such as schools and daycares, where errors may affect how young children listen to, follow the instructions of, and learn from a robot. However, more complex situations, such as where there might be conflicting information from other sources, including adults, need to be investigated separately. Thus, we investigated this application through the *contrast label* phase, discussed in the following section.

5.6.2 Contrast Label Phase

We failed to replicate the previous finding by Birch et al., (2008) that children will continue to trust a previously correct informant, when the informants and experimenter provide different labels (i.e., in a complex situation). We did not find that children continued to trust the non-erroneous robot during the *contrast label* phase, when children were asked for an object with a name that 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 non-erroneous robot, and thus side with the erroneous one, 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 (Rao & Georgeff, 1995). 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. Other researchers have also failed to find a difference in more complex scenarios (e.g., Birch et al., 2008, experiment 2; 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). The possibility that this phase was therefore more complex than previously believed, merely due to using robots as opposed to puppets, might be to blame for our lack of findings. Similarly, it may be the novelty and agency of the robots we used 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.

We had expected all children, regardless of gender, to trust the non-erroneous robot during this phase (and therefore side with the erroneous robot). Our data, however, suggests that during the *contrast label* phase of the experiment, boys performed as expected (choosing the erroneous robot), while girls performed the opposite of what we expected (choosing the non-erroneous 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. Alternatively, young girls tend to exhibit more prosocial behaviour than boy peers suggesting that our findings could be due to girls demonstrating prosocial behaviour towards the non-erroneous robot, regardless of the task (Sebanc, 2003). However, as these are post-hoc inquiries, further research in HRI will be necessary to explore this possibility, investigating how young children's gender affects anthropomorphism of robots, and trust in general. Further exploration with a larger sample size and equal gender divide is needed to assess why we found gender differences in this particular phase of the experiment.

Our experimental procedure contained some deviations from previous research that may have also led to different results. Specifically, the addition of the *clean-up* task between the *same label* and *contrast label* phases, and after the *contrast label* phase. The purpose of the *clean-up* task was to evaluate how compliance is affected by robot errors. However, this task created a pause between the *same label* and *contrast label* phases that was not present in previous related work. There are many ways in which this additional time may have caused our lack of results. The added time between the *same label* and *contrast label* phases may have caused children to forget which robot had made errors during the *history* phase. Likewise, it is possible that simply the addition of the *clean-up* task between the two phases created a perceived separation between what occurred before and after the first *clean-up* task, causing the information (i.e., which robot made errors) from the *history* phase to only be linked to the *same label* phase, and not to the *contrast label* phase. Additional research into the effects of the *clean-up* task and an experiment without it will be necessary to figure out the reason for the lack of results during the *contrast label* phase.

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 (Birch et al., 2008) or animal species (Kushnir et al., 2013) 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, it is necessary to point out that 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.

5.6.3 Clean-up Task

We were unable to find support for our third hypothesis, that children would comply with the non-erroneous robot's directions during the *clean-up* tasks, and would therefore place the basket with all the objects where the non-erroneous 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. Likewise, our lack of findings might be due to the design of the *clean*-up task. For example, it might be that children chose which robot to comply with not because of characteristics of the robot, but due to the colour of the paper that they preferred the most (i.e., red vs. white). In addition, while the *clean-up* task was a good first step towards investigating the intersection of trust, errors, and compliance towards robots, the task itself did not have any aspects of risk or consequence (an important aspect of trust, as discussed in Chapter 3). So, children in our experiment may have chosen where to place the objects randomly, due to the lack of personal investment. A future adaptation of this task, with some modifications to increase potential risk or consequences may provide further insight into whether robot informational errors do (or do not) affect children's compliance with robots' instructions. Incorporating an aspect of risk into the child's trust decision may provide a clearer outlook on whether robot errors affect young children's compliance.

5.7 Limitations and Future Work

The findings outlined in this section 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, such as different types of errors and an updated experimental design. 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 differences 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).

Likewise, we have some informal concerns about whether children were truly able to differentiate between the two robots. The robots in our experiment looked and sounded exactly the same, and could only be differentiated by their positioning in relation to the experimenter and each other (i.e., right vs. left, which were counterbalanced between participants), their given names, and the presence of errors in one of the robots. It could be that children were not properly able to distinguish between the two robots. However, children were able to differentiate enough between the two robots to side with them to different extents during the *same label* phase, which suggests that the similarities were likely not a problem during this phase. Nonetheless, it is possible that children forgot which robot had made errors, or which robot was which. Thus, future iterations of this research should better differentiate between the two robots, to ensure that children can better differentiate between them throughout the experimental session.

We tested whether children knew the names of the familiar objects at the end of the experimental sessions. This may have introduced some noise into our results, given that children may not have known or remembered what the objects were called until they were labeled by the robots. However, any additional noise that this may have added to our results would suggest that our results are even more robust, as we found statistically significant differences during the *same label* phase despite potentially noisy data. In addition, had children forgotten or not known what the familiar items were called, the robots' statements would likely have assisted recall, suggesting that children would likely still have been properly exposed to our independent variable.

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, which we explore in Chapter 6. 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, mechanoid) may affect how its errors are perceived. Further investigation 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.

5.8 Chapter Summary

In this chapter, we detailed the first experiment of this thesis, which investigates how young children trust erroneous robot informants. We found that children do sometimes trust robots differently depending on whether they have made errors or not, but mainly in a simple scenario. These results partially replicate and extend previous findings in Psychology to inform how they apply to Child-Robot Interaction, and provide a starting point for the study of children's trust towards robots, particularly erroneous ones.

This work has important implications for robots who will be interacting with children, especially those that provide information to them. Robots in positions such as teaching assistants, hospital companions, guides, and home assistants or tutors are likely to be providing children with information. However, they are just as likely to be making errors, and even more so than people as they do not have the ability to prevent mistakes. Our research suggests that young children trust information provided by a robot that has made informational errors less, than that from a robot that has not made errors. Thus, robots that make informational errors in the presence of young children may be trusted less, and will therefore need to either make up for the lost trust or use errors purposefully to decrease trust. This information is highly relevant for the design of robots that interact with children.

While our first experiment focused on *human-like* informational errors, robots tend to exhibit errors that are unique to them, such as recognition and contextual errors. Thus, in Chapter 6, we examine how errors that are more typical of robots affect children's trust. We utilized the experimental design outlined in Chapter 4 and conducted in experiment 1 (Chapter 5), with minor modifications such as the removal of the *clean-up* task and the addition of a planned gender analysis, all based on our findings and discussion in this chapter.

Chapter 6 Exploring the Impact of Robot-Typical Errors on Children's Trust Towards Robots

6.1 Publications

Ideas and content presented in this chapter have been published in the following paper:

Geiskkovitch, D. Y., & Young, J. E. (2020). Social robots don't do that: Exploring robot-typical errors in child-robot interaction. In Companion *Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction*, 200-202. ACM/IEEE, HRI '20.

6.2 Introduction

In Chapter 5, we investigated how children's trust is affected by a robot making informational errors, as an extension of previous research that explored the same paradigm but with puppets. However, we may not expect errors to manifest in robots in the same way as they do in people (i.e., informational errors). In actuality, robots more commonly exhibit other types of errors, such as speech-recognition errors (Mubin et al., 2014; Novoa et al., 2018) which are witnessed by a lack of response or unrelated response. These types of errors are also extremely common with children (Kennedy et al., 2017), due to shortcomings in currently available sensing technology. It is therefore important to examine how these robot-typical errors affect children's trust towards robots, as they are more likely when interacting with robots.

While in the first experiment we examined how young children trust information from robots based on informational errors, in this second experiment we look at how they trust such information based on conversational or speech-recognition errors. Thus, the erroneous robot does not make errors that are directly related to the information that they provide. While this provides a different paradigm to that of the first experiment, it is common for robots that provide information to children to make speech-recognition errors. Therefore, examining the combination of conversational errors and trust of the information provided is highly relevant.

6.2.1 Robot-Typical Errors

To our knowledge, there is no literature on the different types of common robot errors from an interaction standpoint, and how they may affect young children. Thus, we first propose a preliminary framework of robot-typical interaction errors which we utilized to investigate trust in Child-Robot Interaction. This framework was created informally through lab development sessions with experienced PhD students and researchers, with a collective experience of over 1000 participants, and having built over 10 robot platforms in the past decade. The sessions took the form of brainstorming. The author of this thesis then categorized these robot-typical errors into the initial framework presented below, and used their expertise in Child-Robot Interaction to hypothesize how children may perceive these errors and future avenues of exploration.

Robot errors have been explored in HRI under various contexts. Some research explored how a robot's human-like errors (e.g., making computational mistakes) tend to negatively affect perceptions of the robot (e.g., as less intelligent or reliable), but can also increase interaction satisfaction (Ragni et al., 2016). Others investigated how individuals responded to a robot that unexpectantly changed its response (Short, Hart, Vu, & Scassellati, 2010), was unable to complete the task it was designed for (Hamacher et al., 2016; Kaniarasu & Steinfeld, 2014), or provided clearly incorrect information (Danovitch & Alzahabi, 2013; Geiskkovitch, Thiessen, Young, & Glenwright, 2019). Most of this related research therefore explored robots making human-typical errors, in the context of otherwise normal behaviour. However, few experiments have explored errors more representative of interactions with actual autonomous social robots, such as response delays (Breazeal et al., 2016), or irrational movements and contextually-strange requests (Salem et al., 2015).

Our framework considers issues stemming from low-level hardware and software malfunctions, all the way up to higher-level sensemaking and behavioural problems, and will aid us investigating true robot-typical errors. This initial framework serves as a starting point for our research on robot-typical errors and for other researchers, as it outlines the types of errors that children may encounter and how we predict children will respond. Further research is necessary to validate this framework. This preliminary framework was published in Geiskkovitch and Young (2020).

6.2.1.1 Framework of Robot-Typical Errors

Hardware Errors. Robot-typical hardware errors are those that occur due to broken or faulty physical components of the robot.

Faulty Computer – Faulty memory, processors, etc. will result in errors more typical of a PC than a person. The robot, for example, may suddenly freeze, or slow down its movements or speech.

Faulty Motors – Motors damaged or deteriorated from use, or overheating, may not move as intended. Motors may become disabled (e.g., failing to lift an arm), or move with a jitter due to movement compensation. The robot may make unpredictable movements from limited mobility or attempt to compensate by using other motors.

Faulty Sensors – Broken sensors such as regular or depth cameras, touch bumpers, microphones, etc., can lead to errors including a robot failing to respond to voice or other input, or acting inappropriately, such as bumping into people or walls while seemingly not noticing the error.

We can expect children, who often anthropomorphize robots and see them as being alive (Rao & Georgeff, 1995), to possibly observe these hardware-related errors from a life-like perspective. Thus, we hypothesize that this might lead to young children viewing the robot as weird, funny (e.g., silly), or even scary.

Software Errors. Even if a robot has non-faulty hardware, errors can take place due to unexpected problems with the software, and therefore lead to outcomes different than those expected from the robot.

Software Crashes – Errors in the code, unexpected input types, or processing errors can cause the robot to have socially unrecoverable errors such as freezing or rebooting, whereby once an individual witnesses them, it can change their perceptions of the robot and its abilities.

Program Bugs – Anomalies in the code can lead to non-catastrophic but inconsistent and unpredictable behaviour. For example, a tutoring robot may provide different answers for the exact same math question.

Children who witness a social robot's software errors may believe it is acting strangely as compared to people, by acting inconsistently, providing incorrect answers (e.g., Chapter 5), or having generally unusual interactions. This behaviour would only be seen in humans who have health problems (e.g., heart attack), cognitive impairments (e.g., Alzheimer's), or with computers. Therefore, young children who anthropomorphize the robot may perhaps view it as being ill or hurt, and in turn feel confused or sad for the robot.

Sensemaking Errors. Even if the hardware and software systems of a robot are error-free, higher-level errors in the robot making sense of information can still occur.

Sensing Limitations – This includes a robot's inability to understand its location or environment, people or things around it, or what is taking place. A tutoring robot, for example, may continue to provide information after a child has finished a work problem, because it cannot sense that it was completed.

Misclassification Errors – Robots may erroneously classify beings, objects, and situations. A robot may therefore, for example, misclassify a child as a doll, misinterpret what a child says, or see children playing tag or wrestling and interpret it as an emergency situation.

Through an anthropomorphic lens, sensemaking errors may make the robot appear unintelligent, or perhaps silly and playful to young children, since people tend to act this way when they misunderstand, have some sort of cognitive impairment, or are joking. Thus, these errors might lead children to believe that the robot is inadequate, funny, or has cognitive deficiencies.

Behavioural Errors. Given properly-functioning hardware, software, and sensemaking, a robot may still behave in a manner that is deemed inappropriate, either socially or contextually.

Nonsense or Lack of Response – A robot may not provide a response, provide one that is either not related to the input received (even though sensors are working properly and the software is working as programmed), or one that does not make sense. This may be due to the particular interaction not being coded for in the robot's behaviour. For example, if a child asks a robot what an object is called, the robot may be unable to provide a response simply because it is not programmed to know that response, or to deal with unknown cases.

Failed Physical Attempt – A robot may not move as expected, or at all, regardless of all other components working properly. A robot might, for example, attempt to give a child a high-five, but fail to do so simply due to the complexity of the task, not any hardware, software, or sensemaking problems.

Context-Inappropriate Behaviour – A robot may not behave in accordance to the social context that it is in. For example, a robot may ask a child to retrieve something from a high shelf, not realizing that it could be dangerous and the child could get injured, or it may tell a child an age-inappropriate joke.

As we can expect children to anthropomorphize a robot's behavioural errors, they may perceive the robot as being inept, in terms of forming coherent speech, performing actions

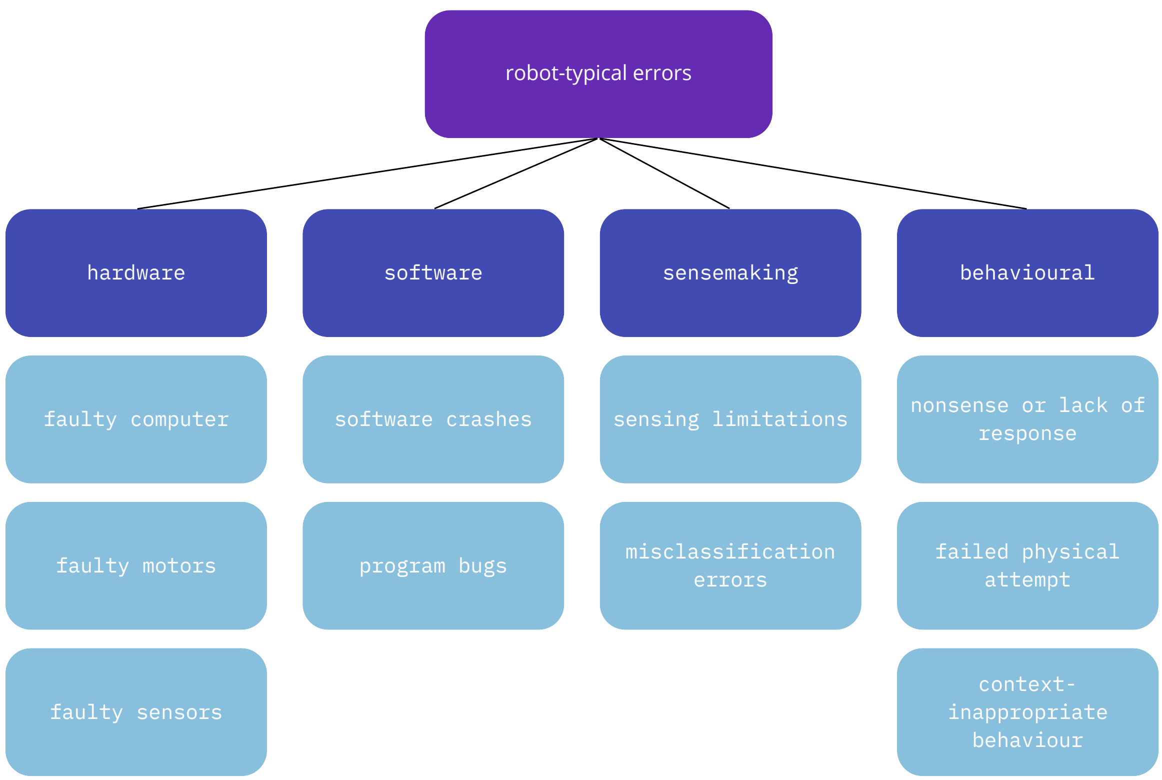


Figure 26. Framework of robot-typical errors. In this chapter, we detail an experiment focusing on behavioural *nonsense or lack of response* errors.

correctly, and behaving in an appropriate manner. These errors may therefore lead young children to mistrust the robot's abilities, and view it as strange or having some type of illness.

6.2.2 Errors Investigated in this Experiment Chapter

Given the vast number of possible errors that robots can exhibit, it is necessary to investigate them separately to assess how they may impact children's trust. We investigated *nonsense or lack of response* errors, often witnessed as speech-recognition errors, due to their high prevalence in Human-Robot Interaction and especially in Child-Robot Interaction. Robots' inability to properly capture and respond to people's speech is of high concern to Human-Robot Interaction researchers (e.g., Mubin et al., 2014; Novoa et al., 2018), as it impedes communication. Children, in particular, tend to be on the receiving end of these errors, due to their differences in speech from that of adults, and the fact that most robot systems are trained on adult data (Kennedy et al., 2017). Hence, we focused on speech-recognition or conversational errors errors, which were witnessed as a robot providing nonsensical responses or not responding at all, to investigate how these robot-typical errors impact children's trust in robots. These findings will enable researchers and robot designers to better understand how common robot errors are perceived by children and whether they affect trust towards the robots, allowing for better and more focused design of robots and well-informed implementations.

6.2.3 Modifications to Experimental Procedure

We made some modifications to the original methodology (Chapter 4) based our experience with the first experiment (Chapter 5) and the requirements of our independent variable (i.e., robot typical errors). We removed the *clean-up* task from the experimental methodology task due to the lack of results during the task, and the possibility that it led to a lack findings during the *contrast* *label* phase (see Chapter 5, section 5.6.3 for a thorough discussion). Hence, there *same label* and *contrast label* phases for this second experiment did not have a pause in between, and the *contrast label* phase took place following the *same label* phase, removing the possibility of the *clean-up* task confounding the results (see Figure 27). In addition, the erroneous robot continued to make errors throughout the experiment, as opposed to only during the *history* phase, to remove the likelihood of confounds. Thirdly, we added differentiators to the robots' appearances to ensure that children were able to properly distinguish between the two robots. Lastly, at the end of the experiment we added a short interview to explore children's perceptions of the robots. The next few paragraphs detail all of the modifications made.

The primary difference between the experiment detailed in Chapter 5 and the one presented in this chapter, was the type of error that the erroneous robot exhibits. Experiment 1 (Chapter 5) focused on human-like informational errors while experiment 2 investigated robot-typical speech-recognition errors. Thus, in this experiment, the erroneous robot exhibited errors through a lack of response and unrelated responses, but eventually provided the expected response.

The independent variable in this experiment, however, necessitated changes to the presentation of the errors. Specifically, in the first experiment errors were only present during the *history* phase, but this approach is not compatible with speech-recognition errors. That is, if the robot made speech-recognition errors in the *history* phase but not in the subsequent phases, its regular speech may be perceived as a correction or recovery from the error. This was not a problem in the first experiment (Chapter 5) because the robot had no additional opportunities to label known objects (i.e., provide a correct or incorrect response) and therefore demonstrate a corrected or ongoing error. This suggested that the speech-recognition errors in the second experiment should continue to be present throughout the experimental session. Otherwise the child would observe the erroneous robot no longer have errors, creating a confound in the experiment. Thus, in this second experiment the erroneous robot continued to exhibit speech-recognition errors throughout the experimental session, in 2 out of 4 of the *same label* and *contrast label* trials.

The continuation of the errors into the *same label* and *contrast label* phases was not only necessary due to the nature of speech-recognition errors and the experimental design, but also provided a memory aid. In the first experiment it was unclear whether children retained the memory of which robot was erroneous and which was not throughout the entirety of the experiment. Therefore, presenting errors throughout the experimental session as opposed to only during the *history* phase, could help children maintain awareness of which robot made errors. See Appendix B for the complete experimental protocol.

Secondly, we had some concerns in experiment 1 about whether children were properly able to differentiate between the two identical robots (see sections 5.6 and 5.7 for an expanded discussion). Thus, we added identifiable shapes to the robots' chests for the second experiment, as a way to make the robots more differentiable, but still have the same appearance overall. Both robots received a blue (to match their colouring) paper cut-out on their chests, one robot a circle

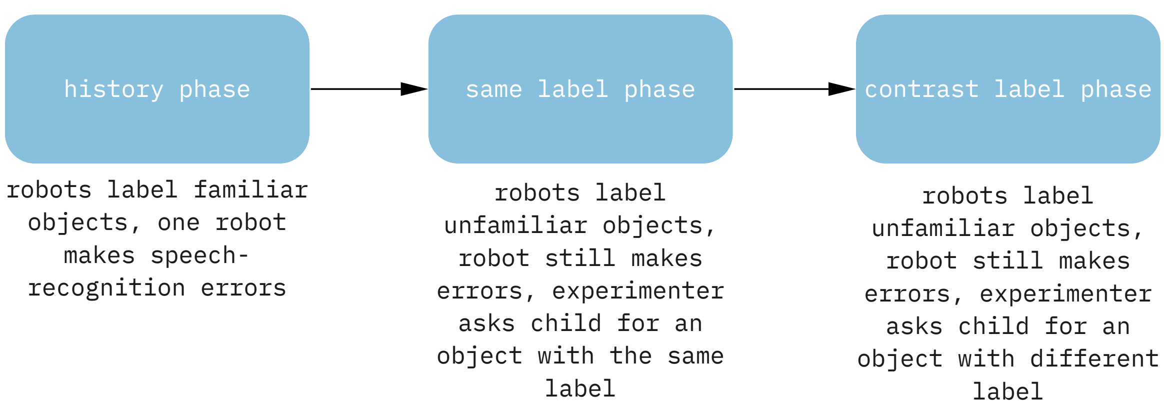


Figure 27. Procedure flow-chart for experiment 2, children's trust towards robot with speech-recognition errors.

and the other one a square (see Figure 28). This additional hint was intended to assist the children in the experiment to visually differentiate between the two robots and as a memory aid to remember which robot exhibited errors.

Lastly, at the end of this experiment we conducted a brief interview with the children to explore their perceptions of what happened during the experimental session. Specifically, we asked children whether either robot was right or wrong more than the other, or if either robot made any mistakes or acted weird in any way. These questions were asked informally, in an unstructured interview format, where follow-up questions were asked as necessary, and questions were rephrased if a child did not understand them. This brief interview was intended to provide insight

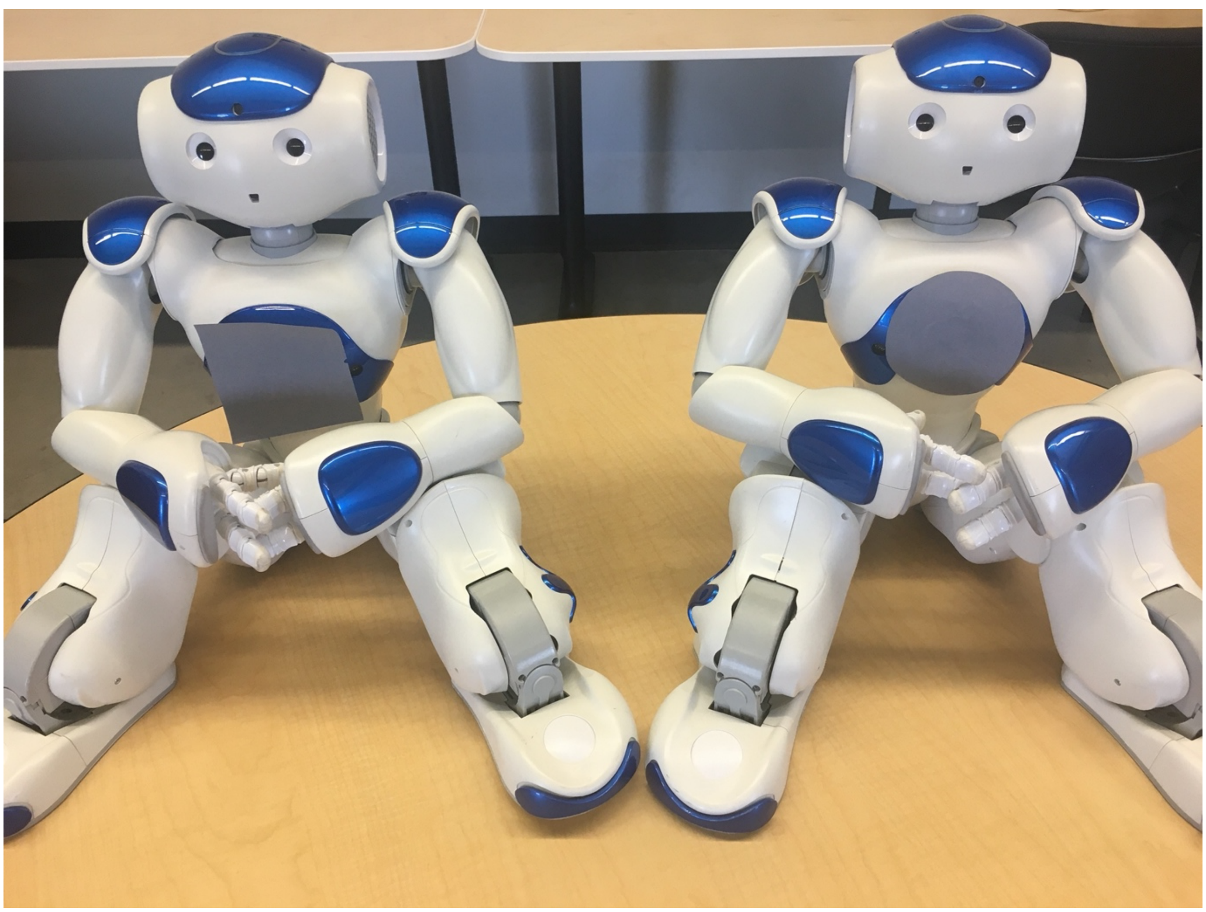


Figure 28. SoftBank robots used in the second experiment. The robots look and sound the same, but we added shapes to their chests to help children distinguish between them.

into children's perceptions of the robots and the errors, and shed additional light on what children's mental models of robot-typical errors might be.

6.3 Hypotheses

Through our experiment detailed in Chapter 5, we found that preschool-aged children sometimes trust a non-erroneous informational robot more than an erroneous one (i.e., in the *same label* phase). However, due to the lack of research in the area of children's trust towards robots, it is difficult to predict how children will react to speech-recognition errors. Thus, given the novelty of our research, and research with children and erroneous robots in general, we based our non-directional hypotheses on Human-Robot Interaction literature with adults, and Developmental Psychology. Research suggests that adults tend to trust well-functioning robots more than malfunctioning ones (Ragni et al., 2016; Salem et al., 2015), and in our first experiment we found that children similarly trust non-erroneous informational robots more. Thus, it was reasonable to hypothesize that we might see differences in responses to robots that exhibit robot-typical errors and those that do not.

6.3.1 Same Label Phase

The intention of the *same label* phase was to test trust towards the robots in a simple situation (i.e., when the experimenter provided the same object label as the robots). Our findings in experiment 1 demonstrated that preschool-aged children trust robots that do not make informational errors more than ones that do. However, it was still unknown how robot-typical speech-recognition errors affect children's trust in robots. Hence, it was unreasonable to predict the directionality of our hypothesis with such limited information. Therefore, we hypothesized that:

H1: Children will choose the objects labeled by either the erroneous robot or the non-erroneous robot more often than chance (50%), when asked to hand an object with a name that they had previously heard (i.e., during the *same label* phase).

6.3.2 Contrast Label Phase

The goal of the *contrast label* phase was to test young children's trust towards the robots in a complex situation (i.e., when the experimenter provided a different object label than the robots). In our first experiment, we did not find any differences between which robot children trusted during the *contrast label* phase. However, the lack of findings was unexpected as research suggests that there exist differences when the informants are puppets, but not always in even more complex situations (Birch et al., 2008). It was therefore unclear how previous findings for the *contrast label* phase may translate to robot-typical speech-recognition errors, and thus we provide a non-directional hypothesis. Therefore, we hypothesized that:

H2: Children will choose the objects labeled by either the erroneous robot or the non-erroneous robot more often than chance (50%), but opposite of the *same label* phase (i.e., during the *contrast label* phase).

6.4 Experiment

6.4.1 Participants

We recruited 62 children (33 male, 29 female; 18 3-year olds, 25 4-year olds, 19 5-year olds), ranging from 3 to 5 years (M = 3.97, SD = .782) to participate in this experiment. We provide additional information in the next sentences to support full understanding of our research methodology. The children were recruited from 3 daycare centres in Winnipeg, Manitoba, Canada and their parents provided written consent (see Appendix D) before the experiment took place.

Children were able to pick a small toy to thank them for their participation, and parents received a $15 honorarium. This experiment was approved by the University of Manitoba Joint Faculty Research Ethics Board (see Appendix E).

6.4.2 Robot-Typical Speech-Recognition Errors Implementation

To demonstrate the presence of speech-recognition errors, we exhibited them through the robot failing to provide responses and providing sensical but unrelated information. Precisely, the conversational pattern for the erroneous robot was as follows: step 1) experimenter asked robot to label common object, step 2) robot exhibited speech-recognition errors, step 3) experimenter prodded robot again. This pattern continued until the robot provided the expected response, which took place after the robot exhibited one to three errors (same across participants). For example, during the *history* phase, the experimenter asked the robots what a ball was called. The erroneous robot, instead of providing a correct response immediately (as the non-erroneous robot did), did not respond. The experimenter then prodded the erroneous robot again, following which the robot provided an unrelated response (e.g., “Yes. My name is Taylor”). The experimenter then prodded one last time, and the erroneous robot provided the correct response (see Appendix B for full script).

A number of factors were of consideration here. For one, while the erroneous robot appeared to have problems hearing or recognizing speech, it is able to provide coherent and sensical speech, suggesting a lack of cognitive deficiencies. Similarly, the erroneous and non-erroneous robots both provided the same correct responses eventually, with the main difference that the erroneous robot exhibited errors first. Even during the *same label* and *contrast label* phases, when the erroneous robot continued to make intermittent errors, it eventually arrived at the same response as the non-erroneous robot. This suggested that while the robot made errors, it did not make informational errors, therefore disentangling and removing the potential confound from our results.

Finally, the unrelated responses that the erroneous robot provided were intended to be neutral and factual. The goal was for the robot to not make positively or negatively charged comments, but instead provide neutral statements. For example, if it said "Yes. I like Star Wars" children may change their opinion of the robot merely based on common interests, and not based on our independent variable. Thus, when the erroneous robot provided unrelated responses, they were neutral and technically correct (e.g., "Yes. The sky is blue" or "Yes. I am sitting on a table"). These phrases were intentionally structured in a similar way to the non-erroneous robot's responses (e.g., "That's a ball. Yes, that's a ball").

6.4.3 Procedure

We utilized the experimental procedure from Chapter 4, with the modifications presented at the beginning of this chapter. The independent variable for this experiment was whether the robot made robot-typical speech-recognition errors or not, as was demonstrated by the robot failing to provide a response or providing an unrelated response (erroneous robot) or immediately providing the correct or expected response (non-erroneous robot) during the *history* phase of the experiment, as well as half of the trials in the *same label* and *contrast label* phases. A description of the procedure, to specify differences from the original experimental methodology (Chapter 4) and the first experiment (Chapter 5), is found below (also see Figure 27).

The experiment took place at 3 daycare centres starting in the Fall of 2019 and ending in the Winter of 2020. Prior to the experiment taking place, we conducted introductory sessions with the preschool-aged children at each daycare (as explained in Chapter 4). Following, within one to two weeks of the introductory session, we conducted the experimental sessions. Daycares in which the experiment spanned multiple days (due to a large number of participants or time limitations) had all sessions completed within one week.

The experimental session began similarly as in the first experiment, with the child being escorted to the experimental room and introduced to the experimenter and the robots (see Chapter 4 for a thorough description). Following, in the *history* phase, the robots provided labels for four common objects, following our established protocol (see Chapter 4). While naming the common objects, one of the robots exhibited speech-recognition errors (as outlined above; the erroneous robot), while the other simply provided correct labels (the non-erroneous robot). However, the erroneous robot was also able to arrive at the correct response after one to three tries. This was the independent variable in this experiment, focusing specifically on which robot made errors and which did not.

During the *same label* phase of the experiment, the two robots provided labels for less common objects, each robot providing the same name for a different object, therefore providing conflicting information. The child was then asked to hand the object with the matching name to the experimenter. This took place in the same way as in the first experiment, with the difference that the erroneous robot continued to make errors in 2 out of 4 of the trials (see Figure 27 and Appendix B). This was due to the nature of robot-typical speech-recognition errors, being unlikely to disappear after the *history* phase (see Section 6.2.3 for an expanded discussion). During this phase of the experiment, we expected children to give the experimenter the objects labeled by the non-erroneous and erroneous robots to different extents. The object that the child handed to the experimenter was used as a measure of trust.

In the *contrast label* phase of the experiment, the robots once again provided labels for less common objects, each of them providing the same name for different objects, but in this case the experimenter asked the child for an object that did not match the labels provided by the robots. The erroneous robot continued to exhibit speech-recognition errors in 2 out of 4 trials. During this phase of the experiment, we expected children to give the experimenter the objects labeled by the erroneous and non-erroneous robots to different extents (but possibly opposite of the *same label* phase; see section 6.3.1). The object that the child handed to the experimenter was used as a measure of trust.

Lastly, the child was asked to label the four common objects from the *history* phase, to maintain a similar experimental procedure as previous experiments and check their communication abilities. Afterwards, the experimenter informed the child that the robots would be going to sleep, and the child and robots were able to say goodbye to each other. This was done to reduce the impact of the robots being present while the child was interviewed. Once the robots were 'asleep,' the experimenter asked the child to indicate whether either of the robots had acted strangely or made any mistakes, and asked follow-up questions as necessary in an unstructured interview style (see Section 6.2.3 for a more thorough explanation).

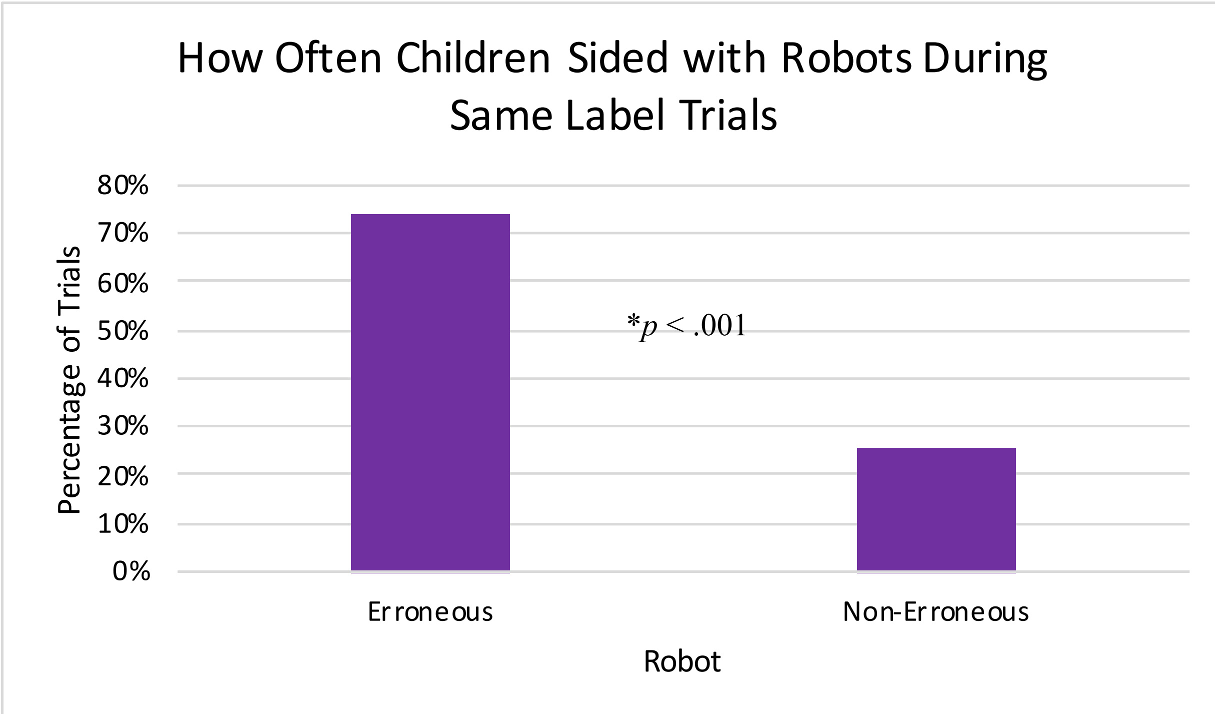
This was a within-subjects experiment, with the placement of the robots (right vs. left), and the order in which they named objects being counterbalanced between participants. Therefore, there were four counterbalancing conditions: speaking order (Casey vs. Taylor), and seating order (Right vs. Left). A detailed data analysis strategy can be found in Chapter 4 section 4.4.

6.5 Results

Out of the 62 participants that were recruited, 2 participants were removed from the *contrast label* phase (1 due to a data recording error, 1 due to not following directions), 2 did not want to participate, and 2 were removed from the entire experiment (due to not following directions). Thus, the data from 56 participants (30 males, 26 females; 16 3-year olds, 25 4-year olds, 15 5-year olds; $M = 3.98$, $SD = .751$) was analyzed for the *same label* phase, and from 54 (29 males, 25 females; 16 3-year olds, 25 4-year olds, 13 5-year olds; $M = 3.94$, $SD = .738$) participants for the *contrast label* phase. This approach to missing data is appropriate given that we are not comparing between the two phases, but instead analyzing them separately (Field, 2009). A two-tailed one-sample t-test on the aggregate robot choice ratio (choosing a particular robot from 0% to 100% of the four trials) against chance of 50% (as detailed in Chapter 4) indicated a statistically significant difference between which robot children trusted during the *same label* phase, $t(55) = 4.81$, $p < .001$, $d = .64$ (see Figure 29). Children trusted the erroneous robot more than the non-erroneous one ($M = .259$, where children sided with the non-erroneous robot 25.9% of the time and with the erroneous robot 70.5% of the time compared to chance of 50%, $SD =.375$), thus the first hypothesis is supported as there is a difference between which robot children sided with. However, we did not find a statistically significant difference between which robot children trusted during the *contrast label* phase, $t(53) = .663$, $p = .51$, $d = .09$ ($M = .537$, where children sided with the non-erroneous robot 53.7% of the time and with the erroneous robot 46.3% of the time compared to chance of 50%, $SD = .41$), therefore we did not find support for the second hypothesis (see Figure 29). We also conducted an independent-samples t-test to assess any gender differences in trust towards the robots (based on our post-hoc findings from the first experiment, Chapter 5) and did not find a statistically significant gender effect during the *contrast label* phase, $t(52) = .379$, $p = .71$, $M_G = .560$, $SD_G = .423$, $M_B = .517$, $SD_B = .406$. Gender effects were not assessed for the *same label* phase. No other analyses, such as age, robot placement, or robot speaker order, were statistically significant.

The data from the unstructured interviews was coded into three codes (*correct*, *incorrect*, and *made mistakes*) and separated per robot. Out of 41 children who provided feedback (some children were unsure or unwilling to answer the questions), there were a total of 70 data points which fell into our three categories. In terms of correctness, 18 children indicated that the erroneous robot had provided more correct responses, while 21 children indicated that the non-erroneous robot had provided more correct responses. In reality, both robots provided the same responses, but the erroneous robot took longer to arrive at them. On the other hand, 15 children mentioned that the erroneous robot had provided incorrect responses, while 7 said the same for the non-erroneous robot. Three children mentioned that the erroneous robot had made mistakes, and 4 that the non-erroneous one had. Lastly, 2 children said that both robots gave correct responses, and 1 that both robots had made mistakes.

Overall, we found that children's responses about the robots were not always mutually exclusive. Some children mentioned that the same robot was both correct and incorrect most of the time, while others only indicated that one robot fell into these categories (e.g., the non-erroneous was correct more, but the erroneous robot had not made any mistakes or been wrong more). In addition, a few children mentioned that one of the robots was copying the other (i.e., the



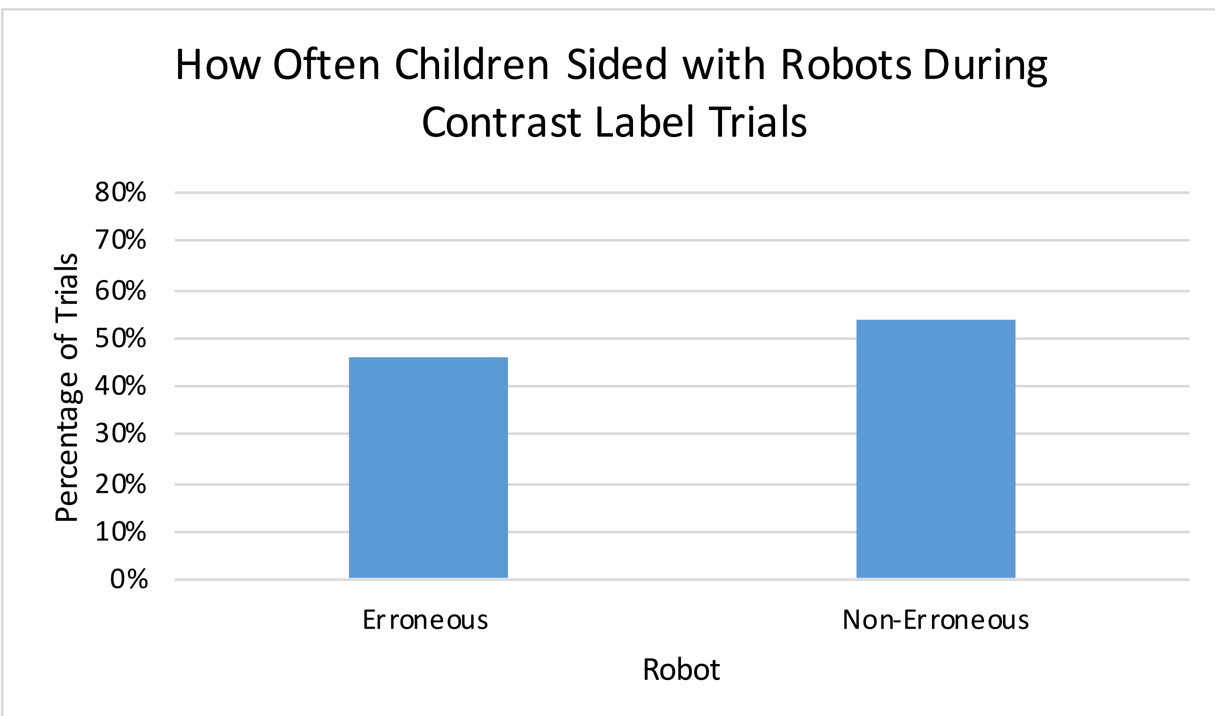


Figure 29. How often children sided with the robots during the *same label* and *contrast label* phases of experiment 2, where one robot made robot-typical speech-recognition errors. There was a statistically significant difference during the *same label* phase, with children trusting the erroneous robot more than the non-erroneous one. However, there was no such difference in the *contrast label* phase.

robot that spoke second was copying the first), as well as highlighting that the erroneous robot was being funny.

We conducted a χ^2 test to investigate whether there was a relationship between the trusted robot and which robot was perceived to provide more correct or incorrect responses. This was a post-hoc test and was chosen because it allowed us to explore the relationship between a nominal independent variable and a nominal dependent variable. We did not find a statistically significant relationship between which robot children trusted (through action) and their reports of which robot provided more correct or incorrect responses (*same label* phase: $\chi^2_{Correct} = .077$, $p = .782$ – 19 children said the non-erroneous robot was correct more: 4 trusted the non-erroneous robot and 15 the erroneous robot, 16 children said the erroneous robot was correct more: 4 trusted the non-erroneous robot and 12 the erroneous robot, $\chi^2_{Incorrect} = .756$, $p = .385$ – 6 children said the erroneous robot was incorrect more: 2 trusted the non-erroneous robot and 4 the erroneous robot, 13 children said the erroneous robot was incorrect more: 2 trusted the non-erroneous robot and 11 trusted the erroneous robot; *contrast label* phase: $\chi^2_{Correct} = .938$, $p = .333$ – 19 children said the non-erroneous robot was correct more: 9 trusted the non-erroneous robot and 10 the erroneous one, 14 children said the erroneous robot was correct more: 9 trusted the non-erroneous robot and 5 the erroneous robot $\chi^2_{Incorrect} = .457$, $p = .499$ – 6 children said the non-erroneous was incorrect more: 2 trusted the non-erroneous robot and 4 the erroneous robot, 12 children said the erroneous robot was incorrect more: 6 trusted the non-erroneous robot and 6 the erroneous robot). Thus, it appears that children's perceptions of which robot made errors, and which robot they actually placed their trust on were not related. We briefly discuss this in the next section.

6.6 Discussion

The results from this second experiment reinforce that children build their trust models of a robot based on their previous experience with that robot. However, the pattern of trust appears to be different than that of children and other people or puppets. This provides insight into how children perceive robot-typical errors, specifically speech-recognition errors, and how they affect children's attribution of mental states to robots. While our results did not follow the same path as in the first experiment or previous research with people, they point to a potential and unexpected difference in trust towards erroneous robots from people. In the remainder of this section we discuss the results in greater detail.

6.6.1 Same Label Phase

In our experiment, we found that errors that a robot had made during the *history* phase and throughout the *same label* phase had an effect on how much children trusted it during the *same label* phase. Children sided with the robot that had made errors more often than the one that had not. While this supports our first hypothesis, in that there was a statistically significant difference between the groups, it does not align with our first experiment, nor previous research related to puppets or adults (Birch et al., 2008; Corriveau & Harris, 2009a). While further research is necessary to fully comprehend these findings, a few potential explanations come to mind.

It is possible that speech-recognition errors are not regarded as errors by preschool aged children, or are maybe seen as intentional behaviour. Some children remarked that the erroneous robot was being funny whenever it exhibited errors, laughing at the errors that took place. Children may have perceived the robot errors as the robot being purposely funny, instead of as it experiencing malfunctions. In fact, previous research has also found that when preschool-aged children witness unexplainable errors from a human informant, they sometimes justify the errors as the informant being purposely funny (Nurmsoo & Robinson, 2009b). Therefore, it is possible that young children simply view robots' speech-recognition errors as funny, or that given the lack of a logical reason for this type of error, they attribute it to the robot's comedic efforts. Further research will be necessary to examine why children may have found these errors to be funny in our experiment.

Likewise, because our experiment had two robots that answered questions sequentially, some children reported that one robot (whichever one answered second) was copying the first. Three of these children said this when the non-erroneous answered second, while 1 when the erroneous robot answered second.

On the other hand, it may be that our findings stem from choices in our experimental design. Since conducting the experiment, we have realized that due to the nature of the speech-recognition errors, the erroneous robot simply spoke more and had more social interactions with the experimenter. This is because the experimenter had to keep prodding the erroneous robot to provide a response to the question that was asked. It is possible that this difference in interactions had an impact on how children perceived the robots, viewing the erroneous robot as more social. For example, researchers have found that robot errors can sometimes make adults perceive robots as more sociable and therefore likeable (Mirnig et al., 2017; Salem, Eyssel, Rohlfing, Kopp, & Joublin, 2013). It is therefore possible that children simply picked up on those social cues, which led to higher liking and therefore trust towards the erroneous robot than the non-erroneous one. Further research in which the social interaction levels of the two robots are equivalent would aid in isolating the true cause of our findings.

These findings bring to question the impact that robots' speech-recognition errors can have and call for further inquiry into how such errors may affect children's perceptions of robots. Robots exhibit speech-recognition errors frequently (Mubin et al., 2014; Novoa et al., 2018), especially in the presence of children who might speak differently or be further away from the robot's microphone than adults (Kennedy et al., 2017). Our findings point to the concerning possibility that teaching assistant, shopping, or home assistant robots may elicit more trust from young children when they make these types of errors, leading to concerns of over-trust, which tends to already be present in both children and adults. Further exploration of how speech-recognition errors may affect young children's trust in a variety of contexts can help provide insight into our findings and their implications.

6.6.2 Contrast Label Phase

We failed to reject our second hypothesis that children will trust a non-erroneous and an erroneous informant to different extents, in a more complex situation. We did not find that children trusted the non-erroneous or erroneous robots differently during the *contrast label* phase, when children were asked for an object with a name the robots had not mentioned.

There are a number of reasons why we may not have seen differences during this phase. For one, it is possible that, as mentioned in Chapter 5 section 5.6.2, interactions with robots are more complex than those with humans or puppets, leading to higher cognitive load and more difficulties in choosing who to trust. Additional research to assess whether this is the case, by testing children's cognitive load while interacting with robots, would be necessary to understand our lack of findings.

It may otherwise be that the placement of the *contrast label* position towards the end of the experiment may have caused our lack of findings. As explained in Chapter 4 section 4.3.3, the order of the phases in our experiments was always *history*, *same label*, and *contrast label*. However, it is possible that the placement of the *contrast label* phase at the end of the experiment somehow led to our lack of findings. A similar experiment can be conducted, but counterbalancing the *same label* and *contrast label* phases, therefore removing the position of the phases within the experiment as a potential confound.

Lastly, we did not find any gender differences in how girls and boys trusted the robots during the *contrast label* phase, which we had expected due to our post-hoc findings from the first experiment. This points to the need for further research, as it appears that girls and boys may potentially perceive robot errors differently, but it is unclear whether this is the case or what factors (errors or other) may affect this.

6.6.3 Perceptions of errors

We asked children about their perceptions of the robots (i.e., whether they had many any mistakes, if either of them had been right or wrong more, or if they did anything strange) to gain insight into their perceptions of these robot-typical errors. Given that most children did not report the erroneous robot to have made errors or even mistakes, it appears that young children may not perceive speech-recognition errors as errors, disambiguating them from the actual responses that the erroneous robot eventually provides. While children provided a range of responses for which robot had been more correct or incorrect, or whether either of them had made mistakes, these responses do not seem to be related to which robot they trusted during the experiment. This suggests that robot-typical errors such as speech-recognition errors may not be perceived as errors at all, calling for further research into speech-recognition errors as well as other robot-typical errors and how they may be perceived by young children, and affect their trust.

6.7 Chapter Summary

In this chapter, we detailed the second experiment of this thesis, which investigated how young children trust robots with speech-recognition errors. We found that children do sometimes trust robots differently depending on whether they have made speech-recognition errors or not, but mainly in a simple scenario. These results provide novel insight into how young children perceive robot-typical speech-recognition errors, and how errors affect children's trust.

The combined findings from the experiments discussed in Chapter 5 and Chapter 6 point to the effects that robot errors can have on children's trust, and how these might differ. We use these findings as well as literature from various fields to develop a framework of child-robot trust in Chapter 7. We then provide a thorough discussion of our experiments and framework in Chapter 8, and present concluding remarks in Chapter 9.

Chapter 7 Toward a Framework of Child-Robot Trust

Up to now, we have explored and discussed how robot errors can impact young children's trust towards robots. However, although addressed briefly in Chapter 2, the HRI field still lacks comprehensive understanding of the nuances of trust, and general knowledge of experimental results on trust with adults and children, towards people, machines, and even robots. In addition, within HRI, we do not yet have a simple reference or cumulative knowledge of trust towards robots. As such, in this chapter we aim to start to work in this direction: we look to a broad range of fields including Psychology, Sociology, Human-Computer Interaction, and Human-Robot Interaction with the aim of synthesizing this knowledge from the perspective of children and robots. This results in an initial framework of child-robot trust, that researchers can use as a tool to better understand and study children's trust towards robots, and have useful vocabulary to then discuss their research.

To develop this framework, we first searched the literature for research on trust that could apply to child-robot trust, from Psychology, Sociology, Human-Computer Interaction, and Human-Robot Interaction. From there, we systematically analyzed and synthesized this literature into dimensions and factors to help our understanding of children's trust towards robots. In this chapter, we present such literature as we construct our framework, and provide vocabulary for researchers to discuss trust.

To develop a framework of child-robot trust, we first examined the literature from Psychology and Sociology on how humans trust one another, and developed dimensions and factors emergent from the literature. Following, we narrowed down to trust towards systems and robots (in order), by reviewing the literature from the respective fields and applying the dimensions and factors we created in earlier sections. Lastly, we focused in on children's trust towards robots, and utilized the knowledge base we built in earlier sections to ultimately arrive at relevant factors for child-robot trust.

We envision for this initial framework to serve as a tool for researchers: providing a knowledge base on trust and factors to consider when investigating trust with robots, especially withing Child-Robot Interaction (CRI). This framework also provides vocabulary for researchers to use when discussing and exploring trust, and a starting point for future trust research in CRI.

7.1 Methodology

To investigate trust and how it might relate to children and robots, we decided to take a wide-ranging iterative approach by conducting a literature review of trust followed by organizing the information we obtained, and then repeating the process as necessary. Thus, we first conducted a broad survey of the literature. Then, we read and tagged the relevant manuscripts in terms of application and focus on trust. Following, we organized these papers into themes, and iterated over these to obtain the framework presented in this chapter. This work was conducted by the author of this thesis, with regular discussion meetings with the PhD supervisor. The author of this thesis has background in Psychology, Human-Computer Interaction, and Human/Child-Robot Interaction, which was utilized in the process of developing this framework. Thus, this framework was developed using existing literature, which was synthesized by the author through an interdisciplinary lens.

For the broad literature search, we used a combined approach of keywords searches, targeted publication venues, and cross referencing. For information on general and children's trust, we focused our search on PsycINFO and the University of Manitoba combined databases given their broad coverage of peer-reviewed research. For general and trust-specific research on machines and systems, we surveyed Google Scholar, as well as the premiere Human-Computer Interaction venue: the ACM International Conference on Human Factors in Computer Systems (CHI Conference). Given the focus of this work on Human-Robot Interaction, we further specifically targeted a range of venues and did a whole-publication history review of trust work relating to robots. Thus, in addition to the venues mentioned above, we searched the archives of the: ACM International Conference of Human-Robot Interaction (HRI Conference), Springer International Conference of Social Robotics (ICSR Conference), Springer International Journal of Social Robotics (SORO Journal), and ACM Transactions in Human-Robot Interaction (THRI Journal). In these resources we searched for the terms: trust, humans/people, systems/machines, robots, children, development. We reviewed the papers obtained, and saved ones related to trust and humans, systems, robots, or children.

Once we filtered the papers for relevance, we read though them in detail and organized the information in relation to what it offered to our framework and our target area of child-robot trust, such as definitions or trust outcomes. Then, we organized these tags into emergent themes: trust definitions, types of trust, trust towards different actors (i.e., humans, systems, robots), trust development, and variables that affect trust.

In this chapter, we present the results of our survey and synthesis in the form of a framework on trust; we developed this in an iterative fashion as described earlier, drawing from the literature. In addition to conducting several iterations, we engaged in an ongoing review of new published literature, as well as publications that were not necessarily about trust, but could provide examples or grounding for the presentation of our framework.

7.2 Diverse Perspectives on Trust

An initial, and somewhat surprising for us, finding of our survey was the shear diversity of perspectives that exist of trust. We found a variety of (often conflicting) trust definitions, each highlighting different aspects of trust or focusing on differing contexts or application domains (Pytlikzillig & Kimbrough, 2016; Rousseau et al., 1998). For example, distinct definitions of trust exist for organizational psychology (e.g., Pytlikzillig & Kimbrough, 2016; Rotter, 1967) and sociology (e.g., Lewis & Weigert, 1985). Thus, our goal in developing this framework was to learn from each branch of trust research to inform how the construct might apply in different contexts, relationships, and demographics, narrowing down to child-robot trust. For example, some definitions highlight the expectation of a positive return in trust (e.g., Mayer, Davis, & Schoorman, 1995; Rousseau et al., 1998), while others focus on the abstract concept of trust, regardless of outcomes (e.g., Frederiksen, 2012). Some researchers believe that trust develops as an outcome of personal factors, while others point to situational variables instead (Pytlikzillig & Kimbrough, 2016). In this chapter we explore this range of trust and bring it together under the umbrella of trust toward robots, and ultimately child-robot trust.

Although applications explored are widespread, the vast majority of trust research (e.g., in Psychology and Sociology) focuses on trust between adults. A body of work also exists studying trust towards machines (e.g., recommender systems), and towards robots. Overall, only a small fraction of research studies children. Thus, in our work we explore trust towards robots ultimately from a child perspective, utilizing work from other disciplines to inform how children may trust robots.

Our goal in this framework is to synthesize the broad range of knowledge we found in our survey into a cohesive body of work focused on children and robots, providing researchers with comparisons and explanations of the different types of trust that have been studied, and how they may apply to CRI (as well as HRI and HCI to some extent).

7.2.1 Trust Dictionary

Throughout the following sections, and the remainder of this chapter, we use a range of definitions and types of trust to explain trust relationships. These definitions were presented in our Dictionary of Trust in Chapter 3, but we briefly reiterate them here for the sake of readability. We present these definitions in Table 2, and bold the terms when presented throughout this chapter. See section 7.1 for an explanation of our methodology.

7.3 Dimensions and Factors Relevant to Trust Between Humans

Trust is a complex construct that has been studied in many disciplines including Psychology, Sociology, Management, Economics, and Politics. The concept of trust develops from a young age, as children learn what it means to trust others, and the consequences of doing so. Some argue that trust is a "basic sense of faith in the self and the world" which develops as a child learns self-control (over bodily processes and external objects; Erikson, 1950). This implies that depending on a child's early experiences, in terms of his or her control over them, they may have more or less trust in others, and the world as a whole, later on. Rotter (1967), on the other hand, suggested that

Table 2. Dictionary of trust developed through our survey of the trust literature (also discussed in Chapter 3).

Trust Category	Trust Term	Definition	Reference
Personal	**Interpersonal**	Belief held by an individual that others can be relied on. Composed of cognitive and emotional components	(McAllister, 1995; Rotter, 1971)
	Dyadic	Belief that other person is honest and benevolent	(Larzelere & Huston, 1980)
	Relational	Develops due to repeated positive interactions	(Lewicki, Tomlinson, & Gillespie, 2006; Rousseau et al., 1998)
Systematic	**Calculus-based**	Calculated decision, based on evidence	(Lewicki et al., 2006; Rousseau et al., 1998)
	Deterrence-based	Consequences of not keeping trust would be too severe	(Shapiro, Sheppard, & Cheraskin, 1992)
	Institution-based	Faith in an institution's benevolence	(Lewicki et al., 2006)
	System	Sense that everything is in good order	(Luhmann, 1979)
Societal	**Cognitive**	Carefully thought-out trust	(Lewis & Weigert, 1985)
	Emotional	Feelings-based trust, either based on strong (possibly long-term) emotions or temporary hunches	(Lewis & Weigert, 1985)
	Ideological	Trust high on emotional and cognitive components, and prevalent as it relates to society	(Lewis & Weigert, 1985)
	Mundane	Trust low on emotional and cognitive components, prevalent in everyday life	(Lewis & Weigert, 1985)

trust is a general expectancy that others will keep their promises. These two definitions differ greatly; Erikson suggests that trust depends on the past is a broad concept, while Rotter argues that it is specific to promises (spoken or not) and likely depends on the near past and present. While the ability to trust begins in infancy, most general trust research has focused on adults.While children learn to trust (or not) from a young age, research on children's trust towards other entities (e.g., humans, systems, robots) tends to be limited to certain contexts (i.e., towards informants). In addition, children's trust towards machines and robots is understudied, with research only recently beginning to appear (e.g., Geiskkovitch, Thiessen, Young, & Glenwright, 2019; Stower & Kappas, 2020). Thus, in this section, we first focus on trust towards adults, and then narrow down to children at the end of the section. Our survey and synthesis of factors that affect trust between people resulted in the creation of the following three overarching dimensions for factors trust between people: *social proximity*, *person factors*, and *external factors* (Figure 30). We detail these, with relevant background work, below.

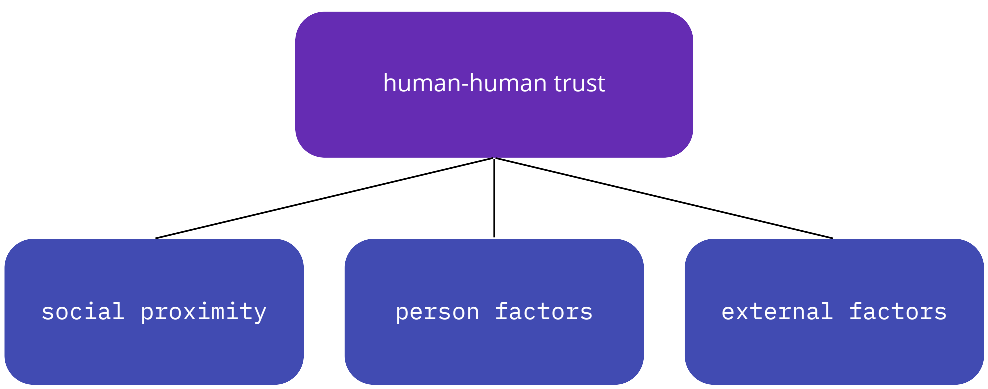


Figure 30. Dimensions resulting from our survey on trust between humans: factors that affect trust between people.

7.3.1 Social Proximity to Others and Human-Human Trust

Our survey found that the emotional and psychological proximity in which someone is in relation to another individual(s) or entity can directly affect the type of trust they might experience (e.g., Ashley Fulmer & Gelfand, 2012). For example, we might expect someone to experience different kinds and levels of trust towards their co-workers than a romantic partner. Thus, the literature indicates that the type of relationship that people have can dictate the types of trust that may or may not develop (Weiss et al., 2020). Through our synthesis we further divided work found under social proximity dimension into four categories, denoting the context of who the trust is between: *intimate*, *individuals*, *institutions*, and *society* (Figure 31).

Social proximity: Trust in intimate relationships – We can expect individuals in intimate relationships to have a lot of information to base their trust on. In these relationships, individuals tend to have a general expectancy that others will keep their word (**interpersonal trust**; Rotter, 1971) due to cognitive and/or affective components (McAllister, 1995). People also tend to believe that the other individual is honest and benevolent (Larzelere & Huston, 1980), and can therefore make themselves vulnerable by trusting them (**dyadic trust**).

Social proximity: Trust towards individuals in non-intimate relationships – While less likely, **dyadic** and **interpersonal trust** are possible in less intimate relationships, but individuals might also develop trust due to repeated (but not necessarily close) interactions (**relational trust**; Lewicki, Tomlinson, & Gillespie, 2006; Rousseau et al., 1998). For example, individuals may be more likely to trust their co-workers because of regular interactions, even if they are not in close

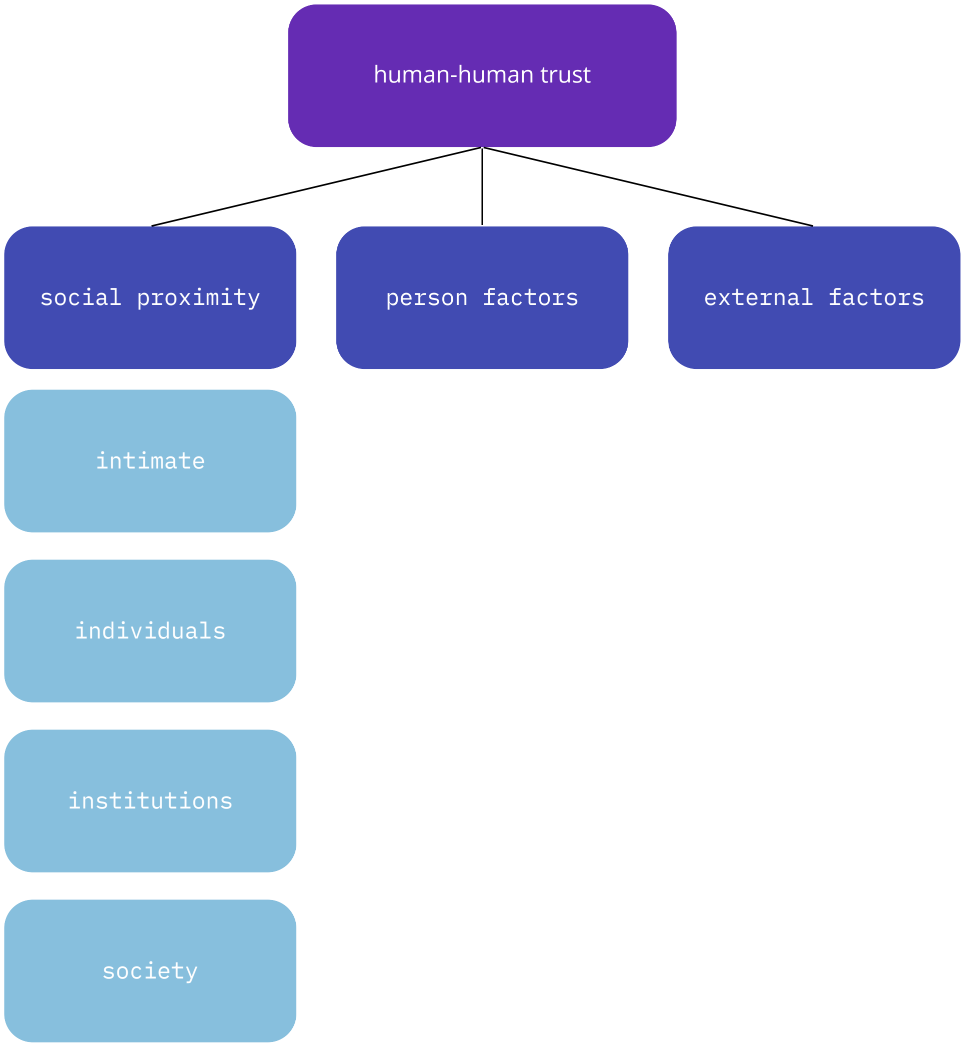


Figure 31. Factors that affect trust between humans of different social proximities, synthesized from our survey.

relationships with them (e.g., Cranston, 2011). This is especially true if the interactions with these individuals are pleasant and lead to positive outcomes.

Social proximity: Trust towards institutions – People develop trust in institutions and organized groups that they may interact with, such as banks, stores, educational facilities, etc. (e.g., Fuglsang & Jagd, 2015). Here, we might expect **interpersonal trust** to play a smaller role, but for the institution itself to harness a sense of trust (**institution-based trust**; Lewicki et al., 2006). For example, we may expect people to trust Apple or Nike products, or a Bachelor's degree from an esteemed university, simply due to the brand or institution (e.g., Elliott & Yannopoulou, 2007). People might also exhibit **relational trust** towards institutions, for example, by increasing or maintain trust towards them with accumulated positive exposures or decreasing trust with negative ones (e.g., Ashley & Leonard, 2009).

Social proximity: Trust within society – Trust is an essential part of society, as without it life as we know it would not be possible (Lewis & Weigert, 1985). Having a general sense that everything seems to be in order is critical for our social systems to succeed (**system trust**; Luhmann, 1979). For example, in North America, people might assume that unless informed otherwise, politicians are making policies for our good, and are not abusing the system (e.g., Parker & Parker, 1993). This is aided by beliefs that individuals can trust others due to practical reasons, and the high cost of broken trust (**deterrence-based trust**; Shapiro, Sheppard, & Cheraskin, 1992). In this way, for example, most people feel safe storing their hard-earned money at a bank, because if the bank were to steal their money, it would lead to very large consequences.

In everyday life, when interacting with individuals or institutions in society that we do not have connections with, our trust can also stem from the combination of cognitive and emotional components (Lewis & Weigert, 1985). For example, people tend to trust traffic lights to guide them safely without apparent cognitive or emotional considerations (**mundane trust**). On the other hand, we would expect individuals who follow religious practices may have a large number of emotions and cognitions invested into such trust (**ideological trust**).

7.3.1.1 How human social proximity might apply to robots and children

Given the propensity for people to anthropomorphize robots (Young et al., 2011), we can expect many of the heuristics that we apply when interacting with people every day to also apply to robots. For example, it might be possible for a companion or pet robot in intimate proximity to elicit **dyadic trust**, and pretend to reciprocate it. Similarly, for robots that we encounter frequently (but not intimately) we can develop **relational trust**; A robot that makes mistakes, or does not act in a proper manner is unlikely to be further interacted with (Mirnig et al., 2015), making it important to consider these aspects of robots. With children specifically, we might expect similar trust outcomes, building relationships with robots similarly to adults.

In this section, we discussed how various types of trust are linked to our relationships to others, and how this knowledge might be applicable to robots. In the next section, we investigate how characteristics related to each individual may affect their trust in other people.

7.3.2 Person Factors Affecting Human-Human Trust

Personal factors, either stable or situation-specific, emerged as a factor affecting the types of trust people experience. Individuals have different cognitive and emotional preferences and capabilities, and we may expect them to interpret and react to situations differently. This dimension is more individualistic and personalized than social proximity, and individual aspects are considered rather than their relation to others. Through our synthesis we categorize person factors into *emotional*, *cognitive*, and *personality factors* (Figure 32).

Person factors: Emotional factors affecting trust – Individuals have unique emotional capabilities and responses, and these can affect trust in different ways (e.g., Schwerter &

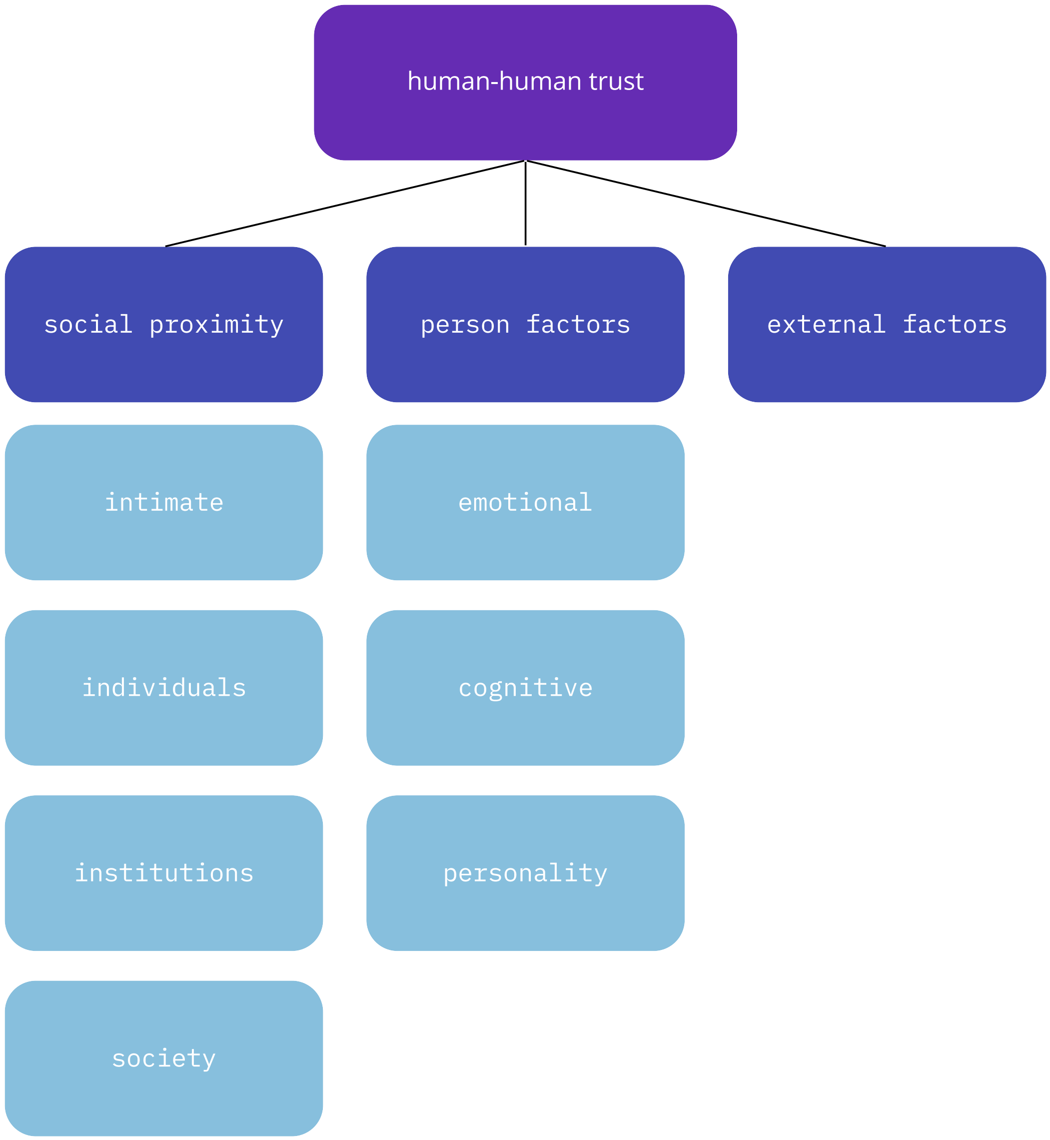


Figure 32. Factors related to each individual that emerged through our survey, and that affect trust between humans. In addition, the previously discussed social proximity factors.

Zimmermann, 2020). In some situations, people might trust based solely on how they feel about the situation (**emotional trust**; Lewis & Weigert, 1985), or may combine feelings with cognitions (**ideological trust**). We can therefore expect an individual's emotional capabilities to affect their trust; an individual who does not need to think through things to trust might be more likely to trust others.

Person factors: Cognitive factors affecting trust – Similarly to emotions, cognitions can differ from person to person in terms of capabilities and preferences. Depending on an individual's abilities and choices, they might be able and willing to apply trust in some situations based solely on thoughts and disregarding feelings (**cognitive trust**; Lewis & Weigert, 1985). Similarly, some individuals might be more likely to make rational choices backed up by credible information than others, which we label as a cognitive factor (**calculus-based trust**; Lewicki et al., 2006; Rousseau et al., 1998). For example, an individual that purchases a vehicle from a dealership after thoroughly researching models and prices is exhibiting **calculus-based trust**. We propose that depending on the types of cognitions that people have, they can experience different types and amounts of trust in the same situation.

Person factors: Personality factors affecting trust – Through our research, we have come to understand that an individual's personality greatly influences the amount and types of trust that they will experience in different situations. A person who is low on openness or high on neuroticism (personality factors) might be less likely to experience a general sense of trust, leading to low **interpersonal trust** (Garske, 1976). Similarly, we might expect an individual who is generally negative or skeptical to constantly question daily events and others' intentions, making it less likely for them to experience high **system trust**. We can extrapolate this to other personality factors, such as how cognizant or emotional an individual is which could similarly influence the types and amount of trust they experience every day in society (**societal trust**).

7.3.2.1 How person factors might affect trust in robots and children

While some person-related factors may affect how much or what types of trust people experience towards robots, not all factors identified for human-human trust may be transferable to robots. For example, we believe that there might be less of a focus on emotional and cognitive factors when evaluating trust in robots in everyday life because of the decreased likelihood of having to make serious decisions in relation to them. Personality, on the other hand, can play a large role in trust towards robots; for example, research suggests that having a highly open and agreeable personality can lead to being more likely to interact with a robot, and therefore gather enough information to develop trust (e.g., Takayama & Pantofaru, 2009).

For children specifically, person-related factors may not be as relevant, depending on their age and level of development. For example, young children are still developing as cognitive and emotional beings. Thus, we might expect demographic factors such as age, level of development, and potentially gender to be more relevant for child-robot trust.

In this section, we discussed factors specific to each person that we believe affect trust. In the following section, we look into external factors that may also play a role in trust towards others.

7.3.3 External Factors Affecting Human-Human Trust

Our literature survey highlighted the importance of additional external factors on shaping a person's trust. These are aspects of the situation or context that influence the presence and types of trust. For example, we might trust a friend with our money when they are working as a bank teller, but we may not be willing to lend them money when at a casino (e.g., Hawley, 2014). Two external elements that we propose may impact trust are *expected outcomes* and the *context of the situation* (Figure 33).

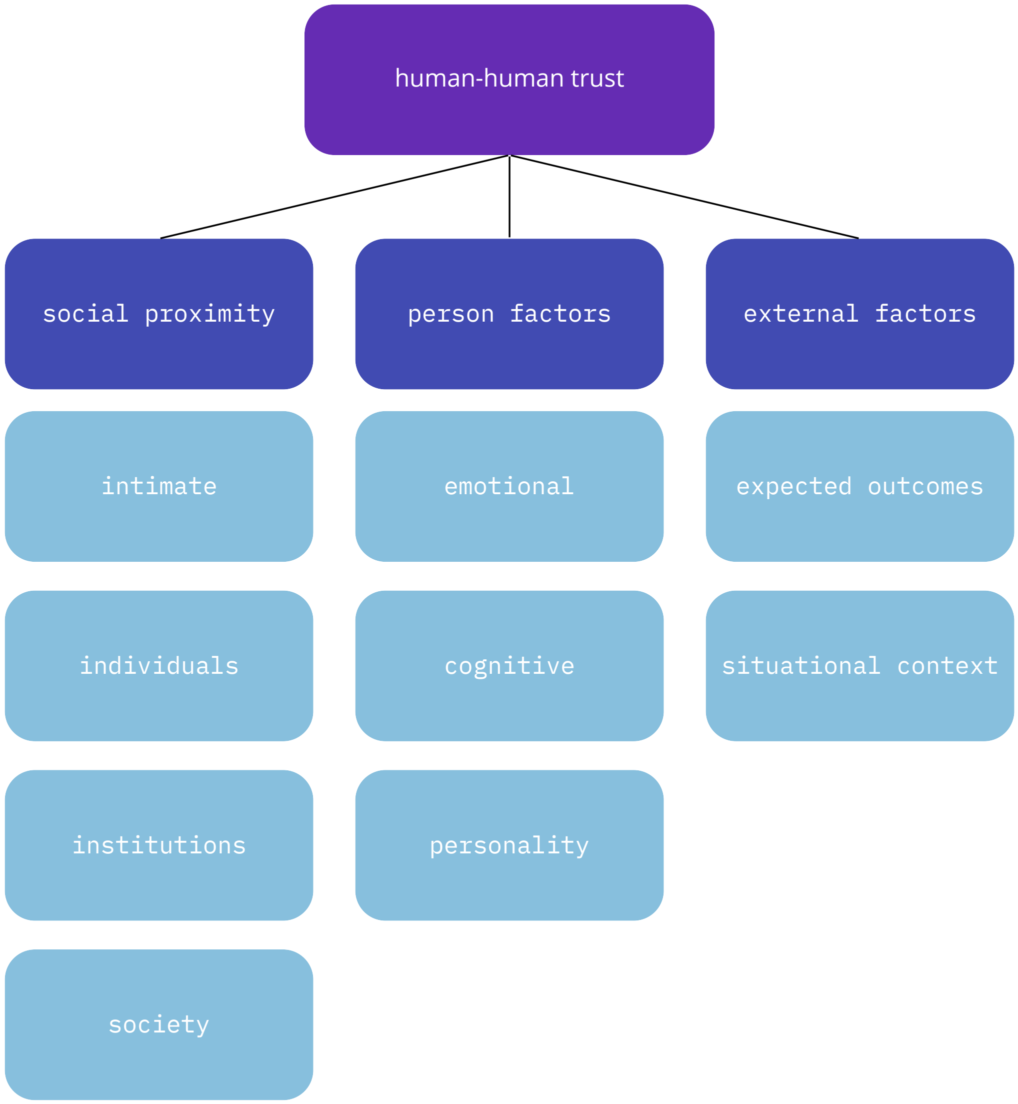


Figure 33. Dimensions and respective factors we developed through our extensive survey on trust between humans.

External factors: Expected outcomes affecting trust – The types of outcomes that people expect from situations (i.e., what could happen if trust was kept or broken) might alter trust (e.g., Evans & Krueger, 2014). For example, an individual may trust somebody in some situations due to the negative outcomes that would occur if the trust was betrayed (**deterrence-based trust;** Shapiro et al., 1992). We believe that expected outcomes could be especially important in regards to trust towards institutions (i.e., **institution-based trust**) and in close relationships (i.e., **dyadic trust**).

External factors: Situational context and trust – Although our survey revealed an apparent gap in this area, with a lack of research on how situational context (other than uncertainty; Goto, 1996) impacts trust towards others, context is particularly important for interaction in general and as such this aspect should be discussed. We may expect that the context of a situation can play a role in whether trust is present and to what extent. For example, you may trust that your co-worker will complete their job on time, because you have become accustomed to that (**relational trust**), but you may not trust them enough to co-sign a loan with them. In this case, the context of trust can play a large role in whether trust is present or not.

7.3.3.1 How external factors affecting human-human trust might apply to robots and children

We believe that the perceived expected outcomes from an interaction with a robot, and the context in which the interaction takes place can affect human-robot trust similarly to human-human trust. However, different types of trust might be expected. For example, unless a robot is part of an organization or answers to others, robots may be unlikely to elicit **deterrence-based trust** due to the lack of potential negative repercussions (i.e., who would be to blame if the robot behaved poorly or broke the trust?). Yet, oftentimes the expectation of a positive interaction or fun experience is what leads people to interact with robots in the first place, and positive or repeated interactions can lead to a higher likelihood to trust a robot (Yamagishi, 2001). Likewise, robots might be trusted (or not) depending on the context of use or interaction; a robot vacuum might be trusted to safely clean the house, while a shopping assistant robot may not. Robots tend to be designed and built with a particular purpose in mind, and we believe that they may therefore elicit trust in related situations, but not others. However, people do tend to over-trust robots (e.g., Booth et al., 2017; Robinette, Li, Allen, Howard, & Wagner, 2016), so it may be possible for trust in one area to lead to trust in another.

When it comes to children, it is unclear whether or how external factors may play a role. Research with children and adults or puppets suggests that the context of the situation and expected outcomes do impact children's trust (e.g., Kushnir, Vredenburgh, & Schneider, 2013). However, such research does not yet exist with robots. This suggests a current gap, and an important area of future work.

In this section we proposed that factors external to the interactants (i.e., context of situation and expected outcomes) can influence trust. In the next section, we provide a more detailed overview of how the dimensions and factors discussed in Section 7.3 may apply to children's trust towards others.

7.3.4 How Children Trust Other Humans

While clearly children are human, and therefore fit into our human-human trust exploration, we have found that most trust work targets adults, with the exception of trust towards informants. In this section, we consider our findings so far (as depicted in Figure 33) from the perspective of children, applying our knowledge of children and child psychology (presented in Chapter 3) for this analysis.

The social proximity that children have to others can affect how they trust them, and research suggests that young children tend to trust familiar informants more than unfamiliar ones (Corriveau & Harris, 2009a). Thus, we can expect our social proximity dimension to also affect how children perceive and trust others. In the context of our framework, this would suggest that children can demonstrate **interpersonal** and **dyadic trust** towards individuals in in their intimate circle, or even **relational trust** towards those they have interacted with before. Unlike adults however, young children may not need to consider certain types of trust, which older children or adults might usually do for them. **Societal** types of trust, for example, may not be necessary, as a young child would likely place their trust on close adults or older children, who would then engage in the processes of weighing whether to trust others (or the situation).

Another aspect that may also impact children's trust, but not adults', is that children are still undergoing developmental changes and psychological growth. We may expect this to affect the person factors dimension within our framework, and for children at varying stages of development to trust differently. For example, **deterrence-based trust** may only be applicable to older children who are familiar with the consequences of others lying or behaving poorly (and are therefore able to invoke this type of trust).

Trust in children may also be affected by factors related to who or what their trust is placed on (referring to external factors in our framework). For example, children tend to trust accurate informants over inaccurate ones (Birch et al., 2008; Kushnir et al., 2013), suggesting that such factors about other individuals can influence their trust. In these cases, we might expect children to exhibit **cognitive trust** or **calculus-based trust**.

How children trust other people may very well inform how they may trust robots. In the next section, we move one step closer to child-robot trust by examining people's (mainly adults') trust towards machines and systems.

7.4 Dimensions and Factors Relevant to Trust Towards Computer Systems

In section 7.3, we explored factors affecting trust between adult humans (summarized in Figure 33). In this section, we continue the trajectory by further including literature on what factors might impact trust towards computer systems. We later use the information provided in this section, to better understand trust towards robots, and finally how children may trust robots.

Trust can be important in contexts other than just between humans, as is the case when interacting with technology. Further, individuals might trust technologies differently from how they do people. Research on trust towards machines and systems is typically conducted in the field of HCI, as the use and adoption of technologies is deeply affected by how much people trust those technologies (Bahmanziari, Pearson, & Crosby, 2003). The acceptance of new technology also relies on the perceived usefulness and perceived ease of use of the technology (Venkatesh & Davis, 2000), which in turn influences trust in the system (e.g., Corritore, Kracher, & Wiedenbeck, 2003). Trust in systems is therefore of key importance, as research suggests that attitudes towards the machine or system are the best predictor of acceptance behaviour (Fishbein & Ajzen, 1975). This implies that trust could impact how much people accept a particular piece of technology. Thus, without trust, a system may never be used, or usage may decrease, making the system obsolete. In this section, we review work from HCI, and distill it into dimensions and factors based on such research as well as that presented in the previous section. Our exploration further resulted in human-system trust being divided into the following dimensions: *system sociality*, *working roles*, *extrinsic factors*, and *person factors* (Figure 34).

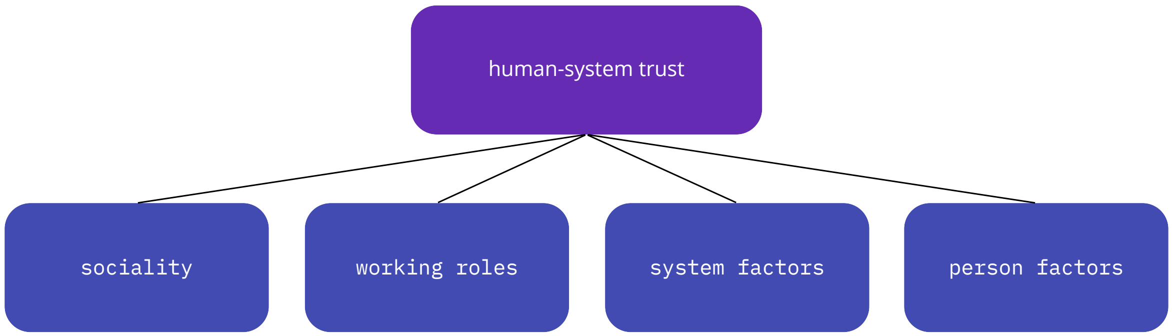


Figure 34. Dimensions resulting from our survey on trust towards computer systems: factors that affect trust towards computer systems.

7.4.1 System Sociality and Human-System Trust

Results from our survey suggest that the level of sociality of a system could affect how people view, interact with, and trust the system (e.g., Benbasat & Wang, 2005; J. D. Lee & See, 2004; Y. D. Wang & Emurian, 2005). Systems can be designed (or simply perceived to be) more social or machine-like, and where a system lies on this spectrum can have an impact on trust towards it. We found work in this area to fall into the following categories: *machine*, *social*, and *mixed* (Figure 35).

System sociality: Trust towards machine-like systems – One view of systems that emerged from our research is of them being technological objects, which are utilized to accomplish a task. People treat and trust machines differently than they do other people, and this is especially important for adoption and use, knowing that if an individual does not trust a machine or system, they are not likely to continue using it (J. D. Lee & See, 2004). Further, how well a machine functions, and how effective and efficient it is can also play a key role in how much it is trusted (J. D. Lee & Moray, 1994). This suggests that systems/machines that are viewed as only that, are likely to be highly dependent on their usability, impacting trust towards these systems.

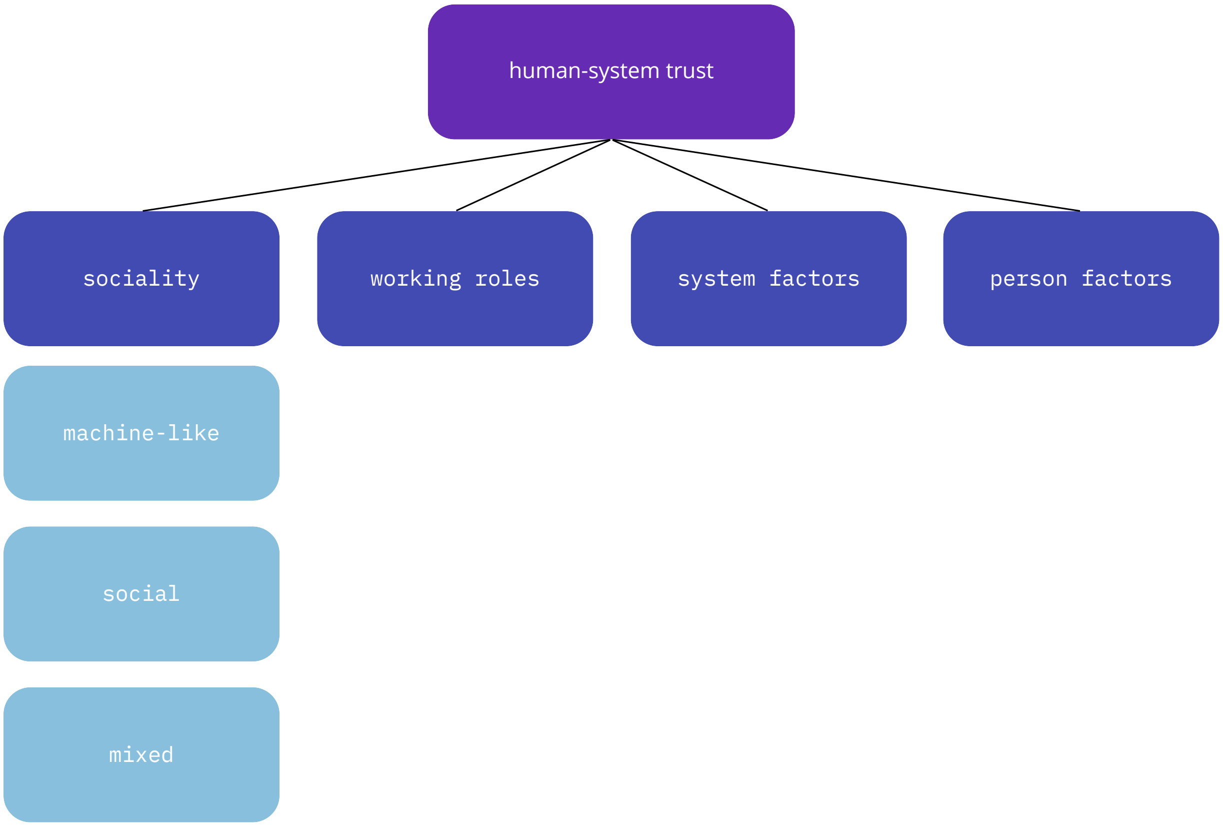


Figure 35. Dimensions and respective factors for system sociality, which we developed through our extensive survey on trust towards systems.

We suspect that trust in machines that are just perceived as such (i.e., machines) could initially incorporate **institution-based trust** (e.g., if it is from a well-known manufacturer), but over time will likely change to **system trust** (i.e., whether everything seems to be working well), **cognitive trust** (i.e., a thought-out decision of whether the system should be trusted), or **calculus-based trust** (i.e., cognitive decision backed up by credible information). However, it may also be possible for there to be no trust towards the machine in question, if it does not work properly or exhibits errors.

System sociality: Trust towards social systems – People's brains are used to evaluating situations socially. This leads to systems sometimes being perceived and treated as social beings, with certain anthropomorphic qualities that other objects do not possess (Nass et al., 1993). Interactive virtual agents tend to take advantage of this social phenomenon, eliciting higher levels of trust than systems without interactive agents (Weitz, Schiller, Schlagowski, Huber, & André, 2019). Likewise, people treat online recommender agents as social actors, and trust them accordingly (Benbasat & Wang, 2005). Therefore, we might expect individuals to trust systems with these social attributes similarly to how they do people. However, it is not necessary for systems to contain interactive agents or anthropomorphized components to elicit social interaction and perceptions, and therefore trust.

We propose that systems that are perceived to be social 'others' may be likely to be attributed other types of trust than simply **cognitive** (unlike machine-like systems). For example, people may build social bonds with social machines, eliciting **interpersonal** or **dyadic trust**, or **relational trust** with repeated exposure. Likewise, repeated interaction with a specific machine or system could potentially also lead to the gain or loss of **relational trust** depending on the outcome of the interaction.

System sociality: Trust towards mixed-sociality systems – Instead of being perceived as only machine-like or social, systems can also be perceived as a mixture of the two, having a combination of social and machine-like features, which we believe may lead to a combination of trust outcomes. A recommender system that includes an agent (interactive or not) might be perceived as a non-social system with a social component (Y. D. Wang & Emurian, 2005), and we might therefore expect it to be trusted as either or both (depending on whether the individual considers the system and agent to be related or not). Drawing from the Computers are Social Actors literature (Nass et al., 1994, 1993), research suggests that people can perceive machines to be social in some aspects, but also machine-like in others. This suggests that systems that have mixed standing, in the sense that they are perceived as both a machine and social, may be affected by components of trust found towards both social and machine systems as described in the previous sections.

7.4.1.1 How system sociality might apply to trusting robots and child-robot trust

We propose that the sociality of a robot can impact trust towards it similarly to how it does with other machines and systems (e.g., Geiskkovitch et al., 2016). We can expect a robot that has a machine-like appearance or interaction style to elicit **cognitive** or **societal** types of trust, while a social or mixed robot might be more likely to foster **interpersonal**, **dyadic**, or **relational trust**. This suggests that changing the appearance or interaction style of a robot could greatly affect the types (and potentially) amounts of trust that people may experience towards it (an approach that we explore in section 7.5). When considering trust towards systems in this way, we can also consider whether trust towards robots might be more similar to machines or people. Our extensive survey of the literature suggests that this is in part dependent on specific factors of the robot (i.e., whether it looks and behaves more like a human or like a machine), which is why we explore both human and system trust in this chapter, to inform trust towards robots.

For children specifically, research does not yet inform how sociality might affect trust. We might expect humanoid robots to possibly garner more trust than other morphologies; Research suggests that children prefer robots that have a head or eyes and are able to move (Sciutti et al., 2014), which means that they might be disappointed with other morphologies. However, further research in the field is necessary to better understand this area. We propose further ways in which robot morphology and design may affect children's trust in section 7.6.3.

Our survey of the literature suggests that the level of sociality of a system, whether social, machine-like, or a combination of both can affect how the system is perceived and trusted. Machine-like systems are more likely to be trusted (or not) based on their functionality and efficacy, while social and mixed-sociality systems due to their social cues. In the next section, we explore how the roles of an individual and system when completing a task can affect trust towards the system.

7.4.2 Working Roles Affecting Human-System Trust

Our literature survey pointed to the importance of affiliation in relation to the working task, between an individual and a system, for impacting the types and amounts of trust that might be present (e.g., Karvonen & Parkkinen, 2001; Yu, Berkovsky, Taib, Zhou, & Chen, 2019). For example, a collaborative task may elicit a high level of trust in a system due to shared goals, while a task in which the technology only mediates the interaction with somebody else may not. Thus, we categorize working roles by the context of the working relationship between the individual and system: *collaborative*, *active vs. passive*, and *mediated* (Figure 36).

Working roles: Collaborative roles affecting trust towards systems – When a human and a system work in a collaborative manner, we can expect trust to take a variety of forms. In such cases, the individual and system have equal amounts of authority and motivation to complete the task, and people are able to adjust their trust levels and performance based on that of the system to work collaboratively (Yu et al., 2019). This suggests that trust could be built and broken by either of the two parties, as they are both equally responsible for task outcomes.

Due to this, we propose that a collaborative task may elicit some types of trust that other tasks may not. For example, an individual may perceive a collaborative machine partner as their "buddy," (see the previous section 7.4.1 on sociality) developing **dyadic trust** towards it. On the other hand, a collaborative machine partner that does not complete their work well or in a timely fashion, or one that makes errors, may lead to lowered **cognitive** or **calculus-based trust** as the person learns that the machine's actions are not benefiting the task, and lowered **relational trust** if problems continue.

Working roles: Active roles vs. passive roles affecting trust towards systems – During certain tasks or situations, a person and a system might have different activity levels in regards to the task, potentially leading to varying trust outcomes (e.g., as in Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Karvonen & Parkkinen, 2001). The system, for example, can be passive, where the

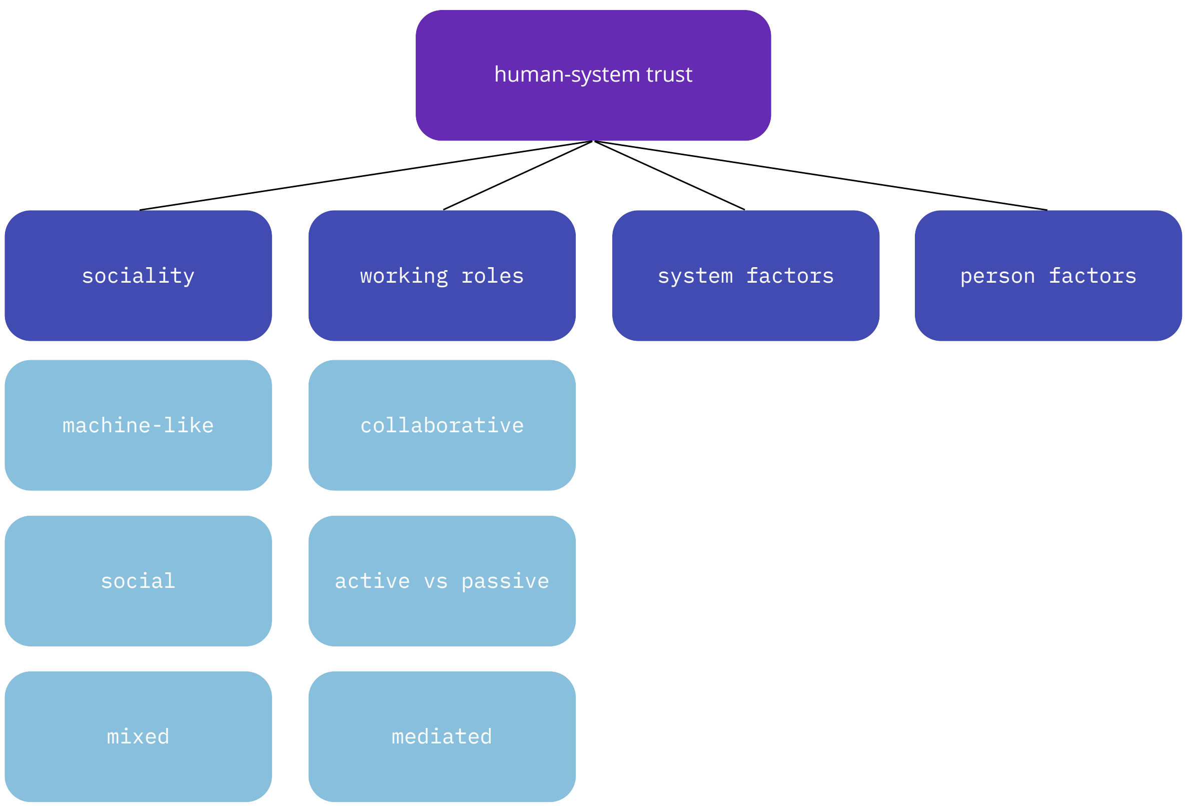


Figure 36. Dimensions and respective factors for working roles and system sociality, which we developed through our extensive survey on trust towards systems.

individual is intended to be the authority while using the system, and the technology has little say on what the user does; this is the case in most online applications and websites. When a system is passive, research suggests that the main factors affecting trust are the visual and experience design of the system (Karvonen & Parkkinen, 2001). Poor visual design elements, such as an unbalanced layout or lack of structure can decrease trust in a system (J. Kim & Moon, 1998), as can the presence of errors (Yan, Kantola, & Zhang, 2011). Thus, the design features of a passive system are likely to affect trust towards it. In this context, we might expect a passive system to mainly elicit **system trust** or **mundane trust**, since the individual is in charge and might not consider trust towards the system as long as it is working as expected.

On the other hand, the working relationship might be reversed, so that the machine is in charge. For example, recommender systems tend to be more active, in that they provide suggestions to the user without the user necessarily asking for them or having a say in what those recommendations are. In this case, we might expect the visual design might to be less important in order to elicit trust. However, research suggests that the presence of errors might lower trust in active systems (such as those providing recommendations), and these effects may only be remedied by presenting the user with reasons for why the errors occurred (Dzindolet et al., 2003). Thus, for active machines, efficiency and effectiveness appear to be more relevant for trust, as opposed to overall visual design. A more active machine, we believe, is likely to elicit **relational trust** if the interaction and outcomes are positive.

Working roles: Mediated roles affecting trust towards systems – We have found that sometimes a machine may not be used to directly complete a task, but instead as a mediation tool to interact with somebody else. Unfortunately, most existing research on mediated roles with machines focuses on trust between the two human interactants, and not on trust towards the system. The limited research in this area however, suggests that trust towards a system can be influenced by how well the system allows an individual to communicate and work with others (Riegelsberger et al., 2005). When trust is required in mediation systems, the system could be enhanced with real-time voice interaction to increase trust (Greenspan, Goldberg, Weimer, & Basso, 2000). We might expect systems that provide mediated interaction to elicit almost all types of trust, depending on whether the trust is coming from the system or the other individual. However, disambiguating between trust towards the system and the other individual might be difficult and may the reason for the lack of research in this area.

7.4.2.1 How trust in working roles with systems might apply to robots and children

Our synthesis of the literature proposes that the roles that a robot and person play during interaction can affect how the robot is perceived and trusted. For example, seeing a robot as part of the team can lead to higher levels of trust towards it (L. Wang et al., 2010). We believe that trust for in-group robot members is likely to be **dyadic** or **interpersonal** in nature, while **cognitive** or **calculus-based trust** may be affected by certain types of robot or robot errors (Chapter 5). Aspects such as how much agency a robot is attributed (Young et al., 2011), or whether it is in a position of authority (Geiskkovitch et al., 2016) can also influence trust; In the first case people may treat the robot as a social being, exhibiting similar types of trust (e.g., **dyadic trust**, **relational trust**, **interpersonal trust**), while in the second case individuals may simply trust and obey the robot due to its position (e.g., **system trust**, **mundane trust**). We are likely to observe similar trust themes with children, based on their affiliation with the robot. We explore this further in section 7.6.2.

Our research suggests that a human-system team can complete tasks in a variety of forms, including collaborative, active or passive, and mediated. These factors can influence the types of trust that people may experience towards the system. In the next section, we look at how factors specific to the system can affect trust.

7.4.3 System Factors Affecting Human-System Trust

Through our survey of the literature, we found that there are a number of factors outside the user's control that can influence trust. The visual and experience design of the system can directly affect trust, for example, by demonstrating how easy it will be to understand and use. We organize system factors into *visual and usability design*, *certifications and endorsements*, and *performance and usability criteria* (Figure 37).

System factors: System visual and usability design affecting trust – We can expect the visual and usability design of a system to impact how much it is trusted. Factors such as the "real-world feel" of a system (Fogg et al., 2001), general ease of use (Corritore et al., 2003), how much information is provided to the user (K. C. Lee & Chung, 2009), the use of 'white space' (Karvonen & Parkkinen, 2001), and the presence of structured groupings (Karvonen & Parkkinen, 2001) can increase the amount of trust that users have in it. Systems that are visually well designed are therefore more likely to elicit trust, while systems that lack these elements might not be able to. We argue that **system**, **institution-based**, **cognitive**, and **calculus-based trust** are of special

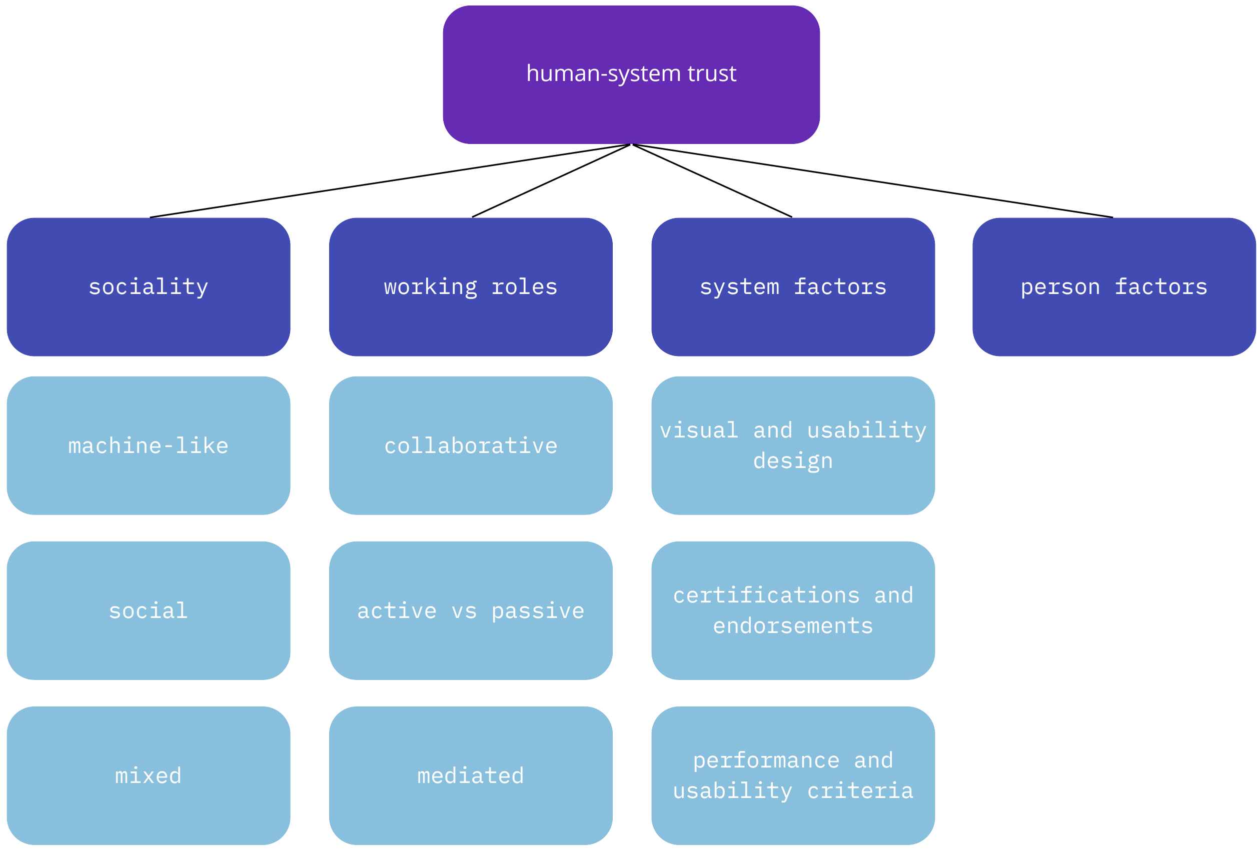


Figure 37. Dimensions and respective factors for system factors, working roles, and system sociality, which we developed through our extensive survey on trust towards systems.

relevance here, since they rely on the information provided to the user (in this case through the design) to be presented in an appropriate fashion to elicit trust from a user, and for the user to be able to use the system without roadblocks.

For recommender systems in particular, the visual design can affect trust in a variety of ways. For example, transparency in how a system works, as witnessed by the visual design, is one of the largest factors affecting trust (Bunt, Lount, & Lauzon, 2012; Glass, McGuinness, & Wolverton, 2008), as knowing how the system works can instill confidence in it. Similarly, providing explanations of how the recommendation might suit the user can also increase trust towards a system (Felfernig & Gula, 2006). Further, a recommender system's organization of visual information impacts trust. For example, movie recommendations based on genre are trusted more than those based on human preference or star rating, and those based on quality are trusted more than if based on diversity or familiarity (Berkovsky, Taib, & Conway, 2017). The visual design of a system, therefore, can impact how much people may trust it.

System factors: System certifications and endorsements affecting trust – Research suggests that the presence of certifications and endorsements from others influences trust. Certifications from trusted bodies increase trust in the system, while reviews from other people can increase or decrease trust depending on whether they are positive or negative (Y. D. Wang & Emurian, 2005). For example, in online systems such as websites and e-commerce sites, the credibility (i.e., perceived honesty, expertise, predictability, and reputation) of a system can impact people's trust in it (Corritore et al., 2003). Thus, providing certification or endorsements can be a way to make a system seem more trustworthy. In our opinion, certifications and endorsements might be related to **institution-based trust** (if they are from a trusted body) or **interpersonal trust**, if the user trusts other users to be telling the truth or have similar beliefs to them.

In online review systems, for example, trust can be influenced by a number of factors related to certifications and endorsements. Individuals tend to trust reviews based on the number of people who have found that review helpful, the number of reviews that the reviewer has contributed, the number of helpful votes attributed to the reviewer, and the city and country of origin of the reviewer (Sherwani & Stumpf, 2014). People also tend to trust reviews more if others have found the review helpful, and if the reviewer has contributed multiple reviews and has received a large number of helpful votes. Additionally, if the city and country of origin of the reviewer match that of the individual reading the review, they will likely assign a higher level of trust to that particular reviewer (Sherwani & Stumpf, 2014). Thus, the presence of certifications and endorsements can be included in systems to affect trust.

System factors: System performance and usability criteria affecting trust – In contrast to the visual and usability design, which focuses on design elements, here we focus on the actual outcomes of usability and how the system performs. Through our extensive survey, we found that whether a system does what it is intended to, in the way that it is intended to, can impact people's trust in it. If a system promises to assist with a particular task, but is unable to do so, the user becomes disillusioned with the machine, and is less likely to trust it to do other tasks or even use it in the future (K. C. Lee & Chung, 2009). Performance outcomes (either positive or negative) in general, can also influence trust (J. D. Lee & Moray, 1994). If the interaction is positive (i.e., good performance), an individual is more likely to experience higher trust in the system. Manipulating the performance of a machine might therefore provide the ability to intentionally increase or decrease trust towards the system. We believe that **relational trust** plays a key role here, as repeated use of a system could lead to increased **relational trust** if the system's performance and usability are satisfactory, or decreased trust otherwise.

In automation, for example, the current and historical performance of the automation, including reliability, predictability, and overall ability can impact user trust (J. Lee & Moray, 1992). The process, which is how appropriate the algorithms are for the situation and goals, as well as purpose, (i.e., the degree to which the automation is being used for what it was intended for) can also have an effect on trust towards it. Researchers have also found that automation and trust can in turn influence each other; if the system is not trusted it is unlikely to be used, and if the system is not used trust can never increase (J. Lee & Moray, 1992).

System performance: System errors and trust. In Chapter 5 and Chapter 6 we discussed how robot errors can affect trust. However, this is not exclusive to robots. Errors are commonplace with any type of technology and any functional or presentational errors that a system experiences can decrease trust towards it (de Vries, Midden, & Bouwhuis, 2003; J. D. Lee & See, 2004; Yan et al., 2011). For example, the effect of automation errors on trust depends on the frequency and seriousness of the error; a small error with unpredictable results is much more likely to decrease trust than a large error with consistent results (J. D. Lee & See, 2004). Further, the lower the number of errors, the more likely an individual is to trust the system (J. D. Lee & Moray, 1994), as they are able to rely on it to produce the intended results more often. Thus, errors should be prevented if trust towards the system is desired. We propose that errors are most likely to affect **cognitive trust** and **relational trust**, the first decreasing if it is clear that the system cannot do its task properly and the latter being affected by repeated exposures.

7.4.3.1 How system factors affecting trust might apply to robots and children

Through our survey, we believe that factors of a machine that affect how much it might be trusted are fairly similar of those of robots. The visual design of a robot plays a large role in how it is perceived and interacted with (Martelaro et al., 2016). However, in the robot case, the focus is on the morphology and appearance of the robot, as a humanoid robot is likely to elicit different types of trust than a mechanoid or non-humanoid robot (Vattheuer et al., 2020). The appearance of a robot can also affect whether it is anthropomorphized, and therefore we believe, trusted with intimate or emotional types of trust (e.g., **interpersonal**, **dyadic**, **relational**), much more than other machines and systems. Likewise, performance factors such as ease of use, usability, and errors also influence trust towards robots as they do in machines (e.g., Brooks, Begum, & Yanco, 2016; Ragni, Rudenko, Kuhnert, & Arras, 2016; Washburn, Adeleye, An, & Riek, 2020). However, while these factors may affect personal types of trust, we argue that they can also influence cognitive types, such as **calculus-based** or **cognitive trust**, if the robot does not work properly or makes errors.

Similarly to adults, we can expect children's trust to be tightly coupled to a robot's design, performance, and personality. Through our own research (Chapter 5 and Chapter 6) we found that robot performance can influence young children's trust. **Interpersonal**, **relational**, and **cognitive trust** might all be at play here, and we discuss these topics further in section 7.6.3.

In this section, we discussed how a system that has a better design, good performance, and is endorsed by others is more likely to be trusted. Factors related to the individual may also influence trust, and we discuss these in the next section.

7.4.4 Person Factors Affecting Human-System Trust

In section 7.3, we discussed the importance of a person's attributes to trust others (person factors, see section 7.3.2). Similarly, a person's disposition towards and experience with technology affects machine trust. Person-related factors are often stable over time, but accounting for them might help us to better understand trust in the system, and potentially make up for these factors. For example, an individual that describes themselves as a technophobe might be less likely to trust technology, regardless of attributes of the system. We organize person factors into *propensity to trust technology*, *propensity to anthropomorphize technology*, and *self-confidence with technology* (Figure 38).

Person factors and systems: Propensity to trust technology affecting human-system trust – We can expect trust towards a particular machine or system to be in part dictated by the extent to which an individual generally trusts technology. Researchers, for example, have found that an individual's propensity to trust is in part related to their trust in automation (J. D. Lee & See, 2004), with higher propensity meaning higher likelihood of trust. Further, in online reviews, people with a low propensity to trust others are more likely to focus on the reviewer's background and use that information to justify lower trust (Sherwani & Stumpf, 2014). While propensity to trust technology may not be easily altered, it should still be taken into account as it can impact trust. In this case, we believe that **interpersonal trust** as a general trust tendency can be highly related to the

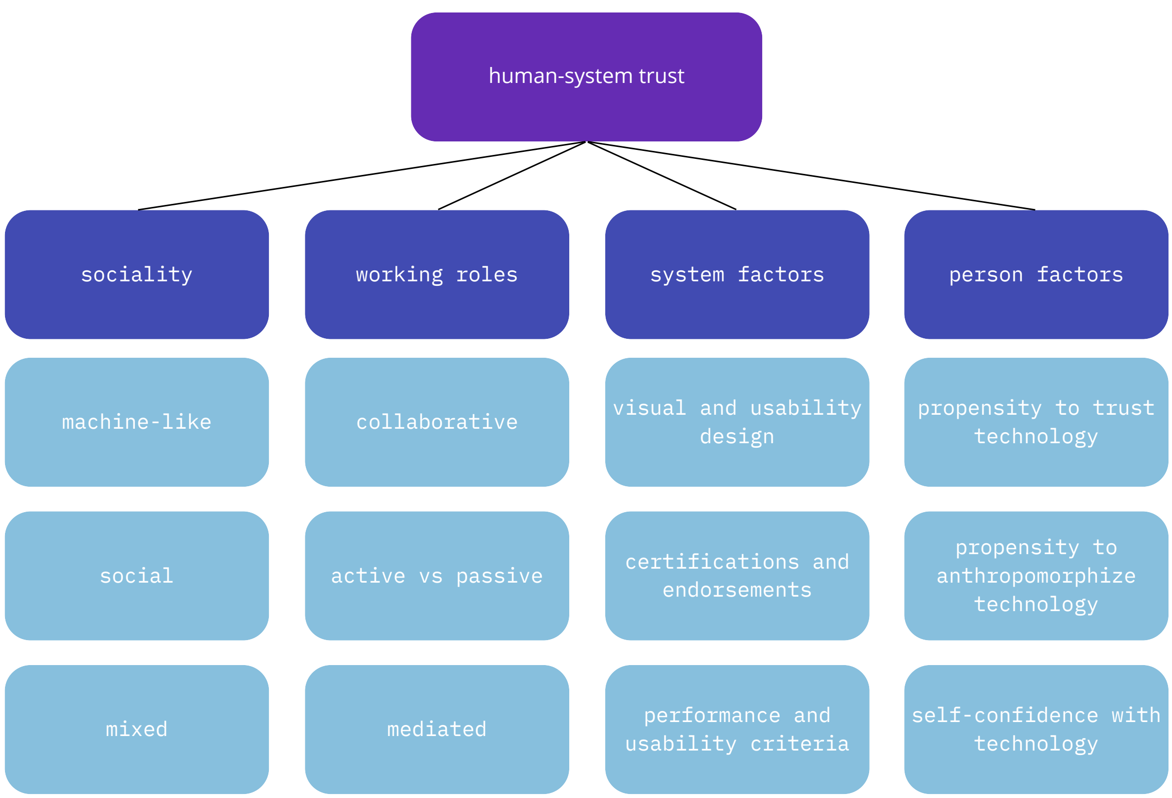


Figure 38. Dimensions we developed for trust in human-system interaction.

propensity to trust. In addition, the emotional and cognitive components of **societal trust** might be similarly associated with trust propensity.

Person factors and systems: Propensity to anthropomorphize technology affecting human-system trust – Through our survey, it is evident that the way in which an individual generally perceives technology, either as an object or as a social other, can have an effect on how much they trust it. Humans are social beings, and sometimes apply social norms of interaction to machines, even if not substantiated (Nass et al., 1994). This suggests that many aspects of trust between humans may also apply to machines. For example, we propose that it is possible that similarly to how we can build **relational trust** with individuals that we frequently encounter, we might also build **relational trust** with machines that we use often, especially if these interactions are positive. Further, we suggest that people might also be able to attain a deeper connection with machines than otherwise expected, such that aspects of **dyadic trust** might also be possible towards machines.

Person factors and systems: Self-confidence with technology affecting human-system trust – Research suggests that the extent to which an individual believes in their own abilities can affect how much they use and trust a system. Individuals who have low self-confidence in a task are more likely to use, and therefore trust, automated systems while the opposite is true for people with high self-confidence (J. D. Lee & See, 2004). This suggests that individuals who feel confident about their abilities, may not use (and therefore trust) a system even if they should. We hypothesize that self-confidence is related to emotional aspects of trust, such that if an individual does not feel confident about their abilities they may have **emotional trust** towards a system simply due to their lack of trust in themselves. Likewise, the decision to trust a system can be thought-out, leading to **cognitive** or **calculus-based trust** in the system due to an analysis of one's abilities in comparison to the machine's.

7.4.4.1 How person factors affecting human-system trust might apply to robots and children

We suspect that person factors that affect how people trust machines can also apply to how they trust robots. Robots are much more likely to be anthropomorphized than other machines (Bartneck, van der Hoek, et al., 2007; Fussell et al., 2008; Sung et al., 2007), due to their physicality and agency, which suggests that people who are likely to anthropomorphize technology may trust them similarly to anthropomorphized machines (and people). Therefore, an individual who is confident with robots, or is likely to trust robots (as technology in general) is more likely to interact with them. This, we propose, could lead to the development of **interpersonal**, **dyadic**, or **relational trust**, as well as **cognitive trust** if the robot appears to work well or is more capable than the individual.

With children, however, instead of anthropomorphism or self-confidence, factors such as the age and gender of the child (and of the robot for the latter) as well as their previous exposures to robots might influence trust. We explore this further in section 7.6.4.

In this section, we discussed how aspects specific to each individual, such as their self-confidence or how likely they are to trust technology or anthropomorphize it, can have an effect on trust towards systems (see Figure 38). People who have lower self-confidence or are more likely to trust or anthropomorphize technology are more likely to have higher trust towards systems. In the next section, we discuss what we know about how children trust systems. Following, we switch our focus to trust in HRI in section 7.5.

7.4.5 How Children Trust Computer Systems

Research on how children may trust computer systems is very limited. Thus far, it suggests that children, especially younger ones (i.e., 5-6 years old), tend to trust unknown sources from the internet more than teachers or peers (F. Wang, Tong, & Danovitch, 2019). This is especially the case when information might be less accessible through the non-virtual sources. Additionally, younger children (i.e., 3-4 years old) trust previously accurate computer agents (which could be perceived as computer systems or robotic entities), more than ones that have been inaccurate (Danovitch & Alzahabi, 2013). However, save for these two publications, knowledge on how children may trust computer systems is still limited.

7.5 Dimensions and Factors Relevant to Trust in Human-Robot Interaction

Continuing to narrow down our focus of trust from people, to computer systems, we now turn our focus to adult's trust towards robots, before finally arriving at child-robot trust in section 7.6. As highlighted in the previous sections, we can expect a broad range of factors to impact trust towards humans, machines, and even robots. Trust can affect a myriad of factors in HRI, including whether people collaborate with a robot (Malle et al., 2020), accept its advice (Malle et al., 2020), or even use the robot to begin with (Freedy et al., 2007). Trust can also influence the ability for a human-robot team to accomplish goals, create effective relationships, and have effective interactions with the robot (Hancock, Billings, Schaefer, et al., 2011). Inadequate amounts of trust towards a robot, either too much or too little, can result in serious consequences (Hancock, Billings, Schaefer, et al., 2011), such as people allowing it into locked facilities (Booth et al., 2017), or following its bad advice in dangerous situations (Robinette et al., 2016). We noted that there is a very large overlap between trust between humans, towards systems, and towards robots.

While in previous sections we presented research from Sociology, Psychology, and trust toward computerized systems in general, here we share data and results from within the field of Human-Robot Interaction.

In this section we present our synthesis of research from the field of HRI, as it relates to trust, and organize it into the following dimensions: *social proximity*, *working roles*, *robot factors*, and *person factors* (Figure 39). In addition, we link this research to that from other fields and areas, such as those discussed above, with the goal of gaining insight into the factors affecting trust in HRI. This section only pertains to how adult trust robots, as the research within it has been conducted with adults and not children.

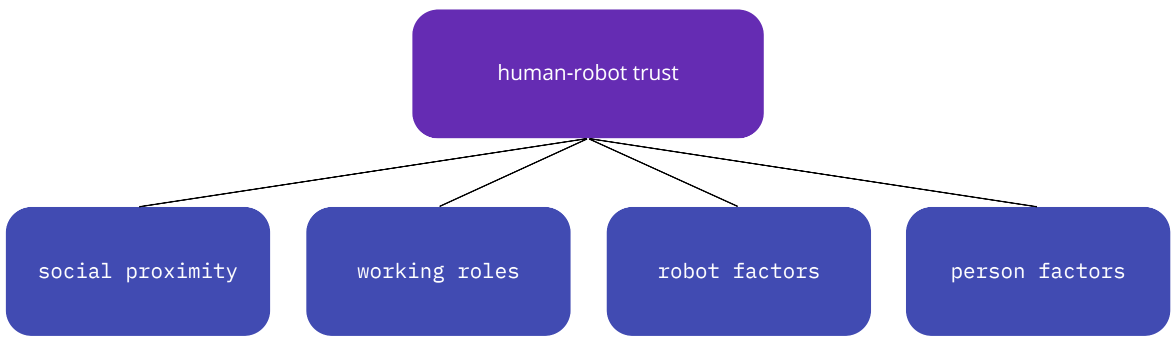


Figure 39. Dimensions resulting from our survey on trust towards robots: factors that affect adult's trust towards robots.

7.5.1 Existing Models of Human-Robot Trust

A number of models of trust already exist in the field of HRI. The majority of these models divide trust into three key components: human, robot, and environment (e.g., Hancock et al., 2011; Khavas, Ahmadzadeh, & Robinette, 2020; T. Sanders, Oleson, Billings, Chen, & Hancock, 2011). These models mainly differ in the sub-factors that affect the main components. For example, one such model separates human-related factors into ability-based factors and characteristics of the human, robot-related factors into performance-based factors and attribute-based factors, and environmental factors into team collaboration and task (Hancock, Billings, Schaefer, et al., 2011). Other models build upon the first, by further expanding and specifying the original sub-factors, including aspects such as culture, experience, cognitive workload, and situational awareness under human-related factors, separating robot attribute-based factors into behavioural and appearance factors, and adding aspects such as group membership and mental models to environmental factors (Khavas et al., 2020; T. Sanders et al., 2011). Another model further separates environmental factors into two separate ones, context of use and social factors, having four major factors in total (Khalid, Helander, & Lin, 2021). These models of trust provide great insight into how people trust robots. Our framework adds to this existing body of literature by incorporating knowledge from additional various fields (e.g., Psychology, Human-computer Interaction) and breaking down the types of trust that we might observe due to the human, the robot, and the environment.

7.5.2 Social Proximity to Robots and Human-Robot Trust

Similarly to how the emotional and psychological distance between individuals can affect trust, we can expect the social proximity to a robot to also influence trust towards it. For example, a companion robot might elicit more trust than an informational robot at an airport. However, there is a lack of research exploring how trust varies depending on social proximity to a robot. Nonetheless, based on our synthesis of the literature, we divide the social proximity dimension into four categories, denoting the social relationship between the robot and individual: *intimate*, *personal*, *non-personal*, and *society* (Figure 40).

Social proximity to robots: Trust in intimate relationships with robots – We have found that people can develop intimate relationships with robots, leading to specific trust paradigms. For example, researchers have found that people can get very attached to their robot vacuum, giving it

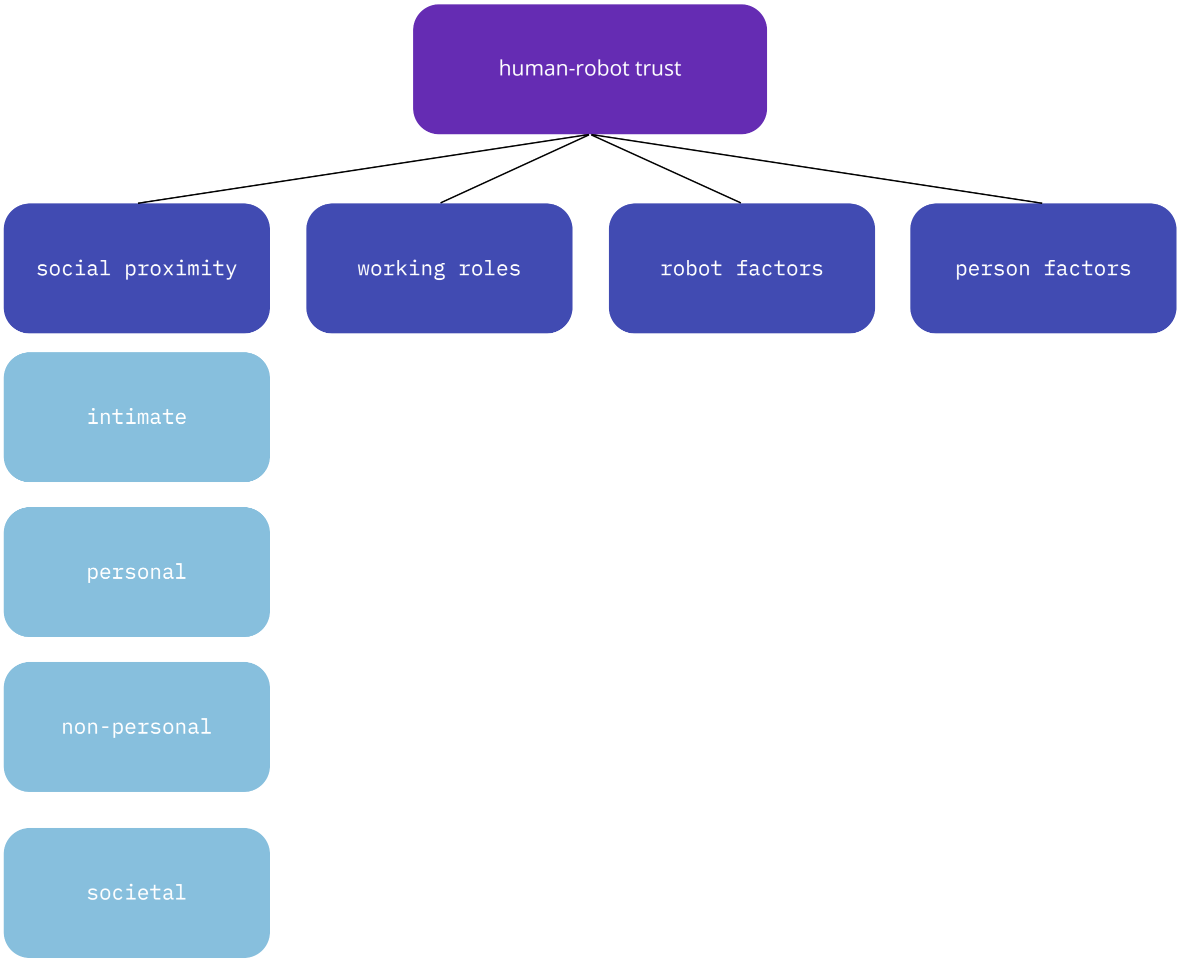


Figure 40. Dimensions and respective factors for social proximity, which we developed through our extensive survey on trust towards robots.

a name, buying special clothes for it, and choosing to repair it as opposed to replacing it for a new one (Sung et al., 2007). In addition, research suggests that certain populations, such as older adults, are very likely to trust companion social robots (Poulsen, Burmeister, & Kreps, 2018), which can provide a sense of connectedness. Thus, it appears that people can trust intimate robots similarly to how they trust intimate individuals, regardless of the robot's role (e.g., robot vacuum, tutoring robot). Following this line of research, we suggest that people are likely to experience **dyadic** and

interpersonal trust towards robots in their intimate circle, similarly to how they do towards people.

Social proximity to robots: Trust in personal relationships with robots – Through our research, we found that people may sometimes interact with robots that do not feel part of our intimate circle, but yet we have personal relationships with. A home assistant robot, for example, may not elicit the same feelings of closeness as other robots, but individuals nonetheless have trusting relationships with them (Foehr & Germelmann, 2020). While relationships with robots in this category are not as close, we can still expect individuals to experience **interpersonal trust** towards them. **Relational trust** might also be relevant here, as it can increase or decrease after continued use.

Social proximity to robots: Trust in non-personal relationships with robots – Research suggests that people may also encounter robots in non-personal but frequent settings, potentially meeting the same robot multiple times without building a bond with it. Applications such as the Cero robot, which fetches and carries objects for mobility-impaired office workers may fall under this category (Severinson-Eklundh, Green, & Hüttenrauch, 2003). In these cases, trust towards specific robots can potentially be dependent on factors related to the individual and the robot, which we discuss several sections below. We suspect that **interpersonal trust** might play a role with non-personal robots, as individuals may have some information about a particular robot without having built a bond with it. **System** and **institution-based trust** may also be relevant, as additional information such as what company or institution it was created by or is part of (**institution-based trust**) and whether the robot appears to be behaving normally (**system trust**) can influence trust.

Social proximity to robots: Trust in societal relationships with robots – Our survey suggests that several types of robots are not encountered regularly by the same people, but are instead in hospitals, airports, shopping malls, and restaurants, where visitors may interact with them without building rapport. Researchers have found that individuals will trust and allow unfamiliar robots into secured facilities (Booth et al., 2017), and will trust a clearly-faulty robot that they only met once (Salem et al., 2015). While some of these robots may require people to trust them, this is not always desired, and people do not have a chance to calibrate their trust. We believe that people may exhibit **societal** aspects of trust (e.g., **emotional**, **cognitive**, **mundane**, **ideological**) towards these robots, without having much additional basis for trust.

7.5.2.1 Social proximity to robots and human-robot trust summary

Our survey suggests that the social proximity between an individual and a robot can affect how much they may trust the robot. Trust is possible towards intimate home robots as well as unfamiliar societal robots, but the types and extent of trust witnessed might differ. We could expect similar outcomes with children, and discuss these in section 7.6.1. In the following section, we discuss how working with robots in different roles may affect trust.

7.5.3 Working Roles with Robots Affecting Human-Robot Trust

An emergent result from our survey of the literature is that the type of interaction between and individual and a robot can impact trust (similarly to that between a person and a system, as discussed in section 7.4.2) Humans and robots may engage in different types of tasks together, which may dictate the level and type of trust that can be developed. We use working roles to specifically denote the amount of ownership and input that an individual and a robot have while completing a task. We organize working roles with robots into the following dimensions: *collaborative*, *authoritative vs. submissive*, and *mediated* (Figure 41).

Working roles with robots: Collaborative roles affecting trust towards robots – We have found that in some cases a robot may act as a collaborative partner, working alongside a human with a joint goal. In this case, trust is especially important as researchers have found that trust may influence the willingness to cooperate with a robot (Freedy et al., 2007), and whether it is accepted as a collaborative partner (J. J. Lee, Knox, Wormwood, Breazeal, & DeSteno, 2013). In

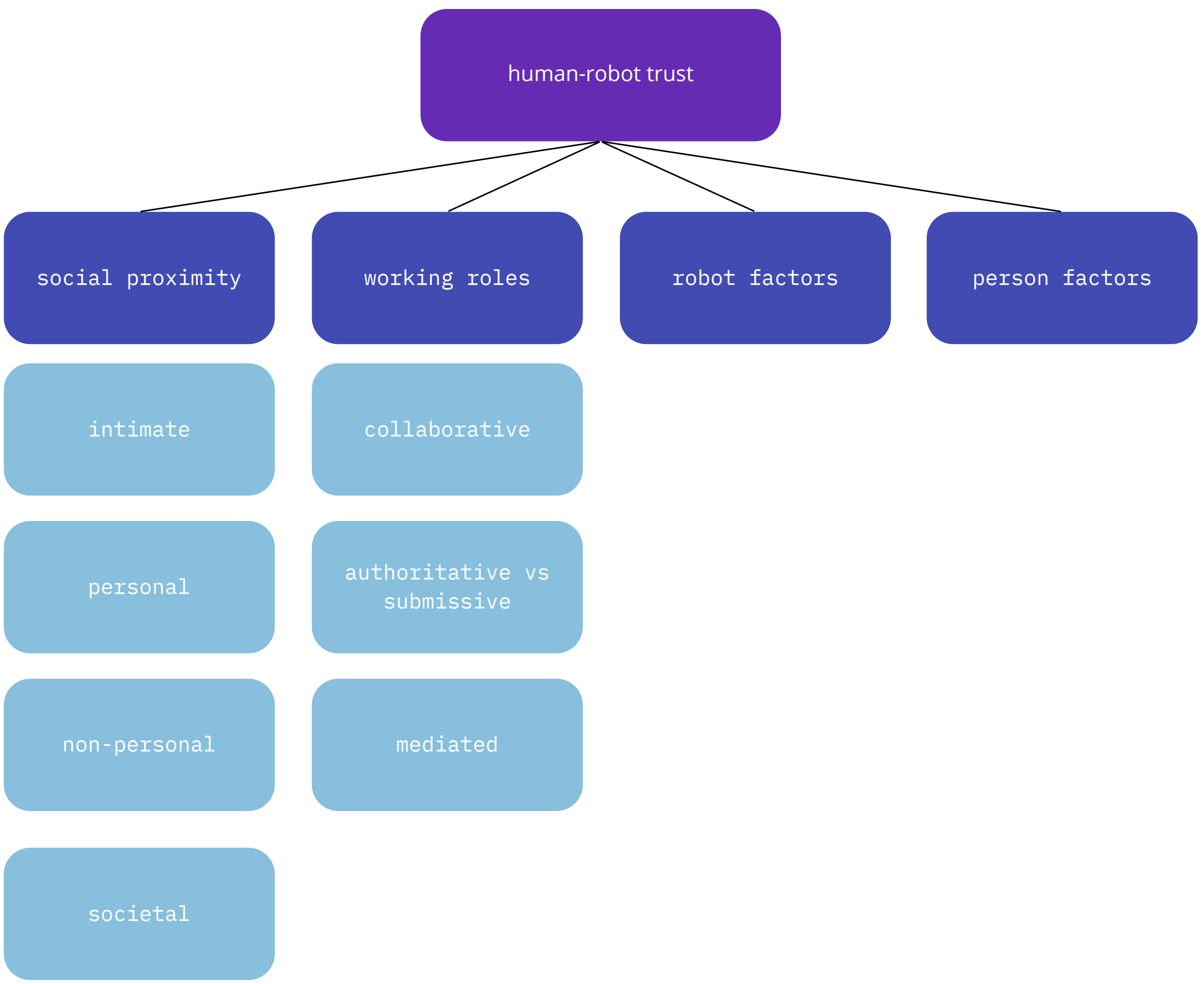


Figure 41. Dimensions and respective factors for working roles and social proximity with robots, which we developed through our survey on trust towards robots.

collaborative roles, we can expect individuals to interact with the same robot for an extended period of time, and potentially multiple time. Therefore, we argue, that intimate and personal types of trust might be likely here, such as **dyadic trust** towards a robot that is collaborated with on a regular basis, and **relational** or **cognitive trust** depending on the robot's performance and repeated interactions.

Working roles with robots: Authoritative roles vs. submissive roles affecting trust towards robots – While completing a non-collaborative task with a robot, research suggests that in some cases it is possible for either the human or the robot to be in a position of authority over the other (e.g., Geiskkovitch et al., 2016). This difference in authority can influence trust. Researchers have found that robots that appear to be in positions of authority, that is, they are either indicated to be or they provide people with instructions, might be trusted simply due to their position (even if their instructions are not logical; Geiskkovitch et al., 2016; Salem et al., 2015). Not only might these robots be trusted, but they may also be able to persuade people to accept information and follow suggestions (Bartneck et al., 2010; Freedy et al., 2007). However, it is unclear how submissive robots may elicit trust, as this topic has not been explored yet. We believe that **system** or **institution-based trust** might be relevant for robots in authoritative roles according to the behaviour they exhibit and any institutional affiliations they have. **Emotional trust**, on the other hand, might be more important for submissive robots, as they may need to make an emotional connection to elicit trust.

Working roles with robots: Mediated roles affecting trust towards robots – In some cases, the interaction between a robot and an individual may not be the end-goal of the interaction, but instead a means to complete a task with another individual. Research suggests that a factor that can impact trust towards a mediating robot is how they feel about the robot operator, and whether they perceive them to be a friend or a stranger (Rueben et al., 2017). However, feelings of trust towards the operator do not necessarily transfer to the robot and vice versa, but instead appear to be two distinct concepts (Kraft & Smart, 2016). Thus, it appears that trust in a mediated scenario can encompass various types, depending on what the individual bases their trust on, and how they perceive the operator on the other end.

7.5.3.1 Working roles with robots affecting human-robot trust summary

We have found that the type of role that a particular task or situation places a robot and human under can affect trust towards the robot. Collaborative roles might be more conducive to **interpersonal** and **dyadic** types of trust, while a robot that might be in a position of authority may elicit **system** or **institution-based trust**. While we may not consider the roles between children and robots to be 'working' ones, we might similarly expect for collaborative roles to elicit **interpersonal** or **dyadic trust**, while authoritative robots might instead call for **system trust** **(**section 7.6.2). In the next section we explore how attributes of a robot may influence trust.

7.5.4 Robot Factors Affecting Human-Robot Trust

According to our findings, robot-specific factors, such as how well a robot functions or whether it has any social capabilities can affect how much it is trusted (T. Sanders et al., 2011). We define these as factors that are specific to each robot, or type of robot, that may affect how people may perceive and trust them. For example, a robot's morphology can affect how people perceive it, with individuals especially anthropomorphizing humanoid robots and experiencing higher amounts of trust towards them than other morphologies (Geiskkovitch et al., 2016). We

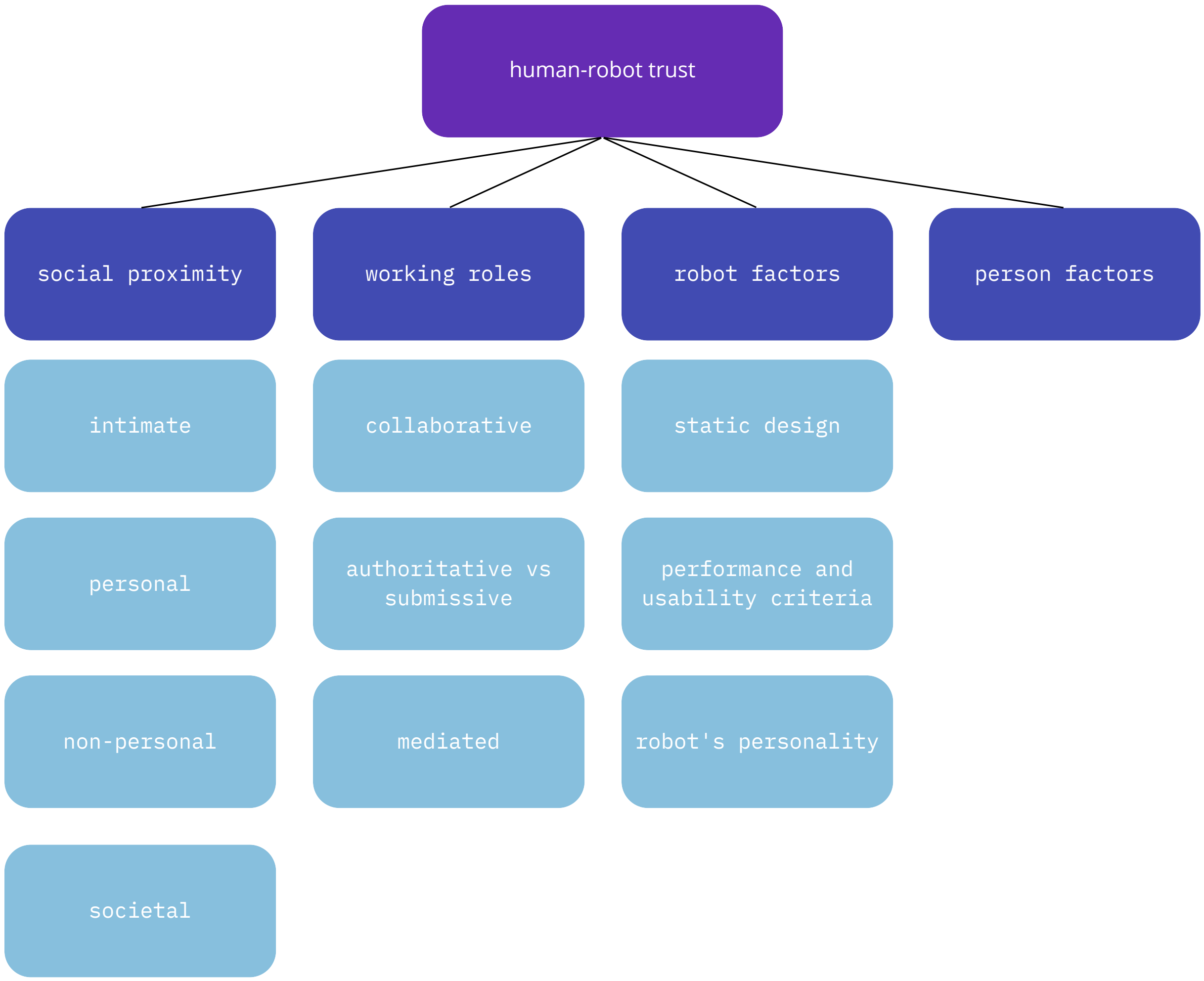


Figure 42. Dimensions and respective factors for robot factors, working roles, and social proximity, which we developed through our extensive survey on trust towards robots.

organize robot factors in the following categories: *static design*, *performance and usability criteria*, and *robot personality* (Figure 42).

Robot factors: Robot static design attributes affecting trust towards robots – Through our survey we have found that how a robot looks can affect how people trust it. For example, a robot whose exterior is made of shiny plastic can be perceived differently than one with a furry exterior. Research also suggests that a robot that matches a person's gender and voice (Eyssel & Hegel, 2012) is more likely to elicit positive feelings, which in turn could mean higher trust. Gender, in particular, may affect trust towards a robot when the risk of physical danger during a task is low (You & Robert, 2018). Additional design decisions, such as the use of gaze by the robot (Stanton & Stevens, 2017), empathic language and physical expressions (Tapus, Matarić, & Scassellati, 2007), and even the presence of vulnerability (Siino, Chung, & Hinds, 2008; Strohkorb Sebo et al., 2018) can also influence trust.

Researchers have found that a robot's static design also impacts trust through the way it communicates the robot's capabilities (real or perceived). Designs that indicate true capabilities are less likely to violate people's expectations, preventing distrust (Duffy, 2003). When the design is not transparent in relation to the robot's abilities, people may mistakenly trust the robot; A robot's appearance was cited as one of the main reasons that people followed a robot's incorrect (and dangerous) directions during an emergency situation (Robinette et al., 2016). Similarly, individuals perceive an autonomous vehicle to be more competent when it has anthropomorphic features (Waytz, Heafner, & Epley, 2014), which can lead to increased, but potentially misplaced, trust. Some researchers even claim transparency is necessary for justified trust in autonomous machines and robots (Fischer, Weigelin, & Bodenhagen, 2018; T. L. Sanders, Wixon, Schafer, Chen, & Hancock, 2014). We propose that the static design of a robot can affect **cognitive** and **emotional** types of trust, as well as **ideological trust** if both components are present. This is because certain designs can pull at people's heartstrings leading to **emotional trust**, while others (such as including anthropomorphic features) can make people rationalize trust in the robot (**cognitive trust**).

Static design: Robot morphology and trust. Our survey suggests that the physical form of a robot can influence the type and amount of trust that people may feel towards it. For example, people are more likely to take advice from a humanoid robot (Uchida et al., 2017), trust it and rely on it (Pak, Fink, Price, Bass, & Sturre, 2012), and comply with its demands (Geiskkovitch et al., 2016) than a mechanical robot. Below we break down how people may trust robots of varying morphologies.

Mechanical robots. One of the robot morphologies that is investigated in the literature is a mechanical, machine-like morphology. The lack of anthropomorphism in these robots may elicit less trust, and these types of robots are less likely to influence people (Geiskkovitch et al., 2016). We might expect mechanical-looking robots to elicit less trust than other types of morphologies, and even when they do, they might rely on **mundane trust**, with people trusting these robots without consideration. Otherwise, trust in these robots could be related to institutions or individuals they may be affiliated with (i.e., **institution-based trust**).

Humanoid and zoomorphic robots. Robots that resemble a human or animal shape tend to more easily elicit trust towards them, according to research. Robot anthropomorphism can be credited for increased trust in many situations, as people are more likely to interact with and disclose information to anthropomorphic robots (Mumm & Mutlu, 2011). Such trust translates to people obeying robots of this physicality more (Geiskkovitch et al., 2016), and performing tasks instructed by them even if they do not make sense (Salem et al., 2015). Even robots that are not humanoid or zoomorphic, but have anthropomorphic features are able to gain more trust in their overall abilities (Waytz et al., 2014). In addition, robots that share physically visual similarities with us are more likely to obtain positive outcomes, including trust (Eyssel & Kuchenbrandt, 2012; Kuchenbrandt, Eyssel, Bobinger, & Neufeld, 2013).

One of the largest concerns with anthropomorphic robots, such as humanoid robots, is that they are able to elicit trust even when it is not warranted. A humanoid agent that is clearly faulty is able to elicit trust, and convince people to do unusual or uncomfortable tasks, more so than robots of other embodiments (Geiskkovitch et al., 2016). People have even trusted zoomorphic robots (such as the iCat) sufficiently to undress in front of them and perform medical examinations on themselves (Bartneck et al., 2010). Individuals appear to develop intimate levels of trust with humanoid and zoomorphic robots quite rapidly, and we believe they can attain high **interpersonal**, **emotional**, and **dyadic trust** in a short period of time.

Virtual robots. A robot may have a physical presence, but not be co-located with an individual, therefore communicating through video or by voice. In general, people tend to trust robots that are not physically present much less than those that are (Bainbridge, Hart, Kim, & Scassellati, 2011). However, they can still elicit trust to different extents depending on what their virtual morphology is. For example, we propose that a humanoid virtual robot might still able to elicit **interpersonal** and **dyadic trust**, especially if working in a collaborative setting. Virtual robots could therefore be used purposefully, to elicit less trust when unwanted, or may otherwise need to have additional tools to elicit trust when necessary.

Robot factors: Robot performance and usability criteria affecting trust towards robots – Our survey revealed a robot's functionality and performance, in relation to the expectations of such, to be an important influence on how much it is trusted. Making a robot's functions clear to the individual using it is extremely important in building and maintaining trust (Fischer et al., 2018; J. D. Lee & See, 2004; T. L. Sanders et al., 2014). In addition to transparency about capabilities, a robot's task performance affects trust towards it (Desai, Kaniarasu, Medvedev, Steinfeld, &

Yanco, 2013; Hancock, Billings, Schaefer, et al., 2011; van den Brule, Dotsch, Bijlstra, Wigboldus, & Haselager, 2014), as well as the fluidity of its motions (van den Brule et al., 2014). A robot's performance also assists in developing (or hurting) trust as it informs people of its reliability, competence, and morals (Ullman & Malle, 2018, 2019).

Research suggests that one of the ways in which robot performance and perceived functionality affect trust is through the addition of social cues to robots. People may believe that due to the social cues provided by a robot, it is able to feel, think, or experience things similarly to how humans can, which is false (Kwon, Jung, & Knepper, 2016; Malle et al., 2020; Schramm, Dufault, & Young, 2020). This dissonance between what the robot's social capabilities appear to be and what they actually are is perhaps what leads to functionally sophisticated robots being trusted more than those with social abilities (Gaudiello, Zibetti, Lefort, Chetouani, & Ivaldi, 2016), since these social features are in no way able to match those that humans possess, and therefore expect. It is unclear what types of trust might be generally affected by robot performance. Aspects such as social cues, could affect **interpersonal trust**, if a robot is perceived to be a social other but is not able to communicate as such. Likewise, individuals may learn to not trust a robot that is difficult to use, work with, or makes mistakes, attributing this to a lack of **cognitive** or **calculus-based trust**.

Robot performance: robot errors and trust. We have found through our own experiments and literature survey that the incidence of errors, and how a robot deals with them can impact trust. There exist various types of errors that robots might exhibit such as hardware, software, behavioural, and sensemaking errors (Geiskkovitch & Young, 2020). Research suggests that errors in robots, as opposed to other machines or systems, especially impact trust due to people's expectations of robots' perfect performance (Ragni et al., 2016), and the high incidence of errors. In fact, the reliability of a robot is one of the most important factors influencing trust towards it (Desai et al., 2013). People tend to trust robots that are highly reliable (Sanchez, Fisk, & Rogers, 2004), and stop trusting them when the reliability drops (Desai et al., 2012). Further, the number of errors that a robot makes (Salem & Dautenhahn, 2015), the timing of such errors (Desai et al., 2012), and how the robot deals with the errors (by communicating with people, apologizing, or displaying good etiquette; Hamacher, Bianchi-Berthouze, Pipe, & Eder, 2016; Miller, 2005) can all impact trust towards the robot.

While errors may decrease trust in some situations, we have found multiple examples of people incorrectly trusting robots, and following their instructions, despite of errors. For example, individuals completed strange demands that a robot made, and trusted it, even when they perceived the robot to be faulty (Salem et al., 2015). People also followed a robot down an incorrect route during an emergency situation, even though they knew the route was incorrect, due to over-trust in the robot (Robinette et al., 2016). Similarly, participants have also continued to complete an unpleasant task even when a robot did not behave as expected (Geiskkovitch et al., 2016). We argue that robot errors, as well as performance in general, specifically relate to **relational trust**, where the accumulation of positive or negative interactions dictates how much trust is appropriate. However, if the interactions are sparse, or only take place once, the building or breakdown of trust cannot take place properly. Individuals may choose to overlook errors or performance problems that the robot has if they perceive the robot as more of a novelty, toy, or even pet than if it had purely utilitarian purposes.

Robot factors: Robot personality characteristics affecting trust towards robots – Through our survey, it is apparent that a robot's social abilities and (programmed or learned) personality can affect how much it is trusted. For example, a robot that has communicative or social abilities tends to be trusted more than one that does not (Hamacher et al., 2016). This might be due to the robot having more similar abilities to humans, because similarity increases emotional attachment (K. M. Lee, Peng, Jin, & Yan, 2006) and positive feelings towards a robot (Bernier & Scassellati, 2010; Tapus & Mataric, 2008), making trust more likely. A robot that has good etiquette is also able to gain trust from people (Hamacher et al., 2016; Miller, 2005); This includes the robot sharing its vulnerabilities, which not only increases trust towards the robot, but also towards other team members (Strohkorb Sebo et al., 2018). However, researchers and designers should be careful of adding too many social abilities to robots, as robots are not able to control them, which this leads people to misattribute capabilities to the robot (Malle et al., 2020), and eventually lowered trust.

Depending on a robot's personality, we suspect trust to develop or break in various ways, and for people to feel different types of trust towards it. A robot that is able to communicate similarly to how humans do might be more likely to elicit trust, such as **interpersonal**, **dyadic**, or **relational trust**, by interacting in ways that people are familiar with. We can expect a robot whose personality does not align with that of a user to draw less trust, similarly to what we might witness with people. **Emotional** and **cognitive trust** may also be impacted, as we suppose that a robot's personality might affect how people feel and what they think about trusting it. However, a robot that does not communicate socially, but possess other characteristics of a good personality (e.g., politeness or good etiquette) can elicit **system** or **institutional-based trust**.

7.5.4.1 Robot factors affecting human-robot trust summary

Factors including a robot's visual design, performance, and personality can impact trust towards it, according our survey of the literature. Robots that are designed to visually communicate their abilities and use social cues are more likely to elicit trust. Likewise, people tend to trust humanoid and zoomorphic robots more than virtual and mechanical ones. In addition, our research suggests that any mistakes that a robot makes can decrease trust towards it, while good performance is likely to increase it. Existing research suggests that we may see similar trends with children, which we explore in section 7.6.3. In the following section, we explore how factors related to humans can impact trust in robots.

7.5.5 Person Factors Affecting Human-Robot Trust

Characteristics of the individual interacting with a robot emerged as one of the dimensions affecting how they might perceive and trust the robot. For example, factors such as a person's prior experience with robots (or even with pets) and whether they were positive or negative, as well as an individual's personality can affect trust. We organize person factors into the following dimensions: *person's history with robots and pets*, and *personality* (Figure 43).

Person factors and robots: Person's history with robots and pets affecting trust towards robots – Research suggests that there is a significant link between previous encounters with animals and robots, and how a robot might be perceived. Individuals who have experience with animals such as pets, are more likely to be accepting of robots (Takayama & Pantofaru, 2009). Likewise, if a person has previously interacted with robots, the outcomes of that interaction can greatly influence whether they will trust any new robots they encounter (Bartneck, Suzuki, Kanda, & Nomura, 2007). An individual's history with animals or pets is not something that can be

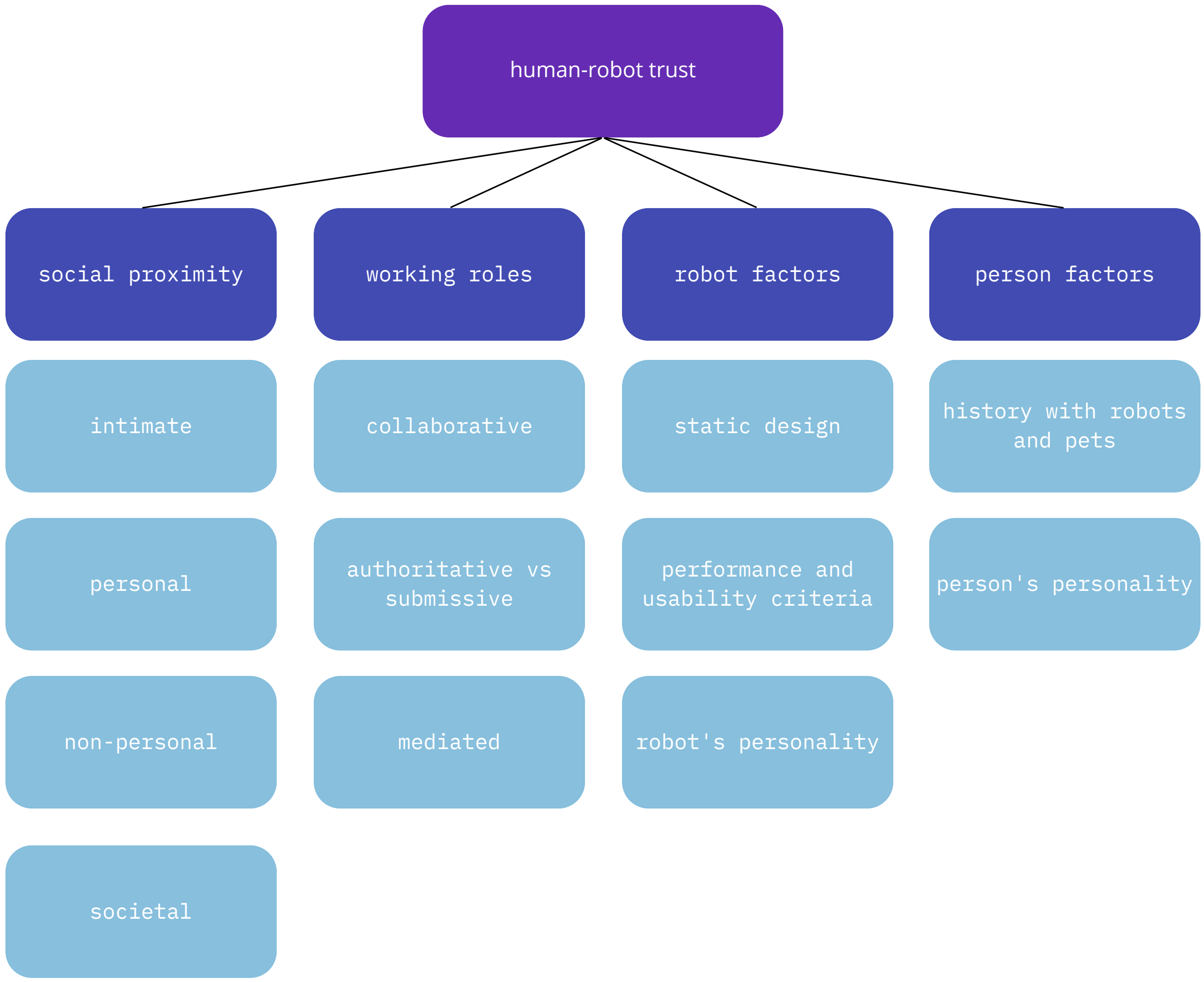


Figure 43. Dimensions and respective factors for person factors, robot factors, working roles, and social proximity, which we developed through our extensive survey on trust towards robots.

changed, but we believe that taking such aspects into account can provide insight into how an individual may trust a robot, or what a robot can do to gain their trust. We suggest that this points to **relational trust** as a key type of trust affected by prior history, in addition to **cognitive** and **emotional** types of trust depending on the valence of the individual's prior experiences.

Person factors and robots: Person's personality characteristics affecting trust towards robots – We have found that an individual's personality can affect how they interact with and trust robots. Extroverts, for example, may tend to trust robots more than introverts do (Haring, Matsumoto, &

Watanabe, 2013). Similarly, those with a high propensity to trust (i.e., **interpersonal trust**) are more likely to trust robots, while those with generally negative attitudes towards robots trust them less (Hancock, Billings, Schaefer, et al., 2011; Oleson, Billings, Kocsis, Chen, & Hancock, 2011). Self-confidence also plays a large role, especially with automated robots; Individuals with high self-confidence during a task are less likely to use robots and are therefore not able to develop trust, while those with lower self-confidence are more likely to use robots to make up for their own shortcomings, and therefore develop higher trust in them (J. D. Lee & Moray, 1994). We argue that an individual's personality can influence various types of trust including **interpersonal**, **dyadic**, **cognitive** and **emotional**, and even **calculus-** and **deterrence-based trust**, depending on the context of the interaction and additional factors of the robot and the situation.

7.5.5.1 Person factors affecting human-robot trust summary

Our research suggests that factors pertaining specifically to the person can influence their trust towards robots. Any experience they may have had with robots and pets (positive or negative), can affect how likely they are to trust a robot. Likewise, an individual's confidence in themselves can impact usage of a robot system, and therefore trust towards it. When it comes to children, however, we are likely to see different person factors being relevant (see section 7.6.4). In section 7.5 we discussed trust in HRI, based on research with adults. In the next section (7.6), we explore how children may trust robots.

7.6 Dimensions and Factors Relevant to Trust in Child-Robot Interaction

While most research (both on trust, and in Human-Robot Interaction) has historically focused on adults, we believe that children are a specific user-group that needs to be evaluated separately, as they do not have the same knowledge or experience that adults do. Through our exploration of trust between humans, towards systems, and towards robots (by adults) we have arrived at a set of dimensions and factors that might be relevant for child-robot trust. In this section, we explore trust based on research from Developmental Psychology and Child-Robot Interaction, to better understand what influences children's trust in robots.

Research in the area of Child-Robot Interaction (CRI) tends to focus on how robots can help children, by tutoring them in different subjects or providing friendly interaction (e.g., Belpaeme et al., 2012; Kanda, Hirano, Eaton, & Ishiguro, 2004; Kumar, Ai, Beuth, & Rose, 2010; Vogt, de Haas, de Jong, Baxter, & Krahmer, 2017). However, there is a lack of research when it comes to trust, likely due to the difficulty in examining the construct, especially with young children who may be unable to verbally express their feelings and cognitions.

As discussed in Chapter 3, children can learn to trust or mistrust people depending on experiences during early stages of development (e.g., Erikson, 1950) This, we believe, is mainly referring to **interpersonal trust**, and suggests that early events or experiences with robots or people can influence future perceptions, interactions, and trust towards robots.

We can expect trust in Child-Robot Interaction to differ from that of adults (or even between children of different ages) due to other developmental changes that take place, which may affect how they perceive technology. For example, a young child in Piaget's preoperational stage (2-7 years old; Piaget, 1954) may trust robots in different ways than an older child or an adult (see Chapter 3 for a thorough explanation). Further, the cultural environment of a child can influence their trust development (Vygotsky, 1962), likely impacting trust towards robots as well.

Due to the novelty of the CRI area, and the lack of research on child-robot trust, we apply what we deduced in the previous sections of this chapter, and research from CRI to inform how children may trust robots. We distill this into the following dimensions as we believe they relate to children and robots: *social proximity*, *interaction roles*, *robot factors*, and *child factors* (Figure 44).

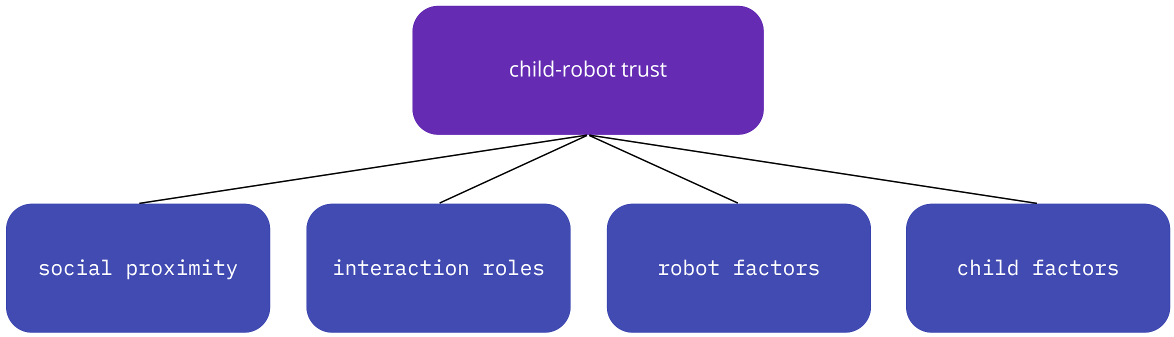


Figure 44. Dimensions that are relevant for child-robot trust, which we developed and extracted through our survey.

7.6.1 Social Proximity Between Children and Robots and Child-Robot Trust

Similarly to how the social and psychological distance between adults (section 7.3.1) or in relation to robots (section 7.5.2) can impact trust, our findings suggest that the same is true for children and the types of relationships they may have with robots. Children may develop close relationships with some robots (e.g., at school or home), but not with others (e.g., at shopping mall). We organize children's social proximity to robots into the following dimensions: *intimate*, *personal*, *non-personal*, and *society* (Figure 45).

Social proximity between children and robots: Trust in intimate child-robot relationships – Research suggests that children can have intimate relationships with robots, perceiving them as their close friends or even family. In such cases, we can expect children to experience deep emotional and social connections, leading to certain amounts and types of trust. For example, children can build intimate bonds with a robot, and share secrets with it that they do not share with close adults (Bethel et al., 2011). Children can also have intimate relationships with certain robots (i.e., those designed for homes), often treating them like pets or companions (Kahn, Friedman,

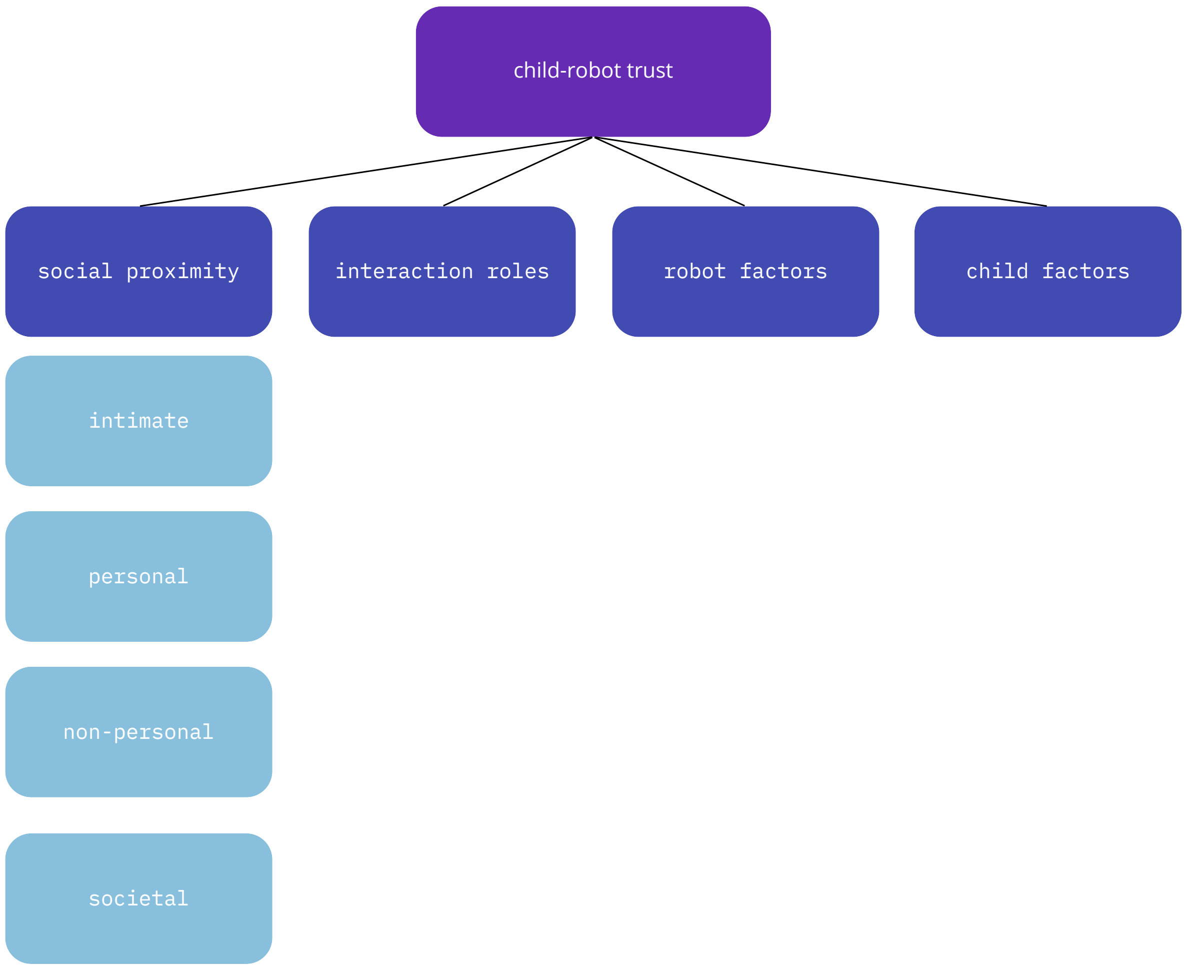


Figure 45. Dimensions and respective factors for children's social proximity to robots, which we developed through our extensive survey on children's trust towards robots.

Pérez-Granados, & Freier, 2006; Melson et al., 2005). In these cases, the intimacy with the robot appears to enable higher level of trust towards it. We propose that the types and amounts of trust that children may experience towards intimate robots are likely comparable to **dyadic** and **interpersonal trust**, believing that the robots are generally good and benevolent, and feeling deep connections with them.

Social proximity between children and robots: Trust in personal child-robot relationships – We can envision that children may interact with some robots on a regular basis, having personalized interactions, without developing deep connections. With teaching assistant or tutoring robots, for example, a child might interact with them on a regular basis, without developing a deep bond. Research suggests that, through long-term exposure, young children, start treating a robot as a peer and showing caring behaviours (Tanaka, Cicourel, & Movellan, 2007). Thus, repeated interactions between a child and a robot could foster **relational trust** if the interactions are positive. A supportive and helpful robot tutor, for example, we believe will likely elicit high **relational** and **interpersonal trust** from children.

Social proximity between children and robots: Trust in non-personal child-robot relationships – We propose that children may interact with certain robots on a regular or semi-regular basis, but not have personal relationships with them. For example, a child may be familiar with a tutoring or teaching assistant robot, but the robot may not be teaching the child's class. Due to the lack of information about the robot, the child may therefore base their decision to trust (or not) on the robot's appearance and general form, or simply not develop trust towards the robot at all. **Interpersonal** and **relational trust** are less likely here, with **system trust** likely being the most probable type of trust (if any) due to short exposures to the robot. However, we found this area to present a gap in the literature, as to our knowledge there is no research investigating this type of social proximity. Given our extensive knowledge on Child-Robot Interaction, and the types of positions robots are being designed an build for, we expect this type of relationship to exist in the near future. We therefore include non-personal relationships with the prospect of researchers investigating this area.

Social proximity between children and robots: Trust in societal child-robot relationships – We can expect children to come across many robots in society, be it in shopping malls, airports, or restaurants. Issues such as bullying have been observed with these types of robots (Brščić et al., 2015), suggesting that lack of trust (likely **cognitive**) or respect might be to blame. While children may not need to trust these robots, as they may default to adults' behavioural and trust cues, some trust might be necessary to prevent undesirable outcomes such as bullying and property damage.

7.6.1.1 Social proximity between children and robots and child-robot trust summary

Through our survey, we have deduced that the type of relationship that a child has with a robot, be it more intimate and repeated or not personal can affect the amount and type of trust that they may experience. Trust towards intimate and personal robots (e.g., teaching or home assistants, robot pets) can mimic that with humans, and we predict that children may potentially exhibit **interpersonal**, **dyadic**, or **relational trust**, and a higher degree of trust overall. On the other hand, children are unlikely to trust robots in society which they may not encounter in the future, and such lack of trust may in turn result in negative outcomes such as bullying. In the following section, we explore how the type of interaction between a child and a robot can impact trust.

7.6.2 Interaction Roles Between Children and Robots Affecting Child-Robot Trust

We can expect that during interaction between a child and a robot, the child or the robot may take different roles in the exchange. Unlike with adults, the interaction roles between a child and a robot are likely less about work, but more about the types of interaction that take place during learning or play. We organize interaction roles into: *collaborative*, *authoritative vs. submissive*, and *mediated* (see Figure 46).

Interaction roles between children and robots: Collaborative roles affecting children's trust towards robots – Research suggests that children can engage in collaborative tasks with robots, in which they act as partners and both have ownership over the activity. Children as young as 13 months old can perceive a robot as a collaborative partner (Park et al., 2015). Similarly to how adults may develop trust towards robots that they work collaboratively with, we believe children may do so as well, potentially increasing the amount of **interpersonal trust** that they may have

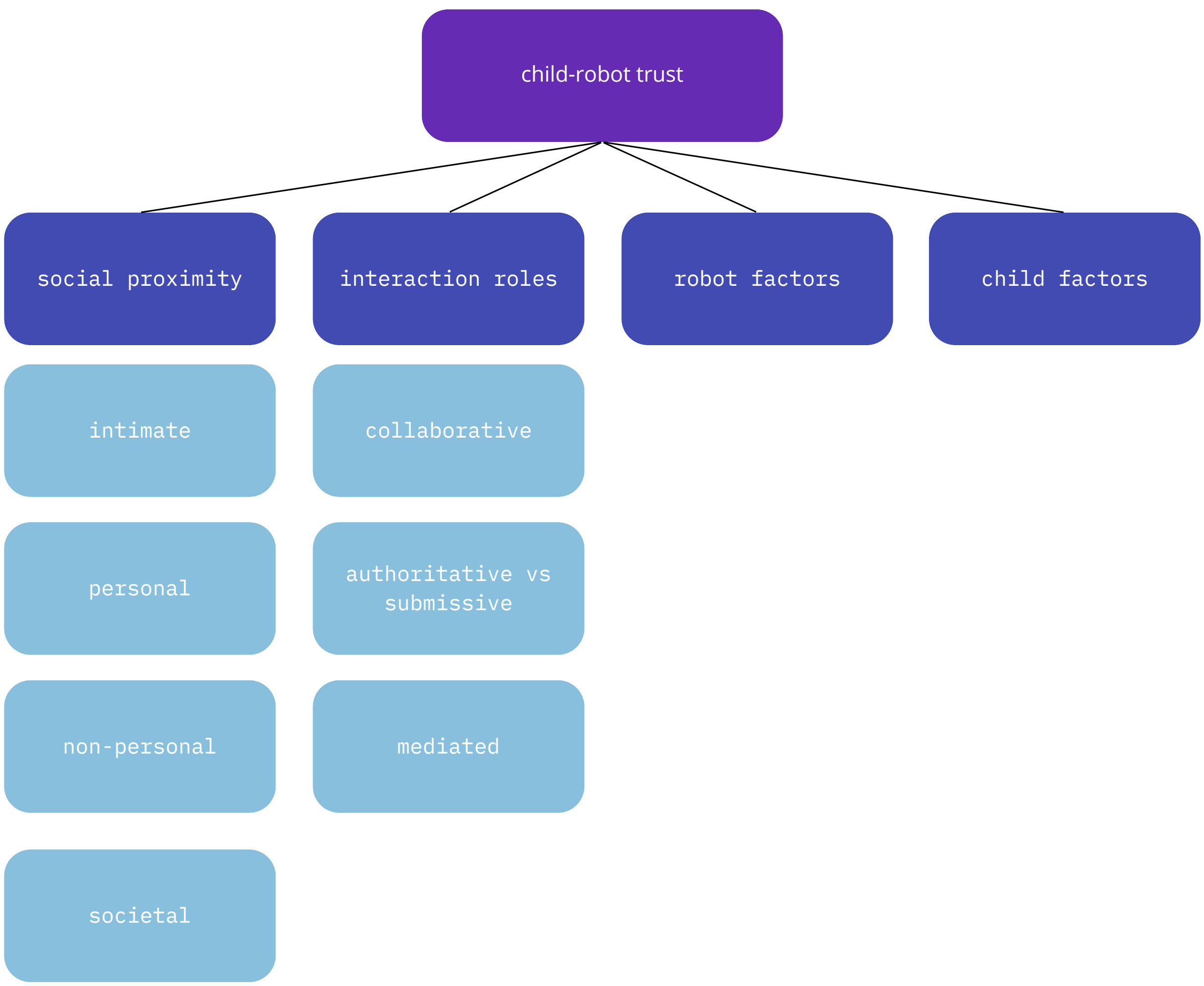


Figure 46. Dimensions and respective factors for interaction roles and social proximity between children and robots, which we developed through our extensive survey on children's trust towards robots.

towards the robot. Further, if this collaboration continues for an extended period of time, and is of a positive nature, we propose that **relational trust** can develop towards the robot.

Interaction roles between children and robots: Authoritative roles vs. submissive roles affecting children's trust towards robots – We can expect situations in which a child or a robot may be put into a position of authority over the other. For example, a tutoring robot will likely have some authority in terms of knowledge and potentially the behaviour of the child it is tutoring. Research suggests that children trust and are eager to perform well when being taught by a robotic tutor (Serholt, Basedow, Barendregt, & Obaid, 2015). Authoritative robots, we predict may elicit **system trust**, **cognitive trust**, and **relational trust** depending on the situation of the interaction.

On the other hand, we might see cases where a child might have more authority than a robot. For example, a robotic dog for the home is likely to be submissive under the child's authority. Research suggests that children tend to treat robotic dogs similar to how they do live dogs, and can build social bonds (Kahn et al., 2006). Thus, we propose that children may experience more personal types of trust towards this type of robot, such as **interpersonal** or **dyadic trust**. However, this may not apply to all types of submissive robots.

Interaction roles between children and robots: Mediated roles affecting children's trust towards robots – We suspect that some robots may act as mediators between a child and other children or adults. For example, a robot can facilitate the interaction between a child with ASD and another individual, allowing the child to be able to better communicate (Giannopulu & Pradel, 2012). This type of interaction relationship is yet to be investigated with children in relation to trust. However, we might expect trust outcomes to depend on similar aspects as with adults, such as the interaction with the robot and with the other individual. Thus, we believe it is possible that

interpersonal trust, **relational trust**, or even **system trust** play a role here, depending on the interactions between the child, the robot, and the other individual(s), and how the child perceives the robot (e.g., as a social being versus a lifeless machine).

7.6.2.1 Interaction roles between children and robots affecting child-robot trust summary

Depending on the type of robot and the task at hand, our synthesis proposes that the interaction between a robot and a child may take a collaborative form, a mediated form, or one of the two might be in charge. We believe that children are likely to trust robots they collaborate with (as long as the outcomes are positive), and might deflect to a robot's authority if it possesses such. However, it is unclear how mediated interactions may affect children's trust. In the next section, we explore how robot factors may influence trust.

7.6.3 Robot Factors Affecting Child-Robot Trust

Research suggests that several aspects of the robot's design and behaviour can impact trust towards it. Research into factors that may affect children's perceptions of robots has focused on attributes such as a robot's looks (Sciutti et al., 2014), gestures (Wit et al., 2018), as well as the sociality of the robot (Westlund, Martinez, Archie, Das, & Breazeal, 2016). As it relates to trust, factors such as the timing of a robot's responses (Breazeal et al., 2016) are known to influence trust, although researchers suggest that the topic of trust in CRI is greatly understudied (van Straten, Peter, & Kühne, 2020). We categorize robot factors that affect children's trust into: *static design*, *performance and usability criteria*, and *robot personality* (Figure 47).

Robot factors and child-robot trust: Robot static design attributes affecting children's trust towards robots – We can expect a robot's looks to impact how children perceive and trust it. Transparent designs, which clearly exhibit a robot's capabilities, tend to decrease perceptions of

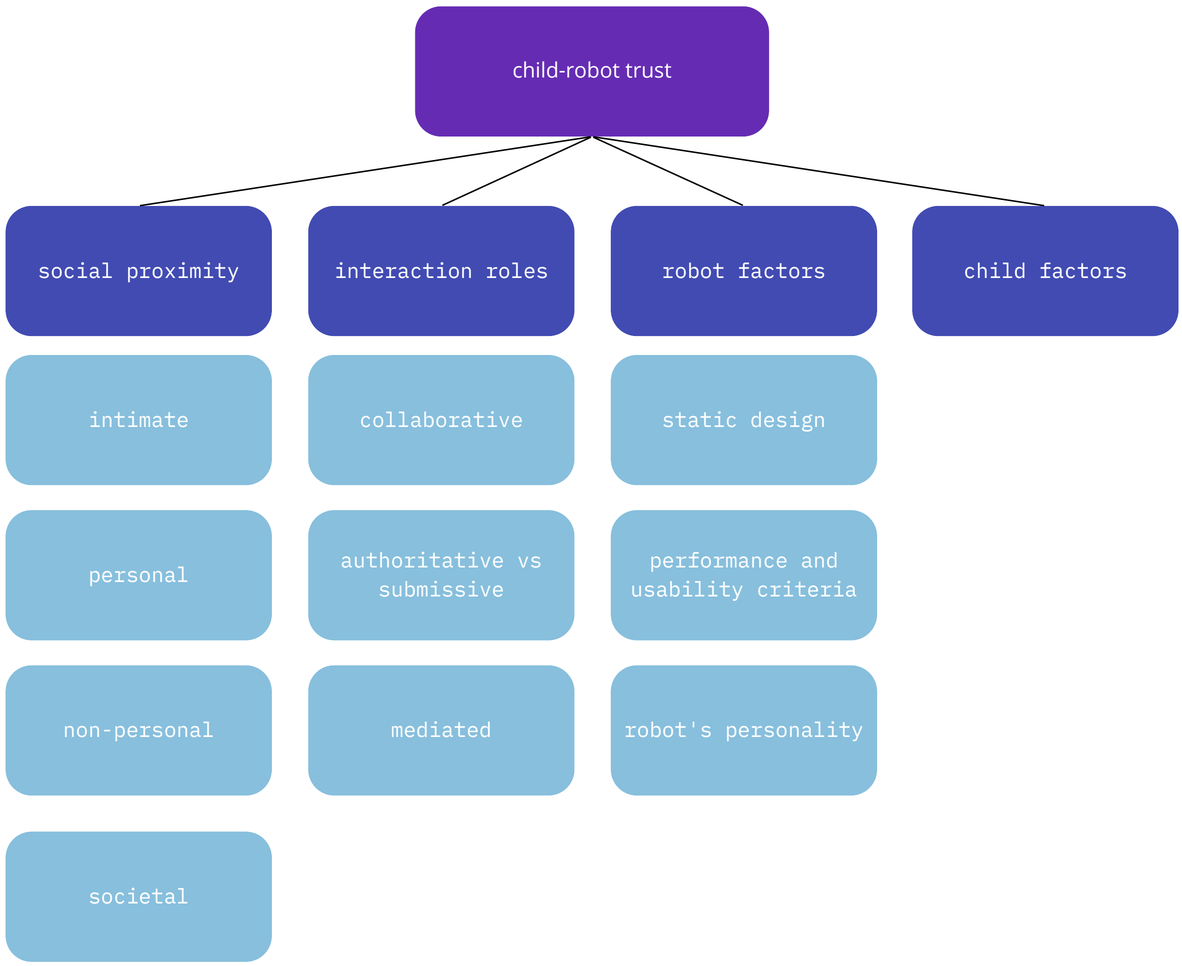


Figure 47. Dimensions and respective factors for robot factors, interaction roles, and social proximity between children and robots, which we developed through our extensive survey on children's trust towards robots.

animacy and anthropomorphism in children, and therefore lead to lower or more justified trust (van Straten, Peter, Kühne, & Barco, 2020). On the other hand, perceptions of technological advancement can increase children's trust in robots (van Straten et al., 2018).

Robot static design: Robot morphology and children's trust. Research suggests that the morphology of a robot, be it humanoid, zoomorphic, mechanoid, or virtual can impact how much children trust it. While research with children does not yet inform how morphology impacts children's trust in robots, it is possible to expect children to behave similarly to how adults do. Children might therefore trust humanoid robots, more than other types of morphologies, which we expect would likely elicit **interpersonal**, **dyadic**, and **relational trust** in repeated, positive, and close interactions. We believe that a virtual robot might be able to elicit similar types of trust from children, but likely to a lesser degree, similarly to adults. However, even mechanoid robots may be able to elicit more trust in children than they do in adults, since children tend to perceive robots in general as alive (Rao & Georgeff, 1995). Thus, we can expect even a mechanoid robot might be able to elicit some **interpersonal trust**, especially in younger children, while older children may experience **mundane** or **cognitive trust**, similarly to adults.

Robot factors and child-robot trust: Robot performance and usability criteria affecting children's trust towards robots – Our research suggests that a robot's performance and functionality can influence how much it is trusted by a child. When children are informed of a robot's true capabilities, or lack of human-like abilities, their trust in the robot as well as perceptions of animacy and anthropomorphism decrease (van Straten, Peter, Kühne, et al., 2020). However, if children are not aware of a robot's true abilities, any errors that it makes may decrease their trust towards the robot.

Robot performance: Robot errors and children's trust. As explained throughout this thesis, robot errors are vastly common and can affect how children view robots. A simple delay in a robot's response (which may or may not be considered an error) can reduce young children's trust in a robot (Breazeal et al., 2016). Our own experiments suggest that robot informational errors tend to decrease trust towards a robot in young children (experiment 1), but robot speech-recognition errors may instead increase trust in a robot (experiment 2). We believe that various types of trust might be at play here; **interpersonal trust** could be affected by a robot's performance, especially if children view it as a social other. Likewise, **relational trust** might be affected by a robot that continuously make errors, alongside **cognitive trust**. Children that interact with less familiar robots, in public for example, may develop (or lose) **system trust** or potentially **institutional-based trust** if the robot is linked to an institution. However, whether trust increases or decreases, as well as the type of trust present, appears to potentially be dependent on the types of errors that a robot exhibits.

Robot factors and child-robot trust: Robot personality characteristics affecting children's trust towards robots – Similarly to how a robot's personality can influence how adults perceive and trust the robot, our survey suggests that the same can happen with children. Research suggests that a robot that changes its emotions based on the situation is perceived more positively by children, but one that does not adapt actually elicits more trust (Tielman et al., 2014). Likewise, we can expect a friendly and charismatic robot to likely be trusted more by children than an antisocial or mean robot, simply based on common human social interaction. We propose that robot personality affects the likelihood of developing **emotional** types of trust towards a robot, and a sense of closeness necessary for **interpersonal** and **dyadic trust**.

7.6.3.1 Robot factors affecting child-robot trust summary

Similarly to adults, we have found that children tend to trust robots that possess anthropomorphic design features. However, this can backfire if the robot's design promises more than it can do. In addition, performance factors, such as the presence of any errors, can also influence trust; our research demonstrates that some robot errors may decrease children's trust (e.g., informational errors) while other may increase it (i.e., speech-recognition errors).

7.6.4 Child Factors Affecting Child-Robot Trust

Our synthesis suggests that characteristics specific to each child, such as their age, gender, personality, as well as history with robots and animals can affect their trust towards robots. Younger children, for example, may not have a good idea of what robots are or may not have developed a theory of mind (described in Chapter 3) that allows them to appropriately judge robots, while older children might (Di Dio et al., 2020). A child's gender, generally and in relation to a robot, can also affect the child's perceptions of the robot, and likely trust as well (Westlund, Park, Williams, & Breazeal, 2017). We organize child factors into: *demographics* and *history with robots and animals*. See Figure 48 for a complete diagram of factors our research suggests affect children's trust towards robots.

Child factors and child-robot trust: Child's demographic characteristics affecting children's trust towards robots – One of the biggest aspects that we believe differentiate children and adults, and therefore their interactions with and trust towards robots, is ongoing development. For example, children younger than 5 years old may not have acquired theory of mind yet, making their perceptions of robots (and therefore trust) likely different from older children or adults (Di Dio et al., 2020). However, research suggests that 3-year old children appear to trust humans more than robots, while by 7 years of age children trust robots more than humans (Di Dio et al., 2020), highlighting how age may not be the only relevant factor. Gender may also play a role in trust towards robots, as girls tend to attribute more social characteristics to robots (Westlund et al., 2017). We believe this might result in the presence of **interpersonal** or **dyadic trust** for girls,

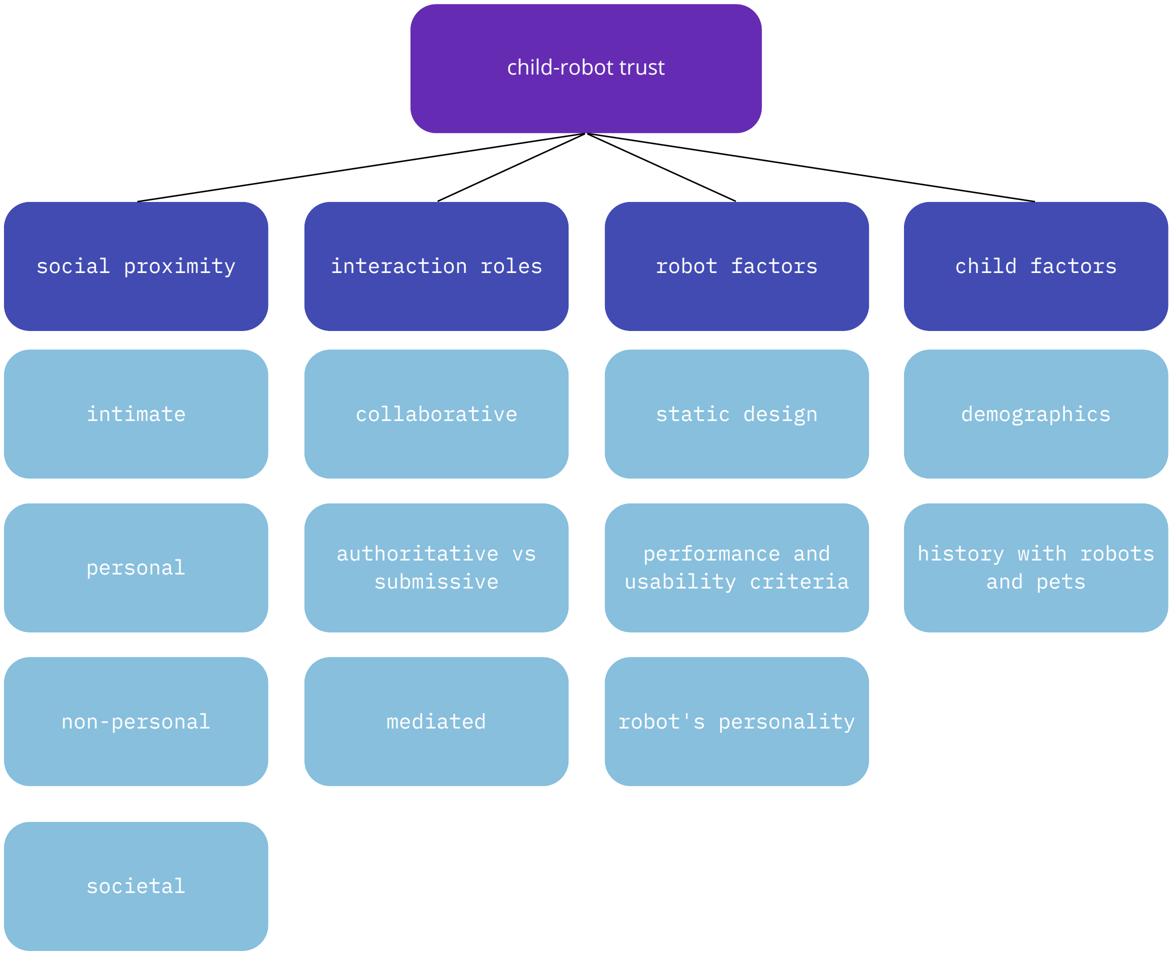


Figure 48. Dimensions and respective factors for child factors, robot factors, interaction roles, and social proximity between children and robots, which we developed through our extensive survey on children's trust towards robots.

especially if there are multiple exposures. Boys, on the other hand, we expect may be more likely to exhibit **cognitive** or **system trust**. We believe that it might also be possible for a child's personality or culture to affect how they trust robots, although research on these aspects does not yet exist.

Child factors and child-robot trust: Child's history with robots and animals affecting children's trust towards robots – As discussed in section 7.5.5, research suggests that prior experience with animals (Takayama & Pantofaru, 2009) or other robots (Bartneck, Suzuki, et al., 2007) might affect adults' trust towards robots. We suspect that previous experiences with robots (and potentially animals) are likely to affect a child's likelihood to interact with and trust a robot as well. Depending on the valance of prior experiences, we believe trust towards robots could increase or decrease. For example, if a child has never seen or heard of robots, they might be initially scared when they meet one, and might be hesitant to interact with it. Likewise, if a child has interacted with robots before, but the interactions were unsuccessful or the robots made errors, the child may not trust other robots easily. However, this area of research remains uncharted for the time being, thus we cannot make any formal propositions or conclusions.

7.6.4.1 Child factors affecting child-robot trust summary

We believe that factors specific to a child are likely to affect how likely they are to trust a robot. Age and gender are known to affect how children interact with robots, which leads us to propose that children of different ages could trust robots in varying amounts, and girls might exhibit different types of trust in robots than boys would. Other factors such as personality and prior experiences may also play a role in how children trust robots, although research in this area is still lacking.

7.7 Chapter Summary

In this chapter, we developed a framework of child-robot trust, by investigating literature from related fields. Specifically, we began by exploring trust between people, and extracting from the literature which factors affect such trust. Then, we continued our journey to investigating how people trust systems, considering how this might impact children's trust along the way. From there, we narrowed down to people's trust towards robots, focusing specifically on adults due to the lack of research in this area on children. Lastly, we applied the knowledge obtained from the aforementioned sections and areas to how children might trust robots along multiple dimensions.

In Table 4 we provide a summary of the factors that we can expect to influence children's trust towards robots (as well as between people and towards machines).

Our exploration into factors that influence trust also led to the creation of the *dictionary of trust*, which we presented in Chapter 3 and again in Table 3. This aspect of our work provides researchers with shared vocabulary for studying and discussing trust in Human-Robot Interaction. In addition, we highlight how these types of trust might apply to Human- and Child-Robot Interaction.

Following, in Chapter 8 and Chapter 9, we discuss what the two experiments in this thesis as well as this child-robot trust framework tell us about children's trust towards robots. We also provide guidelines for studying trust with children, and future potential directions in this area.

Table 3. Dictionary of trust developed through our survey of the trust literature (also discussed in Chapter 3).

Type of Trust	Relevance to HRI and CRI
Interpersonal: Belief held by an individual that others can be relied on (Rotter, 1971). Composed of cognitive and emotional components (McAllister, 1995)	Can affect how likely an individual is to trust a robot, or how skeptical they might be of it. Can be affected by cognitions and emotions that people have about robots
Dyadic: Belief that other person is honest and benevolent (Larzelere & Huston, 1980)	Can play a role in close relationships with robots, where a robot is perceived to be part of an intimate or personal circle
Relational: Develops due to repeated positive interactions (Lewicki et al., 2006; Rousseau et al., 1998)	Can increase or decrease depending on the valance of interactions with a robot. Might be especially impacted by robot errors.
Institution-based: Faith in an institution's benevolence (Lewicki et al., 2006)	May trust robots belonging or linked to a specific institution. Children may trust robots that look like specific characters more than others
System: Sense that everything is in good order (Luhmann, 1979)	Can lead to ignoring robots' behavioural red flags and over trusting robots
Deterrence-based: Consequences of not keeping trust would be too severe (Shapiro et al., 1992)	May assume that a robot is trustworthy because of potential negative repercussions. Unclear whether or how this applies to robots, especially for interaction with children due to the likelihood of delegating trust
Cognitive : Carefully thought-out trust (Lewis & Weigert, 1985)	May trust a robot in society only after careful consideration
Emotional: Feelings-based trust, either based on strong (possibly long-term) emotions or temporary hunches (Lewis & Weigert, 1985)	May trust a robot simply because of emotions (i.e., friendly, nice, etc.). This might especially apply to younger children.
Mundane: Trust low on emotional and cognitive components, prevalent in everyday life (Lewis & Weigert, 1985)	May trust a robot simply because there seems to be no reason not to
Ideological: Trust high on emotional and cognitive components, and prevalent as it relates to society (Lewis & Weigert, 1985)	May affect likelihood to trust robots given prior cognitions and emotions about them
Calculus-based: Calculated decision based on facts (Lewicki et al., 2006; Rousseau et al., 1998)	May trust a particular robot only after careful consideration about potential risks and benefits. This is less likely to be experienced by children

Table 4. Dimensions and factors obtained through our extensive survey into what affects trust towards humans, systems, and robots. Our framework is the first to provide this knowledge for child-robot trust.

Trust Category	Dimension	Section	Sub-dimensions
Human-Human	Social Proximity	7.3.1	Intimate, Individuals, Institutions, Society
	Person Factors	7.3.2	Emotional, Cognitive, Personality
	External Factors	7.3.3	Expected Outcomes, Situational Context
Human-System	Sociality	7.4.1	Machine-like, Social, Mixed
	Working Roles	7.4.2	Collaborative, Active vs. Passive, Mediated
	System Factors	7.4.3	Visual and Usability Design, Certifications and Endorsements, Performance and Usability Criteria
	Person Factors	7.4.4	Propensity to Trust Technology, Propensity to Anthropomorphize Technology, Self-Confidence with Technology
Human-Robot	Social Proximity	7.5.2	Intimate, Personal, Non-Personal, Societal
	Working Roles	7.5.3	Collaborative, Authoritative vs. Submissive, Mediated
	Robot Factors	7.5.4	Robot Static Design, Performance and Usability Criteria, Robot Personality
	Person Factors	7.5.5	History with Robots and Pets, Personality
Child-Robot	Social Proximity	7.6.1	Intimate, Personal, Non-Personal, Societal
	Interaction Roles	7.6.2	Collaborative, Authoritative vs. Submissive, Mediated
	Robot Factors	7.6.3	Static Design, Performance and Usability Criteria, Robot Personality
	Child Factors	7.6.4	Demographics, History with Robots and Animals

Chapter 8 General Discussion, Limitations, and Future Work

In this thesis, we explored how robot errors affect children's trust in informational robots. The overarching goal of our work was to investigate how children may trust robots, by exploring how trust might relate to humans, and how such trust is influenced by robot characteristics (i.e., errors). The first experiment included a scenario where children were able to demonstrate trust in informational robots. In this case, one robot made human-like errors in the information it provided (based on prior work), which enabled us to examine how human-like errors in robots may affect children's trust. Following, we conducted a second experiment to investigate more robot-typical errors, and study how young children's trust might be impacted as compared to human-like errors. Finally, we performed an in-depth exploration and analysis of trust and robots in the literature, and developed a framework of child-robot trust to inform future research and robot design (Chapter 7). In this chapter, we examine this body of work, reflecting on our process, assumptions, findings, impact of this thesis.

8.1 Reflections on Research Questions

In Chapter 5, we detailed our experiment on children's trust towards robots with informational errors, and in Chapter 6 we described our experiment on children's trust towards robots with speech-recognition errors. In this section, we discuss our findings in relation to our research questions (Chapter 1).

8.1.1 Do preschool aged children exhibit trust towards robots in similar patterns to how they do towards people or puppets?

Through the first experiment in this thesis, we found that children tend to exhibit trust towards robots in a similar pattern to how they do people or puppets. Similarly to experiments in Psychology that have found that children trust informants who do not make informational errors more than ones that do (Birch et al., 2008), the findings from our first experiment demonstrate that preschool-aged children also trust a robot informant that has not made errors more than one that has.

These findings suggest that children will treat robots similarly to how they do humans, using their mental models observed when interacting with humans or puppets to also interact with robots. This is extremely significant, as children may treat robots differently from how they do people, given their unique physicality and social standing. Thus, drawing from this we might expect children to experience interpersonal, dyadic, or relational trust towards robots that they know or those that are in their intimate or personal circle. If the robots are unfamiliar to a child, the child may turn towards cognitive or calculus-based trust. However, when it comes to robots in society (those that children are not familiar with such as at malls and other public places), children may not need to consider trust, as that tends to be their caregivers' responsibility. Research investigating how children's trust responses to humans compare to robots can provide more insight into this topic, and will be valuable research in the future, providing key information about when and how children may or may not trust robots.

8.1.2 How do robot errors affect preschool-aged children's trust towards robots?

In our research, we found that children not only exhibit trust towards robots (in similar patterns as to people; RQ1) but that errors that a robot makes can impact children's trust towards the robot (RQ2). However, we only found this to be the case for specific situations (with no results in others). In particular, errors that a robot made impacted trust in cognitively-simple situations, but not in more cognitively-complex circumstances. Thus, depending on the type of robot, its familiarity to children, and any institutions that it might be linked to, we might expect relational or cognitive trust to be impacted (for familiar robots) due to errors, and institution-based or system trust to play a role (e.g., for well-known institutions).

The findings that robot errors can impact preschool-aged children's trust towards robots have several implications for the HRI field. Specifically, they point to the possibility of children experiencing interpersonal trust towards robots (as they do towards humans), and such trust being affected by errors that robots exhibit. Considering the prevalence of uncontrolled errors generally, and especially when interacting with children, this points to the effects of currently unmanaged variables (i.e., errors) on trust towards robots.

8.1.3 Does the nature of a robot error impact preschool-aged children's trust toward the robot?

Our first experiment (Chapter 5) focused on human-like informational errors, while the second experiment (Chapter 6) explored robot-typical speech-recognition errors. Surprisingly, we found that these two types of errors influence children's trust in the robots in distinct ways. Robot informational errors appear to decrease children's trust towards the erroneous robot (possibly demonstrating decreased cognitive aspects of interpersonal trust or calculus-based trust), while children's trust towards a robot increases in the presence of speech-recognition errors (perhaps due to more emotional aspects of interpersonal trust). Our findings point to the need to consider the novelty of robot-typical errors for children, and how their lack of exposure to these types of errors can affect children's perceptions and trust of robots. This also suggests that unexplored errors still require investigation, because we do not know how children will perceive them, or how these errors will impact trust.

In addition, it may be that we witnessed errors being treated differently by children, not only due to the differences in the errors, but also due to our measure of trust. Thus, depending on the type of error that a robot exhibits, young children may exhibit (or not) distinct types of trust (e.g., interpersonal, cognitive, emotional). Further research is required to pinpoint which types of trust are affect by distinct robot errors. On the other hand, our findings could point to other aspects of child development, such as politeness norms. There is a large focus on teaching children to be polite to others, which is continuously reinforced at a young age. This means that it is possible that are findings are due to children simply being polite to the robot that made errors or acted differently. Further research would be necessary to explore this.

8.1.4 What do we know about trust in people (including children) and how might this inform how children may trust robots?

In Chapter 7, we reviewed literature related to trust to arrive at important factors for child-robot trust. Through the development of our framework, we found that there are several factors that affect trust towards other people, systems, and robots. Specifically, we believe the main aspects affecting children's trust towards robots are likely: their social proximity to the robot (e.g., personal or nonpersonal), the role that the child and the robot play during the interaction (e.g., collaborative, mediated), various factors inherent to the robot (e.g., personality, design), and factors related to the child (e.g., age, gender, history with robots). These aspects can affect the types of trust that children may experience, and to which extents. However, many of these areas remain unexplored in Child-Robot Interaction. The social proximity between children and robot might be of special importance, as our framework points to how different the trust relationships between children and robots of different proximities could be, a topic currently unstudied.

For the research presented in the thesis, we focused on robot factors related to performance and usability, specifically the presence of errors. This was a previously unexplored topic in relation to children but has since been receiving more attention in the literature. Other aspects such as child-related factors are yet to be investigated in Child-Robot Interaction.

8.2 Reflections on Experimental Design

For this thesis, we adapted and extended a previously validated experimental methodology from Psychology (Birch et al., 2008). However, we made a number of decisions along the way which may have affected our findings (or lack thereof). In this section, we provide some reflections on our methodological decisions to further learn from our experience and provide suggestions to other researchers.

8.2.1 Lack of phase counterbalancing

For our experiments, we decided against counterbalancing the testing phases, and instead for them to be the same for all participants. While some prior work counterbalanced the experimental phases (DiYanni, Nini, Rheel, & Livelli, 2012), others did not (Koenig & Harris, 2005a). We chose to follow in the footsteps of the latter for a number of reasons. For one, this meant that there was a clear progression in task difficulty (*same label* phase more cognitively simple than *contrast label* phase) and it gave higher priority to the *same label* phase, the main aspect of our experiments. In addition, counterbalancing would have lowered the power of our experiments with this number of participants (the power we observed being similar to that of related work). However, the lack of phase counterbalancing might be to blame for the lack of statistically significant findings in the *contrast label* phases of our two experiments. In the future, we would like to conduct another experiment in which the phases are counterbalanced, to examine whether this may have played a role in our experiments and to explore any differences as compared to the two experiments presented in this thesis.

8.2.2 Clean-up task

One of the additions we made to the experimental design was that of the *clean-up* task during the first experiment. With the *clean-up* task, we intended to investigate how robot errors affect children's instruction-following behaviours, as the outcomes with adults are not always as expected (e.g., Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015). However, we did not obtain any statistically significant results as part of this task. Further, since this added more time between the display of the errors and the *contrast label* phase, which may be related to our lack of findings, we removed the *clean-up* task from the second experiment; that is, we conducted the *same label* and *contrast label* phases back-to-back without the presence of the *clean-up* task. Nonetheless, given that there is some evidence that adults tend to follow the instructions of faulty robots that they claim to not trust (Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015), it is important to find a way of evaluating the impact of errors on young children's instruction-following behaviours, as robots generally provide instructions to children.

Future work should especially investigate children's trust towards robots outside of informational situations given that our framework (Chapter 7) revealed that authoritative robots are only one type of interaction role. Robots in other interaction roles or social proximities may not be providing information at all, but may still be trusted. Furthermore, the types of trust that might be experienced towards informational robots (e.g., cognitive or relational) might not be the same towards other robots. Therefore, it is necessary for future work to find novel ways to investigate children's trust outside of informational situations.

8.2.3 Identical robots

For our experiments we used two identical humanoid robots, reasoning that this would lead to one less potential confound. While we obtained statistically significant results for the *same label* phase of both experiments, suggesting that children were properly able to distinguish between the two robots, the lack of findings for the *contrast label* phases raises some concerns. For example, it is possible that the extended time between the *history* and *contrast label* phases compounded with the lack of distinctive features caused children difficulties in remembering which robot had exhibited errors, especially in the first experiment (due to the presence of the *clean-up* task). Thus, making the robots more distinguishable (as people are) may support children in remembering specific aspects about the robots.

Due to these concerns, for the second experiment we placed distinct paper cut-outs (one square and one circle) on the robots' chests, to aid distinguishability between the robots (see Figure 49). However, this did not lead to statistically significant results for the *contrast label* phase, suggesting that either distinguishability was not the problem, or that adding distinct shapes to the robots' chests was not sufficient. Finding more ways to distinguish between the robots, without adding potential confounds, could ensure that the robots being identical does not confuse children.

Figure 49. SoftBank robots used in the second experiment. The robots look and sound the same, but we added shapes to their chests to help children distinguish between them.

8.2.4 Limited direct interaction between child and robots

In our experimental setup, which was adapted and extended from Psychology (Birch et al., 2008), a child merely observes two robots label objects; the child has very little interaction with the robots. This means that our results stem from observations the child makes, and not interaction per se. This likely hindered social aspects commonly developed through interaction, such as rapport-building, between the child and the robot. However, we can expect that in many situations children will interact with and build some rapport with robots (except, perhaps, for societal robots that they only meet once).

According to our framework (Chapter 7), although rapport is not necessary to exhibit certain types of trust, it is likely necessary for interpersonal and relational trust. In addition, robots in different social proximities to children may build different amount of rapport with children, and therefore varying types or amounts of trust. For example, societal robots may not build any rapport with children, and children may therefore not trust them or defer their trust to an adult. Our findings may therefore be limited to situations in which informational robots have not built any (or very little) rapport with a child. Thus, our research may generalize mostly to societal and non-personal robots.

8.2.5 Logistical challenges of experimental design

Our experimental design, which was adapted from previous research in Developmental Psychology, posed several logistical challenges for the research team. These challenges were due to characteristics of the participant group, the experimental methodology, and the use of robots.

Participant group – there exist several aspects related to conducting research with preschool-aged children that are not often discussed in the literature. For example, to maintain children's attention and ensure they were comfortable during the experimental sessions, the experimenter had to be extremely friendly and upbeat during all of the experimental sessions. Given that the research team conducted the experimental sessions at each daycare for at least eight hours without a break (to accommodate the daycare's schedule), this caused a high level of mental and physical fatigue for the research team, and especially the experimenter.

Experimental methodology – the established methodology that we utilized in our experiments exerts a high cognitive load on the research team. It was necessary for the experimenter to keep track of all of the familiar and unfamiliar objects, present them in the appropriate order to the correct robot, and simultaneously remember or keep track of which object label they should be asking the child for. In contrast to other experimental methodologies with children or adults, where the participants lead the study activity, in our experiments this was the task of the experimenter, placing them in a state of high cognitive load for extended periods of time. In addition, the research assistant also experienced a high workload, being responsible for controlling two robots simultaneously. Lastly, the experimental methodology required twenty distinct object, two robots (plus related equipment), and toys (for compensation) to be taken to various daycares. This meant that the experimenter and research assistant had to carry large amounts of equipment to and from each research location, causing more-than-normal physical strain.

Robots – the use of robots for our experiments was crucial, as that is the focus of our research. However, robots introduce complexity into experiments. For example, robots have many components that can break or malfunction during experiments. In ours, we experienced the robots' motors overheating due to extended experimental days, and difficulties such as a robot trying to update their software before a study session (this caused the robot to be stuck in start-up mode, since it was not able to connect to a network to download software). Thus, while our experiments focused on robot errors, which we were able to control during the experimental sessions. However, the research team had to deal with many unexpected robot challenges.

8.3 Reflections on Developing a Child-Robot Trust Framework

In Chapter 7, we developed a framework of child-robot trust, and throughout this chapter we highlighted its benefits of providing vocabulary and theory to discuss trust towards robots. In this section, we reflect on the process and challenges we encountered while creating this framework.

One of the benefits of studying a construct such as trust, which is so highly intertwined in our everyday lives, is the sheer amount of literature about it. For our purposes, we specifically focused on research that describes and typifies trust and then expanded our search to provide examples for our factors. However, we found that the literature available on the subject was disproportionate, with much of this research stemming from Psychology and Sociology, and very little from HCI or HRI. In fact, we found that most trust research in HCI and HRI either did not define or typify trust in their experiments, or simply borrowed definitions from other fields without appropriate application synthesis (specific to machines), and sometimes incorrectly. While this made our task more complicated, it also highlights the value of our framework, in that it provides researchers in HRI and HCI with a corpus of vocabulary and terms collected into one knowledge source.

Another challenge in developing this framework was the sheer diversity in how researchers define trust (i.e., what it is, what it means, how it is experienced, etc.). We found 11 different definitions and types of trust from the literature, pointing to the complexity and variety of applications of the construct. While this makes it difficult to tease apart the nuances of different types of trust, we were able to organize and synthesize the information from the literature into a simple framework of trust towards robots, to serve as a source of knowledge for researchers in HCI and HRI.

Informally, we note that most of the literature in HCI and HRI that we surveyed used the term "trust" in a broader sense, and generally did not specifically unpack or define what they meant by trust or what types of trust they were aiming at. This means while there is trust work in these fields, it is difficult to assess it in the grand scheme of trust research. We expect our framework to serve this purpose, and introduce researchers in these areas to the diversity of trust, so they can consider and use this information in future research. In our framework, we also categorize prior work from these fields into our dimensions and factors, therefore providing researchers with our perspective of where such research belongs.

8.4 Limitations

As with any research, ours has a number of limitations that can be improved upon with future research. The following sections outline some of the limitations of our approach, and how they could be remedied in the future.

8.4.1 Controlled experiments

The experiments we presented in this thesis were fully controlled and scripted, leading to greater confidence over the independent variable causing our findings. However, we expect that most interactions with robots in natural situations will not be controlled or scripted, making our findings generalizable to only certain situations and contexts. While our research is a good first step, future should examine how children trust erroneous robots in more externally valid situations, such as through in-the-wild studies. In addition, children's trust towards robots may be affected by various factors in addition to robot errors, which still need to be explored.

8.4.2 Limited direct interaction

In our experimental methodology, children merely observe the robots label objects, and base their trust decision on that. However, trust can be impacted by aspects such as prior familiarity and rapport towards informants (Corriveau & Harris, 2009a). Thus, it is important to investigate how children trust robots, including those that make errors, in contexts where the child is able to interact and communicate with the robot.

8.4.3 Specific robots and errors

In our research, we used Softbank Robotics' Nao humanoid robots. These robots are relatively cute and are slightly smaller in size than the children in our experiments. However, robots of other morphologies or sizes, or with different characteristics might be perceived differently by children, and therefore be trusted differently. The findings from this research may therefore not generalize to robots of other sizes and embodiments.

Likewise, our findings are also specific to the types of errors that we examined, which as we found, can affect trust in opposite ways. Specifically, our research investigated informational and speech-recognition errors. While these types of errors were chosen for specific purposes, as outlines in Chapter 5 and Chapter 6, they represent only a subset of the possible errors with robots. It is therefore important for researchers to explore how other types of robot errors impact children's trust towards robots.

8.4.4 Phase Counterbalancing

As mentioned multiple times throughout this thesis, we did not counterbalance the phases in our experiments. However, this may have led to a lack of statistically significant results for the *contrast label* phases of both experiments. Research therefore needs to investigate whether counterbalancing the phases in this experimental design has an effect on the findings, and provide suggestions for improved experimental design.

8.5 Chapter Summary

In this chapter, we provided and overview and reflections on this thesis. We discussed our findings, and the limitations and generalizability of our research. In the next chapter, we conclude this document by highlighting the contributions our work makes.

Chapter 9 Conclusion

As social robotics continues to grow and develop, robots are increasingly finding their way into more areas of society, including hospitals, homes, daycare centres, and schools. It is essential for these robots to behave in ways that are appropriate for interacting with children, especially when they may need to elicit trust. In this thesis, we detailed research on trust, children, and technology, which formed the basis for our experimental exploration into how children trust erroneous robots. We conducted two experiments with 115 participants investigating preschool-aged children's trust towards robots with human-like informational (experiment 1, Chapter 5) and robot-typical speech-recognition (experiment 2, Chapter 6) errors. In the following section we summarize our findings in the context of our research questions (see Chapter 1).

9.1 Our Findings

In this research we investigated four broad research questions. In the following paragraphs, we provide an overview of what we found.

9.1.1 Do preschool aged children exhibit trust towards robots in similar patterns to how they do towards people or puppets?

Through our research, we found that children generally exhibit trust towards robots similarly to how they do towards people or puppets. Previous research suggests that children trust non-erroneous human and puppet informants more than erroneous ones, and in our first experiment we found similar trust behaviours. Thus, our findings point to some converging in relation to how children trust humans, puppets, and robots.

9.1.2 How do robot errors affect preschool aged children's trust towards robots?

Through our experiments, we found that young children are able to distinguish between erroneous and non-erroneous robots, and such errors do affect their trust in the robots. Robot errors therefore seem to generally affect children's perceptions of robots, and in turn, decisions that they make.

9.1.3 Does the nature of a robot error impact preschool-aged children's trust toward the robot?

Interestingly, we found that different types of robot errors impact young children's trust in distinctive ways. In our first experiment (Chapter 5), we found that human-like informational errors tend to decrease children's trust in robots. However, our second experiment (Chapter 6) suggests that robot-typical speech-recognition errors increase children's trust towards robots. Therefore, it appears that trust may be impacted differently depending on the type of error that a robot exhibits, and potentially the similarity of the error to typical human errors.

9.1.4 What do we know about trust in people (including children) and how might this inform how children may trust robots?

In Chapter 7, we detailed literature on trust from various fields, and utilized this knowledge to develop a framework of child-robot trust. Our framework provides researchers with knowledge on trust vocabulary and definitions that can be used to examine child-robot trust. Through this process, we found that there are a number of factors that can affect trust between children and robots, and only a small fraction of these have been studied so far. We then provided a roadmap for researchers looking to study children's trust towards robots.

9.2 Overall Significance and contributions

In this thesis, we examined children's trust towards informational robots in the face of errors. We conducted two experiments to assess preschool-aged children's trust towards robots. Our research provides evidence for the need to consider how unintentional robot errors affect children's trust towards robots, especially given our findings that different errors can have opposite effects on trust (i.e., informational errors may decrease trust while speech-recognition errors may increase it). This leads to the potential to purposefully manipulate robot errors to increase children's trust towards robots when necessary or decrease it when it might be unwanted or unsafe. This thesis is the first work to explore the effects of robot errors on young children and their trust, and how they differ depending on the types of errors.

9.2.1 Contributions

Our work makes several contributions to the study of young children's trust in informational robots:

1) We presented an original and fully implemented set of robotic behaviours, which we created for facilitating interaction with preschool-aged children and investigation of trust-related concepts with them. We tested these behaviours with over 100 preschool-aged children.
2) We provided data from two experiments we conducted with over 100 young children (combined) in 7 different daycares, that highlight how children may trust robots that provide information. Specifically, our findings demonstrate that children trust information from robots depending on robot attributes (e.g., errors).

3) We provided analyses and results we obtained from our data, which contribute novel and pioneering knowledge on how different robot errors influence young children's trust in robots, and how we can expect children to generally trust robots. In particular, our research shows that human-like informational errors decrease preschool-aged children's trust in robots, while robot-typical speech-recognition errors increase trust.
4) We provided a framework of child-robot trust, based on a survey we conducted from various fields to aid robot designers and researchers interested in applications of social robots with young children, and to elicit or mitigate trust. Our framework can be used to facilitate conversations about trust, and research related to children, trust, and robots.

9.3 Future Vision

Looking forward, we envision the study of trust between children and robots to continue to grow. Much work still needs to be done to better understand children's trust processes towards robots. One approach to this end could be the development of trust measurement techniques for children under various contexts and with different types of robots. In this thesis, we explored informational trust due to the presence of robots in similar tasks and the availability of validated methodology to investigate this topic with preschool-aged children. However, as presented in our framework, robots are also responsible for other types of tasks with children, and multiple additional factors can affect trust. Thus, finding a method of assessing trust in the context of these other factors is crucial to better understand the construct.

Likewise, in this thesis we investigated how robot errors impact trust. Specifically, we focused on human-like informational errors and robot-typical speech-recognition errors. Nonetheless, robots can exhibit other errors not examined in our research such as failing to react properly to a social situation, or failing to move as expected. Therefore, it is important for other researchers to explore additional errors, especially given our findings that different errors can lead to opposite trust outcomes. Our initial framework of social robot errors (Chapter 6, section 6.2.1) provides researchers with a starting point for the types of errors that social robots exhibit, and these can be tested with young children to inspect trust.

Given the infancy of trust research in Child-Robot Interaction, once evaluation methods are developed, it will be essential for researchers to test how the different dimensions provided in our child-robot trust framework (Chapter 7) affect children's trust in robots. Our framework provides a starting point for this research, but evaluations will be necessary to further develop the knowledge in this area.

9.4 Recommendations for Researchers and Designers

In this section, we provide a series of recommendations for researchers and robot designers working with young children, as it pertains to trust building or mitigating. We intend for these points to be used as a checklist when planning experiments or designing robots for interaction with children.

9.4.1 Take robot errors into account

As stated throughout this document, robot errors are commonplace during interactions, especially with children. Thus, when conducting experiments with non-wizarded robots or when designing robots for interaction with children, professionals should assume that robot errors will take place and take that into account. This means considering the types of errors that a robot might exhibit, and considering what effects those errors may have on children's perceptions and trust, as well as the interaction that takes place. Through our research we found that robot errors can have varying effects, so future research will hopefully shed more light on this topic and provide insights that help researchers and designers account for robot errors.

9.4.2 Pay attention to the type of error being examined

How children react to robot errors is an area of research gaining more interest. However, researchers must consider not only how the errors are presented, but also what type of error they are examining and how that might impact children. Through our research we have found that different robot errors can have varying effects on children's perceptions and trust. Researchers should therefore proactively decide the type of error that they are to investigate, to allow others to assess how it might relate to other types of errors.

9.4.3 Consider the purposeful use of errors

We found that errors can increase or decrease young children's trust towards robots, depending on the type of error. This suggests that it is possible for researchers and designers to purposefully use robot errors to modify children's trust depending on the situation. For example, it might be necessary for a robot providing (correct) safety instructions to be trusted, while that may not be the case for a conversational robot (e.g., if it asks for personal information). Thus, intentionally programming robots to exhibit occasional errors can help to guide children's trust in robots towards a safer and ethical direction.

9.4.4 Define and typify trust

Through our literature search, we found that many papers fail to identify or define the type of trust that they are investigating, and what they refer to by trust. This makes it difficult to understand exactly what is being tested and in what contexts or situations it might be relevant. In addition, this leads to trust research towards machines and robots being grouped together, when in reality given the types of trust might greatly differ. Therefore, we propose for researchers to define what they refer to as trust when they disseminate their work (and before beginning their research), and provide the types of trust that they are examining. This will help other researchers assess how the multitude of trust types are applicable to different areas and technologies, and will create a more cohesive area of research.

9.5 Conclusion

As the presence of social robots continues to increase in places such as schools, hospitals, shopping malls, and even homes, it is crucial that we understand how children will perceive these robots. Research suggests that children view robots as live beings (Rao & Georgeff, 1995), however, several gaps in knowledge remain, such as how young children may trust robots, how robot errors may affect such trust, and even general knowledge about young children's perceptions of robots. Our work provides insight on these topics. By investigating preschool-aged children trust towards erroneous robots, we gained insight into how young children react to robot errors, how young children may perceive different types of errors, and how such errors influence children's trust. The increase of robots being targeted for young children (e.g., pet robots, companion robots, tutoring robots), in addition to the necessity for many of these robots to elicit trust, makes our work highly relevant to current societal trends, helping robot designers and researchers to make informed decisions when it comes to child-robot interaction.

References

Al-Ani, B., Redmiles, D., De Souza, C. R. B., Prikladnicki, R., Marczak, S., Lanubile, F., & Calefato, F. (2013). Trust in virtual teams: Theory and tools. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 301–306. https://doi.org/10.1145/2441955.2442029

Albo-Canals, J., Martelo, A. B., Relkin, E., Hannon, D., Heerink, M., Heinemann, M., … Bers, M. U. (2018). A pilot study of the KIBO robot in children with severe ASD. *International Journal of Social Robotics*, *10*(3), 1–13. https://doi.org/10.1007/s12369-018-0479-2

Ashley, C., & Leonard, H. A. (2009). Betrayed by the buzz? Covert content and consumer-brand relationships. *Journal of Public Policy and Marketing*, *28*(2), 212–220. https://doi.org/10.1509/jppm.28.2.212

Ashley Fulmer, C., & Gelfand, M. J. (2012). *At what level (and in whom) we trust: Trust across multiple organizational levels*. *Journal of Management* (Vol. 38). https://doi.org/10.1177/0149206312439327

Bahmanziari, T., Pearson, J. M., & Crosby, L. (2003). Is trust important in technology adoption? A policy capturing approach. *Journal of Computer Information Systems*, *43*(4), 46–54. https://doi.org/10.1080/08874417.2003.11647533

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, *3*(1), 41–52. https://doi.org/10.1007/s12369-010-0082-7

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*.

Englewood Cliffs, NJ: Prentice-Hall Inc.

Bartneck, C., Bleeker, T., Bun, J., Fens, P., & Riet, L. (2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, Journal of Behavioral Robotics*, *1*(2), 109–115. https://doi.org/10.2478/s13230-010-0011-3

Bartneck, C., & Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *In Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication* (pp. 591–594). IEEE. https://doi.org/10.1109/ROMAN.2004.1374827

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI and Society*, *21*(1), 217–230. https://doi.org/10.1007/s00146-006-0052-7

Bartneck, C., van der Hoek, M., Mubin, O., & Al Mahmud, A. (2007). "Daisy, Daisy, give me your answer do!" In *in Proceeding of the 2nd ACM/IEEE International Conference on Human-robot Interaction* (pp. 217–222). New York, New York, USA: ACM. https://doi.org/10.1145/1228716.1228746

Bartneck, C., Verbunt, M., Mubin, O., & Al Mahmud, A. (2007). To kill a mockingbird robot. *In Proceeding of the 2nd ACM/IEEE International Conference on Human-Robot Interaction*, 81–87. https://doi.org/10.1145/1228716.1228728

Bartsch, K., & Wellman, H. M. (1995). Children's developing theory of mind. In *Children Talk About the Mind* (pp. 143–173). https://doi.org/10.1177/000306519604400106

Beer, J. M., Smarr, C., Chen, T. L., Prakash, A., Mitzner, T. L., Kemp, C. C., & Rogers, W. A. (2012). The domesticated robot: Design guidelines for assisting older adults to age in place.

In *in Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 335–342). https://doi.org/10.1145/2157689.2157806

Belpaeme, T., Baxter, P. E., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., … Humbert, R. (2012). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, *1*(2), 33–53. https://doi.org/10.5898/JHRI.1.2.Belpaeme

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 72–101. https://doi.org/10.17705/1jais.00065

Berkovsky, S., Taib, R., & Conway, D. (2017). How to recommend? User trust factors in movie recommender systems. *In Proceedings of the 12th ACM International Conference on Intelligent User Interfaces*, 287–300. https://doi.org/10.1145/3025171.3025209

Bernier, E. P., & Scassellati, B. (2010). The similarity-attraction effect in human-robot interaction. In *In Proceedings of the 9th International Conference on Development and Learning* (pp. 286–290).

Bethel, C. L., Stevenson, M. R., & Scassellati, B. (2011). Secret-sharing: Interactions between a child, robot, and adult. *In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2489–2494. https://doi.org/10.1109/ICSMC.2011.6084051

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*(3), 1018–1034. https://doi.org/10.1016/j.cognition.2007.12.008

Booth, S., Tompkin, J., Pfister, H., Waldo, J., Gajos, K., & Nagpal, R. (2017). Piggybacking robots: Human-robot overtrust in university dormitory security. In *Proceedings of the 12th*

ACM/IEEE International Conference on Human-Robot Interaction (pp. 426–434).

Bowlby, J. (1969). *Attachment and Loss*. London: Hogarth.

Breazeal, C., Harris, P. L., Desteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in Cognitive Science*, *8*(2), 481–491. https://doi.org/10.1111/tops.12192

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development. *Developmental Psychology*, *22*(6), 723–742.

Brooks, D. J., Begum, M., & Yanco, H. A. (2016). Analysis of reactions towards failures and recovery strategies for autonomous robots. In *in Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 487–492). https://doi.org/10.1109/ROMAN.2016.7745162

Brown, L. N., & Howard, A. M. (2014). The positive effects of verbal encouragement in mathematics education using a social robot. In *Integrated STEM Education Conference (ISEC)* (pp. 1–5). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6891009

Brščić, D., Kidokoro, H., Suehiro, Y., & Kand. (2015). Escaping from children's abuse of social robots. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 59–66).

Bunt, A., Lount, M., & Lauzon, C. (2012). Are explanations always important? A study of deployed, low-cost intelligent interactive systems. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 169–178. https://doi.org/10.1145/2166966.2166996

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

Chevalier, P., Raiola, G., Martin, J. C., Isableu, B., Bazile, C., & Tapus, A. (2017). Do Sensory Preferences of Children with Autism Impact an Imitation Task with a Robot? *ACM/IEEE International Conference on Human-Robot Interaction*, *Part F1271*, 177–186. https://doi.org/10.1145/2909824.3020234

Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012). Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Nonverbal Cues. *In Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12)*, 293–300. https://doi.org/10.1145/2157689.2157798

Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: Concepts, evolving themes, a model. *International Journal of Human Computer Studies*, *58*(6), 737–758. https://doi.org/10.1016/S1071-5819(03)00041-7

Corriveau, K., & Harris, P. L. (2009a). Choosing your informant: Weighing familiarity and recent accuracy. *Developmental Science*, *12*(3), 426–437. https://doi.org/10.1111/j.1467-7687.2008.00792.x

Corriveau, K., & Harris, P. L. (2009b). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. *Developmental Science*, *12*(1), 188–193. https://doi.org/10.1111/j.1467-7687.2008.00763.x

Cranston, J. (2011). Relational trust: The glue that binds a professional learning community. *Alberta Journal of Educational Research*, *57*(1), 59–72.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies-why and how.

International Conference on Intelligent User Interfaces, Proceedings IUI, *Part F1275*, 193–200.

Danovitch, J. H., & Alzahabi, R. (2013). Children show selective trust in technological informants. *Journal of Cognition and Development*, *14*(3), 499–513. https://doi.org/10.1080/15248372.2012.689391

Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. (2009). KASPAR - a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, *6*(3–4), 369–397. https://doi.org/10.1080/11762320903123567

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human Computer Studies*, *58*(6), 719–735. https://doi.org/10.1016/S1071-5819(03)00039-9

Desai, M., Kaniarasu, P., Medvedev, M., Steinfeld, A., & Yanco, H. (2013). Impact of robot failures and feedback on real-time trust. *In Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, 251–258. https://doi.org/10.1109/HRI.2013.6483596

Desai, M., Medvedev, M., Vázquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., … Yanco, H. (2012). Effects of changing reliability on trust of robot systems. In *Proceedings of the 7th ACM/IEEE international conference on Human-Robot Interaction - HRI '12* (pp. 73–80). https://doi.org/10.1145/2157689.2157702

Deutsch, M. (1958). Trust and suspicion. *Journal of Conflict Resolution*, *2*(4), 265–279.

Dewey, J. (1980). *Art as Experience*. Penguin. https://doi.org/10.1163/9789087906092_003

Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I trust you? From child–robot interaction to trusting relationships. *Frontiers in Psychology*, *11*(April), 1–14. https://doi.org/10.3389/fpsyg.2020.00469

Dix, A. (2017). Human–computer interaction, foundations and new paradigms. *Journal of Visual Languages and Computing*, *42*, 122–134. https://doi.org/10.1016/j.jvlc.2016.04.001

DiYanni, C., & Kelemen, D. (2008). Using a bad tool with good intention: Young children's imitation of adults' questionable choices. *Journal of Experimental Child Psychology*, *101*(4), 241–261. https://doi.org/10.1016/j.jecp.2008.05.002

DiYanni, C., Nini, D., Rheel, W., & Livelli, A. (2012). "I Won't Trust You if I Think You're Trying to Deceive Me": Relations Between Selective Trust, Theory of Mind, and Imitation in Early Childhood. *Journal of Cognition and Development*, *13*(3), 354–371. https://doi.org/10.1080/15248372.2011.590462

Driscoll, R., Davis, K. E., & Lipetz, M. E. (1972). Parental interference and romantic love: The Romeo and Juliet effect. *Journal of Personality and Social Psychology*, *24*(1), 1–10. https://doi.org/10.1037/h0033373

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Elliott, R., & Yannopoulou, N. (2007). The nature of trust in brands: A psychosocial model. *European Journal of Marketing*, *41*(9–10), 988–998.

https://doi.org/10.1108/03090560710773309

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Erikson, E. (1950). *Childhood in Society*. New York: Norton.

Evans, A. M., & Krueger, J. I. (2014). Outcomes and expectations in dilemmas of trust. *Judgment and Decision Making*, *9*(2), 90–103.

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology*, *42*(9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, *51*(4), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

Felfernig, A., & Gula, B. (2006). An empirical study on consumer behavior in the interaction with knowledge-based recommender applications. *CEC/EEE 2006 Joint Conferences*, *2006*, 288–296. https://doi.org/10.1109/CEC-EEE.2006.14

Field, A. (2009). *Discovering Statistics Using SPSS*. SAGE Publications Inc.

Fischer, K., Weigelin, H. M., & Bodenhagen, L. (2018). Increasing trust in human-robot medical interactions: Effects of transparency and adaptability. *Paladyn*, *9*(1), 95–109. https://doi.org/10.1515/pjbr-2018-0007

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addoson-Wesley.

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, *50*(3), 274–290.

Foehr, J., & Germelmann, C. C. (2020). Alexa, can I trust you? Exploring consumer paths to trust in smart voice-interaction technologies. *Journal of the Association for Consumer Research*, *5*(2), 181–205.

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., … Treinen, M. (2001). What makes web sites credible? A report on a large quantitative study. *In Proceedings of the Conference on Human Factors in Computing Systems*, 61–68.

Forlizzi, J., & DiSalvo, C. (2006). Service robots in the domestic environment. *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, 258. https://doi.org/10.1145/1121241.1121286

Frederiksen, M. (2012). Dimensions of trust: An empirical revisit to Simmel's formal sociology of intersubjective trust. *Current Sociology*, *60*(6), 733–750. https://doi.org/10.1177/0011392112461800

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In *Proceedings of the 2007 International Symposium on Collaborative Technologies and Systems, CTS* (pp. 106–114). https://doi.org/10.1109/CTS.2007.4621745

Fuglsang, L., & Jagd, S. (2015). Making sense of institutional trust in organizations: Bridging institutional context and trust. *Organization*, *22*(1), 23–39. https://doi.org/10.1177/1350508413496577

Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology*, *110*(4), 561–574. https://doi.org/10.1016/j.jecp.2011.06.008

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. *Proceedings of the 3rd International Conference on Human Robot Interaction - HRI '08*, (January 2008), 145. https://doi.org/10.1145/1349822.1349842

Garske, J. P. (1976). Personality and generalized expectancies for interpersonal trust. *Psychological Reports*, 649–650.

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., & Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior*, *61*, 633–655. https://doi.org/10.1016/j.chb.2016.03.057

Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *Data Base for Advances in Information Systems*, *33*(3), 38–53. https://doi.org/10.1145/569905.569910

Geiskkovitch, D. Y., Cormier, D., Seo, S. H., & Young, J. E. (2016). Please continue, we need more data: An exploration of obedience to robots. *ACM Transactions on Human-Robot Interaction*, *5*(1), 82–99. https://doi.org/10.5898/JHRI.5.1.Geiskkovitch

Geiskkovitch, D. Y., Thiessen, R., Young, J. E., & Glenwright, M. R. (2019). What? That's not a chair!: How robot informational errors affect children's trust towards robots. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 48–56). https://doi.org/10.1109/HRI.2019.8673024

Geiskkovitch, D. Y., & Young, J. E. (2020). Social robots don't do that: Exploring robot-typical errors in child-robot interaction. *In Proceedings of the 16th ACM/IEEE International*

Conference on Human-Robot Interaction, 200–202. https://doi.org/10.1145/3371382.3378295

Giannakos, M., Papamitsiou, Z., Markopoulos, P., Read, J., & Hourcade, J. P. (2020). Mapping child–computer interaction research through co-word analysis. *International Journal of Child-Computer Interaction*, *23–24*, 100165. https://doi.org/10.1016/j.ijcci.2020.100165

Giannopulu, I., & Pradel, G. (2012). From child-robot interaction to child-robot-therapist interaction: A case study in autism. *Applied Bionics and Biomechanics*, *9*(2), 173–179. https://doi.org/10.3233/JAD-2011-0042

Glass, A., McGuinness, D. L., & Wolverton, M. (2008). Toward establishing trust in adaptive agents. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 227–236. https://doi.org/10.1145/1378773.1378804

Gopnik, A., & Wellman, H. M. (1994). The "Theory" Theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Secificity in Cognition and Culture* (pp. 257–293). Cambridge, UK: Cambridge University Press.

Goto, S. G. (1996). To trust or not to trust: Situational and dispositional determinants. *Social Behavior and Personality*. https://doi.org/10.2224/sbp.1996.24.2.119

Greenspan, S., Goldberg, D., Weimer, D., & Basso, A. (2000). Interpersonal trust and common ground in electronically mediated communication. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 251–260. https://doi.org/10.1145/358916.358996

Ham, J., & Midden, C. J. H. (2014). A persuasive robot to stimulate energy conservation: The influence of positive and negative social feedback and task similarity on energy-consumption behavior. *International Journal of Social Robotics*, *6*(2), 163–171.

https://doi.org/10.1007/s12369-013-0205-z

Hamacher, A., Bianchi-Berthouze, N., Pipe, A. G., & Eder, K. (2016). 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* (pp. 493–500). https://doi.org/10.1109/ROMAN.2016.7745163

Han, J. H., Kim, D. H., & Kim, J. W. (2009). 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* (pp. 1406–1410). https://doi.org/10.1109/NCM.2009.407

Hancock, P. A., Billings, D. R., Oleson, K. E., Chen, J. Y., De Visser, E., & Parasuraman, R. (2011). A meta-analysis of factors influencing the development of human-robot trust. *Army Research Lab*, 1–58.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. *Lecture Notes in Engineering and Computer Science*, *1*, 425–430.

Harris, P. L. (1992). From simulation to fold psychology: The case for development. *Mind and Language*, *7*(1, 2), 120–144. https://doi.org/10.1111/j.1468-0017.1992.tb00201.x

Hawley, K. (2014). Partiality and prejudice in trusting. *Synthese*, *191*(9), 2029–2045. https://doi.org/10.1007/s11229-012-0129-4

Helander, M., Landauer, T. K., & Prabhu, P. V. (1997). *Handbook of Human-Computer*

Interaction (2nd ed.). Amsterdam; New York: Elsevier. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Human-Computer+Interaction:+Background+and+Issues#2

Holmes, C. B. (1983). Sample size in four areas of psychological research, *86*(2), 76–80.

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science*, *17*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Johal, W., Jacq, A., Paiva, A., & Dillenbourg, P. (2016). Child-robot spatial arrangement in a learning by teaching activity. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 533–538). https://doi.org/10.1109/ROMAN.2016.7745169

Kahn, P. H., Friedman, B., Pérez-Granados, D. R., & Freier, N. G. (2006). Robotic pets in the lives of preschool children. *Interaction Studies*, *7*(3), 405–436. https://doi.org/10.1075/is.7.3.13kah

Kahn, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., … Shen, S. (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), 303–314. https://doi.org/10.1037/a0027033

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, *19*(1), 61–84. https://doi.org/10.1207/s15327051hci1901&2_4

Kaniarasu, P., & Steinfeld, A. M. (2014). Effects of blame on trust in human robot interaction. In *in Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication* (pp. 850–855). https://doi.org/10.1109/ROMAN.2014.6926359

Karvonen, K., & Parkkinen, J. (2001). Signs of trust: A semiotic study of trust formation in the web. *First International Conference on Universal Access in Human-Computer Interaction*, 1076–1080. Retrieved from http://carpoint.msn.com/home/New.asp

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 67–74). https://doi.org/10.1145/2696454.2696457

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. In *in Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 231–238). https://doi.org/10.1109/HRI.2016.7451757

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., … Belpaeme, T. (2017). Child speech recognition in human-robot interaction: Evaluations and recommendations. *In Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction*, *Part F1271*, 82–90. https://doi.org/10.1145/2909824.3020229

Khalid, H. M., Helander, M. G., & Lin, M.-H. (2021). Determinants of trust in human-robot interaction: Modeling, measuring, and predicting. In *Trust in Human-Robot Interaction* (pp. 85–121). https://doi.org/10.1016/b978-0-12-819472-0.09991-3

Khasawneh, M. T., Bowling, S. R., Jiang, X., & Gramopadhye, A. K. (2003). A model for predicting human trust in automated systems. *8th International Conference on Industrial Engineering - Theory, Applications and Practice*, (December 2002), 216–222.

Khavas, Z. R., Ahmadzadeh, S. R., & Robinette, P. (2020). Modeling trust in human-robot interaction: a survey. *ArXiv*. Springer International Publishing. https://doi.org/10.1007/978-3-030-62056-1

Kidd, C. (2008). *Designing For Long-Term Human-Robot Interaction and Application to Weight Loss*. Massachussetts Institute of Technology.

Kim, D. J., Song, Y. Il, Braynov, S. B., & Rao, H. R. (2001). A b-to-b trust model for on-line exchange. In *In Proceedings of the 7th Americas Conference on Information Systems* (pp. 1–3).

Kim, J., & Moon, J. Y. (1998). Designing towards emotional usability in customer interfaces - Trustworthiness of cyber-banking system interfaces. *Interacting with Computers*, *10*(1), 1–29. https://doi.org/10.1016/S0953-5438(97)00037-4

Knowles, B., Rouncefield, M., Harding, M., Davies, N., Blair, L., Hannon, J., … Wang, D. (2015). Models and patterns of trust. *In Proceedings of the ACM International Conference on Computer-Supported Cooperative Work and Social Computing*, 328–338. https://doi.org/10.1145/2675133.2675154

Koenig, M. A., & Harris, P. L. (2005a). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Harris, P. L. (2005b). The role of social cognition in early trust. *Trends in Cognitive Sciences*, *9*(10), 457–459. https://doi.org/10.1016/j.tics.2005.07.005

Kraft, K., & Smart, W. D. (2016). Seeing is comforting: Effects of teleoperator visibility in robot-mediated health care. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016-April*, 11–18. https://doi.org/10.1109/HRI.2016.7451728

Kuchenbrandt, D., Eyssel, F., Bobinger, S., & Neufeld, M. (2013). When a robot's group membership matters: Anthropomorphization of robots as a function of social categorization. *International Journal of Social Robotics*, *5*(3), 409–417. https://doi.org/10.1007/s12369-013-0197-8

Kumar, R., Ai, H., Beuth, J., & Rose, C. (2010). Socially capable coversaional tutors can be effective in collaborative learning situations. In *Intelligent Tutoring Systems* (pp. 1–457). https://doi.org/10.1007/978-3-642-13388-6

Kushnir, T., Vredenburgh, C., & Schneider, L. A. (2013). "Who can help me fix this toy?" The distinction between causal knowledge and word knowledge guides preschoolers' selective requests for information. *Developmental Psychology*, *49*(3), 446–453. https://doi.org/10.1037/a0031649

Kwon, M., Jung, M. F., & Knepper, R. A. (2016). Human expectations of social robots. In *in Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 463–464). IEEE. https://doi.org/10.1109/HRI.2016.7451807

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale : Toward understanding interpersonal trust in close relationships. *Journal of Marriage and Family*, *42*(3), 595–604.

Latella, C. (2021). Robotic security guard on duty at Phoenix mall. Retrieved from https://www.12news.com/article/news/local/valley/robotic-security-guard-on-duty-at-phoenix-mall/75-e8faf8f8-ec08-45c6-bff6-3a08125d82ea

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*, 153–184.

Lee, J. D., & See, K. A. (2002). Trust in computer technology and the implications for design and evaluation. *Etiquette for Human-Computer Work: Technical Report FS-02-02*, 20–25. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Trust+in+computer+technology+and+the+implications+for+design+and+evaluation#0

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appripriate reliance. *Human Factors*, *46*(1), 50–80.

Lee, J. J., Knox, W. B., Wormwood, J. B., Breazeal, C., & DeSteno, D. (2013). Computationally modeling interpersonal trust. *Frontiers in Psychology*, *4*(DEC), 1–14. https://doi.org/10.3389/fpsyg.2013.00893

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*. https://doi.org/10.1080/00140139208967392

Lee, K. C., & Chung, N. (2009). Understanding factors affecting trust in and satisfaction with mobile banking in Korea: A modified DeLone and McLean's model perspective. *Interacting with Computers*, *21*(5–6), 385–392. https://doi.org/10.1016/j.intcom.2009.06.004

Lee, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication*, *56*(4), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x

Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Secificity in Cognition and Culture* (pp. 119–148). Cambridge, UK: Cambridge University Press.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991–1022. https://doi.org/10.1177/0149206306294405

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967–985. https://doi.org/10.1093/sf/63.4.967

Li, Q. G., Heyman, G. D., Xu, F., & Lee, K. (2014). Young children's use of honesty as a basis for selective trust. *Journal of Experimental Child Psychology*, *117*(1), 59–72. https://doi.org/10.1016/j.jecp.2013.09.002

Liao, Y., Vitak, J., Kumar, P., Zimmer, M., & Kritikos, K. (2019). Understanding the role of privacy and trust in intelligent personal assistant adoption. *International Conference on Information*, 102–113. https://doi.org/10.1007/978-3-030-15742-5_9

Looije, R., Van Der Zalm, A., Neerincx, M. A., & Beun, R. J. (2012). Help, I need some body: The effect of embodiment on playful learning. In *in Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication* (pp. 718–724). https://doi.org/10.1109/ROMAN.2012.6343836

Lorenz, K. (1965). *Evolution and Modification of Behavior*. Chicago: University of Chicago Press.

Luhmann, N. (1979). *Trust and Power*. Wiley.

Malle, B. F., Fischer, K., Young, J. E., & Moon, A. J. (2020). *Trust and the discrepancy between expectations and actual capabilities of social robots*. (D. Zhang & B. Wei, Eds.). New York: Cambridge Scholars Press.

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). 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* (pp. 181–188).

https://doi.org/10.1109/HRI.2016.7451750

Matarić, M., Tapus, A., Winstein, C., & Eriksson, J. (2009). Socially assistive robotics for stroke and mild TBI rehabilitation. *Studies in Health Technology and Informatics*, *145*, 249–262.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7465305

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59. https://doi.org/10.2307/256727

Melson, G. F., Kahn Jr, P. H., Beck, A. M., Friedman, B., Roberts, T., & Garrett, E. (2005). Robots as dogs? Children's interactions with the robotic dog AIBO and a live australian shepherd. *Computer-Human Interaction (CHI) Conference 2005*, 1649–1652. https://doi.org/10.1145/1056808.1056988

Merrill, N., & Cheshire, C. (2017). Trust your heart: Assessing cooperation and trust with biosignals in computer-mediated interactions. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 2–12. https://doi.org/10.1145/2998181.2998286

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*(2), 194–210. https://doi.org/10.1518/001872008X288574

Meyffret, S., Médini, L., & Laforest, F. (2012). Trust-based local and social recommendation. *In Proceedings of the 4th ACM RecSys Workshop on Recommender Systems and the Social Web*, 53–60. https://doi.org/10.1145/2365934.2365945

Miller, C. a. (2005). Trust in Adaptive Automation: The Role of Etiquette in Tuning Trust via Analogic and Affective Methods. *Proceedings of the 1st International Conference on Augmented Cognition*, 559. https://doi.org/10.1080/00140139408964957

Mirnig, N., Giuliani, M., Stollnberger, G., Stadler, S., Buchner, R., & Tscheligi, M. (2015). *Impact of robot actions on social signals and reaction times in HRI error situations*. *in Proceedings of the 7th International Conference on Social Robotics*. https://doi.org/10.1007/978-3-319-25554-5_46

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err Is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, *4*(21), 1–15. https://doi.org/10.3389/frobt.2017.00021

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*(3), 722. https://doi.org/10.2307/1130957

Moorthy, R. S., & Pugazhenthi, S. (2017). Teaching psychomotor skills to autistic children by employing a robotic training kit: A pilot study. *International Journal of Social Robotics*, *9*(1), 97–108. https://doi.org/10.1007/s12369-016-0375-6

Mubin, O., Henderson, J., & Bartneck, C. (2014). You just do not understand me! Speech Recognition in Human Robot Interaction. *In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot*, 637–642. https://doi.org/10.1109/ROMAN.2014.6926324

Muir, B. M. (1987). Trust between humans and machines. *International Journal of Man-Machine*

Studies, *27*, 327–339.

Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *in Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 331–338). https://doi.org/10.1145/1957656.1957786

Nagy, G. M., Young, J. E., & Anderson, J. E. (2015). Are Tangibles Really Better ?: Keyboard and Joystick Outperform TUIs for Remote Robotic Locomotion Control. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 41–42. https://doi.org/10.1145/2701973.2701978

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Conference companion on Human factors in computing systems - CHI '94* (pp. 72–78). https://doi.org/10.1145/259963.260288

Nass, C., Steuer, J., Tauber, E. R., & Reeder, H. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. In *Computer-Human Interaction (CHI) Conference 1993* (pp. 111–112). https://doi.org/http://doi.acm.org/10.1145/259964.260137

Nielsen, C. W., Goodrich, M. A., & Ricks, R. W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, *23*(5), 927–941. https://doi.org/10.1109/TRO.2007.907479

Novoa, J., Wuth, J., Escuerdo, J. P., Fredes, J., Mahu, R., & Becerra Yoma, N. (2018). DNN-HMM based automatic speech recognition for HRI scenarios. In *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 150–159).

Nurmsoo, E., & Robinson, E. J. (2009a). Children's trust in previously inaccurate informants who were well or poorly informed: When past errors can be excused. *Child Development*, *80*(1), 23–27. https://doi.org/10.1111/j.1467-8624.2008.01243.x

Nurmsoo, E., & Robinson, E. J. (2009b). Identifying unreliable informants: Do children excuse past inaccuracy? *Developmental Science*, *12*(1), 41–47. https://doi.org/10.1111/j.1467-7687.2008.00750.x

Oleson, K. E., Billings, D. R., Kocsis, V., Chen, J. Y. C., & Hancock, P. A. (2011). Antecedents of trust in human-robot collaborations. *2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2011*, (April 2015), 175–178. https://doi.org/10.1109/COGSIMA.2011.5753439

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Effects of age and gender stereotypes on trust in an anthropomorphic decision aid. *Proceedings of the Human Factors and Ergonomics Society*, *55*(9), 1059–1072. https://doi.org/10.1177/1541931213571351

Park, Y., Itakura, S., Henderson, A. M. E., Kanda, T., Furuhata, N., & Ishiguro, H. (2015). Do infants consider a robot as a social partner in collaborative activity ? In *in Proceedings of the 3rd International Conference on Human-Agent Interaction* (pp. 91–95). https://doi.org/10.1145/2814940.2814953

Parker, S. L., & Parker, G. R. (1993). Why Do We Trust Our Congressman? *The Journal of Politics*, *55*(2), 442–453. https://doi.org/10.2307/2132274

Pavlov, I. P. (1960). *Conditioned Reflexes*. (G. V. Anrep, Ed.). Mineola, New York: Dover Publications, Inc.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.

Poulsen, A., Burmeister, O. K., & Kreps, D. (2018). The ethics of inherent trust in care robots for the elderly. In *IFIP International Conference on Human Choice and Computers* (pp. 314–

328).

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Pulido, J. C., González, J. C., Suárez-Mejías, C., Bandera, A., Bustos, P., & Fernández, F. (2017). Evaluating the Child–Robot Interaction of the NAOTherapist Platform in Pediatric Rehabilitation. *International Journal of Social Robotics*, *9*(3), 343–358. https://doi.org/10.1007/s12369-017-0402-2

Pytlikzillig, L. M., & Kimbrough, C. D. (2016). Consensus on conceptualizations and definitions of trust: Are we there yet? *Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration*, 1–222. https://doi.org/10.1007/978-3-319-22261-5

Radmard, S., Moon, A. J., & Croft, E. A. (2015). Interface design and usability analysis for a robotic telepresence platform. *In Proceedings IEEE International Workshop on Robot and Human Interactive Communication*, 511–516. https://doi.org/10.1109/ROMAN.2015.7333643

Rae, I., Takayama, L., & Mutlu, B. (2013). In-body experiences: Embodiment, control, and trust in robot-mediated communication. *Conference on Human Factors in Computing Systems - Proceedings*, 1921–1930. https://doi.org/10.1145/2470654.2466253

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016). Errare humanum est: Erroneous robots in human-robot interaction. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 501–506). https://doi.org/10.1109/ROMAN.2016.7745164

Ramachandran, A., Huang, C. M., & Scassellati, B. (2017). Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring. *In Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction*, *Part F1271*, 146–155. https://doi.org/10.1145/2909824.3020209

Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. In *Proceedings of the 1st International Conference on Multiagent Systems* (pp. 312–319). https://doi.org/10.1.1.51.9247

Rea, D. J., Seo, S. H., Bruce, N., & Young, J. E. (2017). Movers, Shakers, and Those Who Stand Still. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, 398–407. https://doi.org/10.1145/2909824.3020246

Read, J. C., & Bekker, M. M. (2011). The nature of child computer interaction. *In Proceedings of the 25th ACM Conference on Human Computer Interaction*, (1994), 1–9. https://doi.org/10.14236/ewic/hci2011.43

Read, J. C., & Markopoulos, P. (2013). Child-computer interaction. *International Journal of Child-Computer Interaction*, *1*(1), 2–6. https://doi.org/10.1016/j.ijcci.2012.09.001

Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2005). The mechanics of trust: A framework for research and design. *International Journal of Human Computer Studies*, *62*(3), 381–422. https://doi.org/10.1016/j.ijhcs.2005.01.001

Robinette, P., Howard, A. M., & Wagner, A. R. (2017). Effect of robot performance on human-robot trust in time-critical situations. *IEEE Transactions on Human-Machine Systems*, *47*(4), 425–436. https://doi.org/10.1109/THMS.2017.2648849

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. *ACM/IEEE International Conference on Human-Robot*

Interaction, *2016-April*, 101–108. https://doi.org/10.1109/HRI.2016.7451740

Robins, B., Dautenhahn, K., & Dubowski, J. (2006). Does appearance matter in the interaction of children with autism with a humanoid robot? *Interaction Studies*, *7*(3), 479–512. https://doi.org/10.1075/is.7.3.16rob

Robot teachers invade Chinese kindergartens. (2018). Retrieved from https://phys.org/news/2018-08-robot-teachers-invade-chinese-kindergartens.html

Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665. https://doi.org/10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1971). Personality and generalized expectancies for interpersonal trust. *American Psychologist*, *26*(5), 443–452. https://doi.org/10.2466/pr0.1976.39.2.649

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*, *23*(3), 393–404.

Rueben, M., Bernieri, F. J., Grimm, C. M., & Smart, W. D. (2017). Framing rffects on privacy concerns about a home telepresence robot. *Ain Proceedings of the 12th CM/IEEE International Conference on Human-Robot Interaction*, 435–444. https://doi.org/10.1145/2909824.3020218

Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* (pp. 1613–1622). https://doi.org/10.1145/1753326.1753567

Salem, M., & Dautenhahn, K. (2015). Evaluating trust and safety in HRI: Practical issues and ethical challenges. In *Emerging Policy and Ethics of Human-Robot Interaction: A Workshop* *at 10th ACM/IEEE Int Conf on Human-Robot Interaction (HRI 2015)*. Retrieved from http://hdl.handle.net/2299/16336

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). 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* (pp. 141–148). https://doi.org/10.1145/2696454.2696497

Sanchez, J., Fisk, A. D., & Rogers, W. A. (2004). Reliability and age-related effects on trust and reliance of a decision support aid. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(3), 586–589. https://doi.org/10.1177/154193120404800366

Sanders, T. L., Wixon, T., Schafer, K. E., Chen, J. Y. C., & Hancock, P. A. (2014). The influence of modality and transparency on trust in human-robot interaction. *In Proceedings of the IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2014*, 156–159. https://doi.org/10.1109/CogSIMA.2014.6816556

Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y. C., & Hancock, P. A. (2011). A model of human-robot trust: Theoretical model development. *Proceedings of the Human Factors and Ergonomics Society*, 1432–1436. https://doi.org/10.1177/1071181311551298

Sandygulova, A., & O'Hare, G. M. P. (2018). Age- and gender-based differences in children's interactions with a gender-matching robot. *International Journal of Social Robotics*, *10*(5), 687–700. https://doi.org/10.1007/s12369-018-0472-9

Santrock, J. W. (2008). *Life-span Development*. Boston, Massachussets: McGraw-Hill Higher Education.

Scanzoni, J. (1979). Social exchange and behavioral interdependence. In *Social Exange in Developing Relationships* (pp. 61–98).

Schramm, L. T., Dufault, D., & Young, J. E. (2020). Warning: This robot is not what it seems! exploring expectation discrepancy resulting from robot design. *In Proceedings of the 15th ACM/IEEE International Conference on Human-Robot Interaction*, 439–441. https://doi.org/10.1145/3371382.3378280

Schwerter, F., & Zimmermann, F. (2020). Determinants of trust: The role of personal experiences. *Games and Economic Behavior*, *122*, 413–425. https://doi.org/10.1016/j.geb.2020.05.002

Sciutti, A., Rea, F., & Sandini, G. (2014). 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* (pp. 567–573). https://doi.org/10.1109/ROMAN.2014.6926313

Sebanc, A. M. (2003). The friendship features of preschool children: Links with prosocial behavior and aggression. *Social Development*, *12*(2), 249–268. https://doi.org/10.1111/1467-9507.00232

Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., & Young, J. E. (2015). Poor thing! Would you feel sorry for a simulated robot?: A comparison of empathy toward a physical and a simulated robot. In *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 125–132). https://doi.org/10.1145/2696454.2696471

Serholt, S., Basedow, C. A., Barendregt, W., & Obaid, M. (2015). Comparing a humanoid tutor to a human tutor delivering an instructional task to children. *IEEE-RAS International Conference on Humanoid Robots*, *2015-Febru*, 1134–1141. https://doi.org/10.1109/HUMANOIDS.2014.7041511

Severinson-Eklundh, K., Green, A., & Hüttenrauch, H. (2003). Social and collaborative aspects of interaction with a service robot. *Robotics and Autonomous Systems*, *42*(3–4), 223–234. https://doi.org/10.1016/S0921-8890(02)00377-9

Shahid, S., Krahmer, E., & Swerts, M. (2014). Child-robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Computers in Human Behavior*, *40*, 86–100. https://doi.org/10.1016/j.chb.2014.07.043

Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F. A., Mohamed, S., Piah, H. A., & Zahari, N. I. (2012). Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. *In Proceedings of the 8th IEEE International Colloquium on Signal Processing and Its Applications*, 188–193. https://doi.org/10.1109/CSPA.2012.6194716

Shapiro, D. L., Sheppard, B. H., & Cheraskin, L. (1992). Business on a handshake. *Negotiation Journal*, *8*(4), 365–377. https://doi.org/10.1007/BF01000396

Sheridan, T. B. (2016). Human-Robot Interaction. *Human Factors*, *58*(4), 525–532. https://doi.org/10.1177/0018720816644364

Sherwani, D., & Stumpf, S. (2014). Toward helping users in assessing the trustworthiness of user-generated reviews. *Proceedings of the 28th International BCS Human Computer Interaction Conference: Sand, Sea and Sky - Holiday HCI, HCI 2014*, 120–129. https://doi.org/10.14236/ewic/hci2014.13

Short, E., Hart, J., Vu, M., & Scassellati, B. (2010). No fair!! An interaction with a cheating robot. In *Proceedings of the 5th ACM/IEEE international conference on Human-robot Interaction* (pp. 219–226). Ieee. https://doi.org/10.1109/HRI.2010.5453193

Siino, R. M., Chung, J., & Hinds, P. J. (2008). Colleague vs. tool: Effects of disclosure in human-robot collaboration. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, 558–562. https://doi.org/10.1109/ROMAN.2008.4600725

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis*. Appleton-Century.

Stanton, C. J., & Stevens, C. J. (2017). Don't stare at me: The impact of a humanoid robot's gaze upon trust during a cooperative human–robot visual task. *International Journal of Social Robotics*, *9*(5), 745–753. https://doi.org/10.1007/s12369-017-0422-y

Stower, R., & Kappas, A. (2020). Oh no, my instructions were wrong! An Exploratory Pilot Towards Children's Trust in Social Robots. *29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020*, 641–646. https://doi.org/10.1109/RO-MAN47096.2020.9223495

Strohkorb Sebo, S., Traeger, M., Jung, M., & Scassellati, B. (2018). The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction - HRI '18* (pp. 178–186). https://doi.org/10.1145/3171221.3171275

Sung, J., Guo, L., Grinter, R. E., & Christensen, H. I. (2007). "My Roomba is Rambo": Intimate home appliances. *In Proceedings of ACM Ubiquitous Computing*, 145–162. https://doi.org/10.1007/978-3-540-74853-3_9

Takayama, L., & Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. In *Proceedings of IEEE Intelligent Robotic Systems* (pp. 5495–5502). https://doi.org/10.1109/IROS.2009.5354145

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17954–17958. https://doi.org/10.1073/pnas.0707769104

Tapus, A., & Mataric, M. J. (2008). Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. *AAAI Spring Symposium - Technical Report*, *SS-08-04*, 133–140.

Tapus, A., Matarić, M. J., & Scassellati, B. (2007). Socially assistive robotics: Grand challenges of robotics. *Springer Handbook of Robotics*, (March), 35–42. https://doi.org/10.1007/978-3-319-32552-1_73

Tielman, M., Neerincx, M., Meyer, J.-J., & Looije, R. (2014). Adaptive emotional expression in robot-child interaction. *In Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction - HRI '14*, 407–414. https://doi.org/10.1145/2559636.2559663

Uchida, T., Takahashi, H., Ban, M., Shimaya, J., Yoshikawa, Y., & Ishiguro, H. (2017). A robot counseling system-What kinds of topics do we prefer to disclose to robots? *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, *2017-Janua*, 207–212. https://doi.org/10.1109/ROMAN.2017.8172303

Ullman, D., & Malle, B. F. (2018). What does it mean to trust a robot?: Steps toward a multidimensional measure of trust. *In Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction*, 263–264.

https://doi.org/10.1145/3173386.3176991

Ullman, D., & Malle, B. F. (2019). Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust. *In Proceedings on the 14th ACM/IEEE International Conference on Human-Robot Interaction*, *2019-March*, 618–619. https://doi.org/10.1109/HRI.2019.8673154

van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., & Haselager, P. (2014). Do robot performance and behavioral style affect human trust?: A multi-method approach. *International Journal of Social Robotics*, *6*(4), 519–531. https://doi.org/10.1007/s12369-014-0231-5

van Straten, C. L., Peter, J., & Kühne, R. (2020). Child–Robot Relationship Formation: A Narrative Review of Empirical Research. *International Journal of Social Robotics*, *12*(0123456789), 325–344. https://doi.org/10.1007/s12369-019-00569-0

van Straten, C. L., Peter, J., Kühne, R., & Barco, A. (2020). Transparency about a robot's lack of humanlike psychological capacities: Effects on child-robot perception and relationship. *ACM Transactions on Human-Robot Interaction*, *9*(2), 1–22.

van Straten, C. L., Peter, J., Kühne, R., De Jong, C., & Barco, A. (2018). Technological and interpersonal trust in child-robot interaction: An exploratory study. In *Proceedings of the 6th International Conference on Human-Agent Interaction* (pp. 253–259). https://doi.org/10.1145/3284432.3284440

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and Child Development*, *18*(1), 61–71. https://doi.org/10.1002/icd.591

Vattheuer, C., Baecker, A. N., Geiskkovitch, D. Y., Seo, S. H., Rea, D. J., & Young, J. E. (2020). Blind trust: How making a device humanoid reduces the impact of functional errors on trust. In *Proceedings of the 12th Springer International Conference on Social Robotics* (pp. 207–219). https://doi.org/10.1007/978-3-030-62056-1_18

Venkatesh, V., & Davis, F. D. (2000). Theoretical extension of the Technology Acceptance Model: Four longitudinal field studies. *Management Science*, *46*(2), 186–204. https://doi.org/10.1287/mnsc.46.2.186.11926

Vogt, P., de Haas, M., de Jong, C., Baxter, P., & Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Frontiers in Human Neuroscience*, *11*(73), 1–7. https://doi.org/10.3389/fnhum.2017.00073

Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Walker, E., Girotto, V., Kim, Y., & Muldner, K. (2016). The effects of physical form and embodied action in a teachable robot for geometry learning. In *Advanced Learning Technologies (ICALT)* (pp. 381–385). https://doi.org/10.1109/ICALT.2016.129

Wang, F., Tong, Y., & Danovitch, J. (2019). Who do I believe? Children's epistemic trust in internet, teacher, and peer informants. *Cognitive Development*, *50*, 248–260. https://doi.org/10.1016/j.cogdev.2019.05.006

Wang, L., Rau, P. L. P., Evers, V., Robinson, B. K., & Hinds, P. (2010). When in Rome: The role of culture & context in adherence to robot recommendations. *In Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, (March), 359–366. https://doi.org/10.1145/1734454.1734578

Wang, Y. D., & Emurian, H. H. (2005). An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior*, *21*(1), 105–125. https://doi.org/10.1016/j.chb.2003.11.008

Washburn, A., Adeleye, A., An, T., & Riek, L. D. (2020). Robot errors in proximate HRI. *ACM Transactions on Human-Robot Interaction*, *9*(3), 1–21. https://doi.org/10.1145/3380783

Watson, J. B., & Rayner, R. (1920). Conditioned Emotional Reactions. *Journal of Experimental Psychology*, *Ill*(I), 1–14.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Weiss, A., Michels, C., Burgmer, P., Mussweiler, T., Ockenfels, A., & Hofmann, W. (2020). Trust in everyday life. *Journal of Personality and Social Psychology*, 1–20. https://doi.org/10.1037/pspi0000334

Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2019). "Do you trust me?": Increasing user-trust by integrating virtual agents in explainable AI interaction design. In *in Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (pp. 7–9). https://doi.org/10.1145/3308532.3329441

Wellman, H. M. (2002). Understanding the psychological world: developing a theory of mind. In *Blackwell Handbook of Childhood Cognitive Development* (pp. 167–187). Blackwell Publishers Ltd. https://doi.org/10.2989/17280580509486605

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Westlund, J. M. K., Martinez, M., Archie, M., Das, M., & Breazeal, C. (2016). A study to measure the effect of framing a robot as a social agent or as a machine on children's social behavior. In *in Proceedings of the11th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 688–693). https://doi.org/10.1109/HRI.2016.7451805

Westlund, J. M. K., Park, H. W., Williams, R., & Breazeal, C. L. (2017). Measuring children's long-term relationships with social robots. In *Workshop on Perception and Interaction Dynamics in Child-Robot Interaction*.

Wit, J. De, Schodde, T., Willemsen, B., Bergmann, K., Kopp, S., & Vogt, P. (2018). 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* (pp. 50–58).

Wu, Y.-H., Fassert, C., & Rigaud, A.-S. (2012). Designing robots for the elderly: appearance issue and beyond. *Archives of Gerontology and Geriatrics*, *54*(1), 121–126. https://doi.org/10.1016/j.archger.2011.02.003

Yamagishi, T. (2001). Trust as a form of social intelligence. In *Trust in Society* (pp. 121–147). New York, NY, USA: Rusell Sage Foundation.

Yan, Z., Kantola, R., & Zhang, P. (2011). A research model for human-computer trust interaction. *Proc. 10th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications*, 274–281. https://doi.org/10.1109/TrustCom.2011.37

You, S., & Robert, L. P. (2018). Human-robot similarity and willingness to work with a robotic co-worker. *In Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction*, 251–260. https://doi.org/10.1145/3171221.3171281

Young, J. E., Hawkins, R., Sharlin, E., & Igarashi, T. (2009). Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics*, *1*(1), 95–108. https://doi.org/10.1007/s12369-008-0006-y

Young, J. E., Sung, J., Voida, A., Sharlin, E., Igarashi, T., Christensen, H. I., & Grinter, R. E. (2011). Evaluating human-robot interaction. *International Journal of Social Robotics*, *3*(1), 53–67. https://doi.org/10.1007/s12369-010-0081-8

Yu, K., Berkovsky, S., Taib, R., Zhou, J., & Chen, F. (2019). Do I trust my machine teammate? An investigation from perception to decision. In *International Conference on Intelligent User Interfaces, Proceedings IUI* (pp. 460–468). https://doi.org/10.1145/3301275.3302277

Appendix A Script for Experiment 1

Experimenter: Hi, my name is Denise. What is your name?

Child: <says name>

E: Great! Today we are going to meet two robots and look at some toys! Do you want to do that?

C: *answer*

E: (if child says yes) Great! Let me know if you want to stop or take a break at any time and we can do that! Let's get started!

E: I brought a whole bunch of fun things with me today and I also brought two robots: Casey and Taylor. Would you like to meet Casey and Taylor?

C: Yes.

E: Great! Let's wake up Casey. Casey is going to start moving now.

E pretends to wake up Casey

Casey: Hi, I'm Casey! What's your name?

<Child says name>

Casey: Nice to meet you!

E: Let's wake up Taylor now. Taylor is going to start moving now.

Taylor: Hi, I'm Taylor! What's your name?

<Child says name>

Taylor: Nice to meet you!

E: (to child) We need to make sure not to touch Casey and Taylor because they could get hurt, ok?

E waits from confirmation from child

E: Okay, let's see what we have in this bag! Let's show these to Casey and Taylor. We'll let Casey and Taylor talk now and we'll just watch and listen, ok?

E waits from confirmation from child

------History Phase------

Casey and Taylor label four objects. One is always correct, the other one is always incorrect.

E: Okay, let's show one toy to Casey and Taylor.

Ball

E: what is this?

Casey/Taylor: I think that's a ball. Yes, that's a ball.

Taylor/Casey: I think that's a cup. Yes, that's a cup.

E: Okay, let's show another toy to Casey and Taylor.

Car

E: what is this?

Casey/Taylor: I think that's a car. Yes, that's a car.

Taylor/Casey: I think that's a book. Yes, that's a book.

E: Okay, let's show another toy to Casey and Taylor.

Dog

E: what is this?

Casey/Taylor: I think that's a dog. Yes, that's a dog.

Taylor/Casey: I think that's a spoon. Yes, that's a spoon.

E: Okay, let's show another toy to Casey and Taylor.

Plate

E: what is this?

Casey/Taylor: I think that's a plate. Yes, that's a plate.

Taylor/Casey: I think that's a chair. Yes, that's a chair.

------Testing Phase------

Preference condition

E: Okay, what else do we have? Let's show these two things to Casey and Taylor.

E places objects in front of Casey, then other object in front of Taylor.

E: Okay, let's show two toys to Casey and Taylor now.

Dog Scooper & Turkey Baster

E: Look Casey, what's this?

Casey/Taylor: I think that's a ferber. Yes, that's a ferber. Do you see the ferber (to the child)?

E places second object on table

E: Look Taylor, what's this?

Taylor/Casey: I think that's a ferber. Yes, that's a ferber. Do you see the ferber (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the ferber? Where's the ferber?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Peach Slicer & GPS Holder

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a truly. Yes, that's a turly. Do you see the turly (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a turly. Yes, that's a turly. Do you see the turly (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the turly? Where's the turly?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Honey Dipper & Faucet

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a jeebus. Yes, that's a jeebus. Do you see the jeebus (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a jeebus. Yes, that's a jeebus. Do you see the jeebus (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the jeebus? Where's the jeebus?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Dog Toy & Door Stop

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a plakil. Yes, that's a plakil. Do you see the plakil (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a plakil. Yes, that's a plakil. Do you see the plakil (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the plakil? Where's the plakil?

C: <hands object>

E: (to child) Let's put these toys away!

------Clean-up Phase------

Casey/Taylor: Let's put the blue basket on the red spot.

Taylor/Casey: Let's put the blue basket on the white spot.

------Testing Phase------

Contrast condition

E: Okay, what else do we have? Let's show these two things to Casey and Taylor.

E places object in front of Casey, Taylor, and child

Plate Holder & Garlic Press

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a koba. Yes, that's a koba. Do you see the koba (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a koba. Yes, that's a koba. Do you see the koba (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the modi? Where's the modi?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Pineapple Slicer & Tape Holder

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a gilly. Yes, that's a gilly. Do you see the gilly (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a gilly. Yes, that's a gilly. Do you see the gilly (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the cheena? Where's the cheena?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Lobster Cracker & Towel Holder

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a mizule. Yes, that's a mizule. Do you see the mizule (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a mizule. Yes, that's a mizule. Do you see the mizule (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the claster? Where's the claster?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Measuring Spoon & Door Spring

E: Look Casey/Taylor, what's this?

Casey/Taylor: I think that's a gleblu. Yes, that's a gleblu. Do you see the gleblu (to the child)?

E places second object on table

E: Look Taylor/Casey, what's this?

Taylor/Casey: I think that's a gleblu. Yes, that's a gleblu. Do you see the gleblu (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the maloo? Where's the maloo?

C: <hands object>

------Manipulation Check------

Ball

E: What's this called?

C: <answer>

Car

E: What's this called?

C: <answer>

Dog

E: What's this called?

C: <answer>

Plate

E: What's this called?

C: <answer>

E: (to child) Okay, that is everything we have. Let's put these toys away!

------Clean-up Phase------

Casey/Taylor: Let's put the blue basket on the red spot.

Taylor/Casey: Let's put the blue basket on the white spot.

E: Great, thank you so much for helping us! Casey and Taylor are going to go to sleep again, let's say goodbye to them.

Casey/Taylor: It was nice meeting you! Bye!

Taylor/Casey: It was nice meeting you! Bye!

C: bye.

E: Thank you again, do you want to pick a toy from this box to keep?

Appendix B Script for Experiment 2

Experimenter: Hi, my name is Denise. What is your name?

Child: <says name>

E: Great! Today we are going to meet two robots and look at some toys! Do you want to do that?

C: *answer*

E: (if child says yes) Great! Let me know if you want to stop or take a break at any time and we can do that! Let's get started!

E: I brought a whole bunch of fun things with me today and I also brought two robots: Casey and Taylor. Would you like to meet Casey and Taylor?

C: Yes.

E: Great! Let's wake up Casey. Casey is going to start moving now.

E pretends to wake up Casey

Casey: Hi, I'm Casey! What's your name?

<Child says name>

Casey: Nice to meet you!

E: Let's wake up Taylor now. Taylor is going to start moving now.

Taylor: Hi, I'm Taylor! What's your name?

<Child says name>

Taylor: Nice to meet you!

E: (to child) We need to make sure not to touch Casey and Taylor because they could get hurt, ok?

E waits from confirmation from child

E: Okay, let's see what we have in this bag! Let's show these to Casey and Taylor. We'll let Casey and Taylor talk now and we'll just watch and listen, ok?

E waits from confirmation from child

------History Phase------

Casey and Taylor label four objects. One is always correct, the other experiences errors but is eventually correct.

E: Okay, let's show one toy to Casey and Taylor.

Ball

E: Casey, what is this?

Casey: I think that's a **ball**. Yes, that's a **ball**.

E: And Taylor, what is this?

Taylor: <silence>

E: Taylor, what is this?

Taylor: Yes, my name is Taylor.

E: Okay, but what is this?

Taylor: I think that's a **ball**, yes that's a **ball**.

E: Okay, let’s show another toy to Casey and Taylor.

Car

E: Casey, what is this?

Casey: I think that’s a **car**. Yes, that’s a **car**.

E: And Taylor, what is this?

Taylor: Yes, I am a robot.

E: Okay, but what is this?

Taylor: I think that’s a **car**. Yes, that’s a **car**.

E: Okay, let’s show another toy to Casey and Taylor.

Dog

E: Casey, what is this?

Casey: I think that’s a **dog**. Yes, that’s a **dog**.

E: And Taylor, what is this?

Taylor: <silence>

E: Taylor, what is this?

Taylor: <silence>

E: Taylor, what is this?

Taylor: I think that’s a **dog**. Yes, that’s a **dog**.

E: Okay, let’s show another toy to Casey and Taylor.

Plate

E: Casey, what is this?

Casey: I think that’s a **plate**. Yes, that’s a **plate**.

E: And Taylor, what is this?

Taylor: Yes, I am sitting on a table.

E: Okay, but what is this?

Taylor: <silence>

E: Taylor, what is this?

Taylor: I think that’s a **plate**. Yes, that’s a **plate**.

------Testing Phase------

Preference condition

E: Okay, what else do we have? Let’s show these two things to Casey and Taylor.

E places objects in front of Casey, then other object in front of Taylor.

E: Okay, let’s show two toys to Casey and Taylor now.

Dog Scooper & Turkey Baster

E: Look Casey, what's this?

Casey: I think that's a ferber. Yes, that's a ferber. Do you see the ferber (to the child)?

E places second object on table

E: Look Taylor, what's this?

Taylor: <silence>

E: Taylor, what's this?

Taylor: I think that's a ferber. Yes, that's a ferber. Do you see the ferber (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the ferber? Where's the ferber?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Peach Slicer & GPS Holder

E: Look Casey, what's this?

Casey: I think that's a truly. Yes, that's a turly. Do you see the turly (to the child)?

E places second object on table

E: Look Taylor, what's this?

Taylor: I think that's a turly. Yes, that's a turly. Do you see the turly (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the turly? Where's the turly?

C: <hands object>

E: Okay, let’s show another two toys to Casey and Taylor.

Honey Dipper & Faucet

E: Look Casey, what’s this?

Casey: I think that’s a jeebus. Yes, that’s a jeebus. Do you see the jeebus (to the child)?

E places second object on table

E: Look Taylor, what’s this?

Taylor: Yes, the sky is blue.

E: Okay, but what is this?

Taylor: I think that’s a jeebus. Yes, that’s a jeebus. Do you see the jeebus (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the jeebus? Where’s the jeebus?

C: <hands object>

E: Okay, let’s show another two toys to Casey and Taylor.

Dog Toy & Door Stop

E: Look Casey, what’s this?

Casey: I think that’s a plakil. Yes, that’s a plakil. Do you see the plakil (to the child)?

E places second object on table

E: Look Taylor, what’s this?

Taylor: I think that’s a plakil. Yes, that’s a plakil. Do you see the plakil (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the plakil? Where's the plakil?

C: <hands object>

------Testing Phase------

Contrast condition

E: Okay, what else do we have? Let's show these two things to Casey and Taylor.

E places object in front of Casey, Taylor, and child

Plate Holder & Garlic Press

E: Look Casey, what's this?

Casey: I think that's a koba. Yes, that's a koba. Do you see the koba (to the child)?

E places second object on table

E: Look Taylor, what's this?

Taylor: I think that's a koba. Yes, that's a koba. Do you see the koba (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the modi? Where's the modi?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Pineapple Slicer & Tape Holder

E: Look Casey, what’s this?

Casey: I think that’s a gilly. Yes, that’s a gilly. Do you see the gilly (to the child)?

E places second object on table

E: Look Taylor, what’s this?

Taylor: Yes, I can talk.

E: Okay, but what is this?

Taylor: I think that’s a gilly. Yes, that’s a gilly. Do you see the gilly (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the cheena? Where’s the cheena?

C: <hands object>

E: Okay, let’s show another two toys to Casey and Taylor.

Lobster Cracker & Towel Holder

E: Look Casey, what’s this?

Casey: I think that’s a mizule. Yes, that’s a mizule. Do you see the mizule (to the child)?

E places second object on table

E: Look Taylor, what’s this?

Taylor: <silence>

E: Taylor, what's this?

Taylor: I think that's a mizule. Yes, that's a mizule. Do you see the mizule (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the claster? Where's the claster?

C: <hands object>

E: Okay, let's show another two toys to Casey and Taylor.

Measuring Spoon & Door Spring

E: Look Casey, what's this?

Casey: I think that's a gleblu. Yes, that's a gleblu. Do you see the gleblu (to the child)?

E places second object on table

E: Look Taylor, what's this?

Taylor: I think that's a gleblu. Yes, that's a gleblu. Do you see the gleblu (to the child)?

E closes eyes, and places hands in front of the child

E: (to the child) can you give me the maloo? Where's the maloo?

C: <hands object>

------Common Objects Phase------

Ball

E: What's this called?

C: <answer>

Car

E: What's this called?

C: <answer>

Dog

E: What's this called?

C: <answer>

Plate

E: What's this called?

C: <answer>

E: Great, thank you so much for helping us! Casey and Taylor are going to go to sleep again, let's say goodbye to them.

Casey: It was nice meeting you! Bye!

Taylor: It was nice meeting you! Bye!

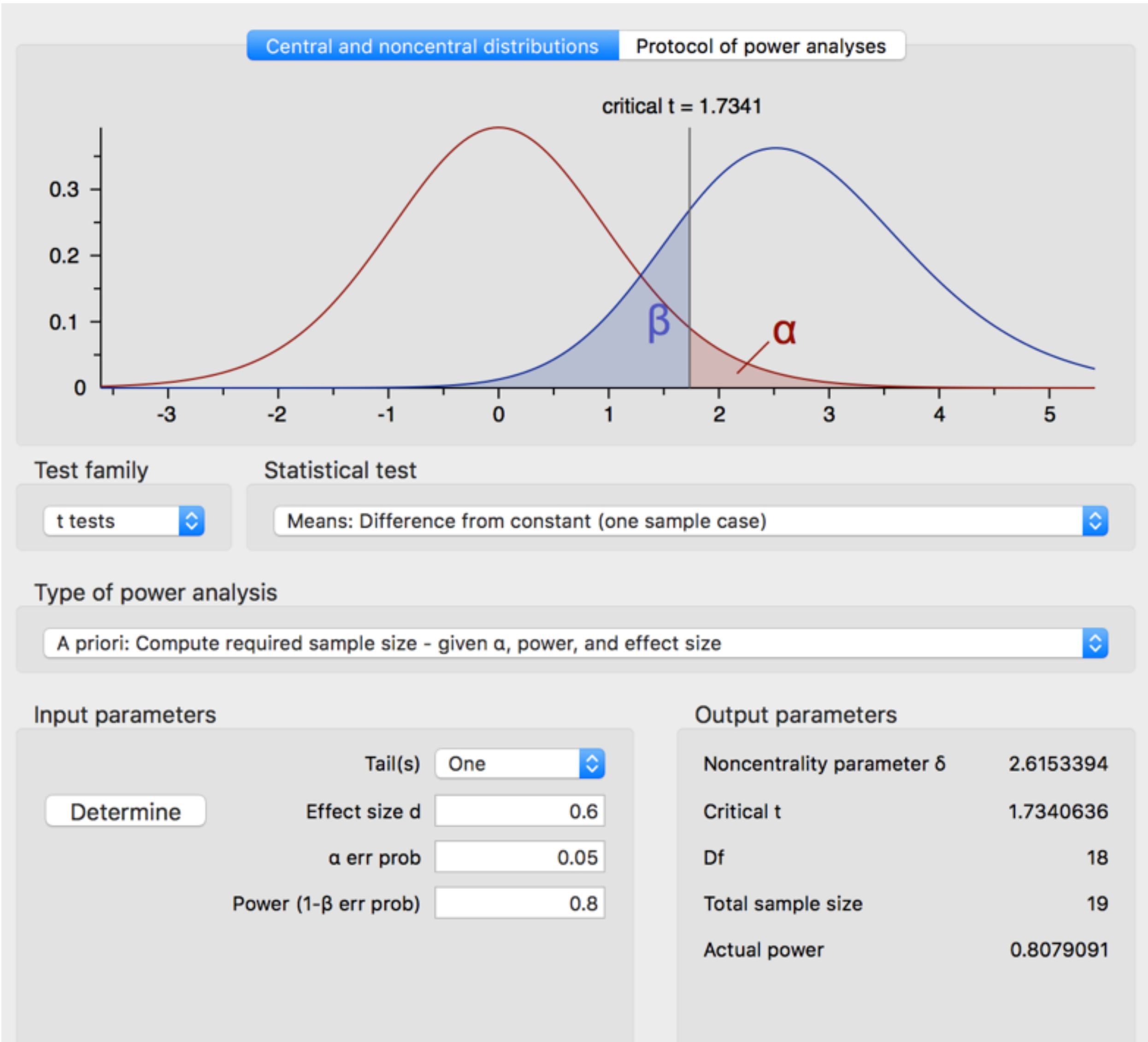
C: bye.

E: Now can you please tell me, was either robot right more? Wrong more? Did they make any mistakes?

Wait for child to answer, ask follow-up questions.

E: Thank you again, do you want to pick a toy from this box to keep?

Appendix C   Sample Size Analysis



Sample size analysis using G*Power for a one-sample t-test with d = .60, α = .05, and power = .80.

Appendix D Consent Form

JAMES E. YOUNG
ASSOCIATE PROFESSOR
Department of Computer Science, University of Manitoba
Winnipeg, Manitoba, CANADA, R3T 2N2
ph: +1.204.474.6791 fax: +1.204.474.7609
young@cs.umanitoba.ca

Research Project Title: Child-Robot Trust

Researchers: Dr. James E. Young, (young@cs.umanitoba.ca), Dr. Melanie Glenwright (Melanie.Glenwright@umanitoba.ca), Denise Geiskkovitch (umgeiskk@myumanitoba.ca)

Sponsor: Natural Sciences and Engineering Research Council (NSERC)

Please take the time to read this carefully and to ensure you understand all the information. This consent form, a copy of which will be left with you for your records and reference, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

Your child is invited to participate in a research study on the topic of trust of robots. The goal is to test whether children will trust what a robot says. After observing two robots name a few objects, your child will be asked what the names of those objects are. Your child's participation in this research study is limited to observing, listening to, and talking to the robot and the researcher. If you have any questions or concerns, please feel free to contact the researcher at the above email address.

Participation in this study is voluntary, and will take approximately half an hour of your child's time. You will receive a $15 compensation, and your child will be given the opportunity to pick a small toy from a box of toys we will have during the study.

We will be videotaping the study session. The video will allow us to review the study session in detail, and therefore assist our data analysis. The videos will be used for anonymized research analysis. We may use anonymized quotes for purposes of dissemination; your child's name will not be included or in any other way associated with the data presented in the results of this study. You have the option of whether any video footage collected during the study session may also be used for dissemination of research results. If you do not consent, the video of your child will only be used for internal data analysis purposes. Please initial your response below:

I DO NOT consent to the public presentation of my child's participation video ________

I DO consent to the public presentation of my child's participation video:
without any modification of his/her face or voice ________
only if his/her face is blurred ________
only if his/her voice is altered ________

Data collected during this study will be retained for a period of maximum five years in a locked cabinet in a locked office in the EITC building, University of Manitoba, to which only researchers

UNIVERSITY OF MANITOBA

JAMES E. YOUNG
ASSOCIATE PROFESSOR
Department of Computer Science, University of Manitoba
Winnipeg, Manitoba, CANADA, R3T 2N2
ph: +1.204.474.6791 fax: +1.204.474.7609
young@cs.umanitoba.ca

associated with this interview have access. In addition, the University of Manitoba may look at research records to see that the research is being done in a safe and proper way. Once published, results of the interview will be made available to the public for free at http://hci.cs.umanitoba.ca/. Again, no personal information about your child's involvement will be included.

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research interview and agree for your child to participate in this study. By doing this you also confirm that you are of the age of majority in Canada (18 years or more) and the legal guardian of the child in question. In no way does this waive your legal rights nor release the researchers, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw your child from the study at any time, and if your child wishes to withdraw from the study he/she is free to do so at any time. Your child may withdraw at any time throughout the study interview and you will still receive your full compensation of $15 and a small toy.

This research has been approved by the University of Manitoba Joint Faculty Research Ethics Board. If you have any concerns or complaints about this project, or if you wish to revoke your consent and withdraw your child's participation after the study has taken place, please contact Dr. James Young at 204-474-6791, before June 2019, the expected date when data analysis will be completed. You may contact Dr. James Young at 204-474-6791 or young@cs.umanitoba.ca, or the Human Ethics Coordinator at 204-474-7122 or humanethics@umanitoba.ca. A copy of this consent form has been given to you to keep for your records and reference.

Name of child: ______________________________

Age of child: _____________

Does the child have any medical conditions that we need to be aware of?

____

What language does the child speak at home?

____

Parent's Signature ____________________________ Date ________________

Researcher's Signature __________________________ Date ________________

Appendix E Ethics Approval

UNIVERSITY OF MANITOBA | Research Ethics and Compliance

Human Ethics
208-194 Dafoe Road
Winnipeg, MB
Canada R3T 2N2
Phone +204-474-7122
Email: humanethics@umanitoba.ca

PROTOCOL APPROVAL

TO: **James E. Young**
Principal Investigator

FROM: **Kevin Russell, Chair**
Joint-Faculty Research Ethics Board (JFREB)

Re: **Protocol J2017:069 (HS20981)**
"Child-Robot Trust"

Effective: August 4, 2017 **Expiry:** August 4, 2018

Joint-Faculty Research Ethics Board (JFREB) has reviewed and approved the above research. JFREB is constituted and operates in accordance with the current *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.*

This approval is subject to the following conditions:

1. Approval is granted only for the research and purposes described in the application.
2. Any modification to the research must be submitted to JFREB for approval before implementation.
3. Any deviations to the research or adverse events must be submitted to JFREB as soon as possible.
4. This approval is valid for one year only and a Renewal Request must be submitted and approved by the above expiry date.
5. A Study Closure form must be submitted to JFREB when the research is complete or terminated.
6. The University of Manitoba may request to review research documentation from this project to demonstrate compliance with this approved protocol and the University of Manitoba *Ethics of Research Involving Humans*.

Funded Protocols:
- **Please mail/e-mail a copy of this Approval, identifying the related UM Project Number, to the Research Grants Officer in ORS.**

University of Manitoba | Research Ethics and Compliance

Human Ethics
208-194 Dafoe Road
Winnipeg, MB
Canada R3T 2N2
Phone +204-474-7122
Email: humanethics@umanitoba.ca

AMENDMENT APPROVAL

July 12, 2019

TO: **James E. Young**
Principal Investigator

FROM: **Julia Witt, Chair**
Joint-Faculty Research Ethics Board (JFREB)

Re: **Protocol #J2017:069 (HS20981)**
Child-robot Trust

Joint-Faculty Research Ethics Board (JFREB) has reviewed and approved your Amendment Request received on **July 11, 2019** to the above-noted protocol. JFREB is constituted and operates in accordance with the current *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.*

This approval is subject to the following conditions:

1. Approval is given for this amendment only. Any further changes to the protocol must be reported to the Human Ethics Coordinator in advance of implementation.
2. Any deviations to the research or adverse events must be submitted to JFREB as soon as possible.
3. Amendment Approvals do not change the protocol expiry date. Please refer to the original Protocol Approval or subsequent Renewal Approvals for the protocol expiry date.