T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	TRUST IN ROBOTS: Studying the Effect of Social Robots ’ Behaviors on Improving Trust in Human-Robot Interactions
著者(和文)	ANZABI Naeimeh
Author(English)	Naeimeh Anzabi
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12712号, 授与年月日:2024年3月26日, 学位の種別:課程博士, 審査員:梅室 博行,顧 秀珠,SEABORN KATIE,福田 恵美子,松井 知己
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12712号, Conferred date:2024/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Powered by T2R2 (Science Tokyo Research Repository)

PhD Dissertation

TRUST IN ROBOTS

Studying the Effect of Social Robots' Behaviors on Improving Trust in Human-Robot Interactions

NAEIMEH ANZABI

Supervised by

Professor Hiroyuki Umemuro

東京工業大学
Tokyo Institute of Technology

TRUST IN ROBOTS:

STUDYING THE EFFECT OF SOCIAL ROBOTS' BEHAVIORS ON IMPROVING TRUST IN HUMAN-ROBOT INTERACTIONS

by

NAEIMEH ANZABI

Supervised by Professor Hiroyuki Umemuro

A dissertation submitted in fulfillment of the requirements
for the degree of Doctor of Engineering
in the Department of Industrial Engineering and Economics
Graduate School of Engineering
Tokyo Institute of Technology

February 2024

ABSTRACT

With the increased use of social robots in prominence and beyond functional performance, they are expected to foster trust and confidence in people. Various factors involve providing social robots with more trustworthy behavior. This dissertation aimed to investigate the effectiveness of social robots' different behaviors on perceived trust in human-robot interaction. Three behaviors were studied including; 1) listening behaviors, 2) benevolence and competence attributes, and 3) self-backstory disclosure. Three studies were conducted to inspect the effect of these behaviors on different types of general, cognitive and affective trust.

The first study investigated whether the listening behaviors of social robots could affect the perception of trustworthiness in human-robot interaction. The results indicated that "active empathic listening behavior" provided the participants with the highest impression of trustworthiness, compared to other forms of listening such as "active listening", specifically in affective trust. For nonverbal and verbal dimensions of listening behaviors, it was confirmed that nonverbal behaviors such as nodding, body movement, and eye gaze along with verbal behaviors, had a significant effect in eliciting higher affective trust in human-robot interaction. Moreover, the social robot with active empathic listening behavior was perceived as more alive, likable, and intelligent. Therefore, it could enhance the anthropomorphic attributes of social robots.

The second study investigated how the competence and benevolence characteristics of social robots affected the perception of trustworthiness and were integrated in human-robot trust relations. The results revealed the effectiveness of both benevolence and competence attributes in human-robot trust, as the social robot that behaved as competent-benevolent, competent-nonbenevolent, and benevolent-noncompetent was assessed to have higher general trust than that of the noncompetent-nonbenevolent robot. Furthermore, the results confirmed the primacy of benevolence in fostering affective trust, and modulating general and cognitive trust. However, the perceived competence of the social robot did not significantly influence cognitive trust.

The third study examined the influence of a social robot which disclosed its own backstories and experiences, on the development of trust in human-robot interaction. The

results indicated that the social robot disclosing a happy backstory provided the participants with higher impression of trustworthiness in general and affective trust compared to the social robot telling no backstory. However, the social robot disclosing sorrowful backstory was not evaluated to lead to higher trustworthiness than the social robot with no backstory. Furthermore, the social robot with happy backstory scored higher than the one with sorrowful backstory in all types of general, affective and cognitive trust.

Overall, the results indicated the importance of social robots' behaviors in human-robot trust, and confirmed that social robots' behaviors influence people's perceived trust in different ways. It was appeared that influential factors and behaviors in interpersonal trust could be comparable to human-robot trust, and participants treated social robots in a manner similar to human beings. Moreover, empathic and emotional behaviors of social robots played a significant role in enhancing trustworthiness of robots, specifically in terms of affective trust, which emphasizes the importance of developing emotional robots in future.

ACKNOWLEDGMENTS

All praise is rightfully directed to Allah, the Most Compassionate, the Most Merciful, the All-Knowing. He has consistently provided me with unwavering support and guidance throughout every moment of my life, offering profound kindness during challenging times. I am deeply grateful to Him and extend my heartfelt appreciation.

I would never have been able to finish my thesis without the guidance of my promoter and committee members, help from friends, and support from my family. It is not possible to thank everyone; however, I would like to specially thank the greatest helps.

First, I wish to extend my sincere gratitude to my supervisor, Prof. Hiroyuki Umemuro, for his continuous support, guidance, invaluable advice, and profound insights throughout my academic journey. His extensive knowledge and continuous motivation have been instrumental in my development and success. His patience and kind demeanor have enabled me to learn and grow during the course of my research, ultimately allowing me to successfully complete this journey.

My sincere thanks also go to Prof. Tomomi Matsui, Assoc. Prof. Katie Seaborn, Assoc. Prof. Emiko Fukuda, and Assoc. Prof. Xiuzhu Gu for their time and effort to provide constructive suggestions to improve the quality of this dissertation.

I would also like to thank my former and current fellow lab mates for their kindness, warmth, support and fabulous friendship experience. I also thank Ms. Matsubara for her great assistance. With a special mention to my friends in Japan, Mahsa Jamshidi, Ladan Mirchegini, Kosar Khajeh, Sarina Hosseini, Fereshte Azade, Helen Etemda, and many others who have not been mentioned.

A very special gratitude goes out to Dr. Ehsan Javanmardi (Assistant Professor at Tokyo University) for his invaluable help and unwavering assistance throughout my study. This project owes its completion to his generous support and guidance.

And finally, last but by no means least, I am grateful to my deer parents who were always supporting me and encouraging me with their best wishes along the way.

Thanks for all your encouragement and support!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF WRITING TERMS

Social robots: Type of robots designed to interact with human users in a natural, and interpersonal manner, and to operate in human environments alongside people.

Human-robot interaction (HRI): A research field to understand, design, and evaluate robot systems used by humans.

Human-robot trust (HRT): A measure of how much users rely on an automated system or robots to achieve a goal.

General trust: The attitude that an agent will help to achieve an individual's goals in a situation characterized by uncertainty and vulnerability

Affective trust: Type of trust based on emotions, feelings and moods

Cognitive trust: Type of trust based on rational decision-making, knowledge and good reasons

Active listening (AL): A form of carefully and attentively listening and responding verbally and nonverbally to the speaker, to achieve a deeper understanding of the message and context.

Active empathic listening (AEL): A form of listening in which conventional AL is combined with empathy to achieve a superior and more effective form of listening.

Benevolence: An attribute which a person shows positive intention toward others, without any bad intentions behind it.

Competence: An attribute which refers to someone's ability to provide accurate information, including characteristics such as intelligence, capability, expertise, knowledge, skill, and efficiency.

Self-backstory: A story is told about oneself including experiences, capabilities, limitations, and any other personal information that might be disclosed.

LIST OF ABBRIVATIONS

HRI: Human-Robot Interaction

HRT: Human-Robot trust

AL: Active Listening

AEL: Active Empathic Listening

NAL: nonactive-nonempathic listening

AEL$_{VO}$: Verbal-Only Active Empathic Listening

NB-NCc: NonBenevolent-NonCompetent condition

NB-Cc: NonBenevolent-Competent condition

NC-Bc: NonCompetent-Benevolent condition

B-Cc: Benevolent-Competent condition

H-Bc: Happy Backstory condition

S-Bc: Sorrowful Backstory condition

N-Bc: No Backstory condition

CHAPTER 1

INTRODUCTION

1.1 TRUST IN HUMAN-ROBOT INTERACTION

Robots emanate from the field of automation, and the definitions range from programmable automation to automated systems that acts on its own decisions with input from a human operator, and may have human characteristics such as complex actions and anthropomorphic features (Schaefer, 2013). Robots development and utilization is expanding from the traditional and passive tool-based robots to that of active social robots that interact and integrate with humans in various contexts (Chen & Terrence, 2009). Indeed, we see robots becoming social entities, designed to serve as our companions, aid us in various tasks, and even provide entertainment. These robots, commonly referred to as **social robots** are designed to interact with human users in a natural, and interpersonal manner, and they communicate and coordinate their behavior with humans through verbal, nonverbal, or affective modalities (Breazeal et al., 2016; Mahdi et al., 2022). Social robots represent a special technology, as they are purposefully designed to accompany and assist individuals in their daily lives, and become a ubiquitous aspect of modern society. Consequently, they are integrated in various domains as close to people as possible, and currently are becoming popular in several contexts, such as healthcare, education, service, entertainment, and research (see Mahdi et al., 2022).

Furthermore, Scholtz (2003) has suggested that the type of interaction with robots, beside other qualities such as physical nature of the robot, differentiate human-robot interaction from other human-computer or human-machine interactions. For human-robot cooperation to be successful, it must not only be physically safe, but also psychologically comfortable for the person and socially acceptable (David et al., 2022). Meanwhile, as the applications of robots evolve toward greater social roles, it becomes imperative to explore how humans perceive these social robots with whom they are meant to interact with. Moreover, if social robots are to be used in everyday environments, they will have to encounter and interact with people, and they must be capable of operating in chaotic and unpredictable environments, ensuring safety and efficiency while performing tasks

(Lum, 2020). Thus, the field of human-robot interaction **(HRI)** strives to enhance the quality of interactions between social robots and humans across multiple dimensions, encompassing the social, behavioral, emotional, cognitive, and psychological aspects. Furthermore, HRI has attempted to design social robots that are useful, intuitive, and user-friendly to interact and collaborate with humans (Beer et al., 2017), though HRI research is still relatively new in comparison to traditional robotics fields.

However, social robots are being introduced more and more into our daily lives, it does not necessarily guarantee that they are well accepted. Meanwhile, the willingness of people to accept robot-generated information and the robot's suggestions are strongly associated with the perceived trust of the robot, and people intervene the robots sooner if they are not trustworthy (Stanton & Stevens, 2017). Human's trust of robots will influence whether they are accepted to engage in social interactions with humans. It is also argued that trust mediates the relationship between people and technology, just as it mediates relationships between people (Lee & See, 2004). Trust plays a pivotal role in overcoming the cognitive complexity individuals encounter when interacting with sophisticated robot technologies. People tend to place their trust in social robots they find reliable, while often rejecting social robots they don't trust. Ultimately, trust in robots shape how much humans rely on robots in various domains. Therefore, trust is crucial in maintaining the utilization of robot systems by users and for fostering effective relationships with robots (Kim et al., 2020). In HRI, despite the obvious need to investigate the trust, the primary focus was placed on the technical design of robots and recently, the concern for this issue is increasing with advancements in robot functionality (Hancock et al., 2011; Schaefer, 2013).

Trust is a topic in human communication (Robbins, 2014), and the foundation of interpersonal cooperation. Many interpersonal relationships in marriage, friendship, and management, as well as important personality traits and survival of social groups, depend on the presence of trust (Simpson, 2007). The significance of trust has been a subject of inquiry in various contexts throughout the history (see Robbins, 2014). Some common areas of research on trust include trust in interpersonal relationships, trust in organizations and institutions, and trust in government. There are also many studies that explore the antecedents and consequences of trust, as well as how trust can be measured, developed, and maintained (see Hasnain, 2018). While there is a significant body of research on human-human trust, the study of human-robot trust is a relatively recent and emerging field, and the earliest empirical research did not occur until the latter half of the 1990s (Schaefer, 2013). Therefore, there remains a substantial opportunity for further investigation into the antecedents of trust in social robots. This is particularly pertinent when considering the behavioral aspects of social robots, that possess the capacity and probability of dissatisfaction and harmful usage of technology and social robots. Thus, the primary focus of my dissertation has been to research trust in human-robot interactions, with the aim of establishing and enhancing trustworthiness in humans' interactions with future social robots.

1.1.1 Two approaches in human-robot trust

Trust in technology, and specifically in HRI is a relatively new field of study with ongoing research. Indeed, trust in HRI was derived from human communication and interpersonal trust. However, there are several controversial debates as to whether or not interpersonal trust can be useful and applicable to technological and robotic domain. Muir (1987) stated that individuals’ trust in machines can be affected by similar factors in interpersonal trust; therefore, the model of trust between humans can be useful and valuable to designers of HRI in general. Certain researches have indicated that the social norms and social psychological principles governing human-human interaction may also apply to human-computer interaction, as users tend to respond to machines as independent entities (see Madhavan & Wiegmann, 2007). Sheridan (1975) argued that trust can mediate relations between people and automation, as it does between individuals. However, other researchers such as Lewandowsky et al. (2000) believe that individuals trust to technology different than people and it is unclear whether findings related to trust in human communication can be transferred and applied to automation and HRI field. Jian et al. (2000) described that people are convinced differently of human-automation interaction from human relations because the negative outcome of distrust in humans seems larger than non-human entities. Furthermore, Madhavan and Wiegmann (2007) revealed that while humans may react socially to machines, there are still subtle differences in the manner in which humans perceive automated aids vs human advisors.

Furthermore, human-robot trust **(HRT)** have been considered an extension of the factors affecting human-automation trust and share similar characteristics. However, Hancock et al. (2011) stated that robots differ from most of other automated systems in that they are interactive, mobile, designed to effect at a distance, and approximate human and animal forms, which suggests a different paradigm for HRT compared with other forms of automation. Unlike automated systems and computer interfaces used in industrial fields, social robots are supposed to have a characteristic in that it is designed by imitating the tasks or behaviours performed by humans. Moreover, robots are perceived to respond to situations self-governed and not anticipated in contrast to automation which is supposed to be pre-programmed (Natarajan & Gombolay, 2020). Therefore, the type of interactions with robots requires different issues of trust than the general domain of human-automation trust.

Thus, the domain of HRT is still demanding and it needs further research about the factors and behaviors causing interpersonal trust and human-automation trust, and examining those factors and behaviors in HRI applications across multiple dimensions (i.e., cognitive, affective, physical, and behavioral). Improving trustworthiness of social robots is necessary to design reliable social robots with beneficial roles and as contributors to human decisions.

1.2 TRUST DEFINITIONS

Trust has been conceptualized in various ways across different fields of study (see Robbins, 2014). According to Schaefer (2013) there are over 300 distinct definitions of trust that have been documented across different research domains, with 32 definitions specifically related to the domain of HRI. In the field of psychology, the definition by Rotter (1967) has been predominantly used, which defines trust as "*an expectancy held by an individual or a group that the word, promise, verbal, or written statement of another individual or group can be relied upon*" (p. 651). The primary definition used in the sociology is from Baeber (1983), who defines trust as two expectations: "(that) *of technically competent performance and of fiduciary obligation and responsibility"* (p. 165). Among the scholars, the definitions of Moorman et al. (1993) and Morgan and Hunt (1994) have been the most influential (Castaldo et al., 2010). The first authors conceived trust as a "*willingness to rely on the other party in whom one has confidence*" (Moorman et al., 1993). Morgan and Hunt (1994) defined trust as "*confidence"* in the other party's "*reliability"* and "*integrity*."

In automation and HRI, trust is defined in various ways (Table 1.1). In one of the most cited and thorough definitions, trust is defined as "*the attitude that an agent will help to achieve an individual's goals in a situation characterized by uncertainty and vulnerability*" (Lee & See, 2004). In this definition, an agent can be automation or another entity that actively interacts with the environment on behalf of the person. This definition addressed the importance of vulnerability similar to one of the most accepted and used definitions by Mayer, Davis and Schoorman (1995) in interpersonal trust, which defined trust as "*the willingness of a party to be vulnerable to the outcomes of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party*" (Mayer et al., 1995). This view suggested that vulnerability is significant for trust and without it, trust becomes unnecessary for the trustor because outcomes seem to be inconsequential. Lee and See's definition also suggested that uncertainty is critical to trust, because trust loses its necessity if the trustor can control the trustee's actions or has complete knowledge about them (Moorman et al., 1993). Furthermore, Sheridan (2002) who argued about the similarities of trust in automation and interpersonal trust, stated that "*trust is an effect or outcome of certain automation characteristics, and trust is a cause of operators' behaviour when utilizing automation*" (Madhavan & Wiegmann, 2007). This definition focused on the attributes of an automated system and considered trust as an outcome of its characteristics. This implied that certain characteristics of an automated system, such as robots, can significantly influence its trustworthiness. This definition closely aligned with the core objective of my dissertation, as we aimed to explore and understand the factors contributing to the perceived trustworthiness of social robots.

Overall, trust has been predominantly conceived in different ways in definitions: as a reliance, a belief, a willingness, an expectation, a confidence, and an attitude, and it is distinguished by specific characteristics such as benevolence, competencies, honesty and etc., which establish future actions in situations of consistent perceived risk and vulnerability (Castaldo et al., 2010). Moreover, there are two main approaches which consider trust as a belief or a behavior (action) (Reiersen, 2017). Belief-based approach that originates from Hardin (2001) in which trust and distrust are seen as beliefs about others' trustworthiness (Reiersen, 2017). Actions are not trust or mistrust, but only measurable indicators of these. Hence, trust is not a behavior but rather something that underlies a particular behavior and is characterized by what people think or know, not what they do. On the other side, behavior-based approach is characterized by what people do. Typically, when trust is regarded as a learned behavior, it is developed through a gradual process over time, evolving from past experiences or interactions rather than being an inherent trait (Hupcey et al., 2001). In this study, we mainly focused on evaluating participants' trust beliefs or perceptions of social robots, as participants' trust actions are diverse and change according to the context and the duration of interaction.

Table 1.1 Some of most cited and important definitions of trust in automation (Schaefer, 2013)

Trust definition	Definition citation
An expectation or mental attitude an agent maintains regarding its social environment.	Barber (1983)
Trust is a particular level of subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects his own action.	Gambetta (1988)
Trust as an effect or outcome of certain automation characteristics (e.g. reliability) and trust as a cause of operators' behaviour when utilizing automation.	Sheridan (2002)
System automation trust is defined as having confidence in and entrusting the system automation to do the appropriate action.	Biros, Daly, and Gunsch (2004)
The attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability.	Lee and See (2004)
The reliance by an agent that actions prejudicial to their well-being will not be undertaken by influential others.	Hancock, Billings, & Schaefer (2011)

1.3 TYPES OF TRUST

There are different categories and types of trust in human communication (e.g., Johnson-George & Swap, 1982; Rousseau et al., 1998; Yamagishi & Yamagishi, 1994). Each type of trust relates to a belief about a specific configuration of trust-warranting properties (Riegelsberger et al., 2005). Table 2.1 lists some of the most cited and important categories of trust. One of the most valuable, beneficial and the most frequently used categories is **affective** and **cognitive trust** established by Lewis and Weigert (1985), and extended by McAllister (1995). According to the classification of trust into cognitive and affective dimensions, trust can be based on rational decision-making or an emotional, affective foundation. In cognitive trust, "*we cognitively choose whom to trust in which respects and under which circumstances, and we base the choice on what we take to be 'good reasons', constituting evidence of trustworthiness*" (McAllister, 1995, p. 970).

The amount of knowledge necessary for trust is somewhere between total knowledge and total ignorance. Lewis and Weigert (1985) stated that "*if one were omniscient, actions could be undertaken with complete certainty, leaving no need, or even possibility, for trust to develop. On the other hand, in the case of absolute ignorance, there can be no reason to trust. When faced by the totally unknown, we can gamble but we cannot trust*" (p. 970). McAllister (1995) defined knowledge and good reasons as the basis for trust decisions. Cognitive trust is performance-based, and strongly relates to the trustee's expertise and competence. It results from accumulated knowledge, and warrants trusting the trustee with certain level of confidence. Affective trust, which is complementary to its cognitive type, consists of the emotional bonds between individuals in the relationship that reciprocally express care and concern. Affective trust relies on a partner's emotions and perceived similarity with the interaction partner, and it may go beyond the available knowledge. It is characterised by the perceived strength of the relationship, the sense of security felt in the relationship, a high level of emotional investment, mutual caring, and a willingness to be vulnerable with one another. Affective trust is more difficult to quantify and assess, as it is based on the feelings, emotions, and moods of the other. The partner usually acts with benevolence to elicit an emotional bond of trust. Regarding the relation of cognitive and affective trust, cognitive trust is considered one of the antecedents of affective trust, and emotions, on the other side, do influence the perception and cognitive evaluations, even after the dissipation of emotions (Bernotat et al., 2019; Duncan & Barrett, 2007; Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995; Zur et al., 2012). Therefore, the reciprocity of cognitive and affective trust is intertwined and they cannot be counted individually.

A number of studies have considered cognitive and affective trust in the context of HRI (e.g., Bernotat et al., 2019; Gompei & Umemuro, 2018). Affective trust can be established through various means, such as personalization of the robot's appearance, using natural language processing to enable more natural and engaging conversations, or by providing emotional feedback through facial expressions and body language. Cognitive trust, on the other side, can be developed by providing evidence of the robot's performance, reliability, safety, and by designing the robot to be transparent about its decision-making process. Gompei and Umemuro (2018) found that factors such as security, teammate, and performance related with cognitive trust, while factors such as teammate, performance, autonomy, and friendliness appeared to relate with affective trust. Cognitive trust refers mostly to people's confidence in robot's performance, whereas affective trust involves robot-related attributes, such as personality traits, robot's motives, benevolence, and user's attitudes and emotions towards HRI (Schaefer, 2013).

Table 2.1 Some of most cited and important types of trust (Hasnain, 2018)

Citation	Types of trust	Definition
Shapiro, Sheppard and Cheraskin (1992)	- Deterrence based trust - Knowledge based trust - Identification based trust	- related to fear of punishment - based on the prediction on the behaviour of another party - based on shared values, and empathy and considerations for each other
Lewicki and Bunker (1996)	- Calculus base trust - Knowledge based trust - Identification based trust	Same to previous classification
Kramer (1999)	- Dispositional trust - Category-based trust - Rule- based trust	- common belief about other people - based on information regarding a trustee's membership in a social or organisational category - based on a conscious calculation of consequences and shared understandings
Zucker (1986)	- Characteristic based trust - Process based trust - Institutional trust	- based on the person and their character, qualities, background. - develops on the past and expected interactions - attached to the formal social structure
Castaldo (2003)	- Target based trust - Content based trust - Strength based trust	- distinguishes the application of trust - is antecedent based, calculative, rational, and cognitive - based on the quality and consistency of trust
Lewis and Weigert (1985)	- Cognitive trust - Emotional trust - Behavioural trust	- based on good reason and knowledge - based on emotions - based on the expected actions of the parties
McAllister (1995)	- Affect based trust - Cognition based trust	- based on perceived emotions and affection - based on knowledge and competence

1.4 TRUST ANTECEDENTS

Several factors from cognitive to emotional, physical, and behavioral incorporate the building of trust. Rempel et al. (1985) introduced three basic factors, including predictability, dependability, and faith, that promote the growth of interpersonal trust. According to this model, the most concrete component of trust is based on the predictability or the stability of an individual's performance over a period of time. The second component of trust, dependability, is built upon a person's innate dispositional characteristics, which is appeared in the level of confidence one has in the trustee.

Finally, faith is based on beliefs about the future behavior or accuracy of an information source, which can be reflected in a person's willingness to trust the trustee again or to use a particular support in the future (Madhavan & Wiegmann, 2007). Similar to Rempel's model, Lee and See (2004) identified three general bases of trust in human-automation field which includes performance, process and purpose. Performance refers to the current and past operation of automation and includes dimensions such as reliability, predictability and ability. Process, which corresponds to Rempel et al.'s component of dependability, primarily describes how the automation operates and which algorithms of automation are appropriate for a situation. Finally, purpose refers to the extent to which automation operates based on the designer's intent and describes why the automation was designed. Purpose roughly corresponds to the component of faith and benevolence in the Rempel et al.'s model.

Mayer et al. (1995) concluded three bases of trust: ability, integrity, and benevolence. Ability is the group of skills, competencies, and characteristics that enable the trustee to influence the domain. Integrity explains if the trustee adheres to a set of accepted principles by trustor. Benevolence is the extent to which the intents and motivations of the trustee are in alliance with those of the trustor. These factors have been examined in different ways in human-automation and HRI domain (Chapter 3).

Finally[1], Hancock et al. (2011) provided a meta-analysis of factors affecting trust in HRI and concluded three main categories each consist of two subcategories that affect the process by which trust develops in HRI: 1) human-related factors (ability and human characteristics), 2) environmental factors (team collaboration and task-based factors), and 3) robot-related factors (performance-based, attribute-based factors). (Schaefer et al., 2016) revealed a revised three-factor model of Hancock et al.'s HRT model in which the robot-related factors were considered in terms of features and capability. They revealed that robot characteristics, and in particular, performance-based factors are predominant in perceived trust in HRI. Attribute-based factors covers various behaviors originating from human communications to human-machine interactions that affect trust (Sanders et al., 2011), and studies in this context are restricted, as the primary focus of

[1] Different trust antecedents in HRI have been discussed by other researchers. We only mentioned the main and the most cited ones. More researches can be found in related literature, e.g., (Nam & Lyon, 2021).

trust development in HRI has been on enhancing functional features. Finally, Khavas et al. (2020) updated the previous models and introduced a model with three categories: 1) robot-related factors, 2) human-related factors and 3) task and environment-related factors. The authors classified robot related factors under three subcategories including; robot-performance, robot-appearance, and robot-behaviors. The last category was the basis in this dissertation, with a primary main focus on robot behaviors.

1.4.1 Robot-related antecedents of HRT: Related work

As mentioned in the part 1.4, robot-related antecedents of trust have broadly categorized into performance-based and attribute-based factors. Performance-based factors refer to how the robot functions, while attribute-based factors associate with how the robot looks, feels and behaves (Kaplan et al., 2021). Robot-related antecedents, specifically those related to performance, have been extensively studied and found to be strongly related to the development of trust. there is a limited and insufficient body of research on factors related to robots' personality and behaviors which do paly a very significant role in establishing trust. Thus, my primary focus centered on robot attributes, encompassing three subcategories: appearance, behavior, and emotion.

1.4.1.1 Physical factors

Pak et al. (2012) found that increasing the physical anthropomorphic appearance of an automated system significantly impacted trust development through subjective assessment as well as objective/behavioral assessment. In HRT, Schaefer et al. (2016) examined that physical form of robots could impact on initial evaluation of users toward robot's trustworthiness. The physical presence of robots is another significant factor influencing trust and the overall quality of interaction between humans and robots. Embodied robots have been proven to have stronger influence on user performance and also improve the quality of interaction (e.g. Powers et al., 2007; Wainer et al., 2006). However, some studies showed contrary results and did not approve the relation of robot embodiment and trust (e.g. Herse et al., 2018; Natarajan & Gombolay, 2020).

Certain number of studies have explored the role of gender and appearance of robots in HRT; shape of robot, facial characteristics, gender similarity and gender stereotypes (e.g., Bernotat et al., 2019; Ghazali et al., 2018; Kraus et al., 2018). Furthermore, John and Catherine (2017) investigated how gaze can influence trust, likability, and compliance, and indicated that participants would be most likely to be persuaded by a robot with situational gaze which means only gazed during disagreements. Finally, You and Robert (2018) showed how facial similarity leads to increased attributions of trustworthiness in a robot and willingness to work with robots. Although, further studies have been conducted considering other influential factors like body posture or gestures of robots in HRT (see Khavas et al., 2020), researches in this area are still ongoing and demands further studies.

1.4.1.2 Behavioral factors

Behavioral and emotional factors are the most complicated and neglected factors that are currently developing with new advancements in robot’s intelligence and autonomy. People tend to consider the intention of a robot's behavior more when interacting with robots that resemble humans or exhibit more human-like characteristics. This makes behavioral factors especially important in trust calibration. Natarajan and Gombolay (2020) investigated how the type of robot, robot presence and different forms of behavior; apologetic (the agent apologizes after making a mistake), accountable (holds the human liable for the error), and indifferent (does not provide any feedback on whose side the error was committed), can impact overall trust. They found significant influence of robot’s behavior on trust, however the results did not prove statistical significance for robot presence on trust. Salem et al. (2015) studied whether participants follow instructions given by a faulty robot in a home environment. Similarly, the results by Robinette et al. (2016) showed that people over-trusted a robot in fire emergency evacuation scenarios, although the robot was shown to be defective in various ways.

Martelaro et al. (2016) focused on vulnerable behavior of robots and indicated that students had more trust and feelings of companionship with a vulnerable robot, and reported disclosing more with an expressive robot. Herse et al. (2018) explored whether robots showing preference elicitation feature could develop trust and the results revealed a positive and significant relation. Sebo et al. (2019) evaluated different trust repair behaviors by robot including apology or denial in repairing a human’s trust to a robot after the violation framing (competence or integrity). They concluded that participants interacting with a robot employing the integrity trust violation framing and the denial trust repair strategy are significantly more likely to exhibit behavioral retaliation toward the robot. Moreover, extroverted social robots are perceived more trustworthy than introvert ones (Kaplan et al., 2021). In a study, Lee et al. (2021) explored whether polite behavior of robot could provide more social and enhanced interaction experience for children. They manipulated different behaviors in the robot, and the results showed improvement in the children's evaluation of the robot for sociality, likability, perceived ease of use, trust and perceived usefulness.

Furthermore, the effects of robots' other behaviors and attributes on trust, such as cooperation and altruistic behaviors (Kumar et al., 2020), reciprocity (Sanders et al., 2011), apologizing (Xu & Howard, 2022), proximity (e.g., physical and physiological proximities) (Kory-Westlund & Breazeal, 2019), speech entrainment and self-disclosing (Kory-Westlund & Breazeal, 2019), turn-taking, and emotional expressions (van Pinxteren et al., 2019) have been examined in prior studies.

Finally, one specific aspect of both robot appearance and behavior is the degree of anthropomorphism in social robots. Anthropomorphism of the agent or the robot is perceived one significant factor in establishing trust between humans and robots, as human's perception of physical or virtual robots differ (see Powers et al., 2007). Anthropomorphism refers to the attribution of a human form, human characteristics, or human behavior to nonhuman things such as robots, computers, and animals (Bartneck et al., 2009). Prior work has shown that increased anthropomorphism of a robot or a software agent leads to a more positive interaction experience, and increased empathy (see Natarajan & Gombolay, 2020). Human-likeness of robots have a positive effect on trust and people trust a more humanoid robot over a less humanoid robot (Kim et al., 2020).

1.4.1.3 Emotive factors

New advancements in AI and robotic technologies support creating robots that display overt signs of empathy. In order to emulate empathy, a robotic system should be capable of recognizing and understanding the user's emotional state, behaving as if the others' emotions affect it, displaying emotion, mimicking emotions and conveying the ability of taking perspective. People experience emotions and express those emotions to communicate to others. Thus, scholars assume that the understanding and exchanges of emotional expressions between people and robots could reinforce the formation of friendly, effective and more trustworthy communication (Tapus et al., 2007).

Few studies exist in this domain and the opportunity remains for new ideas in future researches, as expressions and demonstrations of openness, empathy, and goodwill may increase trust towards robots. Generally, Leite et al. (2013) argued that the robots which behaved empathically and could recognize another's affect and respond appropriately, are more successful at establishing and maintaining a positive relationship with users and are perceived as friendlier. Mathur and Reichling (2016) stated that people tended to trust robots that showed more positive emotion. Likability and warmth had a positive effect on intentions to trust the robot and people tended to trust robots that had likeable features (Oksanen et al., 2020). Social robots exhibited empathy, kindness and a social personality are deemed more trustworthy (Kaplan et al., 2021). In a pilot study, Charrier et al. (2018) explored how an artificial empathy module could affect HRIs in different metrics, and according to the results participants in the empathic condition felt the robot to be more trustworthy. Tapus et al. (2007) also suggested that empathic language and physical expression enable more trustworthiness in robots. Martelaro et al. (2016) discussed that people tend to disclose about themselves, and feel companionship toward robots with expressive behaviors.

1.5 PURPOSE OF DISSERTATION

In various research fields, including HRI, a significant portion of trust-related studies hinges on evaluating the influential factors and determinants of trust. This exploration is essential for gaining a profound understanding of the concept of trust, knowing how to establish, enhance, and repair trust, as well as to prevent distrust. Understanding the antecedents of trust provides better prediction of when humans will accept and or reject new social robots. As it was discussed in previous sections, robot-related factors have been recognized as highly influential factors in HRT, with robot attributes standing to be a relatively recent and less-explored area within this domain. A number of studies aimed to explore trust-related robot attributes and characteristics in terms of appearance, behavior and emotions, as mentioned in section 1.4.1. Moreover, there are numerous influential behaviors in interpersonal trust that have not been studied and analyzed in the context of HRT. As discussed in section 1.1.1, many principles and behaviors related to trust in interpersonal relationships apply to HRT, as individuals tend to interact with social robots as if they were social entities.

Thus, the primary aim of this dissertation was **"to identify some of trust-related behaviors from the interpersonal domain, apply them to social robots and assess their impact on human's perceived trust of social robots in HRIs."** Figure 1.1 shows the current research position in relation to previous studies on HRT. There were three major components necessary to meet this end:

- determining the underlying theory of trust and influential factors as it applies to human-human relationships
- designing social robots' behaviors based on the identified factors in interpersonal trust as it coordinates to human-robot interactions
- evaluating the influence of social robots' behaviors on different types of trust

Social robots' behaviors may be specific in automation and robotics domain such as level of automation, or resemble human-like behaviors. The main scope of this dissertation was related to behaviors of social robots and their contribution to HRT. To accomplish this objective, we observed numerous behaviors in human communication which are crucial in building interpersonal trust. We identified several behaviors in interpersonal trust including; **"listening behavior", "benevolence and competence characteristics" and "self-disclosing personal stories and experiences"** as three candidates to be implemented and analyzed in social robots and HRT. We designed and implemented each behavior for a social robot and evaluated participants perceived trust of the social robot in terms of general, affective and cognitive trust. We provided a real interaction environment between participants and social robot which could warrant the feasibility of results in real social situations and outside of the experimental conditions.

In conducting this research, we expanded the scope of investigating robot-related factors influencing trust between humans and social robots. This exploration opens up new possibilities for designing human-like and novel behaviors for future social robots. While previous studies have explored numerous behaviors and attributes in HRT, there remain unexplored behaviors that can enhance the anthropomorphic features of social robots, making them more trustworthy. Enhancing the trustworthiness of social robots is crucial for their improved acceptance, operation, interaction, and engagement by humans. If we envision integrating social robots into our daily lives in the future, ensuring their trustworthiness becomes a necessary pursuit.

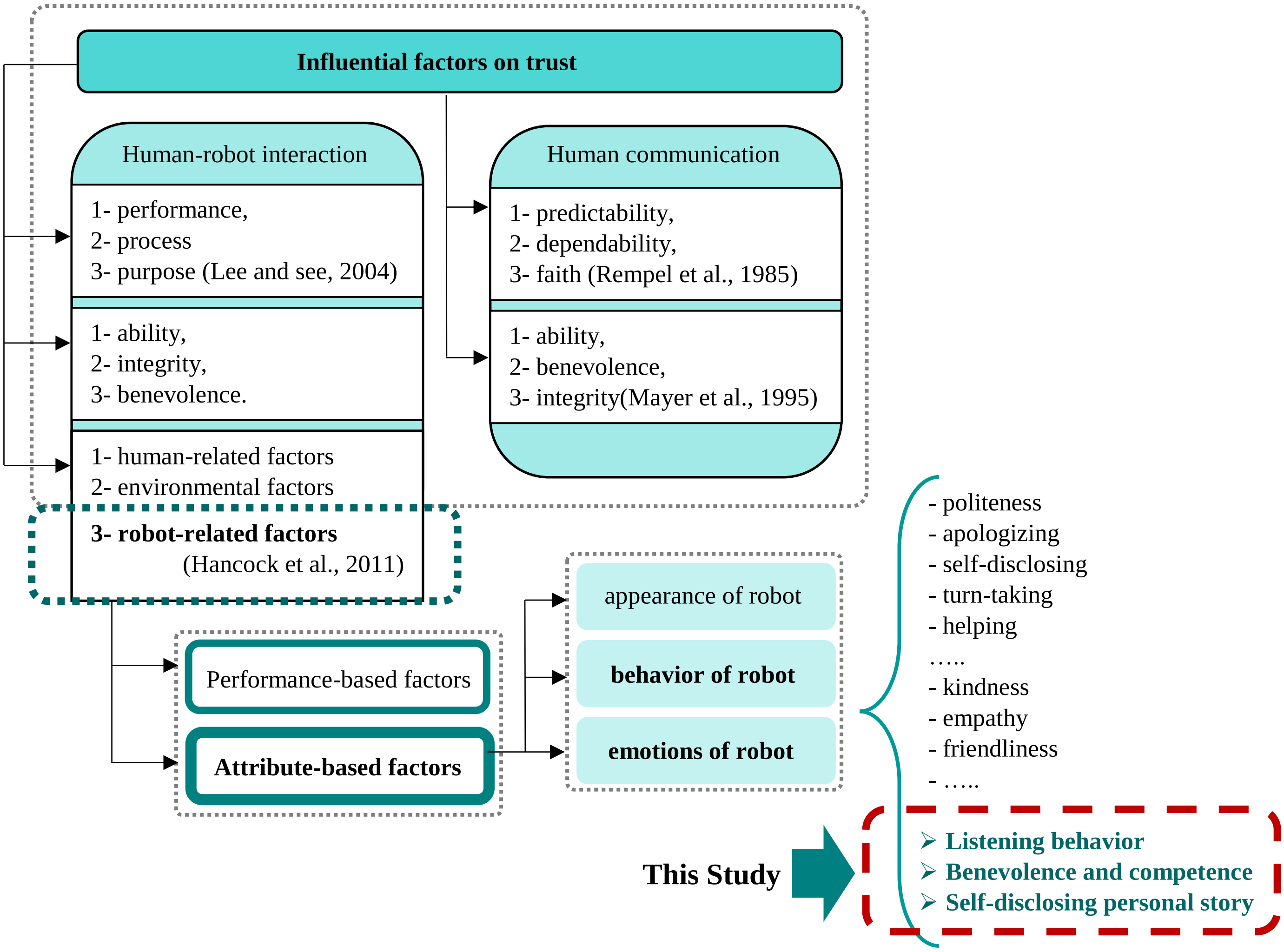


Figure 1. 1 The position of current study in the literature of human-robot trust

1.6 STRUCTURE OF DISSERTATION

This dissertation is divided into three main parts: the first part introduced the theoretical background of trust in HRIs and influential factors (Chapter 1), the second part included three studies conducted to evaluate the influence of social robots' behaviors on HRT (Chapter 2-4), and the third part discussed results of studies in general and concluded the dissertation (Chapter 5). Figure 1.2 depicts linkages of each chapter of this dissertation and the detail is described as follows.

Chapter 1 introduced the background of trust research in HRI and provided some crucial information in this domain. Two different approaches to HRT have been discussed in this part and the most important and known trust definitions were reviewed. an important contribution of this chapter is allocated to review the influential factors in HRT and find the possible research area respectively. Different robot-related factors to enhance trust in social robots have been discussed and the purpose of study have been concluded. The scheme of the dissertation as well as the contribution of the research is enclosed in this chapter.

Chapter 2 aimed to explore the influence of listening behaviors of social robots on trust. It presented the concept of different listening behaviors in human relations and their relation to trust. Thereafter, listening behaviors' components and structure were discussed to implement them in social robot. Finally, the influence of social robots' listening behaviors on human's trust evaluations were investigate and the results have been reported.

Chapter 3 examined the second set of social robots' behaviors in this dissertation including benevolence and competence attributes. The main definitions of two concepts and their contribution to interpersonal trust as well as HRT have been considered. The experiment conducted during the study was introduced and concluded results were presented and discussed.

Chapter 4 reported the results of the third experiment conducted to explore whether a social robot which disclose self stories is more trustworthy. The idea of telling story by social robots while interacting with humans were exchanged. Different types of stories including happy and sad stories were evaluated in association with trust in HRI, and the results were reported.

Chapter 5 discusses general findings in this dissertation by considering the three presented studies. Among the discussion points are the importance of social robots' behaviors and the necessity of improving human-like and emotional behaviors for future robots. It deduces the dissertation with a summary of outcomes, possible implications and limitations of the research.

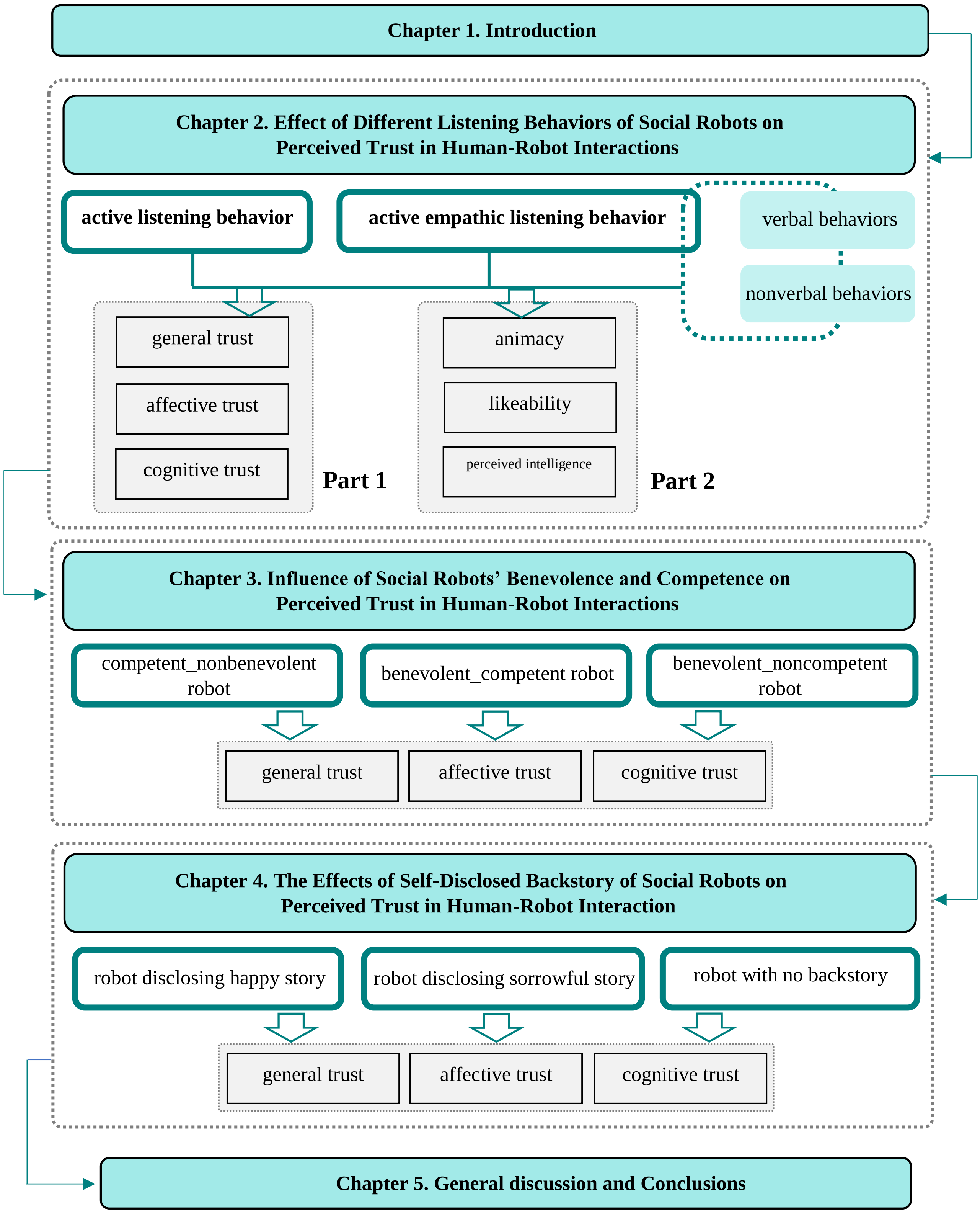


Figure 1. 2 Structure of dissertation

1.7 REFERENCES

Baeber, B. (1983). The Logic and Limits of Trust. In *Rutgers University Press*. https://doi.org/10.5840/bpej1983237

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Beer, J. M., Liles, K. R., Wu, X., & Pakal, S. (2017). Affective Human–Robot Interaction. In *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 359–381). Elsevier Inc. https://doi.org/10.1016/B978-0-12-801851-4/00015-X

Bernotat, J., Eyssel, F., & Sachse, J. (2019). The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-019-00562-7

Breazeal, C., Dautenhahn, K., & Kanda, T. (2016). Social Robotics. In B. Siciliano & O. Khatib (Eds.), *Springer Handbook of Robotics.* (2nd ed., pp. 1935–1971). Springer International Publishing. https://doi.org/10.1007/978-3-319-32552-1_72

Castaldo, S., Premazzi, K., & Zerbini, F. (2010). The Meaning(s) of Trust. A Content Analysis on the Diverse Conceptualizations of Trust in Scholarly Research on Business Relationships. *Journal of Business Ethics*, *96*(4), 657–668. https://doi.org/10.1007/s10551-010-0491-4

Charrier, L., Galdeano, A., Cordier, A., Lefort, M., Charrier, L., Galdeano, A., Cordier, A., Lefort, M., Display, E., Charrier, L., Galdeano, A., Cordier, A., & Lefort, M. (2018). Empathy Display Influence on Human-Robot Interactions : a Pilot Study. *Workshop on Towards Intelligent Social Robots: From Naive Robots to Robot Sapiens at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)*.

Chen, J. Y. C., & Terrence, P. I. (2009). Effects of imperfect automation and individual differences on concurrent performance of military and robotics tasks in a simulated multitasking environment. *Ergonomics*, *52*(8), 907–920. https://doi.org/10.1080/00140130802680773

David, D., Thérouanne, P., & Milhabet, I. (2022). The acceptability of social robots: A scoping review of the recent literature. *Computers in Human Behavior*, *137*(August). https://doi.org/10.1016/j.chb.2022.107419

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and Emotion*, *21*(6), 1184–1211. https://doi.org/10.1080/02699930701437931

Ghazali, A. S., Ham, J., Barakova, E. I., & Markopoulos, P. (2018). Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers Robotics AI*, *5*(JUN), 1–16. https://doi.org/10.3389/frobt.2018.00073

Gompei, T., & Umemuro, H. (2018). Factors and Development of cognitive and affective trust on social robots. *Social Robotics. ICSR 2018. Lecture Notes in Computer Science*, *11357*, 45–54. https://doi.org/10.1007/978-3-030-05204-1_5

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

Hasnain, S. S. (2018). How do You Measure Trust?: Trust Taxonomies & Measurement Tools Revisited. *International Journal of Innovative Research and Development*, *7*(9), 54–59. https://doi.org/10.24940/ijird/2018/v7/i9/sep18050

Herse, S., Vitale, J., Tonkin, M., Ebrahimian, D., Ojha, S., Johnston, B., Judge, W., & Williams, M. A. (2018). Do You Trust Me, Blindly? Factors Influencing Trust Towards a Robot Recommender System. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, 7–14. https://doi.org/10.1109/ROMAN.2018.8525581

Hupcey, J. E., Penrod, J., Morse, J. M., & Mitcham, C. (2001). An exploration and advancement of the concept of trust. *Journal of Advanced Nursing*, *36*(2), 282–293. https://doi.org/10.1046/j.1365-2648.2001.01970.x

Jian, J.Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

John, C., & Catherine, S. (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

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, *43*(6), 1306–1317. https://doi.org/10.1037/0022-3514.43.6.1306

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

Kaplan, A. D., Kessler, T. T., Sanders, T. L., Cruit, B. J., Brill, J. C., & Hancock, P. . (2021). A time to trust: Trust as a function of time in human-robot interaction. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction* (pp. 143–157). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-819472-0.00006-X.

Khavas, Z. R., Ahmadzadeh, S. R., & Robinette, P. (2020). Modeling Trust in Human-Robot Interaction: A Survey. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12483 LNAI*, 529–541. https://doi.org/10.1007/978-3-030-62056-1_44

Kim, W., Kim, N., Lyons, J. B., & Nam, C. S. (2020). Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach.

Applied Ergonomics, *85*(103056). https://doi.org/10.1016/j.apergo.2020.103056

Kory-Westlund, J. M., & Breazeal, C. (2019). Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning. *Frontiers in Robotics and AI*, *6*(July). https://doi.org/10.3389/frobt.2019.00054

Kraus, M., Kraus, J., Baumann, M., & Minker, W. (2018). Effects of gender stereotypes on trust and likability in spoken human-robot interaction. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 112–118. https://aclanthology.org/L18-1018

Kumar, A., Capraro, V., & Perc, M. (2020). The evolution of trust and trustworthiness. *Journal of the Royal Society Interface*, *17*(169). https://doi.org/10.1098/rsif.2020.0491

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, J. gil, Lee, J., & Lee, D. (2021). Cheerful encouragement or careful listening: The dynamics of robot etiquette at Children's different developmental stages. *Computers in Human Behavior*, *118*(November 2020), 106697. https://doi.org/10.1016/j.chb.2021.106697

Leite, I., Pereira, A., Mascarenhas, S., Martinho, C., Prada, R., & Paiva, A. (2013). The influence of empathy in human-robot relations. *International Journal of Human Computer Studies*, *71*(3), 250–260. https://doi.org/10.1016/j.ijhcs.2012.09.005

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123. https://doi.org/10.1037/1076-898X.6.2.104

Lewis, J. D., & Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, *63*(4), 967–985.

Lum, H. C. (2020). The role of consumer robots in our everyday lives. In R. Pak, E. Rovira, & E. J. de Visser (Eds.), *Living with Robots: Emerging Issues on the Psychological and Social Implications of Robotics* (pp. 141–152). Elsevier Inc. https://doi.org/10.1016/B978-0-12-815367-3.00007-4

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Mahdi, H., Akgun, S. A., Saleh, S., & Dautenhahn, K. (2022). A survey on the design and evolution of social robots — Past, present and future. *Robotics and Autonomous Systems*, *156*(104193). https://doi.org/10.1016/j.robot.2022.104193

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more: Designing HRI to encourage more trust, disclosure, and companionship. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016-April*, 577. https://doi.org/10.1109/HRI.2016.7451864

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. *Cognition*, *146*, 22–32.

https://doi.org/10.1016/j.cognition.2015.09.008

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McAllister, D. J. (1995). Affect and CognitionBased Trust As Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, *38*(1), 24–59. https://doi.org/10.2307/256727

Moorman, C., Deshpande, R., & Zaltman, G. (1993). Factors Affecting Trust in Market Research Relationships. *Journal of Marketing*, *57*(1), 81–101. https://doi.org/10.2307/1252059

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, *58*(3), 20–38. https://doi.org/10.2307/1252308

Muir, B. M. (1987). Trust between humans and machines. *International Journal of Man-Machine Studies*, *27*, 327–339. https://doi.org/10.1016/S0020-7373(87)80013-5

Nam, C. S., & Lyon, J. (2021). Trust in human robot interaction. In C. S. Nam & J. Lyon (Eds.), *IEEE Spectrum* (1st ed.). Academic Press. https://doi.org/https://doi.org/10.1016/C2018-0-04443-6

Natarajan, M., & Gombolay, M. (2020). Effects of anthropomorphism and accountability on trust in human robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction*, 33–42. https://doi.org/10.1145/3319502.3374839

Oksanen, A., Savela, N., Latikka, R., & Koivula, A. (2020). Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human–Technology Interactions Online. *Frontiers in Psychology*, *11*(December), 1–13. https://doi.org/10.3389/fpsyg.2020.568256

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, *55*(9), 1059–1072. https://doi.org/10.1080/00140139.2012.691554

Powers, A., Kiesler, S., Fussell, S., & Torrey, C. (2007). Comparing a computer agent with a humanoid robot. *HRI 2007 - Proceedings of the 2007 ACM/IEEE Conference on Human-Robot Interaction - Robot as Team Member*, 145–152. https://doi.org/10.1145/1228716.1228736

Reiersen, J. (2017). Trust as belief or behavior? *Review of Social Economy*, *75*(4), 434–453. https://doi.org/10.1080/00346764.2016.1269939

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in Close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95

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

Robbins, B. G. (2014). *On the Origins of Trust*. University of Washington.

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

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

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Towards safe and trustworthy social robots: Ethical challenges and practical issues. *In: Tapus A., André E., Martin JC., Ferland F., Ammi M. (Eds) Social Robotics. ICSR 2015. Lecture Notes in Computer Science*, *9388*, 584–593. https://doi.org/10.1007/978-3-319-25554-5_58

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

Schaefer, K. E. (2013). *The perception and measurement of huamn-robot trust* [University of Central Florida Orlando, Florida]. http://purl.fcla.edu/fcla/etd/CFE0004931

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems. *Human Factors*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Scholtz, J. (2003). Theory and evaluation of human robot interactions. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003*. https://doi.org/10.1109/HICSS.2003.1174284

Sebo, S. S., Krishnamurthi, P., & Scassellati, B. (2019). "I Don't Believe You": Investigating the Effects of Robot Trust Violation and Repair. *ACM/IEEE International Conference on Human-Robot Interaction*, *2019-March*, 57–65. ttps://doi.org/10.1109/HRI.2019.8673169

Sheridan, T. B. (1975). Considerations in Modeling the Human Supervisory Controller. *Proceedings of the IFAC 6th World Congress*, 1–6. https://doi.org/10.1016/s1474-6670(17)67555-4

Simpson, J. a. (2007). Foundations of interpersonal trust. In A. W. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (Issue January 2007, pp. 587–607). The Guilford Press. http://psycnet.apa.org/record/2007-11239-025

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

Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics [Grand challenges of robotics]. *IEEE Robotics and Automation Magazine*, *14*(1), 35–42.

https://doi.org/10.1109/MRA.2007.339605

van Pinxteren, M. M. E., Wetzels, R. W. H., Rüger, J., Pluymaekers, M., & Wetzels, M. (2019). Trust in humanoid robots: implications for services marketing. *Journal of Services Marketing*, *33*(4), 507–518. https://doi.org/10.1108/JSM-01-2018-0045

Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Matarić, M. J. (2006). The role of physical embodiment in human-robot interaction. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 117–122. https://doi.org/10.1109/ROMAN.2006.314404

Xu, J., & Howard, A. (2022). Evaluating the Impact of Emotional Apology on Human-Robot Trust. *RO-MAN 2022 - 31st IEEE International Conference on Robot and Human Interactive Communication: Social, Asocial, and Antisocial Robots*, 1655–1661. https://doi.org/10.1109/RO-MAN53752.2022.9900518

Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, *18*(2), 129–166. https://doi.org/10.1007/BF02249397

You, S., & Robert, L. P. (2018). Human-Robot Similarity and Willingness to Work with a Robotic Co-worker. *ACM/IEEE International Conference on Human-Robot Interaction*, 251–260. https://doi.org/10.1145/3171221.3171281

Zur, A., Leckie, C., & Webster, C. M. (2012). Cognitive and affective trust between Australian exporters and their overseas buyers. *Australasian Marketing Journal*, *20*(1), 73–79. https://doi.org/10.1016/j.ausmj.2011.08.001

CHAPTER 2

2.1 Effect of Different Listening Behaviors of Social Robots on Perceived Trust in Human-Robot Interactions

2.1.1 INTRODUCTION

Listening behavior is one of the influential factors in interpersonal trust, and numerous studies have approved it. For example, McGarvey (1996) stated that trust is an important element of listening and an outcome of good active listening behavior (Brunner, 2008). Ramsey and Sohi (1997) counted effective listeners as more trustworthy. Among the different listening behaviors, *active listening* **(AL)** and *active empathic listening* **(AEL)** are more closely related to trust (Figure 2.1). AL builds trust by showing attention and confirming the emotions and experiences of the speaker (Lester, 2002). When AL is accompanied by empathy, it results in AEL and acts even more powerfully than AL, enhancing trust, because it is equipped with empathic reactions. Studies in different fields showed that AL and AEL build up trust, for instance, between patients and their psychotherapists, in education and formation of trust between students and supervisors, in management and business, or controlling pessimism in marital conflicts between couples (e.g., Aggarwal et al., 2005; Gordon, 1975; Lloyd et al., 2015; Nadler & Simerly, 2006; Weger et al., 2014).

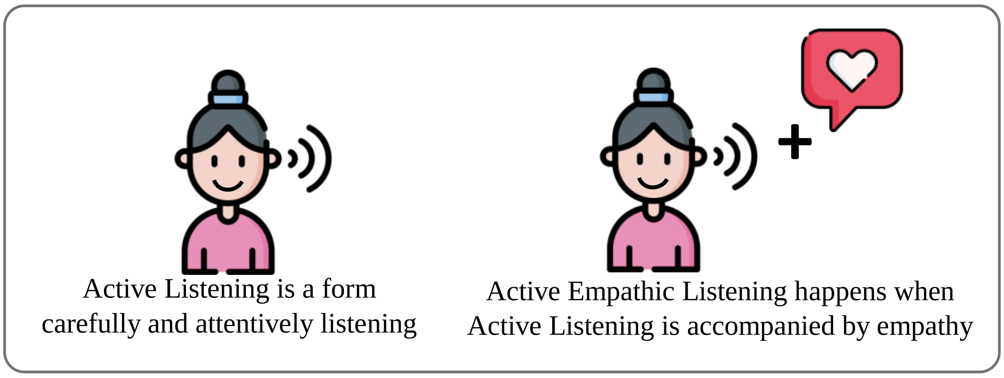


Figure 2.1 Two types of listening

AL and AEL are emerging as novel concepts in HRI. Recently, research has been conducted on the listening behavior of social robots. Some researchers tried to simulate natural listening in robots (Ee et al., 2008; Mohammad & Nishida, 2008; Ogasawara et al., 2005; Ogawa & Watanabe, 2000), while others have emphasized on more AL in robots (Kobayashi et al., 2010). In a valuable study, Inoue et al. (2020) implemented attentive listening system for the android ERICA and the proposed system generated several types of listener responses: backchannels, repeats, elaborating questions, assessments, generic sentimental responses, and generic responses. They compared the performance of the system with a "Wizard-of-Oz" (WoZ)[1] operated system and found that the generated system was comparable in terms of perceived basic listening skills with the WoZ system, but still primitive for sophisticated skills such as displaying empathy.

Certain studies have focused only on the verbal part of listening behaviors using backchannels or fillers to conduct attentive listening and produce coherent dialogues during conversations (Johansson et al., 2016; Lala et al., 2017), or in large-scale projects such as SimSensei (DeVault et al., 2014). Others have searched for both verbal and nonverbal components of listening behaviors to evaluate body movements and utterances simultaneously (Kanda et al., 2007). Some studies explored the outcome of robots' listening approaches in improving human's interactions with robots. For example, Lee et al. (2021) tested different behavioral approaches in child-robot interaction and predicted careful listening to increase older children's sociality perception of robots. Another study by Nakamura and Umemuro (2022) aimed to clarify the effect of robot's listening attitude on elderly's self-disclosing and they found that a robot listened with neutral behavior first and subsequently with positive behavior, could encourage the participants to self-disclose more. Another interesting study by Nishio et al. (2021) developed talking twin robots with a previously proposed active listening function that could converse with elderly people, and the authors found that elderly subjects talked more with the proposed robot system than that with the control system. Yonezawa et al. (2010) developed an assistant robot showing AL by changing robot's eye gaze, nodding and back-channel feedback, and evaluated its effect on video communication. The results showed the effectiveness of

[1] WoZ is a widely accepted and used method for experiments in HRI (see Riek, 2012). The method is specially used when the robot is required to demonstrate a certain level of intelligence or consistent behavior (Hinwood et al., 2018). Usually, a human operator in a laboratory setting, simulates the behavior of a theoretical intelligent system (often by going into another room and intercepting all communications between participant and system), and remotely operates the intelligent system like a robot and controls its movements, speech utterances, gestures, etc., while the participants are not aware of it (Rietz et al., 2021). Considering the complexity of natural dialogue, the WoZ technique gives subjects more freedom of expression or constrains them in more systematic ways (Dahlbäck et al., 1993).

robot's attitude on the impression of the robot itself and improvement of video communication. Sundar et al. (2017) compared a companion and assistant robot's different speech style (playful, serious), and the results indicated that the robot's roles and speech styles interact with each other in promoting user experience; A companion robot that spoke in a calm tone was perceived to be the most attractive, while a serious assistant was the least attractive. Moreover, participants moderately liked playful robots regardless of their roles. Listening behaviors of social robots beside other cues like asking a user's emotional state, appropriate rephrasing of user's emotional disclosure, and a communicated sense of understanding, create the robots to be perceived as empathetic by users (Johanson et al., 2021). However, we could not find a thorough research which has investigated the role of AL or AEL behaviors of social robots in establishing trust.

Therefore, during the first study, we attempted to investigate the effect of different listening behaviors of social robots on people's perceived trust. We conducted an experiment, asking participants to have a conversation with a social robot while showing AL and AEL behaviors. Different types of trust, including general, cognitive and affective trust were measured as variables to capture the perceived trustworthiness of participants. With a better understanding of the effect of listening behavior of social robots on trust, we can design more trustworthy, cooperative, and friendly robots and effectively enhance HRI.

2.1.2 DEVELOPING HYPOTHESES

2.1.2.1 AL and AEL behaviors and trust

Listening is a part of interactive process and the foundation of effective communication in all social environments, which enriches the relations and improves morale and productivity of the communication. It is a very complex process that requires both reactive and deliberative procedures, and involving cognitive, affective and behavioral activities (Johansson et al., 2016). Listening includes at least three components: sensing, evaluating and responding (Ramsey & Sohi, 1997). Sensing refers to the actual receipt of messages, evaluating refers to activities that take place in the mind of the listener, while responding involves acknowledging the receipt of messages. Table 2.1 shows the description of different dimensions of listening behavior.

Table 2. 1 Components of three levels of listening behavior (adopted from Comer & Drollinger, 1999)

Sensing	
Verbal component: Receiving all verbal cues including words, inflection, paralanguage. Involves hearing, noting inflection, attending to the message, sensing the tone of the message.	*Non-verbal component:* Receiving all non-verbal cues (body language, facial expressions, proxemics). Receiving by all senses, not just hearing.
Processing	
Understanding: Ascribing accurate meaning to the words, meaning of messages underlying the words including emotions, thoughts, feelings. Interpreting: Assessing the implications of the messages. Comparing incoming messages to those in memory. Being alert to discrepancies. Evaluating: Assessing appropriateness, placing value, and prioritizing importance of messages. Determining key messages. Remembering: Committing new material to memory. Updating knowledge structures.	
Responding	
Verbal component: Giving appropriate verbal responses (e.g., acknowledging, agreeing), short prompts, paraphrasing, questioning. Using appropriate tone of voice, inflection and using familiar terminology.	*Non-Verbal component:* Maintaining appropriate eye contact, facial expressions, head nods, and body language.

Several modes of listening have been studied in the literature (e.g., Bauer & Figl, 2008; Weissglass, 1990; Worthington & Bodie, 2018), and we recognized two notable concepts of AL and AEL in relation to interpersonal trust. AL goes back to Thomas Gordon (1975) and Carl Rogers (Rogers & Farson, 1957). AL is *a form of carefully and attentively listening and responding, to achieve a deeper understanding of the speaker's message and context* (Bauer & Figl, 2008; Browning & Waite, 2010). AL is not limited to aural messages, but includes receipt of nonverbal messages as well. Although active listeners try to be attentive to verbal and nonverbal cues, they sometimes appear mechanically and fail to project emotions to the speaker (Comer & Drollinger, 1999). Therefore, *a listener sometimes acts with empathic tendencies* to make the functional components of AL more emotional and insightful (Gearhart & Bodie, 2011), resulting in the behavior of AEL. AEL was originally defined in the context of product sales as a form of listening practiced by salespeople in which conventional AL is combined with **empathy** to achieve a superior and more effective form of listening (Drollinger et al., 2006, p. 162). Thus, in AEL, the listener attempts to evaluate the underlying meaning of the message intuitively and empathize with the person, in addition to responding verbally and nonverbally. It is common in counseling, therapy, and marketing literature to help them better understand their clients and customers.

For AL and trust incorporation, evidence shows the effectiveness of AL on trust in HC and its impact on being friendly (Bodie et al., 2012) or socially attractive (Weger et al., 2010). Nugent and Halvorson (1995) stated that AL builds trustworthy relationships with clients during therapy. Fassaert et al. (2007) showed that good listeners are more liked and trusted, and Lasky (2000) established AL as an important first step in communication for developing trust. It is widely recognized that trust plays a vital role in seller and buyer relations, as Ramsey and Sohi (1997) proved in their research that salespersons with better listening behavior are considered more trustworthy. Additionally, AEL is positively related to salespersons' trust in marketing, and the findings support the notion that salespeople with higher levels of AEL would have higher quality relationships and are regarded as trustworthy (Drollinger & Comer, 2013).

AL and AEL for social robots, are both new concepts in HRT and their effect on trust have not been previously investigated and confirmed. Therefore, we proposed the first hypotheses considering the impact of AL and AEL behaviors of social robots on trust, to know if these listening behaviors can contribute toward enhancing HRT. It should also be emphasized that it is important to know the impact of AL on trust individually, as sometimes we may only rely on AL, because of limitations in making empathic responses between humans and robots. To investigate the influence of AL and AEL on trust perception of robots, we assumed comparing them with *nonactive-nonempathic listening behavior* **(NAL)**, as control condition, that is "just listening" without showing attention or empathy to the speaker. Thus, the following hypotheses were derived.

H1a: Social robots with the behavior of AL results in more (general) trust perceived by the user toward the robot than social robots with the behavior of NAL.

H1b: Social robots with the behavior of AEL results in more (general) trust perceived by the user toward the robot than social robots with the behavior of NAL.

Although, some authors have used the terms AL and AEL synonymously and interchangeably (Nugent & Halvorson, 1995; Weger et al., 2014), AEL is considered superior to AL in selling (Comer & Drollinger, 1999), as it is empowered with empathy. Empathy is perceiving another's emotional state and its meanings as if one were the other person, and is likely to play a particularly important role in earning trust in medical services, marketing environment, individual relations and etc. (Aggarwal et al., 2005). Comer and Drollinger (1999) admitted that, as empathy increases in salesperson behavior, the level of listening increases. Aggarwal et al. (2005) also established a strong positive correlation between a salesperson's empathy and listening behavior with trust and satisfaction. In HRI, empathy is a demanding and growing field of research, and it is proposed that people would perceive their relationship with empathic robots differently. Number of studies presented the effectiveness of empathy in trustworthiness, human-likeness, attractiveness, likeability, and familiarity of the robots (see (Park & Whang, 2022)). Therefore, based on the aforementioned arguments, the next hypothesis was added to the comparison of AL and AEL of social robots, to measure how much empathy in listening behavior of social robots can be influential in making higher trust perception.

H1c: Social robots with the behavior of AEL results in more (general) trust perceived by the user toward the robot than social robots with the behavior of AL.

2.1.2.2 Listening behaviors and types of trust

As discussed in section 1.3, various types of trust are identified, with affective and cognitive trust being among the most frequently referenced categories in the literature. Considering the definitions of AL and AEL, it was mentioned that AEL is characterized by the inclusion of an empathetic and emotional overlay. The empathetic component of AEL constructs an emotional link between individuals, as Floyd (2014) argued that empathic listening operates as a form of indirect nonverbal affection and conveys a message of care, love, and tenderness for the partner. However, AL attempts to deeply understand the sender's point of view by confirming, rationalizing, or seeking more details (Bauer & Figl, 2008). Thus, considering the relationship between affective trust and emotional clues, as well as the potential of AEL in conveying emotions and empathy, we proposed the next hypothesis.

H2a: Affective trust is significantly higher for social robots exhibit the behavior of AEL than social robots with the behavior of AL.

Regarding the relation of cognitive and affective trust, cognitive trust is considered one of the antecedents of affective trust in some studies (e.g., Johnson & Grayson, 2005;

Lewis & Weigert, 1985), and emotions, on the other side, do influence the perception and cognitive evaluations, even after the dissipation of emotions. Therefore, the reciprocity of cognitive and affective trust is intertwined and they are highly interdependent. Additionally, according to modern psychology, affection involves several basic cognitive functions and appears to be necessary for normal conscious experiences. It influences, modulates, and mediates basic cognitive processes (Duncan & Barrett, 2007). Furthermore, empathy, has both cognitive and affective components which are relevant and influential to each other. Cognitive component of empathy involves understanding, while affective component consists of an internal emotional reaction that results in understanding of the other's feelings (Comer & Drollinger, 1999). Therefore, considering the cognitive and affective components of empathy in AEL, and additionally, due to the substantial influence of affection on the cognitive processes in humans, we concluded that AEL would be more influential in eliciting cognitive trust as well. Based on the above arguments, the next hypothesis was derived.

H2b: Cognitive trust is significantly higher for social robots exhibit the behavior of AEL than social robots with the behavior of AL.

2.1.2.3 Verbal and nonverbal dimensions of listening behaviors

Listening behaviors have two components which provide **verbal** and **nonverbal** feedback to the speaker (Comer & Drollinger, 1999; Robertson, 2005). Different aural techniques such as paraphrasing, restating a version of the speaker's message, asking clarifying questions, and nonverbal messages such as nodding when talking, proper posture and body positioning, facial expressions and eye contact, as well as summarizing, paying attention, or encouraging and balancing are used to fully catch the speaker (Bauer & Figl, 2008; Browning & Waite, 2010; Buschmeier et al., 2014; Comer & Drollinger, 1999; Weger et al., 2010). Table 2.2 summarizes some examples of verbal and nonverbal components of listening behaviors.

Some researchers claimed that regardless of the specific verbal responses of listeners, showing concern and attention to a person produces positive perceptions (Libow & Doty, 1976). Furthermore, concise findings showed that 55% of the total impact of a message is associated with nonverbal aspects and only seven percent is due to verbal segments (Mehrabian, 1971). Specifically, many studies have confirmed the incorporation of nonverbal behaviors and effective empathic listening (Jones & Guerrero, 2001). As Weger et al. (2010) stated that listeners receive more care and concern from nonverbal behaviors of listening than specific verbal behaviors such as paraphrasing, questioning, giving advice, or reflecting the emotional content of messages. Additionally, Floyd (2014) argued that listeners who use nonverbal behaviors convey empathy and support more effectively than those who do not. In relation to nonverbal behavior and trust, it was also confirmed that, although verbal messages of physicians influence patients' interpersonal trust, nonverbal behavior is most likely to have a more crucial influence on trust (Hillen et al., 2014). Thus, owing to nonverbal behaviors' empathic effect, the importance of these behaviors in enhancing trust should be particularly dominant in AEL.

Table 2. 2 Some of verbal and nonverbal components of listening behavior

Verbal
1. Backchanneling (e.g., *Hmm, ah ha, Uh-huh...,*) (Gearhart & Bodie, 2011; Lala et al., 2017) 2. Affirmative statements (e.g., *OK, yes, I understand, sure, Keep going..., I see, I know, good, Really?, That's very interesting, Then?, Sounds good, right, please go on*) (Gearhart & Bodie, 2011; Lala et al., 2017) 3. Making encouraging comments (Bauer & Figl, 2008; Browning & Waite, 2010; Weger et al., 2010) 4. Paraphrasing (both content and feelings) (Bauer & Figl, 2008; Robertson, 2005; Weger et al., 2010) 5. Summarizing (Bauer & Figl, 2008; Kobayashi et al., 2010; Robertson, 2005) 6. Asking questions (open ended questions) (Bauer & Figl, 2008; Browning & Waite, 2010; Weger et al., 2010) 7. Making inferences (Browning & Waite, 2010) 8. Making associations and analogies (Robertson, 2005) 9. Calling by name (Robertson, 2005)
Nonverbal
1. Facial expressions, e.g., smiling, furrow-brow, frown, and lower eyes, raising eyebrows (Bodie et al., 2012; Comer & Drollinger, 1999; Robertson, 2005) 2. Voice, e.g., pace, loudness, tone (Floyd, 2014; Hadar et al., 1985) 3. Eye contact (Comer & Drollinger, 1999; Floyd, 2014; Robertson, 2005) 4. Head movement, e.g., nodding, tilted head (Buschmeier et al., 2014; Hadar et al., 1985) 5. Gestures (Buschmeier et al., 2014; Floyd, 2014) 6. Body posture and movement (Comer & Drollinger, 1999; Floyd, 2014; Robertson, 2005) 7. Personal space and territoriality (Bodie et al., 2012) 8. Sitting position (Bodie et al., 2012; Floyd, 2014) 9. Physically displaying emotion (Kobayashi et al., 2010)

Moreover, several studies have considered the influence of nonverbal behaviors of social robots in different HRI fields (see Rifinski et al., 2021) ; for example, some studies demonstrated how a nonverbal cue, such as the robot's gaze, could influence participants' roles, turn-taking and people's interpersonal evaluation on robot in human-robot conversation. Furthermore, responsive robots with gestures were more attractive and liked. Although, the inclusion of nonverbal behaviors is confirmed as essential, it is not clear how effective the nonverbal component of AEL behavior of social robots contributes to HRT. The importance of this issue increases when considering the limitations of robots' nonverbal behaviors, as they mostly lack facial expressions, as an important source of conveying emotions. Therefore, for the third set of hypotheses, we considered the comparison of *AEL behavior of social robots without nonverbal components* **(AEL_{VO} = AEL-verbal only)** with the behavior of AEL which includes both verbal and nonverbal components, to assess the power of nonverbal behaviors of AEL of social robots in enhancing HRT. Therefore, considering the abovementioned aspects, the following hypotheses were developed.

H3a: Trust is significantly higher for social robots with the behavior of AEL than social robots with the behavior of AEL_{VO}.

H3b: Affective trust is significantly higher for social robots with the behavior of AEL than social robots with the behavior of AEL_{VO}.

H3c: Cognitive trust is significantly higher for social robots with the behavior of AEL than social robots with the behavior of AEL_{VO}.

2.1.3 METHOD

2.1.3.1 Overview

According to the derived hypotheses, we concluded four listening behaviors for the robot. For the set of hypotheses 1 and 2, three listening behaviors were considered: 1) *nonactive, nonempathic listening behavior* **(NAL)**, 2) *active, nonempathic listening behavior* **(AL)**, and 3) and, *active, empathic listening behavior* **(AEL)**. For hypotheses 3, we considered 4) *verbal-only active, empathic listening behavior* **(AEL_{VO})** compared to active, empathic listening behavior (both verbal and nonverbal). Therefore, we developed four different listening behaviors for the robot. Table 2.3 lists the components of each listening behavior, and we explain design of each behavior in section 2.3.6 in more details. Testing the hypotheses, a between–subject experiment with four conditions was designed. The **four experimental conditions** were defined regarding the listening behaviors of the robot.

NAL condition: Participants talk with the robot acting with the behavior of NAL.

AL condition: Participant talk with the robot acting with the behavior of AL.

AEL condition: Participant talk with the robot acting with the behavior of AEL.

AEL$_{VO}$ condition: Participant talk with the robot acting with the behavior of AEL$_{VO}$.

Table 2. 3 Four listening behaviors of the social robot and their components

	Listening behavior components			
Listening behavior	attention(active)	empathy	verbal	nonverbal
NAL	none	none	low	low
AL	high	none	high	high
AEL	high	high	high	high
AEL$_{VO}$	high	high	high	none

The interaction between the robot and participants was designed by conducting a conversation and questionnaire-based evaluation. The idea was to simulate a dyadic conversation between participants and the robot, in which the robot showed different listening behaviors. Most of previous researches and studies have considered active listening behavior in conversations (see Weger et al. 2010; Bodie, 2018), as a good listening behavior is the process of listening and responding, and it is basically a matter of better communication. The conversation was about *living experiences in Japan*, which was a familiar topic for all international students in Japan. The topic was chosen from various topics based on familiarity, popularity, and security of information regarding potential participants. During the experiment, the robot tried to speak with the participant asking questions and showing different listening behaviors through the conversation. The conversation between the robot and participants, consisted of greeting in addition to 9 questions: 1) "Why did you decide to come to Japan for studying?", 2) "Have you travelled in Japan?", 3) "Do you speak Japanese or like to learn?", 4) "What did you find most interesting about Japan?", 5) "Do you miss your family or friends here?", 6) "What did you find most difficult in Japan?", 7) "How did you find Japanese people and culture different with your country?", 8) "Any plan for near future in Japan?", and 9) "How long have you been here?". According to four experimental conditions, appropriate verbal statements along with nonverbal behaviors was designed for each question.

"Wizard-of-Oz" (WoZ) methodology was employed in the experiment, through which the participants believed that they were interacting directly with the robot by utilizing a natural language interface; however, in reality they were communicating with an operator. The experiment was conducted in English.

2.1.3.2 Participants

A total of 120 *international students*, aged between 19 and 47 years (male: 54, female: 66; M_{age} = 25.71, SD_{age} = 4.28) from *Tokyo Institute of Technology* in Japan, participated in the experiment. Participants were recruited by distributing flyers in the library and various laboratories of Tokyo Institute of Technology. Additionally, we approached international students in person on the Ookayama campus and in the Senzokuike dormitory of the university. Moreover, online platforms, such as LINE and Telegram, was utilized to share digital flyers in various international groups. We also sought assistance from friends, encouraging them to share information about the experiment with their respective networks and friends as extensively as possible. Any particular educational or cultural background requirements was not specified for participants. As a result, the final participants were diverse, representing various majors and countries. Figure 2.2 displays information regarding the participants' home countries.

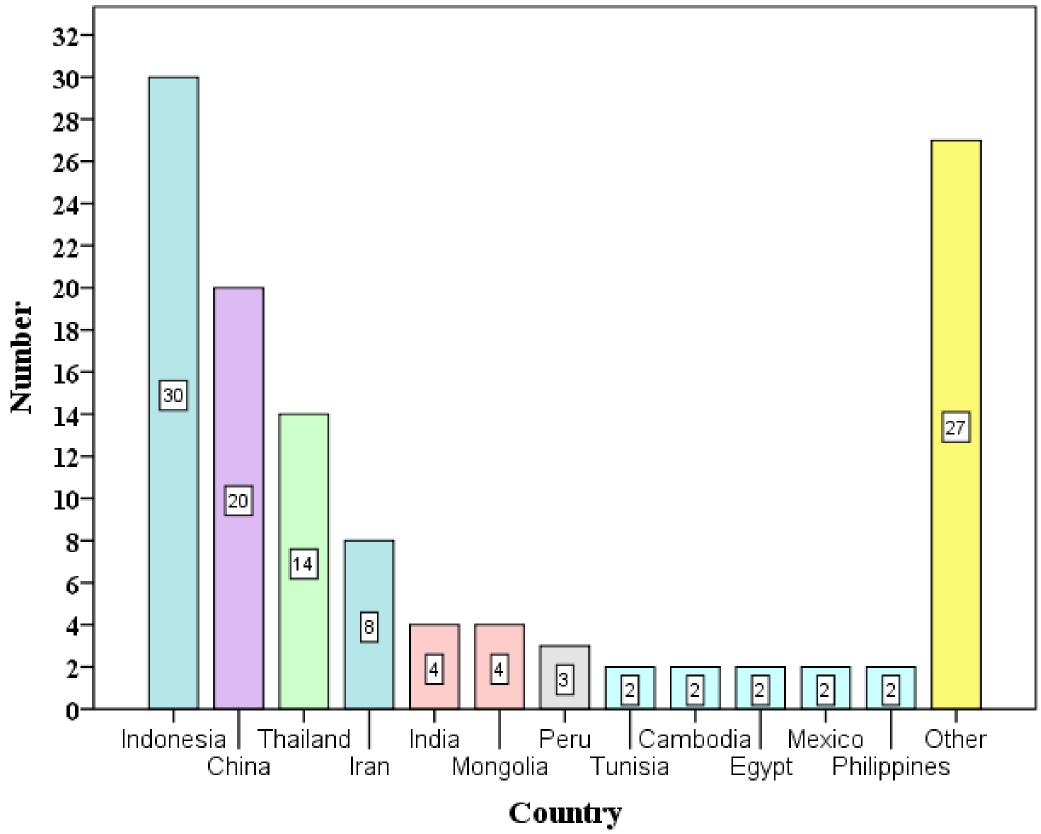


Figure 2. 2 Country disposition of participants during the experiment. (Total number of participants=120)

The participants were assigned to four experimental conditions. The participants were initially assigned through random assignment. However, we endeavored to counterbalance gender distribution across conditions as much as possible. Consequently, we aimed to assign an equal number of female participants to each group as the experiment progressed. Although the number of female participants could not be predicted in advance, so equal numbers of men and woman could not be completely achieved. Finally, thirty participants were allocated to each of the four experimental conditions. Table 2.4 listed the number and gender disposition for each of the four conditions. Because the experiment was conducted in English, a high level of fluency was required; thus, the participants were asked about their level of English proficiency (native speaker: n = 5, non-native speaker: n = 115; fluent professional = 55.8%, working or studying English = 44.2%). Thirty-five participants (29.2%) had seen the robot previously. Among the participants who knew the robot, only a few participants (n = 7, 5.8%) had prior interaction with the robot, similar to talking with the robot in exhibitions or attending other experiments.

Table 2. 4 Number of participants for the four experimental conditions.

Condition	Age Mean (SD)	Male	Female	Total
NAL	25.40 (3.59)	13	17	30
AL	25.17 (4.21)	11	19	30
AEL	26.30 (5.22)	14	16	30
AEL_{VO}	25.97 (4.06)	14	16	30

2.1.3.3 Ethical considerations

The experiment received the approval of all ethical and experimental procedures and protocols from Human Subjects Research Ethics Review Committee of Tokyo Institute of Technology (Project No. A19145). All participants submitted a written consent form and were informed of the experimental procedures and ethical concerns prior to the experiment. The experiment was anonymous and the participants were identified by IDs.

2.1.3.4 Equipment

For the experiment, we used the humanoid robot, **NAO**, previously was developed by Aldebaran Robotics and now by Softbank Robotics. NAO is a small (58 cm tall) programmable and popular robot in education and research. For this study, we required a robot that could talk and move. Considering the abilities of NAO and its ease of programming, it was selected for the experiment.

Four listening behaviors consisting of verbal and nonverbal components were designed for NAO. During the conversation, NAO was controlled by an experimenter using the WoZ method through an interface developed in the HTML. However, a small greeting and conversation were programmed to be executed autonomously prior to the main conversation between the robot and participants, to be realized as intelligent and autonomous robot for participants. NAO communicated verbally using NAO's default text-to-speech settings. Body movements involving head, arms, eyes, and other parts were designed and developed using Choregraphe 2.1.4, based on the behaviors needed for each listening behavior. The behaviors were then attached to specific texts for each listening behavior. The speech rate was set at 100% to represent a normal speaking speed. Several pauses were incorporated into its speech to make it more understandable and natural. NAO can be operated in two sitting or standing situations. The latter was chosen in this study owing to the size of the robot and the variety of behaviors in a standing situation. NAO was placed on a table in front of the participant to make an almost parallel viewpoint between the robot and the participants.

2.1.3.5 Common experimental setup

The experiment was performed in a room separated into three parts: 1) where the experimenter controlled NAO; 2) where participants answered questionnaires before and after conversation; and 3) where NAO was set down and a conversation was held. Figure 2.3 and 2.4 shows the experimental setup. Sitting behind a partition, the wizard team could observe and hear the participants interacting with NAO and control it by considering what participants asked or answered to the robot. The experiment was recorded by using a camera placed on the side of the robot. NAO was positioned in the user's personal space (between 0.3 and 1 m) according to Rossi et al. (2017).

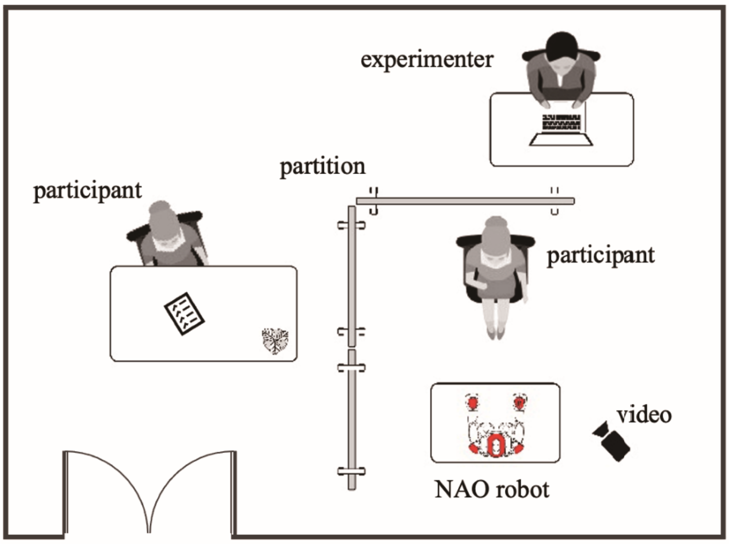


Figure 2. 3 Experimental setup, where the user talked with NAO and the experimenter controlled NAO behind the partition

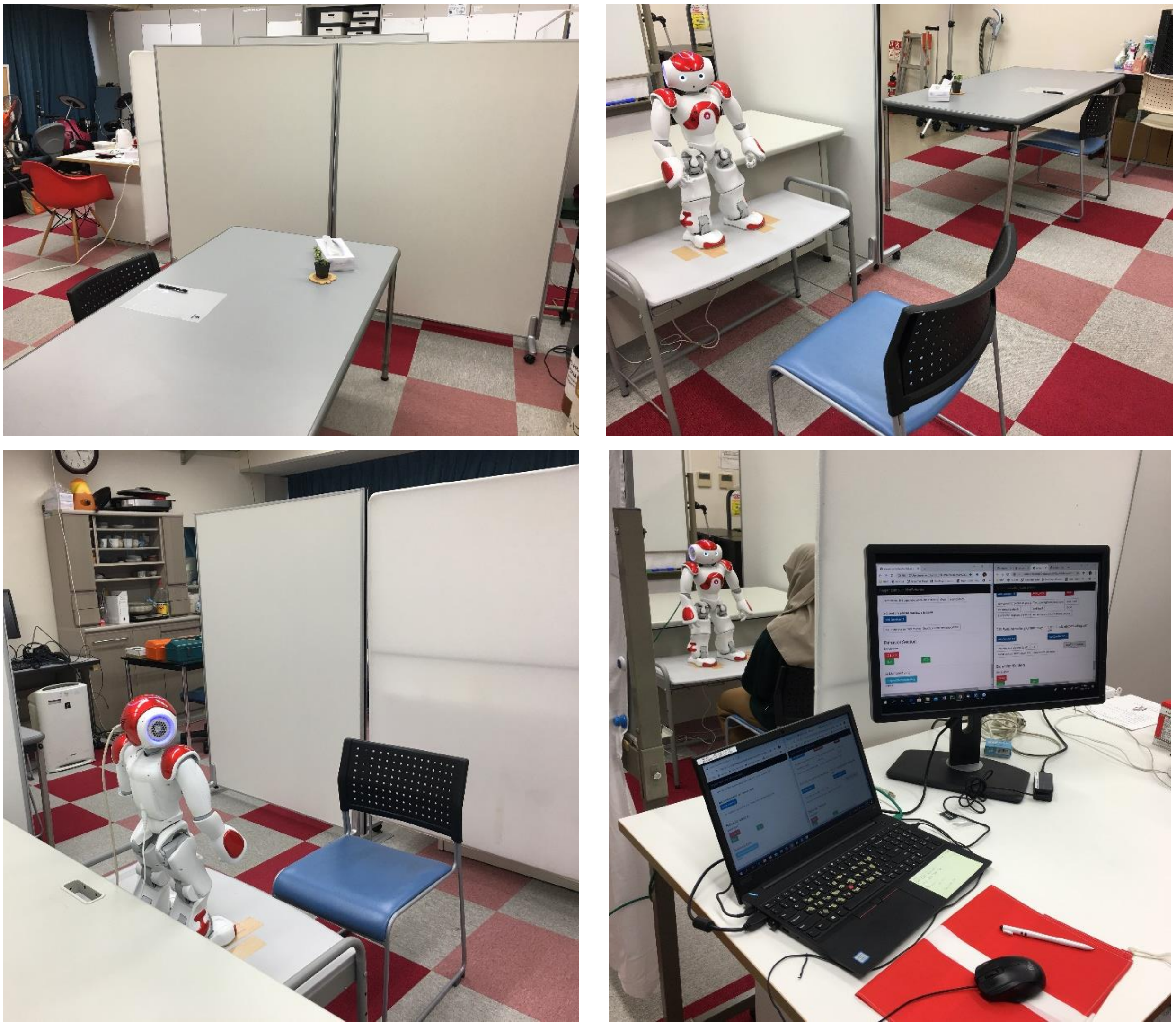


Figure 2. 4 Real experiment environment

2.1.3.6 Procedure

The experiment was held during three months between January-March 2020 at the Ookayama campus of Tokyo Institute of Technology. Before initiating the experimental session, participants were given documents explaining the experiment to review freely and decide voluntarily if they desired to participate in the experiment. The experimenter then explained the procedure and highlighted important points, such as the video recording of the session, to ensure that everything was clear to participant prior to signing the consent form. None of participants opted out of the experiment. Each participant was assigned to one of the four conditions corresponding to NAO's different listening behavior styles.

Participants were asked to fill out their demographic information (i.e., gender, age, nationality, English proficiency, and prior interactions with NAO robot). First, participants learned about NAO through a picture in the explanation form and demographical questionnaire. After completing the primary questionnaire, the experimenter introduced NAO to participant, and NAO shortly greeted autonomously to create first impression before initiating the main conversation. The short greeting of NAO was; "*Hello, I am NAO, nice to meet you*." Before starting the main conversation with NAO, participants rated their trust in the robot on 56-item pre-trust questionnaire, which included 40 items on robot-human trust, 9 items on cognitive trust, and 7 items on affective trust. Thereafter, they were asked to sit in front of NAO to begin the conversation.

Before starting the conversation, the experimenter solicited NAO to start the conversation, and NAO replied to the experimenter autonomously. Thereafter, NAO was left alone with participant. The first short greeting and brief talk before starting the conversation was programmed to be done autonomously to ensure that the participants considered NAO as an intelligent and autonomous robot. The conversation lasted approximately 8–12 min. After completing the conversation as planned, participants were asked to rate their trust in the robot on post-trust questionnaires. The entire experiment was completed in an average of approximately 30-40 min for each participant. Figure 2.5 and 2.6 illustrates the experimental procedure.

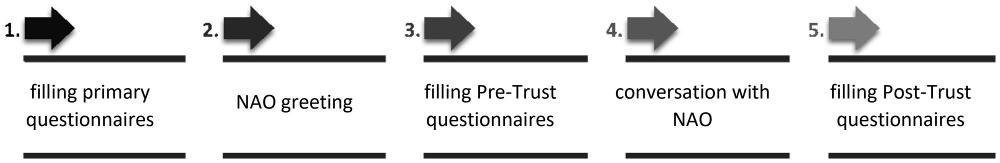


Figure 2. 5 Experimental procedure

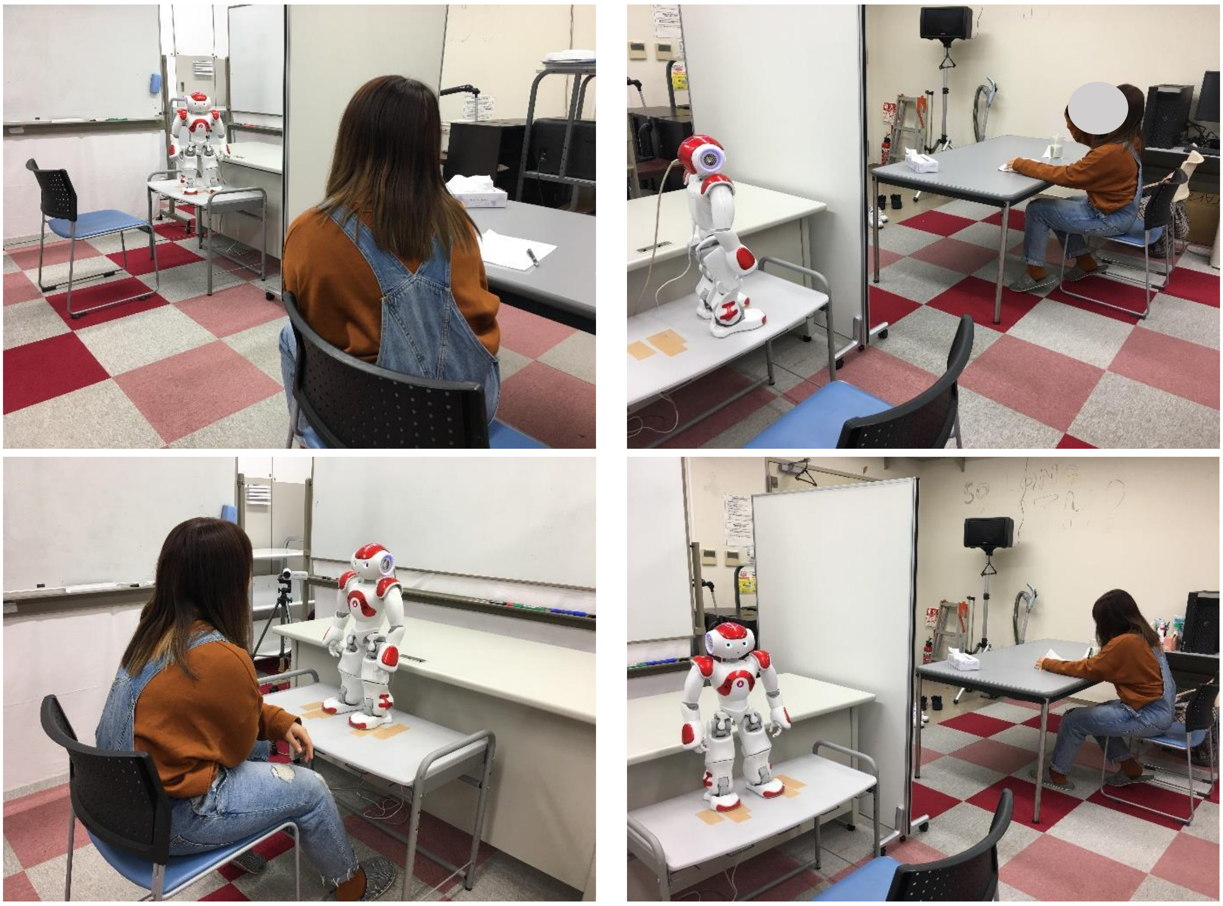


Figure 2. 6 Experiment procedure; (left-top) robot initial greeting, (right-top) answering pre-trust questionnaires after initial robot greeting, (left-bottom) talking with the robot, (right-bottom) answering post-trust questionnaires.

2.1.3.7 Measurement

The quantitative measures of this study included the **general trust scale** for evaluating participants' general trust perception of the robot; the **affective trust scale** for assessing participants' affective trust of the robot; the **cognitive trust scale** for rating participants' cognitive trust of the robot.

2.1.3.7.1 General Trust

There is a scarcity of validated measures to evaluate HRT and none of the existing measures covers all extant measures of HRT. The measurement tools currently available for HRT are heavily skewed toward performance trust and do not consider the emotional or moral aspects. Schaefer (2016) offered the more comprehensive developments of a trust measure that is widely used to evaluate HRT in previous studies (e.g. (Beelen et al., n.d.; Bernotat et al., 2021; Correia et al., n.d.; Savery et al., 2019; Volante et al., 2019; Xie et al., 2019)). The scale offered by Schaefer (2016) covers different features of the robot, the human and the environment in HRT (Chita-Tegmark et al., n.d.; Malle & Ullman, 2021), that represents accuracy and wide applicability of the scale. Therefore, Shaefer's human-robot trust scale was chosen to be used in this study as it is specific to HRT field, and it thoroughly measure the human's trust perceptions in robots in general (Chita-Tegmark et al., n.d.). (Note; we used the term of "general trust" for Schaefer's human-robot trust scale during these studies to differentiate it from other affective and cognitive trust scales). Schaefer's human-robot trust scale consists of 40 items, measures a human's general trust in robots without identifying any specific types of trust. Participants rated their trust in the robot in the range of no trust (0 %) to complete trust (100 %) according to each of the 40 items. The sum of the responses for the items yielded scores on the human-robot trust scale. This score was used as perceived trust to test **H1a, H1b, H1c, and H3a**. The Cronbach's alpha value of the questionnaire was 0.947 which showed high reliability.

2.1.3.7.2 Affective and cognitive trust

Human-robot trust measure of Schafer (2016) has no reference to cognitive or affective trust. Moreover, there are no other measures to evaluate cognitive and affective dimensions of HRT. Thus, we adopted affective and cognitive scales from interpersonal trust as another individual measure. The measure proposed by Johnson and Grayson (2005) and Mcallister (1995) are the most cited and used measures to evaluate affective and cognitive trust in different fields. This measure has also been modified and used in other HRT studies (Gompei & Umemuro, 2018). Therefore, we modified Johnson and Grayson, and McAllister's affective and cognitive trust scale in this study due to the validity and popularity of the them. The descriptions of the original question items referring to a person, were modified to refer to a robot. The final scale consisted of 16 items, 9 items for cognitive trust, and 7 questions for affective trust. Participants evaluated their feelings and perceptions of trust using a seven-point Likert scale (1= "strongly disagree," 7= "strongly agree"). The averages of the responses for the corresponding items provided the scores for cognitive and affective trust. Hypotheses **H2a, H2b, H3b, and H3c** were tested using these scores. The Cronbach's alpha value of the affective trust questionnaire was 0.934 and of cognitive trust was 0.817.

2.1.3.8 Manipulation of the robot

As described in section 2.3.1, we concluded to design four listening behaviors for NAO. Listening behaviors included active or empathic behavior together with verbal and nonverbal components. Therefore, we required to know how active and empathic listening behaviors are different in verbal and nonverbal components. To accomplish this requirement, we first described about the types of verbal and nonverbal behaviors of listening behaviors which were applicable to the robot. Table 2.5 lists the summary of four experimental conditions in terms of listening behaviors**.** Figure 2.7, 2.8 and 2.9 shows the model of NAO's behavioral mode for each listening behavior.

2.1.3.8.1 Nonverbal components of listening behaviors

Several nonverbal behaviors contribute to conveying message and empathy in listening behaviors (refer to Table 2.2). However, due to the limitations of NAO, we relied on certain nonverbal behaviors: eye contact, head and body movements (gestures). Although, facial expression is universal means of communicating and expressing emotions, because NAO has an inelastic rigid face, the display of facial expressions was unfeasible.

Eye contact. Eye contact is a powerful cue for effective listening behavior through which we transfer a large part of the message. Different types of eye contact have been established depending on duration, direction, frequency and timing. Too little or averted gaze is associated with ignoring, lying and not listening, and evokes negative feelings in conversations (Ho et al., 2015). Eye gaze is also one of the most studied social behaviours in HRI research. The robots with appropriate eye gaze are more likely to be interacted by people and affected verbal comprehension, turn-taking, and joint attention (Johanson et al., 2021). Stanton and Stevens (2017) found that situational gaze (robot face to people and make eye contact to show disagreement), acted as averted gaze and was not trustworthy in HRI. Other research replicated that individuals hold constant gaze on their partners, with few averted gazes, while active listening. Furthermore, constant gaze was associated with increased positive valence in the positive and neutral conditions and with increased positive empathy ratings (Ho et al., 2015). Therefore, averted and situational gaze are associated with NAL, and constant gaze contributes to AL and AEL. Accordingly, we differentiated four experimental conditions in terms of eye contact as following:

NAL condition:

averted gaze= NAO did not look at participant while he/she was talking, and turned its head to left, right or top and down.

Situational gaze= NAO gazed at participants when expressed disagreement or negative statement.

AL condition:

constant gaze= NAO built constant and direct eye gaze with participants during conversation.

AEL condition:

constant gaze= NAO built constant and direct eye gaze with participants during conversation.

AEL_{VO} condition:

constant gaze= NAO built constant and direct eye gaze with participants during conversation.

Head movements. Head movements, mainly nodding and shaking, which signals "yes" and "no" answers or showing interest, confirmation, and attention, are a major mode of effective listening (Hadar et al., 1985). Turning head to different directions to avert eyes during conversation shows ignorance (McClave, 2000). Studies have found that head movements encode emotions during conversation. For example, downward-tilted rotational head to show sadness, head shaking to show regret or sadness, upward-nodding head to show happiness and surprise (Hadar et al., 1985). Hence, different head movements were designed for NAO according to McClave (2000), and their relation to listening behaviors. Head movements were also incorporated into suitable verbal expressions. Figure 2.10 shows different head movements of NAO in the experiment.

NAL condition:

NAO turned its head to right and left, top and down to show ignorance.

AL condition:

NAO nodded when saying "yes", "I see" and similar backchannels, and while participants talked, to show attention. NAO showed head shaking while saying "no" or similar disagreement statements.

AEL condition:

In addition to AL condition head movements, turning head down, and shaking downward-head to show sadness, upward-tilted and repetitive nodding to show happiness.

AEL_{VO} condition: No specific head movements.

Body movements. Except for the head, we could move other parts of the NAO, such as arms, hands, fingers, and kneels, to obtain an extensive number of movements. Body movements and postures can represent certain emotions and they are widely researched for human-like robots. A forward lean body posture, for example, is a salient way to show attention towards a speaker during social interactions, and is often recommended for use by clinicians to demonstrate attention or active listening towards patients (Johanson et al., 2021). Furthermore, certain nonverbal behaviors are positively associated with empathy including; head nodding, smiling, and inappropriate use of eye contact, body orientation away from the patient, and a backward lean indicates a lack of empathy (see (Johanson et al., 2021)). Different kinds of robot gestures have been shown to have different effects. Some studies (e.g., (Salem et al., 2015)) found that a robot using incongruent hand gesturing behaviour with its speech were rated higher by users in terms of likability, human-likeness, shared reality, and future contact intentions, compared to a robot using gestures congruent with its speech. We modulated the body movements of NAO according to the works of McColl and Nejat (2014) and Thimmesch-Gill et al. (2017), designing movements such as opening and closing arms, crossed arms, opening palms, and pointing finger. Some movements were synchronized with utterance, whereas others were free movements in reaction to participants to show the activeness and emotional state of the robot. Figure 2.11 and 2.12 shows different body movements of NAO for AL and AEL conditions in the experiment.

NAL condition:

NAO pretended to look at watch or playing with hands to show ignorance.

AL condition:

Appropriate movements along with verbal components: for example, NAO opened arm and hand when asking a question, pointed fingers toward itself when expressing its opinions.

AEL condition:

In addition to AL condition movements, NAO showed additional body movements in accordance with emotional statements: for example, NAO opened its arms together or moved arms up and down to show happiness and surprise.

AEL_{VO} condition: No specific body movements.

2.1.3.8.2 Verbal components of listening behaviors

Verbal components of listening behaviors in this study included two parts: 1) the main content of conversation including questions, answers, comments, attentive and emotional statements, 2) specific verbal techniques to show attention or emotions. Following is the main verbal techniques were used:

Backchannels. Active listeners give more feedback to speakers, which coordinates the speaker's narratives with what the listener needs to know. A common form of feedback in such settings is backchannels (Kraut et al., 1982), which is short acknowledgement utterances, such as "uh-huh" or "ummm," or nonverbal gestures, such as head nods or brief smiles (Johansson et al., 2016). Using backchannels encourages the speakers to continue, but this technique should be followed by asking other questions (Kobayashi et al., 2010). We used backchannels for AL, AEL and AEL_{VO} conditions.

Affirmative statements. Good listeners also use short positive statements to express agreement and understanding of the speaker's content and encourage them to continue their speech. These statements affirm or assert a particular idea, belief, or state of affairs. They are opposite to negative statements and are often used to convey approval, support, agreement, or confirmation. Affirmative statements contribute to a positive and constructive communication style. We used short affirmative statements like "yeah," "OK," "I see," "Really?", "oh," and "good." for AL, AEL and AEL_{VO} conditions.

Paraphrasing. Paraphrasing shortens and clarifies the speaker's statements (both content and feelings) by restating the information received in another from (Bauer & Figl, 2008; Ivey & Daniels, 2016; Kobayashi et al., 2010; Weger et al., 2010). Paraphrasing was used for AL, AEL and AEL_{VO} conditions. An example from AL condition is as follows:

NAO: *What did you find most difficult in Japan?*
Participant: *Maybe Japanese Language ....*
NAO: *Yeah, Japanese language specifically kanjis are difficult.*

Summarizing. Summarization is a concise overview of several statements, the content, or even feelings of a conversation (Ivey & Daniels, 2016) to bring together important ideas and establish a basis for further discussion (Bauer & Figl, 2008). Summarization was used for AL, AEL and AEL_{VO} conditions. An example from AEL condition is as follows:

NAO: *Have you travelled in Japan?*
Participant: *Yeah, I went to Kyoto last year. I wore traditional clothing of Japan and ate many foods.*
NAO: *So, it seems you enjoyed a lot.*

Asking. Clarifying questions that can be open or closed is part of good listening (Kobayashi et al., 2010). Most active listening treatments suggest that the listener asks questions to encourage the speaker to elaborate on his or her beliefs or feelings (Weger et al., 2010). We prepared some questions for NAO, for example like, "what did you do then?" for AL condition, or "did you like it?" for AEL condition. Other verbal techniques, such as *encouraging statements*, *demonstrating concern*, and *sharing similar emotion*s, were also used to design the verbal behaviors of NAO.

Moreover, according to the literature (see (Johanson et al., 2021)), verbal behaviors such as; expressions of understanding, statements of support, encouraging questions, positive reinforcement, humour, and information giving, show empathy. However, interrupting the speaker, chatting with other people, and ignoring questions are some of verbal behaviors that indicate a lack of empathy.

Table 2. 5 Summary and examples of verbal and nonverbal components of listening behaviors of robot in experimental conditions

behavior	Verbal		Nonverbal	
NAL	Limited verbal responses in addition to statements to show that NAO is not listening carefully, such as: - Sorry, I was distracted - I didn't get your point - Sorry I didn't understand what you said - Would you please repeat once again - I misunderstood		- Averted, situational gaze - Head down, right or left - Looking at watch - Playing with hands	
	active	**empathic**	**active**	**empathic**
AL	Attentive verbal responses and statements to show that NAO is actively listening, such as: - Yes, I see. - Yes, I agree with that - I know about it - I understand your point - Good	none	- Constant gaze - Nodding - Shaking head - Some arm and hand movements to ask questions or express opinions	none
AEL	Verbal responses to show understanding such as: - Yeah, that is true - Ok, great - Yes, I see. - Yes, I agree with that - Cool - I know about it - I understand your point - Interesting - Good	Emotional verbal responses and statements to show empathy, such as: - I also feel like you - Then, you enjoyed being here - Amazing, I know your feelings - Glad to hear that - You must feel so happy of it - You are emotionally strong - Don't feel regret about it	- Constant gaze - Nodding - Shaking head - Some arm and hand movements to ask questions or express opinions	- Turning head down, and shaking downward-head to show sadness - Upward-tilted and - Repetitive nodding to show happiness. - Open arms - Joint hands - Leaning forward - Swing arms - Bowed, wrap arms - Nudging
AELvo	Same as AEL condition	Same as AEL condition	Constant gaze	none

All video of the experiment are accessible at: https://osf.io/pqf69/?view_only=fa34adaae24343ef9431b04b7b8daee8

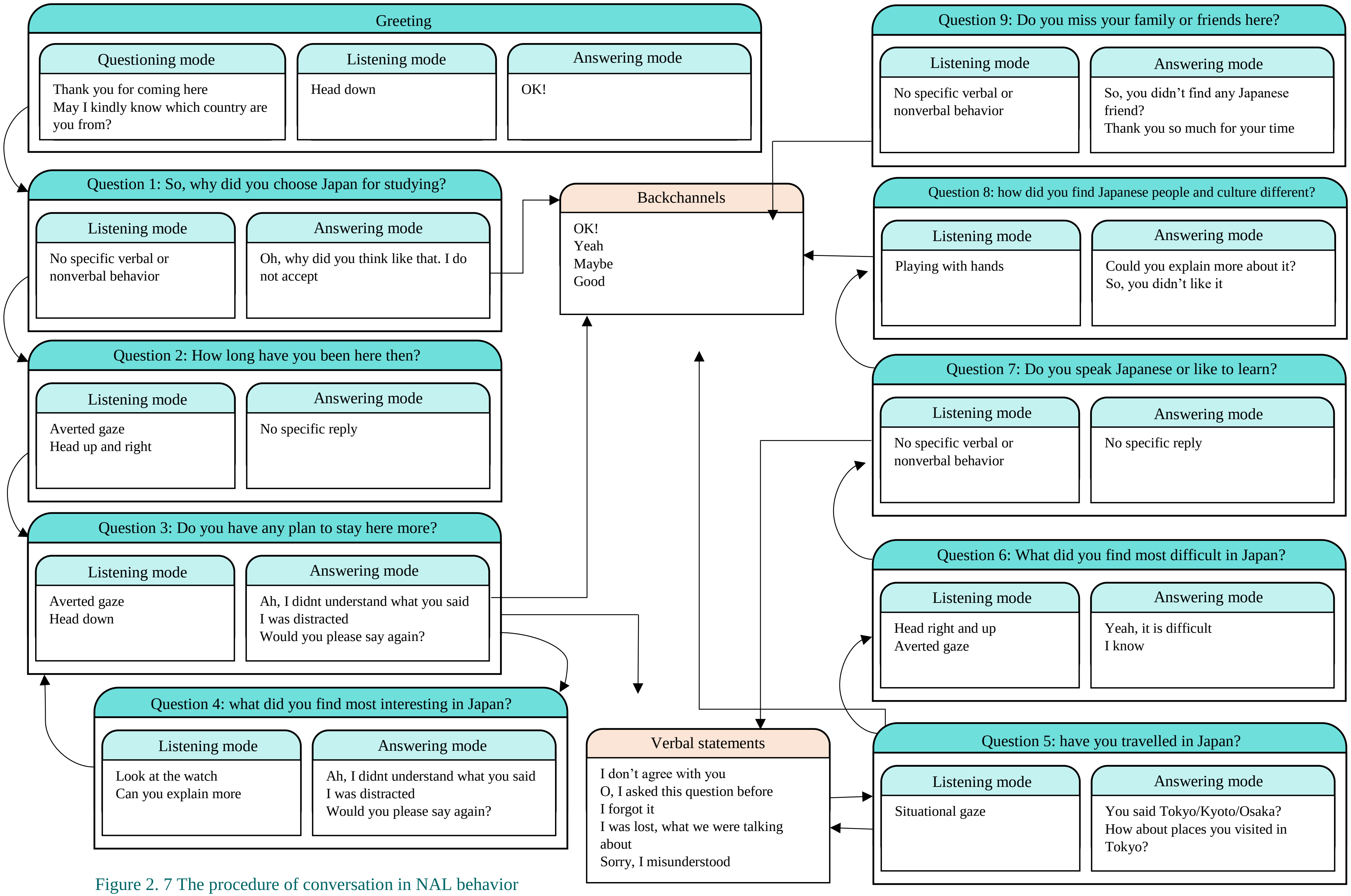


Figure 2. 7 The procedure of conversation in NAL behavior

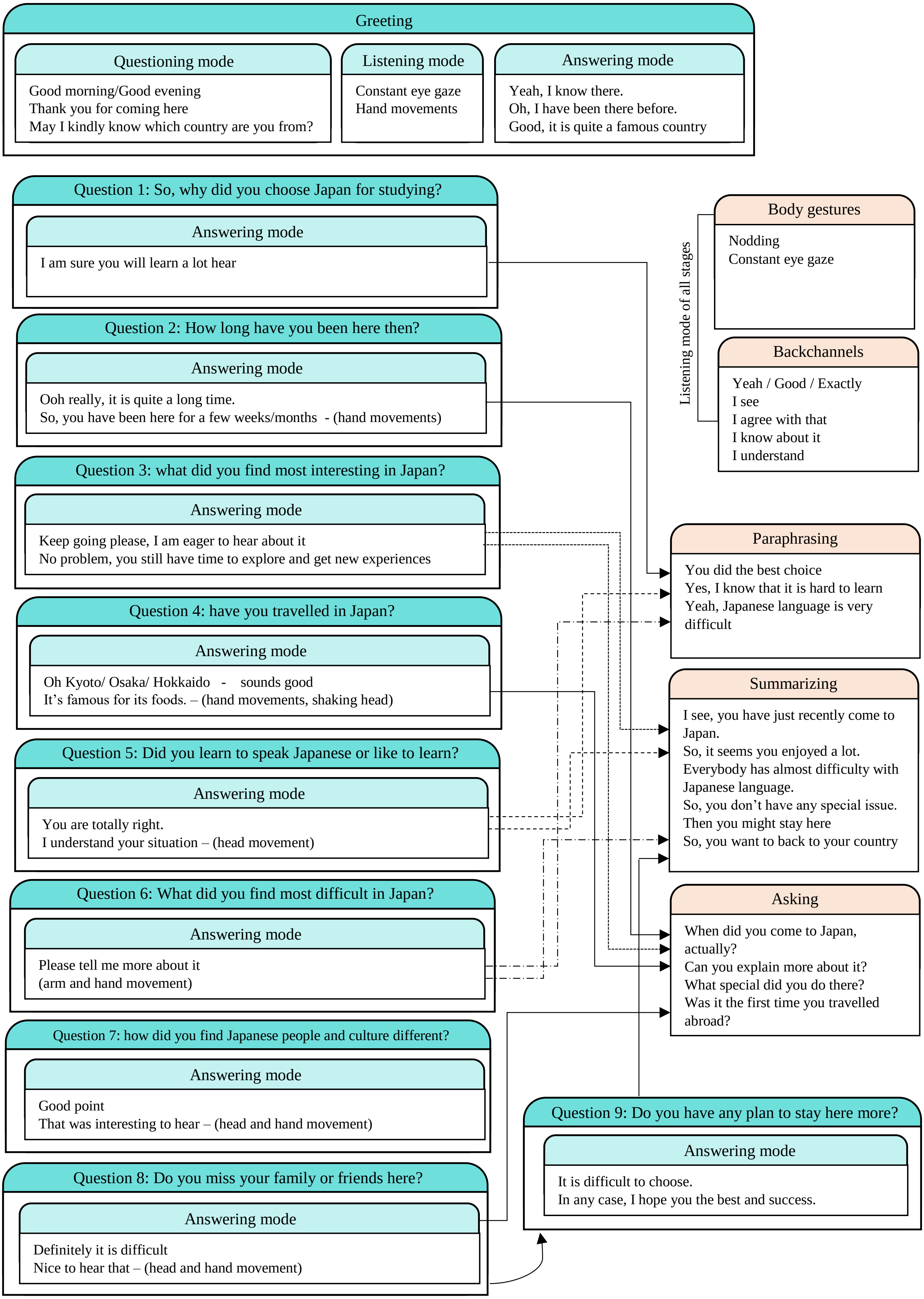


Figure 2. 8 The procedure of conversation in AL behavior

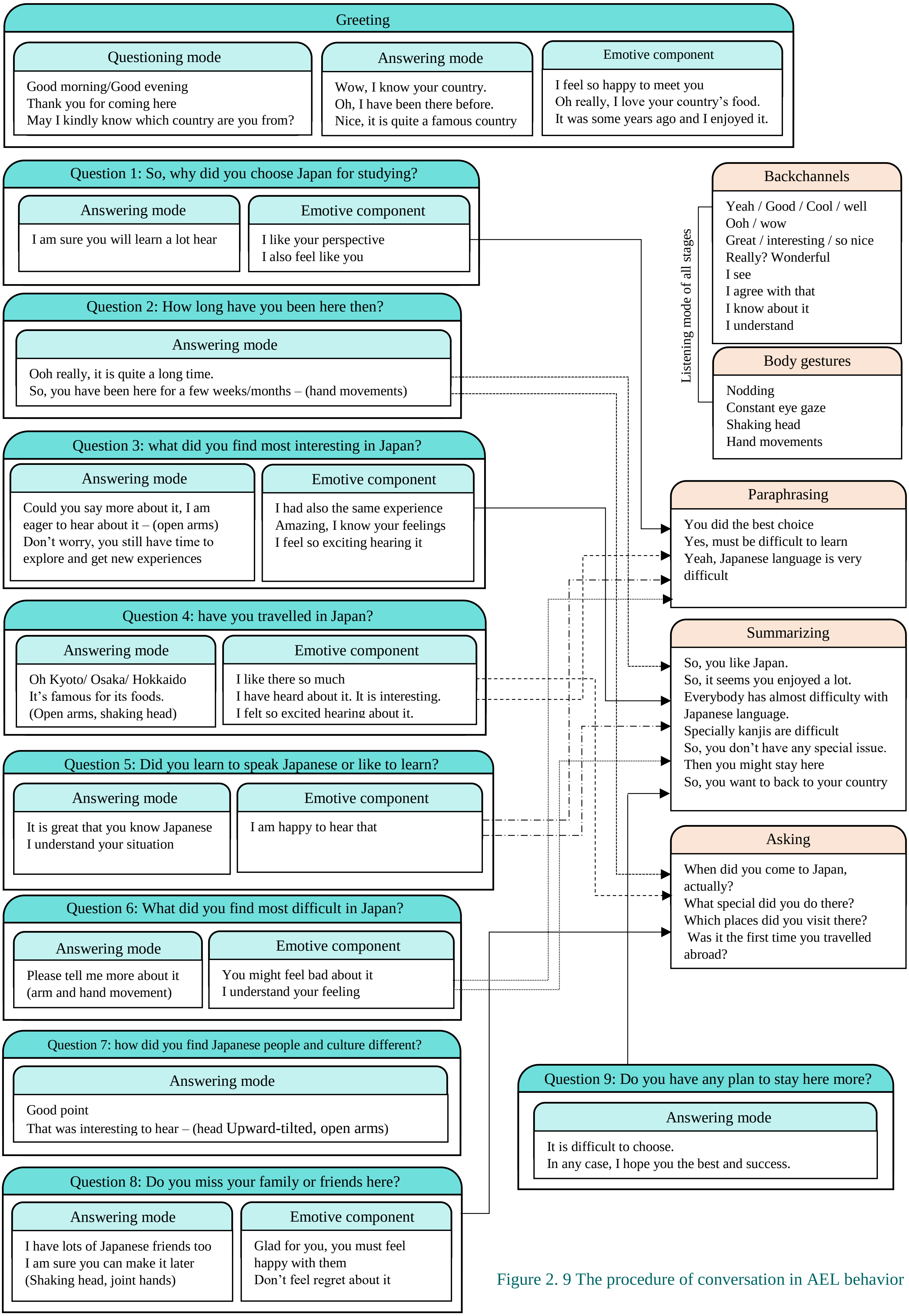


Figure 2. 9 The procedure of conversation in AEL behavior

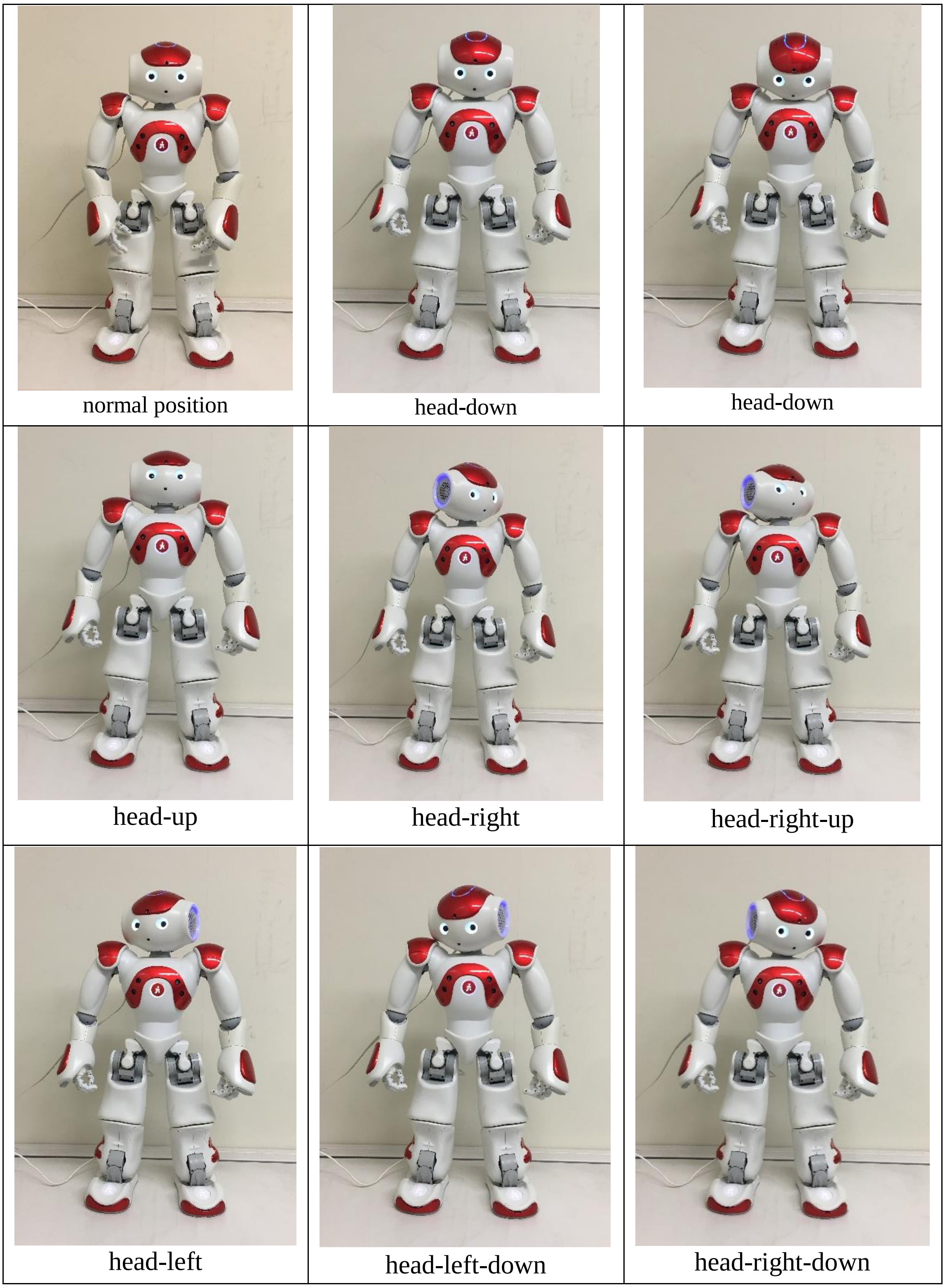


Figure 2. 10 Head movements and positions of NAO in the experiment.

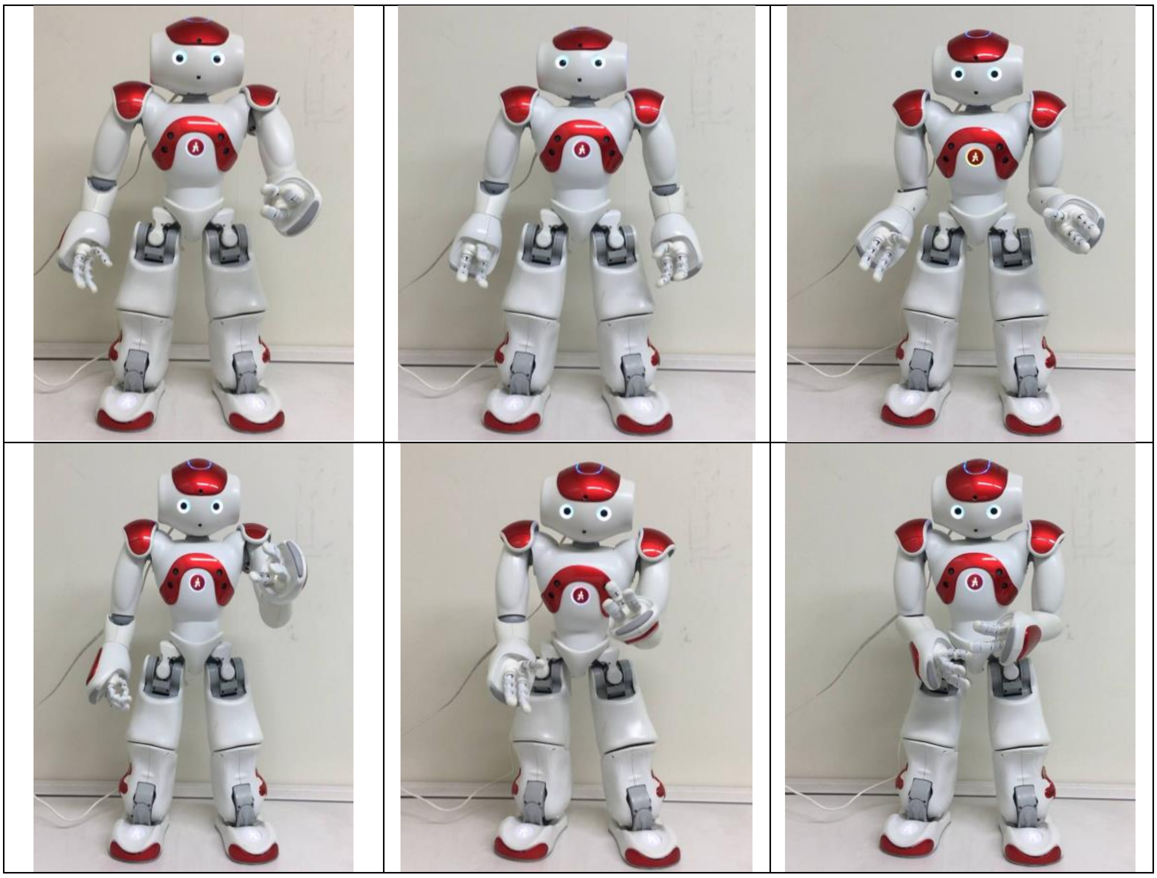


Figure 2. 11 Some examples of NAO body movements for AL condition.

Figure 2. 12 Some examples of NAO body movements for AEL condition.

2.1.4 RESULTS

2.1.4.1 Data analysis

After collecting data, we initially filtered the dataset to identify and address any suspicion of unreliable data. Fortunately, no unreliable data was found, and all data obtained from the 120 participants were used for subsequent statistical analysis. We completed a normality test and the results showed that sample data has been drawn from a normally distributed population. To identify the hidden dimensions of independent variables and to explain their relationships and effect on dependant variables, we conducted factor analysis. For the main analyses, a multivariate analysis of variance (MANOVA) with post-hoc analysis was conducted for the obtained variables from factor analysis to test the statistical significance of the effect of robot's different listening behaviors on different types of trust. Considering the mean differences, the scores for changes between before and after the interaction was compared for dependent variables. SPSS 19 was used to conduct statistical analyses.

2.1.4.2 Factor structure of trusts

To examine the influence of difference in the listening behaviors of social robots on perceived trust, we first conducted exploratory factor analysis to aggregate the 40 items in general trust, 9 items in cognitive trust, and 7 items in affective trust scale into discrete dimensions. First, the factorability of the 40 items was examined. The Kaiser–Meyer–Olkin measure of sampling adequacy was 0.908, above the recommended value of 0.6, and Bartlett's test of sphericity was significant ($\chi^2(780) = 5329.844$, $p < 0.001$). Therefore, factor analysis was considered suitable. The maximum likelihood method with promax rotation revealed an eight-factor structure. The cumulative contribution of these eight factors was 62.72%. These factors were considered to represent the dimensions of people's general trust in robots. Four of the 40 items were eliminated because they did not meet the minimum criteria of having a primary factor loading (how much a factor explains a variable) of 0.35 or above. Table 2.6 shows the factor loading matrix for the final items.

Variables with high loadings on the first factor contained responses to items such as "act as part of the team," "work best with a team," "be considered part of the team," "be a good teammate," "protect people," and "be reliable." These items represented people's expectations for robots to work in collaboration with people and work as a team; thus, they were labeled as "team working."

The second factor contained items such as "be lifelike," "possess adequate decision-making capability," "be conscious," "know the difference between friend and foe," "perform a task better than a novice human user" and "provide feedback," as well as "be autonomous." The variables of this factor are mostly related to how much consciousness the robot has and can act in an intelligent way. Thus, the factor was labeled as "Intelligence"

The third factor included responses to items such as "malfunction (reversed)," "have errors (reversed)," "require frequent maintenance (reversed)," "be unresponsive (reversed)," "be incompetent (reversed)," and "be led astray by unexpected changes in the environment." This factor represented people's expectations that robots should work free of troubles; thus, it was labeled as "Trouble-free"

Items loaded for the fourth factor included "be pleasant," "be friendly," and "be supportive," and it represented the measure of positive feelings and feedback the users received from the robot. Thus, the factor was labeled as "Likeability."

Items loaded for the fifth factor included "act consistently," "openly communicate," "clearly communicate," "function successfully," and "communicate with people." These items represent people's expectations regarding the robot's appropriate function and communication; thus, it was labeled as "Function."

The sixth factor contained items such as "tell the truth," "warn people of potential risks in the environment," "keep classified information secure," and "be responsible." These items represent the extent to which the robot is safe in maintaining information and exhibits reliable behavior. Thus, it was labeled as "Reliability."

Variables that had high loadings on the seventh factor included "perform exactly as instructed," "follow directions," and "be predictable", which were about instructing the robot. Thus, this factor was labeled as "Control."

The eighth factor contained two items, "make sensible decisions" and "possess adequate decision-making capability", which represented people's expectations of robots making proper decisions. Thus, the factor was labeled as "Decision making."

The internal reliability of the eight factors was tested by calculating the Cronbach's alpha indices. Alpha values for the eight factors were 0.91 (team working), 0.79 (intelligence), 0.85 (trouble-free), 0.82 (likeability), 0.81 (function), 0.67 (reliability), 0.83 (control), and 0.82 (decision-making), which indicated high reliability.

In addition, factor analysis was conducted for responses on the cognitive trust scale using the maximum likelihood method with promax rotation. A two-factor structure was revealed, with a cumulative contribution of 51.99%. The Kaiser–Meyer–Olkin measure of sampling adequacy was 0.810, above the recommended value of 0.6, and Bartlett's test of sphericity was significant ($\chi^2(36) = 591.17$, $p < .001$). Table 2.7 shows the factor loading matrix for final items of cognitive trust.

These two factors represented the dimensions of people's cognitive trust in social robots. Item 16 "If people knew more about this robot, they would be more concerned and monitor its performance more closely" was eliminated because it did not meet minimum criteria of having a primary factor loading of 0.35 or above. There were no items with a cross-loading of 0.35 or above.

The first factor contained responses to items of "other people who must interact with NAO consider it to be trustworthy," "most people, including those who are not familiar with NAO, trust and respect NAO," "I can rely on NAO to undertake a thorough analysis of the situation before advising me," "this robot approaches its duty with professionalism and dedication," and "when interacting with NAO, I have no reservation about acting on its advice." This factor represented people's understanding and perception of the robot. Thus, this factor was labeled as "Social reputation."

The second factor contained responses to items such as "I have to be cautious about acting on the advice of NAO because its opinions are questionable (reversed)," "I cannot confidently depend on NAO because it may complicate my affairs by careless behavior (reversed)," and "when interacting with NAO, I have good reason to doubt its competence (reversed)." This factor represented personal trust in the robot and was labeled as "Personal credit." Cronbach's alpha values for these two factors were 0.87 (social reputation) and 0.80 (personal credit), which indicated good internal reliability.

Finally, factor analysis for responses for the affective trust scale with the maximum likelihood method and promax rotation revealed a single–factor structure, in which all seven questions met the minimum criteria of having a primary factor loading of 0.35 or above; thus, no question was eliminated. The one-factor cumulative contribution was 61.79%, and Cronbach's alpha value of 7 questions was 0.89.

Table 2. 6 Factor loadings for 36 items of human robot scale (N = 120)

	Factor							
	F1	F2	F3	F4	F5	F6	F7	F8
Q3 act as part of the team	0.840							
Q18 work best with a team	0.799							
Q26 be considered part of the team	0.798							
Q39 be a good teammate	0.746							
Q2 protect people	0.634							
Q32 be reliable	0.451	0.352						
Q38 be lifelike		0.764						
Q13 possess adequate decision-making capability		0.699						0.355
Q37 be conscious		0.651						
Q11 know the difference between friend and foe	0.380	0.521						
Q10 perform a task better than a novice human user		0.478						
Q12 provide feedback		0.463						
Q35 be autonomous		0.421						
Q5 malfunction (reversed)			0.760					
Q9 have errors (reversed)			0.718					
Q7 require frequent maintenance (reversed)			0.684					
Q34 be unresponsive (reversed)			0.618					
Q29 be incompetent (reversed)			0.577					
Q40 be led astray by unexpected changes in the environment			-0.505					
Q33 be pleasant				0.831				
Q31 be friendly				0.807				
Q28 be supportive				0.615				
Q1 act consistently					0.657			
Q8 openly communicate					0.608			
Q6 clearly communicate					0.515			
Q4 function successfully					0.481			
Q17 communicate with people					0.352			
Q23 tell the truth						0.718		
Q14 warn people of potential risks in the environment						0.423		
Q19 keep classified information secure						0.423		
Q27 be responsible						0.360		
Q20 perform exactly as instructed							0.722	
Q25 follow directions							0.708	
Q36 be predictable							0.431	
Q21 make sensible decisions		0.375						0.553

Note. Factor loadings with absolute value <.35 are suppressed

Table 2. 7 Factor loadings for 8 items of cognitive trust scale (N = 120)

	Factor	
	F1	F2
Q15 Other people who must interact with NAO, consider it to be trustworthy	0.879	
Q14 Most people, even those who are not familiar with NAO, trust and respect NAO	0.788	
Q10 I can rely on NAO to undertake a thorough analysis of the situation before advising me	0.462	
Q13 This robot approaches its duty with professionalism and dedication	0.440	
Q8 Interacting with NAO, I have no reservation about acting on its advice	0.399	
Q11 I have to be cautious about acting on the advice of NAO because its opinions are questionable (reversed)		0.923
Q12 I cannot confidently depend on NAO since it may complicate my affairs by careless behavior (reversed)		0.600
Q9 Interacting with NAO, I have good reason to doubt NAO's competence (reversed)		0.469

Note. Factor loadings with absolute value <.35 are suppressed

Hence, for the rest of analyses, eight variables resulted from factor analysis, in addition to sum of 40 items as overall general trust score was used to evaluate the general trust of participants to the robot. Two variables resulted from factor analysis of cognitive trust scale, in addition to average of responses to 9 items (overall cognitive trust) was used to measure the cognitive trust toward the robot. The average of responses to 7 items was used to measure the affective trust toward the robot.

2.1.4.3 Descriptive statistics and test of results

To investigate the effect of different listening behaviors of the robot on the perceived trust of participants, changes in the trust scores before and after conversation were calculated and compared across four experimental conditions. Table 2.8 shows the means and standard deviations (SDs) of changes before and after conversation in all thirteen trust scores for four experimental conditions. AEL condition had the highest score than other conditions in all trust scores. NAL condition had the lowest score, and the mean score of AL condition was higher than that for NAL condition and moderately lower than that for AEL_{VO} condition in most trust scores. Therefore, AEL was the most stimulating behavior for participants to deduce the trustworthiness of the robot's behavior.

Table 2. 8 Summary of means and standard deviations (SDs) of all trust scores (changes before and after conversation) for four experimental conditions.

General trust scores	NAL Condition		AL Condition		AEL Condition		AEL_{VO} Condition	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Team working	-0.73	1.11	0.01	0.66	0.57	0.75	0.37	0.86
Intelligence	-0.46	1.09	0.29	0.75	0.81	0.78	0.62	0.92
Trouble-free	-0.71	0.99	0.32	0.73	0.45	0.93	0.48	0.61
Likeability	-0.51	1.10	0.51	0.84	0.90	0.79	0.49	0.78
Function	-0.71	1.08	0.65	0.92	0.82	0.89	0.42	0.92
Reliability	-0.66	0.95	-0.15	0.86	0.46	0.96	0.15	0.75
Control	-0.90	1.02	-0.06	0.84	-0.003	1.03	-0.17	0.63
Decision making	-0.002	0.85	0.29	0.93	0.56	0.81	0.50	0.93
Overall general trust	-43.10	64.67	17.33	42.31	48.80	45.08	25.90	38.32
Cognitive and Affective trust scores								
Social reputation	-0.45	1.22	0.31	0.79	0.70	0.81	0.32	0.71
Personal credit	-0.37	1.35	0.14	1.00	0.42	0.84	0.41	0.75
Overall cognitive trust	-0.52	0.97	0.12	0.75	0.48	0.62	0.31	0.55
Affective trust	-0.57	1.37	0.63	1.02	1.63	1.15	0.69	1.41

For testing the hypotheses, a one-way multivariate analysis of variance (MANOVA) was conducted with four experimental conditions (NAL, AL, AEL, AEL_{VO})

as independent variables and thirteen trust scores as dependent variables, including: teamworking, intelligence, trouble-free, likeability, function, reliability, control, decision-making, overall general trust; social reputation, personal credit, overall cognitive trust; and affective trust, resulted from factor analysis. The results indicated a significant difference in trust scores among different listening behaviors of the robot [F (39,305) = 2.60, $p < 0.001$, Wilk's $\Lambda = 0.42$, $\eta^2 = 0.246$]. The multivariate effect size was estimated at 0.246, which implies that almost 25% of the variance in the canonically derived dependent variable was accounted for by listening behaviors. The next step was to determine the source of the differences.

MANOVA was followed by analysis of variance (ANOVA) on each of the thirteen dependent variables, and with Bonferroni correction measured the individual mean difference comparisons across the conditions, and effect sizes were measured by Cohen's *d*. As can be seen in Table 2.9, all of the ANOVA results, except for decision-making, were statistically significant, with effect sizes (partial η^2) ranging from the lowest one 0.09 (personal credit) which indicated a medium effect, to the highest ones of 0.33 (overall general trust) and 0.30 (affective trust) which indicated a large effect.

Table 2. 9 ANOVA results for all trust scores among four experimental conditions.

	df	*F*	partial η^2	*p*
Team working	3	13.24	0.25	0.000***
Intelligence	3	11.68	0.23	0.000***
Trouble-free	3	14.08	0.27	0.000***
Likeability	3	13.66	0.26	0.000***
Function	3	15.47	0.28	0.000***
Reliability	3	8.65	0.18	0.000***
Control	3	6.49	0.14	0.000***
Decision making	3	2.52	0.06	0.061
Overall general trust	3	18.88	0.33	0.000***
Social reputation	3	8.45	0.18	0.000***
Personal credit	3	3.87	0.09	0.011*
Overall cognitive trust	3	10.46	0.21	0.000***
Affective trust	3	16.86	0.30	0.000***

Statistically significant difference: ***$p < 0.001$, *$p < 0.05$

2.1.4.4 Test of hypotheses

2.1.4.4.1 Effect of NAL, AL and AEL behavior on general trust in HRI

The first set of hypotheses explored the difference among NAL, AL and AEL behaviors of the robot in general trust. Table 2.10 indicates where the significant group differences for general trust scores reside. Bonferroni post-hoc test results revealed that the score for overall general trust, was significantly different between NAL (M = -43.10, SD = 64.67) and AL conditions (M = 17.33, SD = 42.31, $p < 0.001$), and between NAL and AEL conditions (M = 48.80, SD = 45.08, $p < 0.001$). However, the difference in the score of overall general trust between AEL and AL conditions was not significant ($p = 0.108$). Therefore, **H1a** and **H1b** were fully supported and **H1c** was not supported. A very large effect sizes were observed for both comparisons of AL and NAL conditions and AEL and NAL conditions**.** Figure 2.13 panel A shows the means and SDs of changes in overall general trust score before and after the conversation in NAL, AL and AEL conditions.

Table 2. 10 Bonferroni post-hoc test results and effects size by Cohen's d for group differences between NAL, AL and AEL conditions.

	AL versus NAL condition		**AEL versus NAL condition**		**AEL versus AL condition**	
	p	Cohen's *d*	*p*	Cohen's *d*	*p*	Cohen's *d*
Team working	0.006**	0.821	0.000***	1.381	0.087	0.790
Intelligence	0.009**	0.800	0.000***	1.346	0.163	0.684
Trouble-free	0.000***	1.651	0.000***	1.208	1.000	0.155
Likeability	0.000***	1.050	0.000***	1.469	0.602	0.471
Function	0.000***	1.351	0.000***	1.545	1.000	0.193
Reliability	0.172	0.557	0.000***	1.175	0.052	0.674
Control	0.003**	0.894	0.001**	0.877	1.000	0.070
Decision making	1.000	0.330	0.088	0.683	1.000	0.311
Overall general trust	0.000***	1.105	0.000***	1.625	0.108	0.693
Social reputation	0.009**	0.739	0.000***	1.110	0.577	0.493
Personal credit	0.302	0.433	0.026*	0.682	1.000	0.271
Overall cognitive trust	0.006**	0.742	0.000***	1.230	0.383	0.520
Affective trust	0.001***	0.997	0.000***	1.722	0.011*	0.903

Statistically significant difference: ****p < 0.001*, ***p < 0.01*, **p < 0.05*

Additionally, Bonferroni post-hoc analysis revealed significant differences between NAL and AL behaviors of the robot for six general trust sub-dimensions with medium to very large effect sizes: teamworking ($p < 0.01$), intelligence ($p < 0.01$), trouble-free ($p < 0.001$), likeability ($p < 0.001$), function ($p < 0.001$), and control ($p < 0.01$). Although participants evaluated AL as higher for reliability and decision-making factors than NAL, the difference was not statistically significant for reliability (p = 0.172), or decision-making (p = 1.00).

For difference among the general trust factors between NAL and AEL conditions, Bonferroni post-hoc test indicated significant difference in seven sub-dimensions of general trust with very large effect sizes in most of factors: teamworking (p < 0.001), intelligence (p < 0.001), trouble-free (p < 0.001), likeability (p < 0.001), function (p < 0.001), reliability (p < 0.001), and control (p < 0.01). However, the results did not show significant difference for decision-making (p = 0.088), between the robot with NAL and AEL. Thus, participants perceived AEL behavior of the robot to be more acceptable and effective in most of general trust factors than NAL.

Finally, Bonferroni post-hoc evaluation revealed only a moderate significant difference for reliability (p = 0.052) between AEL and AL conditions. However, the difference in other sub-dimensions of general trust, including team working (p = 0.087), intelligence (p = 0.163), trouble-free (p = 1.00), likeability (p = 0.602), function (p = 1.00), control (p = 1.00), and decision-making (p = 1.00) was not significant, despite showing a higher mean for AEL. Figure 2.13 panels B-I shows the means and SDs of changes in factor scores of the eight sub-dimensions of general trust before and after the conversation in NAL, AL and AEL conditions.

To summarize, the robot with the behavior of AEL and AL was perceived more trustworthy than a robot with the behavior of NAL almost in all trust dimensions, and the potency of AEL and AL in HRT was proved. However, empathic behaviors of AEL condition were not influential enough enhancing general trust in comparison with AL condition.

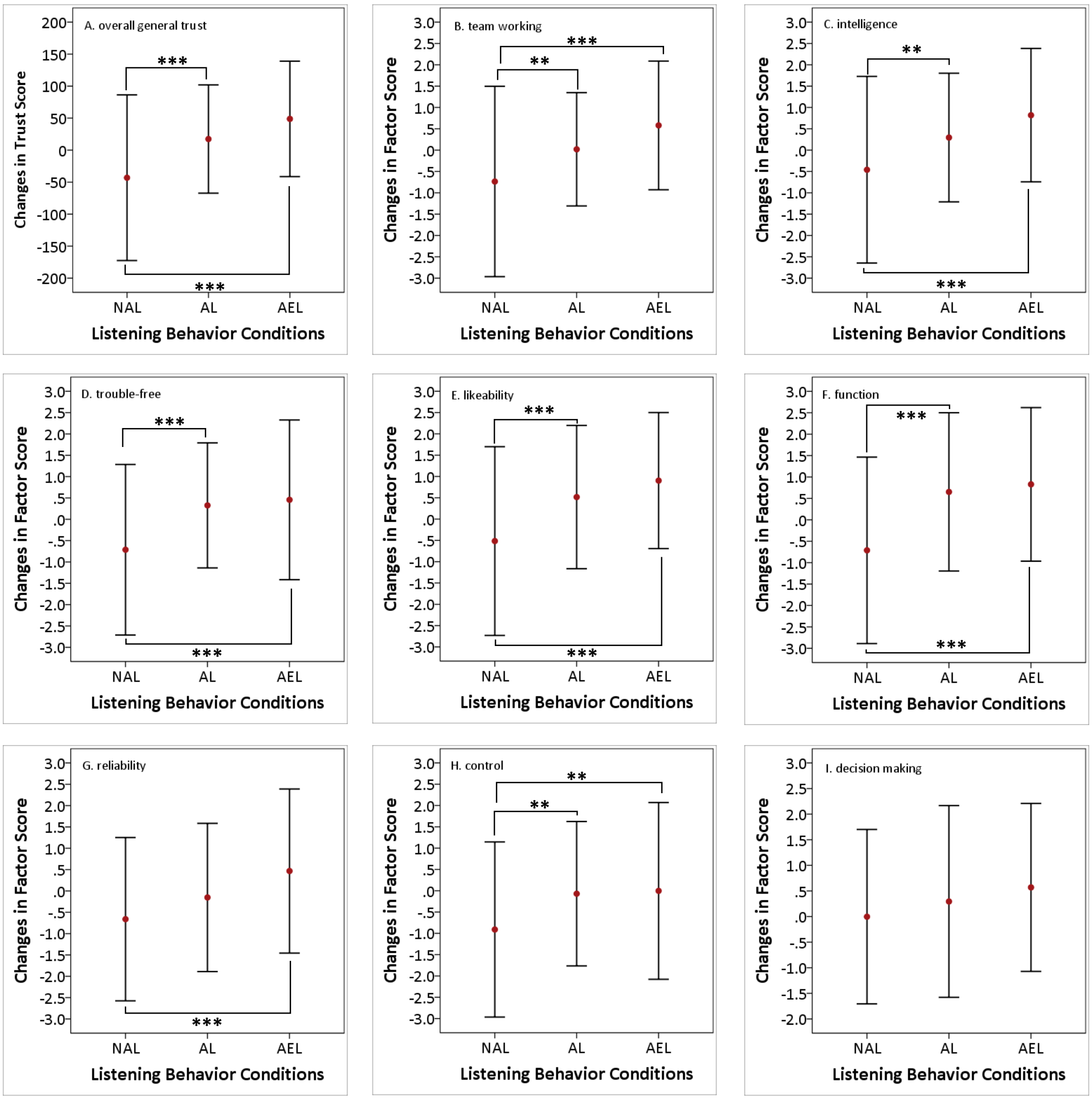


Figure 2. 13 Changes in overall general trust score (panel A), and factor scores (panel B-I) for the eight sub-dimensions of general trust in NAL, AL and AEL conditions. Dots indicate means and vertical lines indicate standard deviations. (****p< 0.001*, ***p < 0.01*)

2.1.4.4.2 Effect of AL and AEL behavior on cognitive and affective trust in HRI

The second set of hypotheses explored the relationship between AL and AEL behaviors of the robot on cognitive and affective trust. As shown in Table 2.10, Bonferroni post-hoc test indicated that there was a significant difference in the score of affective trust between AL and AEL conditions, with AEL condition (M = 1.16, SD = 1.15) scoring discernibly higher than AL condition (M = 0.63, SD = 1.02, p < 0.05). Therefore, **H2a** was fully supported. The cohen's *d* indicated a large effect size in affective trust between

AL and AEL conditions. Figure 2.14 shows the mean and SD of affective trust for AL and AEL conditions.

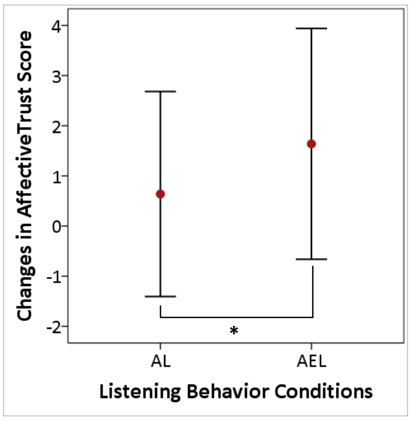


Figure 2. 14 Changes in affective trust score for AL and AEL conditions. Dots indicate means and vertical lines indicate SDs. (*$p < 0.05$)

For overall cognitive trust, although AEL behavior of the robot resulted in a higher mean score than AL, the results of post-hoc analysis indicated that there was not a significant difference comparing AEL condition (M= 0.48, SD = 0.62) and AL condition (M = 0.12, SD = 0.75, p = 0.383). Factor analysis elicited two factors for cognitive trust, namely social reputation, and personal credit. Participants measured AEL condition (M = 0.42, SD = 0.84) higher than AL condition (M= 0.14, SD = 1.00) in personal credit, and similarly they scored AEL condition (M = 0.70, SD = 0.81) higher than AL condition (M = 0.31, SD = 0.79) in personal credit. However, there was not any significant difference for social reputation (p = 0.577) or personal credit (p = 1.00) between AEL and AL conditions. Thus, **H2b** was not supported. Table 2.10 shows the results for cognitive trust scores between AEL and AL conditions. In summary, the difference of AEL and AL conditions only in affective trust revealed the effectiveness of empathic behaviors of the robot on affective trust, but not on general or cognitive trust.

2.1.4.4.3 Effect of AEL and AEL$_{VO}$ behavior on HRT

According to H3a, AEL behavior which consisted of both verbal and nonverbal components could result in higher trust than the behavior of AEL$_{VO}$, which only included verbal components. As shown in Table 2.11, Bonferroni post-hoc test results indicated that there was not a significant difference in the score of overall general trust between AEL condition (M = 48.80, SD = 45.08) and AEL$_{VO}$ condition (M =25.90, SD =38.32, p = 0.523), even though AEL condition was scored considerably higher than AEL$_{VO}$

condition. Therefore, **H3a** was not supported and nonverbal behaviors could not shape a large difference for participants' general trust perception of the robot. Furthermore, no significant difference was found between the AEL and AEL_{VO} conditions, when individually considering the sub-dimensions of general trust.

Table 2. 11 Bonferroni post-hoc results and effects sizes by Cohen's d for group differences between NAL, AL, AEL and AEL_{VO} conditions.

	AEL_{VO} versus NAL condition		**AEL_{VO} versus AL condition**		**AEL versus AEL_{VO} condition**	
	p	Cohen's d	*p*	Cohen's d	*p*	Cohen's d
Team working	0.000***	1.115	0.673	0.465	1.000	0.248
Intelligence	0.000***	1.077	0.911	0.340	1.000	0.220
Trouble-free	0.000***	1.446	1.000	0.238	1.000	-0.037
Likeability	0.000***	1.047	1.000	-0.031	0.481	0.519
Function	0.000***	1.126	1.000	-0.243	0.675	0.439
Reliability	0.003**	0.946	1.000	0.377	1.000	0.360
Control	0.012*	0.863	1.000	-0.137	1.000	0.196
Decision making	0.166	0.571	1.000	0.299	1.000	0.069
Overall general trust	0.000***	1.297	1.000	0.212	0.523	0.520
Social reputation	0.007**	0.777	1.000	0.023	0.667	0.492
Personal credit	0.021*	0.715	1.000	0.296	1.000	0.013
Overall cognitive trust	0.000***	1.066	1.000	0.295	1.000	0.281
Affective trust	0.000***	0.989	1.000	0.055	0.020*	0.791

Statistically significant difference: ****p < 0.001*, ***p < 0.01*, **p < 0.05*

H3b predicted that affective trust was significantly higher when a robot showed AEL behavior than AEL_{VO} behavior. The results of post-hoc analysis, as shown in Table 2.11, indicated that the participants assessed the robot with AEL behavior (M = 1.63, SD= 1.15) as having higher affective trust than AEL_{VO} behavior (M= 0.69, SD= 1.18, $p < 0.05$). Thus, the difference between AEL and AEL_{VO} behaviors of the robot for affective trust was significant, and nonverbal behaviors in the AEL condition elicited higher affective trust. Figure 2.15 shows the mean and SD of affective trust for AEL and AEL_{VO} conditions. Thus, **H3b** was verified based on these results.

For overall cognitive trust, however, no significant difference was detected when comparing AEL condition (M =0.48, SD =0.62) and AEL_{VO} condition (M =0.31, SD =0.55, p = 1.00). Similarly, there was not significant difference for cognitive trust factors (social reputation and personal credit) between AEL and AEL_{VO} conditions. Therefore, the evidence was not statistically efficient in supporting **H3c**. Table 2.11 shows the scores of cognitive trust for the comparison between AEL and AEL_{VO} conditions.

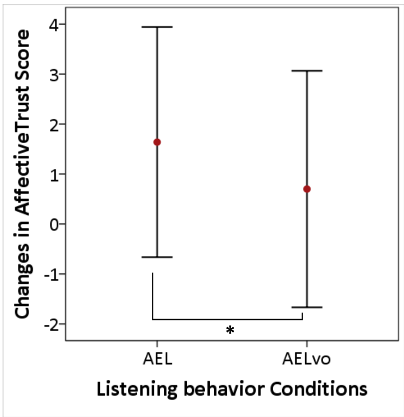


Figure 2. 15 Changes in affective trust score for AEL and AELVO conditions. Dots indicate means and vertical lines indicate SDs. (*$p < 0.05$)

Further analyses have been done comparing AEL_{VO} condition with NAL and AL conditions to ensure the effectiveness of nonverbal behaviors on trust perception in HRI. According to the results shown in Table 2.11, the behavior of AEL_{VO} was significantly different from the behavior of NAL in all trust scores with large effect sizes, except for decision-making. However, there was not any significant difference in trust scores comparing AEL_{VO} and AL, even for affective trust. Thus, the result could confirm the impact of nonverbal behaviors of AEL on affective trust, more confidently.

2.1.4.5 Qualitative findings

The participants were asked to describe their overall experience and perception of NAO robot's listening behavior, immediately after the interaction through open questions. The aim was to capture self-reported evaluations of the robot's behaviors which may not be achievable through questionnaires.

Regarding the robot's NAL behavior, participants expressed feeling bored during conversations with NAO. Its interactions appeared mechanical and failed to emulate human-like conversation.

> *"The questions NAO asked I was bored, and I felt it was static and mechanical."*
>
> *"He is very artificial and still felt like mechanical, so in order to share feelings, he needs to be upgraded."*
>
> *".... It was a one way talking, so I was not comfortable to talk with NAO."*

For the AL robot, participants primarily highlighted NAO's positive behavior in listening and conveying a sense of reliability and friendliness. However, participants did not feel any emotional connection or sharing with the robot.

"NAO was responsive toward the conversation and he was polite and well-mannered."

"… I do not really accept that a robot could share feelings but he was responsive and friendly."

"NAO was listening to my experiences and conversation. I think if NAO become more realistic it will be easier to talk and share personal feelings."

Regarding the AEL robot, the majority of participants found the robot communicated very effectively;

"Talking with NAO made me feel relaxed, …."

However, some participants doubted if the robot truly understood what they were saying. This indicates that, despite robots behaving like attentive listeners, there may still be a sense of artificiality. People need to acknowledge that robots can genuinely understand and perceive their messages.

"He seemed like caring about my words but I really do not know if he really understood what I was saying."

"… I'm not sure if NAO's responses and shared feelings are independent or if they're just programmed to follow instructions."

Participants had also positive perception of NAO's gestures and nonverbal behaviors;

"I liked NAO's gestures which looked like human beings. He was responsive and caring."

"…. His gestures are cute and natural."

However, some participants felt uncomfortable with NAO's body movements;

"Sometimes, NAO lowered his head with a specific gesture, I felt he was not friendly and even invasive. I just felt that he wanted to attack me, …."

Finally, for the AEL_{VO} robot, most participants agreed that NAO listened carefully and was responsive, effectively sharing feelings. However, participants did not care about NAO's static nonverbal behaviors and primarily focused on the content of the conversation.

"NAO was a good listener, analyzed my situation carefully and gave good advice."
"NAO was clearly emotionally involved in the conversation, it was compassionate and seemed to care about what I said."
".... It was interesting that it gave the speaker more priority when it comes to listening"

2.1.5 DISCUSSION

Designing trustworthy social robots using listening behavioral strategies has not been widely discussed in HRI. There is insufficient empirical evidence on how social robots' listening behaviors can shape their trustworthiness. Therefore, we investigated the effect of social robots' listening behaviors on the perception of trust. It was suspected that AEL behavior of a social robot could elicit higher trust than other types of listening behaviors. Additionally, AL behavior of social robot was more effective in evoking trust than NAL. Moreover, it was expected that a social robot with dual communication of verbal and nonverbal listening behavior would be perceived as more trustworthy than a robot that contains the same importance of verbal communication.

2.1.5.1 Human-robot trust and listening behaviors

As expected, AL and AEL behaviors of social robots were both successful in fostering trust in various domains of general, affective and cognitive trust. AL was found to affect the evaluation of general trust during interaction with the robot compared with NAL (H1a). The difference in participants' trust perception of the robot before and after interaction was more toward the robot with the behavior of AL than the robot with the behavior of NAL. This result is consistent with psychological evidence on interpersonal trust (Lasky, 2000; Lester, 2002; Nugent & Halvorson, 1995), and assures us that to make a more trustworthy robot, we can associate it with AL. Trust construction is based on showing the competency of other parties (Mayer et al., 1995), and as AL is attributed to care and attention to the speaker, it leads to trustworthiness. The results confirmed that users act toward the robot as they behave with their fellows. AL was also found to affect all general trust factors, except decision-making and reliability. For participants' evaluations of decision-making, the mean scores were in line with our predictions. However, these differences were not statistically significant. Decision-making did not result in significant difference in any of comparisons. It could be because, decision-making is the process of selecting from alternatives to be acted upon in the future to attain certain goals and it is always related to place, situation, and time (Slovic et al., 2012). It needs a problem or goal to make certain decisions based on values and weighing the evidence. The interaction between the robot and user in the current study was limited to a short conversation, and it did not support any circumstances to make critical decisions that could be a possible reason for the ineffectiveness of listening behaviors on decision-making factor. The factor reliability in this study referred to questions about keeping information safe and warning people of risks. No secure or individual information was exchanged between participants and robot in this study, that explains why significant difference was not found for this factor. Function and trouble-free were evaluated with the highest effect size in comparison between AL and NAL (Table 2.10). As it was mentioned in section 3.7.1, the general trust scale developed by Schaefer (2016), is mostly performance based and these two factors also referred to more functional attributes of the social robot, that could be the reason behind of receiving high effectiveness in trust perception.

The results also confirmed that AEL behavior of social robot was considerably more trust-evoking during interaction than NAL behavior of the social robot (H1b). In the field of HC, effective listeners that show empathy and friendship generally project more positive impressions, as they are perceived to be more trustworthy, friendly, or attractive (Weger et al., 2014). Emotional sharing and responsiveness both verbally and nonverbally advance the formation of interpersonal trust and even rebuild the damaged trust because emotions guide people's behavioral propensities (Ma et al., 2018). Therefore, the behavior of AEL, which is ascribed to emotional expressions, was found to improve trust more in interpersonal trust, and our results followed the same findings. Although participants knew that the robot was not alive and it was programmed to be intelligent, they accepted emotional expressions by the robot and believed in its behaviors and reactions. This provides an opportunity for further consideration and research to develop trustworthy robots equipped with AEL or other affective behaviors. Furthermore, AEL behavior of the robot was perceived higher in most of general trust factors except for decision-making, that was explained in previous part. This result suggested that AEL could be a powerful and successful behavior for social robots in establishing trust, and it covers the attributes required to be trustworthy by users.

However, considering the difference of AL and AEL behaviors of the social robots in HRT, thought-provoking results were revealed. Contrary to hypotheses, the difference in terms of general trust and cognitive trust between AL and AEL was not significant (H1c). However, the results admitted a significant difference with a large effect size for affective trust. The difference between AL and AEL in this study was the emotional statements of the robot in accompany with nonverbal behaviors. The results suggested that AEL was highly influential in fostering affective trust. However, general trust which is mostly performance-based in HRT, was less associated with emotional behaviors of the robot in AEL condition. It means that empathic behaviors of the robot in AEL condition could not improve general trust more than AL condition. It seems logical as participants evaluated the empathic behaviors of the robot effecting affective trust rather than general trust. The findings suggested that people distinguished affective behaviors of robots and associated them with emotional assessments than performance-based evaluations. Moreover, Gompei and Umemuro (2018) found that affective trust occurs at early stages of the interaction than cognitive trust, that could explain the obtained results. In fact, people perceive and evaluate affective trust during the earlier stages of interaction, whereas cognitive trust takes longer to develop. This is again because emotions and affection influence cognition. Furthermore, there was not statistically significant difference for general trust factors between AL and AEL, although the mean score for all these factors was higher in AEL. Emotional experiences enrich the quality of communication, and if AEL could be conceived as an expression of affection (Floyd, 2014), it would improve diverse relationships considerably. Thus, speakers with more emotional expressions had a higher perception of being trustworthy than speakers who cared for and supported other parties rationally (Aggarwal et al., 2005; Comer & Drollinger, 1999; Drollinger & Comer, 2013). It is evident that robots are still far away from ways in which humans convey emotions, which could be the possible reason why AEL did not surpass in all features from AL.

2.1.5.2 Correlation of listening behaviors, and affective, and cognitive trust in HRI

For affective trust, as it was expected in H2a, AEL behavior of the robot received a higher score and it was evaluated by users to be perceived as affectively more trustworthy than an active listener robot. Based on the features of affective trust, it is constructed on emotional experiences and feelings between partners, and as emotional connections deepen, trust goes beyond available knowledge and rational judgments (Johnson & Grayson, 2005). AEL creates an affection exchange between the listener and speaker as it conveys a message of love, kindness, and care (Floyd, 2014). Therefore, it evokes and is positively correlated with affective trust in HCs. This study showed that this conviction exists in HRI field, and social robots are able to elicit affective trust if their listening behavior is engaged in emotion and affectionate behaviors, and users appreciate a robot that demonstrates empathy in relation. This is because empathy is a big driver and it is directly associated with trust (Plank & Reid, 2010). Several designers and researchers have prompted robots with anthropomorphic characteristics, such as sociability, passionability, and intelligence (Beer et al., 2017), to suggest robots as living entities for users. AEL as an anthropomorphic behavior seems to be effective and believable in social robots, and it can help them make more realistic interactions with users and enrich affective trust. More interestingly, the difference of AEL and AL in affective trust was noticeably higher than that for general trust. This finding indicates that empathic behaviors of social robots are powerful in conveying affective messages and building affective trust.

On the other side, however, the study revealed that AEL behavior was not significantly different from AL behavior, regarding cognitive trust (H2b). There is a narrow border between emotion and logic, and they are closely associated with each other. Although some emotions are generated by the rationalization process, unconscious thoughts also lead to emotions about things (Lewis & Weigert, 1985; Ma et al., 2018), and emotions often outweigh logic. Therefore, it was supposed that the affective behaviors of the robot could result in more cognitive trust as well. However, the findings did not cover this hypothesis and AEL behavior of the robot could only lead in higher affective trust. AEL behavior of the robot involved the same rational statements of AL plus sending affectionate messages and emotional body language. Thus, it could be the explanation why the evaluations were similar considering cognitive trust. Furthermore, the mean score difference of AEL and AL behavior of the robot in affective trust was greater than the difference in cognitive trust. This result indicated that AEL and AL were perceived as more similar and closer to each other on the cognitive aspect. However, the emotional understanding of users was discernibly higher toward AEL, which suggests that the emotional manipulation of robots can be successful and consequential. As shown in Table 2.10, AEL and AL were also significantly higher than NAL in both cognitive and affective trust which admitted the superiority of them to NAL.

2.1.5.3 Impact of verbal and nonverbal components of listening behaviors in HRT

This study discussed the effect of verbal and nonverbal aspects of social robots' listening behaviors on trust. Previously, the results indicated that AEL behavior of the robot was outstanding in provoking trust. Because AEL consisted of verbal and nonverbal communication aspects, we determined their impact on trust. Although, participants expressed more general trust in the robot with nonverbal reactions in its AEL behavior beside the utterance, the difference with verbal AEL was not significant (H3a). Similar to previous findings in H1a and H1b, as shown in Table 2.11, AEL_{VO} was also not scored differently from AL in general trust (p = 1.00), but it was evaluated higher than NAL (p < 0.001). In HC, nonverbal behaviors are claimed to be effective means to indicate attention, interest, understanding, satisfaction, and many other social information (see Thepsoonthorn et al., 2018). The robots with both verbal and nonverbal behaviors are also rated as being more natural and engaging (Johanson et al., 2021). Nonverbal communication strategies are rich and compelling means of communication, and in listening behavior, nonverbal aspects convey considerable parts of messages, as many scholars confirm that effective empathic listening incorporates nonverbal immediacy behaviors. However, the current results of this study could not approve the effectiveness of nonverbal behaviors of social robots in HRT and it needs more research. NAO provided limited nonverbal behaviors such as head and arm movements, which were insufficient conveying emotional reactions. Furthermore, some participants got uncomfortable and distressed at the movements of the robot, due to its mechanical voices or the possibility of falling down. Thus, the participants did not generally differentiate AEL_{VO} behavior of the social robot from AEL behavior of the social robot in trustworthiness.

Regarding affective and cognitive trust, participants evaluated AEL behavior of the robot higher in affective trust than AEL_{VO} condition (H3b). People often indicate their emotions in myriad verbal expressions and nonverbal communication in their relationships. However, nonverbal immediacy has an advantage over utterance, as it is associated with messages of positive feelings, intimacy, and affection (Floyd, 2014). Emotions change our voice, body movements, and facial expressions unconsciously; for instance, smiling is a way of communicating happiness to others or waving the arms is a symbol of excitement. Thus, when communication is supported by nonverbal immediacy, affectionate messages are transmitted more easily and faster, and this affects the growth of affective trust in HRI as well. However, interestingly enough, AEL_{VO} resulted in a moderate significant difference (p = 0.055) with AL in affective trust.

Thus, considering the results, we could conclude the superior effectiveness of nonverbal behaviors on affective trust. It is because that eliminating the nonverbal behaviors in AEL_{VO} condition decreased its contribution to affective trust and situated it more similar to the behavior of AL. However, on the side of cognitive trust, the results did not reach a scientifically significant difference, although the mean score was higher for AEL. This could be because both AEL and AEL_{VO} behaviors of social robot demonstrated exactly similar verbal content, which mostly incorporates cognitive-driven trust. Verbal communication is highly language-based and depends on the understanding of the meaning of words. It is found that words are stronger engagement skills, whereas body language affects social-emotional concepts (Thepsoonthorn et al., 2018). Thus, a robot with AEL behavior that consists of both verbal and nonverbal means of communication can be the best option for designers to develop a trustworthy robot with the optimum outcome.

2.2 Effect of Social Robots' Listening Behaviors on Animacy, Likeability and Perceived Intelligence

During the first study, the influence of social robots with different listening behaviors on people's perception was also analyzed in terms of animacy, likeability and perceived intelligence. A separate experiment was not conducted for that, but some further measurements were added to evaluate people's perception of social robots.

2.2.1 INTRODUCTION

The field of HRI attempts to build effective and successful interactions between people and robots. One approach to enhance people's interaction of robots is the attempt to increase a robot's familiarity by using anthropomorphic (humanlike) design and human social characteristics (Fink, 2012). People respond more positively to anthropomorphic social robots that possess humanlike appearance, behaviors or interaction. Accordingly, anthropomorphism is broadly recommended to support an intuitive and responsive interaction between social robots and humans. To enhance human-robot interaction, it is necessary to investigate and understand how people perceive such humanistic behaviors.

Anthropomorphism entails ascribing humanlike properties, characteristics, intentions, motivations, emotions, or mental states to real or imagined nonhuman agents and objects. Various aspects of robot anthropomorphism have been investigated to date, in particular, the extent to which the attribution of human characteristics to robots may influence observers' subsequent behavior towards them, or how the mere presence of anthropomorphic robots may affect socio-cognitive processes (Spatola & Wudarczyk, 2021). In the design of socially interactive robots, anthropomorphism is reflected in the robot's form (appearance), behavior (e.g. motion), and interaction (e.g. modality) (Fink, 2012). While research on anthropomorphic robots equipped with expressive faces, emotional voices and body language (see Tsiourti et al., 2019) has been extensive, the exploration in the context of social robots' behaviors and interactions is still in its early stages, and there are many humanistic attributes have not been studied in HRI.

During the first study, we considered listening behavior as a humanistic behavior for social robots. Despite the growing number of studies exploring the listening behavior of social robots as an anthropomorphic attribute, the specific influence of different listening behaviors and their individual components in HRI remains ambiguous. Therefore, we investigated the influence of social robots' listening behaviors on people's perception, as well as its impact on people's perceived trust.

2.2.2 DEVELOPING HYPOTHESES

2.2.2.1 Social robots' emotional behaviors and perception of robots

Perception of anthropomorphism were influenced by the robots' capability to express emotional state. Most works done in HRI regarding emotional expression focused on face-based expressions, emotional speech cues or gestures (see Destephe et al., 2014). Number of studies presented the effectiveness of emotions in human-likeness, attractiveness, likeability, trustworthiness, and familiarity of the robots (see Park & Whang, 2022). The integration of dynamic emotional expression and movements made the humanoid robot more attractive, favorable, useful, and less mechanical-like (Hosseini et al., 2017). Previous work also suggested that endowing robots with emotions facilitated HRI, and the absence of emotional expressions could be interpreted as indifference toward the human (Eyssel et al., 2010). Wei and Zhao (2016) concluded that when the robot displayed positive emotions, subjects were more inclined to prolong the conversations, and their perception was significantly influenced by the robot's expressed emotions rather than the manipulation of the robot. In terms of companionship, the results of the study by Hosseini et al. (2017) showed that participants accepted the robots' suggestions and selected them as companions that could express emotions. Furthermore, in the listening behaviors, the empathetic component of AEL constructs an emotional link between individuals, and conveys a message of care, love, and friendliness for the partner. Therefore, to evaluate the influence of emotional aspects of social robots' listening behaviors on humans' perception of robots, we considered to compare AL, AEL and AELvo behaviors of the robot;

- **H1.** Social robots with the behavior of AEL have a positive effect on people's perception of the robots compared to those with the behavior of AL.
- **H2.** Social robots with the behavior of AEL_{VO} have a positive effect on people's perception of the robots compared to those with the behavior of AL.

2.2.2.2 Social robots' nonverbal behaviors and perception of robots

As discussed in section 2.1.2.3, listening behaviors comprise two components: verbal and nonverbal. Evidence showed that nonverbal aspects of communication surpass verbal segments in conveying message (Mehrabian, 1971). In HRI, several studies have considered the effectiveness of nonverbal behaviors of social robots (see (Rifinski et al., 2021); for example, Salem et al. (2013) revealed that participants perceived the robot used co-verbal gestures as more likable, reported greater shared reality with it, and showed increased future contact intentions than when the robot gave instructions without gestures. Furthermore, robots that use gestures along with speech elicit more attention and better recall, as well as higher ratings than robots that do not show co-verbal gestures (Admoni et al., 2016). Further research indicated that if the robot used gestures, it was evaluated significantly better in terms of anthropomorphism, animacy, perceived and social intelligence and emotions compared with a no movement condition (Striepe et al., 2021). Moreover, some studies demonstrated how a non-verbal cue, such as the robot's gaze, could influence participants' roles, turn-taking and people's interpersonal evaluation on robot in human-robot conversation (Rifinski et al., 2021).

Furthermore, Bartneck (2001) stated that humans express emotions through body language, with emotions being conveyed through gestures, facial expressions, and physiological responses. Emotional gestures not only amplify the impact of nonverbal behaviors but also influence the state of mind or reaction of their human partner. Given that some social robots are designed to resemble human features, this allows for more natural interactions by incorporating body language. Accordingly, Eyssel et al. (2010) investigated the effect of robots' nonverbal-emotional expression on anthropomorphism, finding that participants rated the emotion-displaying robot as more humanlike and likable. Meanwhile, Weger et al. (2010) stated that listeners receive more care and concern from nonverbal behaviors of listening than specific verbal behaviors, and Floyd (2014) argued that listeners who use nonverbal behaviors convey empathy and support more effectively. Therefore, we accessed the effect of nonverbal components of social robots' listening behaviors on people's perception of animacy, likeability and perceived intelligence. To do so, we compared the AEL and AEL_{VO} behaviors of a social robot. The hypothesis is as follows;

H3. Social robots with the behavior of AEL have a positive effect on people's perception of the robots compared to those with the behavior AEL_{VO}.

2.2.3 METHOD

This study was part of the main study explained in Part 1, where we incorporated additional measurements to assess how people perceive social robots' listening behaviors in terms of animacy, likeability and perceived intelligence.

2.2.3.1 Measurement

For measuring the perception of the robot, we used the English version of **Godspeed questionnaire** (Bartneck et al., 2009), which is one of the most used survey for perception of robots (Thunberg et al., 2022). Although this questionnaire is not validated, we utilized it because it is widely employed in various fields, including human-robot interaction. It's essential to note that this measurement wasn't the primary focus of our study; rather, we conducted it as an additional aspect of our research. The measurement has five subscales, which three relevant subscales of *animacy, likability, and perceived intelligence* were chosen for this study. Animacy measures the tendency of human users to consider the robot alive and to attribute intentions to it, and includes six variables: alive, lively, organic, lifelike, interactive, and responsive. Likeability evaluates the tendency of human users to attribute desirable characteristics to a robot, and depicts the positive impression about others people might have, and includes five variables: like, friendly, kind, pleasant, nice. Perceived Intelligence explores the tendency of human users to consider the behavior of a robot intelligent, and represents the expected capabilities the robot has, and includes five variables: competent, knowledgeable, responsible, intelligent and sensible (Destephe et al., 2014). The measurement employs a semantic differential method where respondents mark which out of two opposite pairs of adjectives best describes the robot. Thus, participants evaluated their perception of the robot with 16 items and using a five-item Likert scale (1 = 'not at all' and 5 = 'very much'). The Cronbach's alpha value was 0.882 for animacy, 0.907 for likeability, and 0.870 for perceived intelligence.

2.2.4 RESULTS

2.2.4.1 Descriptive statistics

The data was analyzed to see the influence of different listening behaviors of social robots on participants' perception of the robot in terms of animacy, likability and perceived intelligence. Table 2.12 shows the means and standard deviations of the variables with their individual items for three listening behaviors of the robot. Robot with AEL behavior showed the higher score in all variables and their individual items, and NAL received the lowest sores. Moreover, the robot with AEL_{VO} behavior stood higher than the robot with AL behavior in perceived animacy, likability and perceived intelligence. Therefore, the robot showing AEL behavior was perceived to be more alive, desirable and intelligent. Interestingly, for the animacy variable, being *responsive* received the highest score for AEL condition which showed that participants perceived the robot with the behavior of AEL as having the best listening behavior. Among the items of likeability, being *nice* and *friendly* were the highest items for the robot with AEL behavior. Furthermore, the individual items assessing perceived intelligence showed similar scores between AEL and AEL_{VO} conditions. In AEL_{VO} condition, the perception of being *intelligent* received higher scores compared to AEL condition, suggesting a possible correlation between perceived intelligence and verbal behaviors rather than nonverbal behaviors.

Table 2. 12 Means and standard deviations of animacy, likeability and perceived intelligence of NAL, AL, AEL and AELvo conditions.

	NAL	AL	AEL	AEL_{VO}
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Animacy	**2.51(0.75)**	**3.26 (0.71)**	**3.72 (0.71)**	**3.37 (0.73)**
dead-alive	2.63 (0.99)	3.47 (0.97)	3.83 (1.05)	3.57 (0.93)
stagnant-lively	2.63 (0.92)	3.63 (1.03)	4.10 (0.99)	3.53 (1.04)
mechanical-organic	1.87 (0.81)	2.37 (0.85)	2.87 (1.07)	2.43 (1.07)
artificial-lifelike	1.97 (0.85)	2.43 (1.07)	3.13 (1.04)	2.60 (1.16)
inert-interactive	2.97 (1.09)	3.83 (1.02)	4.17 (0.79)	3.97 (0.85)
apathetic-responsive	3.03 (1.06)	3.87 (0.97)	4.27 (0.78)	4.13 (0.81)
Likability	**3.54(0.95)**	**4.15 (0.65)**	**4.52 (0.48)**	**4.37 (0.49)**
dislike-like	3.40 (0.96)	3.93 (0.94)	4.40 (0.72)	4.17 (0.79)
unfriendly-friendly	3.67 (1.06)	4.30 (0.70)	4.60 (0.62)	4.40 (0.67)
unkind-kind	3.47 (1.07)	4.17 (0.83)	4.57 (0.62)	4.47 (0.73)
unpleasant-pleasant	3.40 (1.19)	4.17 (0.83)	4.47 (0.57)	4.37 (0.55)
awful-nice	3.80 (0.99)	4.20 (0.88)	4.60 (0.49)	4.47 (0.50)
P-Intelligence	**3.02(0.80)**	**3.35 (0.68)**	**3.79 (0.68)**	**3.78 (0.72)**
Incompetent-competent	2.93 (0.94)	3.10 (1.12)	3.63 (0.80)	3.63 (0.85)
Ignorant-knowledgeable	2.90 (1.06)	3.27 (0.82)	3.80 (0.96)	3.77 (1.07)
Irresponsible-responsible	3.13 (1.04)	3.60 (0.77)	3.93 (0.69)	3.80 (0.76)
Unintelligent-intelligent	3.03 (1.06)	3.23 (1.04)	3.83 (1.02)	3.93 (0.86)
Foolish-sensible	3.10 (0.80)	3.57 (0.81)	3.77 (0.81)	3.77 (0.97)

(P-Intelligence; Perceived intelligence)

2.2.4.2 Effect of social robots' listening behaviors on perception of robots

To test the hypotheses, a series of one-way ANOVA tests was conducted with three listening behaviors as independent variable, and animacy, likability and perceived intelligence as dependent variables. ANOVA results indicated that the main effects of listening behaviors were significant for animacy ($F(2,87) = 3.36$, $p < 0.05$), likeability ($F(2,87) = 3.46$, $p < 0.05$), and perceived intelligence ($F(2,87) = 3.85$, $p < 0.05$). Post-hoc analysis revealed that the all three variables of animacy ($p < 0.05$), likeability ($p < 0.05$), and perceived intelligence ($p < 0.05$) was significantly higher for AEL condition than AL condition. Therefore, **H1** was fully supported and the robot exhibiting AEL behavior was perceived more positively than AL behavior. Furthermore, post-hoc analysis indicated a significant difference for the score of perceived intelligence ($p < 0.05$) between AL and AEL_{VO} conditions. However, AEL_{VO} condition was not significantly higher than AL condition in terms of likeability ($p = 0.12$), and animacy ($p = 0.57$). Thus, **H2** was supported for perceived intelligence, but not for animacy and likeability. Moreover, the results did not show any significant difference for the variables of likeability ($p = 0.28$), and perceived intelligence ($p = 0.94$) comparing AEL and AEL_{VO} conditions. There was only marginally significant difference for animacy ($p = 0.059$) between AEL and AEL_{VO} conditions, suggesting the influence of nonverbal behaviors on the robot's perceived aliveness and lifelikeness. Thus, nonverbal behaviors of AEL could not induce the participants' perception of robot more than AEL_{VO} and **H3** was not supported. Table 2.13 and Figure 2.16 shows the results of post-hoc analysis.

Table 2. 13 Post-hoc results for pairwise comparisons of animacy, likeability and perceived intelligence.

	AEL vs. AL	**AL vs. AEL_{VO}**	**AEL vs. AEL_{VO}**
	p	*p*	*p*
Animacy	0.015*	0.572	0.059
Likability	0.010*	0.127	0.285
P-Intelligence	0.017*	0.020*	0.941

*$p < 0.05$, (P-Intelligence; Perceived intelligence)

2.2.5 DISCUSSION

For this part, we analyzed the participants' perception of robots in terms of animacy, likeability, and perceived intelligence in relation to different listening behaviors of social robots.

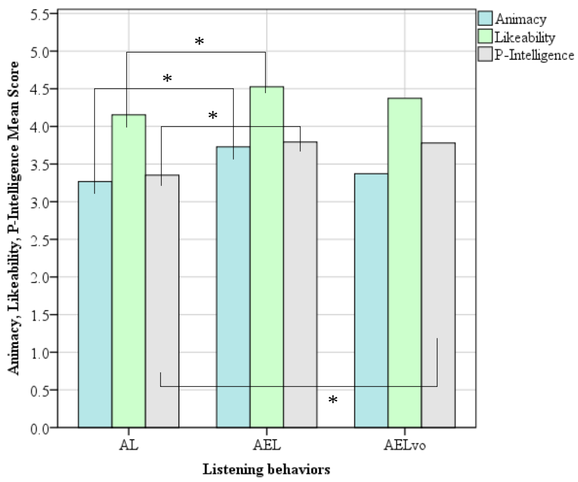


Figure 2. 16 Means of animacy, likeability and perceived intelligence scores for three listening behaviors. (*$p < 0.05$)

As expected, social robot's AEL behavior, which involved providing emotional-verbal and nonverbal responses, built a better impression among the participants. This result is in line with the prior works (Eyssel et al., 2010; Hosseini et al., 2017; Striepe et al., 2021) as anthropomorphic robots enhanced positive relations, and a robot used gestures and emotional expressions is perceived lively, friendly and sensible (Tapus et al., 2007). Several designers and researchers have prompted robots with anthropomorphic characteristics, such as sociability, passionability, and intelligence (Beer et al., 2017), to suggest robots as living entities for users. AEL as an anthropomorphic behavior seems to be effective and believable in social robots, and it can help them make more realistic interactions with users and enrich the perception of social robots.

2.2.5.1 AEL behavior enhance people's perception of social robots

Although, the robot with AL behavior indicated an attentive and caring approach toward the participants, empathetic behaviors of AEL was more influential and succeeded in building a friendly, alive and responsive interaction between the robot and participants. Expressive behavior such as showing emotions is a crucial aspect of human interaction. Emotions can allow humans to establish a relationship with others as well as the environment (e.g., objects, machines). In this sense, some scholars believed that interaction with technology and robots should be guided by the same principle (Ayanoğlu & S.Sequeira, 2019). Emotions could be expressed verbally or nonverbally. Prior studies highlighted the importance of communicating the affective state of a social robot through facial expressions, gaze, and affective speech (Appel et al., 2021), along with verbal expressions. Our results indicated a positive achievement providing social robots with emotional-verbal and nonverbal behaviors, specifically in listening context. Currently, robots are designed to become a part of people's lives and creating emotional engagement with robots can shape the involvement of people during an interaction.

The social robot with AEL behavior was perceived more alive than that of with AL behavior. Regarding the results, being responsive, lively and interactive had the highest means for AEL among the subscales of animacy, compared to other items of being mechanical, artificial and dead. These results revealed that although participants perceived the robots as a man-made and artificial object, they could believe and accept the robots' behaviors to be humanlike and evaluated them as sensitive and responsive partners. Furthermore, the scores for the likeability were the highest among the three variables. Among the five items of likeability, participants evaluated the robot with AEL behavior with the highest scores in terms of being friendly and nice. Finally, emotive factors influenced the perceived intelligence as well. There is a narrow border between emotion and logic, and they are closely associated with each other, and emotions outweigh logic and influence it (Duncan & Barrett, 2007). Therefore, emotional expressions influenced the perception of intelligence as well, although it appears that emotions stem from distinct sources of knowledge and competence.

2.2.5.2 Primacy of verbal components in the perception of social robots' listening behaviors

Contrary to our expectations, participants' perception of AEL was not significantly different from AEL_{VO}, although AEL behavior of the robot was assessed higher in the scores of animacy, likeability and perceived intelligence. This is an interesting finding that does not match with previous works (Eyssel et al., 2010), which showed that people rate a robot as more sympathetic and anthropomorphic when it expressed emotions non-verbally as a reaction to the participants' speech compared to a robot that shows no emotions. Moreover, Salem et al. (2013) indicated that virtual agents are perceived more positive when they produce co-verbal gestures rather than using speech as the only modality. However, our results did not support it in HRI. One possible explanation could be that the experiment was based on dyadic conversation, where participants primarily focused on the content of the speech. Speech is the most natural method of communication between human and the same method can apply to human and machine (Ayanoğlu & S.Sequeira, 2019). As a result, the impact of nonverbal behaviors may have been less pronounced. Montepare et al. (1999) also emphasized that research on the communication of emotion has generally supposed that the perception of emotion is more engaged with facial or vocal expressions than with body movements. Meanwhile, excessive movements of NAO robot caused mechanical sounds and some participants perceived it as weird. In addition, some participants were scared of certain movements of NAO and felt interrupted during conversation. Therefore, our results indicated that despite the significant importance of nonverbal behaviors, we should be careful in using them appropriately based on the context, as researchers suggested that in the design of robots, a balance needs be found between robot-ness and humanness so that the user will feel comfortably engaging the robot (Fink, 2012). When social robots are expected to act as good listeners and convey a sense of understanding and empathy, excessive nonverbal behaviors might be inappropriate and hurtful, especially for robot therapist and counsellor.

The results also revealed that AL and AEL_{VO} were perceived differently in terms of perceived intelligence, but not in animacy and likeability. This is a valuable finding and indicated the correlation of emotional-verbal behaviors with perceived intelligence, as well as the association of nonverbal behaviors with animacy and likability. Empathy influences, modulates, and mediates basic cognitive processes, and could impact the intelligence. Emotional speech conveys a higher level of engagement and thoughtfulness, leading others to perceive the speaker as more intelligent. However, AEL_{VO} had fewer nonverbal behaviors compared to AL, which decreased the participants perception of animacy and likeability. Because, nonverbal behaviors makes a robot humanlike, pleasant and interactive (Wei & Zhao, 2016). Finally, considering the comparison of AEL with AL and AEL_{VO}, the results indicated the contribution of verbal aspects of listening behaviors in enhancing animacy and likeability as well. These findings revealed that both verbal and nonverbal behaviors of social robots contribute to anthropomorphism.

2.3 CHAPTER SUMMARY

This section concludes the results of sections 2.1 and 2.2. The first experiment focused on building different listening behaviors for social robots and exploring their impact on people's trust perception of robots. The results indicated that the behavior of AEL was the most trust-eliciting among all other listening behaviors specifically in affective trust. Thus, manipulating the robots in this way, we hope that we can improve HRT in different applications in the field of healthcare, business, and education. However, we did not define any special situation in our experiment, which could be an opportunity for future studies. Furthermore, AL behavior of the robot was accepted as more trustworthy than NAL behavior of the robot in all aspects of trust, and less trustworthy than AEL behavior of the robot in affective trust. This result indicated that adding emotional behaviors to social robots and supporting users by affectionate companions improved the level of affective trustworthiness and made the robots more believable and benevolent for users. If we cannot provide a robot with emotional aspects because of any limitations, AL can be considered as a second option to reach trustworthiness in HRI.

However, more research is needed considering gender or cultural differences because these factors are influential in interpersonal trust. Additionally, the results indicated the effectiveness of nonverbal behaviors over utterances in affective trust, which admitted the power of nonverbal behaviors in conveying emotional messages. When the robot indicated nonverbal communication such as head nodding, eye gaze, body movement, and gestures, users evaluated it as more reliable and affectively impressive. This result indicated the use of proper nonverbal behaviors for robots to obtain better outcomes in HRI, specifically trust formation. However, we were not able to simulate nonverbal behaviors in all dimensions because of limitations of the robot in terms of facial expressions or body movement, and this provided another possibility for further investigations in the future. In general, we provided a research in relation to the listening behavior and trust of social robots to narrow the gap to develop trustworthy robots in order to design succeeding robots.

Additionally, we considered the influence of social robots' listening behaviors on people's perception of robots in terms of animacy, likeability and perceived intelligence. Results showed that AEL behavior of the robot which provided emotional-verbal and nonverbal behaviors, was perceived more positively in terms of animacy, likeability and perceived intelligence compared to AL behavior of the robot which only exhibited attentive behavior. The findings acknowledged the significant contribution of both verbal and nonverbal behaviors in shaping the anthropomorphic attributes of social robots. However, in comparison with AEL_{VO}, the nonverbal aspects of AEL did not significantly enhance the humans' perception of robots. This finding reminds us to consider the context and specific use cases when designing nonverbal behaviors for social robots. Furthermore, we concluded that verbal behaviors tend to be associated with perceived intelligence, whereas nonverbal aspects are linked to animacy and likeability in listening behaviors.

2.4 REFERENCES

Admoni, H., Weng, T., Hayes, B., & Scassellati, B. (2016). Robot nonverbal behavior improves task performance in difficult collaborations. *ACM/IEEE International Conference on Human-Robot Interaction*, *2016-April*, 51–58. https://doi.org/10.1109/HRI.2016.7451733

Aggarwal, P., Castleberry, S. B., Ridnour, R., & Shepherd, C. D. (2005). Salesperson empathy and listening: Impact on relationship outcomes. *Journal of Marketing Theory and Practice*, *13*(3), 16–31. https://doi.org/10.1080/10696679.2005.11658547

Appel, M., Lugrin, B., Kühle, M., & Heindl, C. (2021). The emotional robotic storyteller: On the influence of affect congruency on narrative transportation, robot perception, and persuasion. *Computers in Human Behavior*, *120*(February), 106749. https://doi.org/10.1016/j.chb.2021.106749

Ayanoğlu, H., & S.Sequeira, J. (2019). *Emotional Design and Human-Robot Interaction : Theory, Method and Application*. http://www.springer.com/series/6033

Bartneck, C. (2001). Affective expressions of machines. *Conference on Human Factors in Computing Systems - Proceedings*, 189–190. https://doi.org/10.1145/634067.634181

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bauer, C., & Figl, K. (2008). Active listening in written online communication - A case study in a course on Soft Skills for Computer Scientists. *Proceedings - Frontiers in Education Conference, FIE*. https://doi.org/10.1109/FIE.2008.4720282

Beelen, T., Velner, E., Ordelman, R., Truong, K. P., Huibers, T., & Evers, V. (n.d.). Children 's Trust in Robots and the Information They Provide. *2023 CHI Con- Ference on Human Factors in Computing Systems (CHI EA '23)*. https://doi.org/10.1145/3544549.3585801

Beer, J. M., Liles, K. R., Wu, X., & Pakal, S. (2017). Affective Human–Robot Interaction. In *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 359–381). Elsevier Inc. https://doi.org/10.1016/B978-0-12-801851-4/00015-X

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (Fe)male Robot: How Robot Body Shape Impacts First Impressions and Trust Towards Robots. *International Journal of Social Robotics*, *13*(3), 477–489. https://doi.org/10.1007/s12369-019-00562-7

Bodie, G. D., St. Cyr, K., Pence, M., Rold, M., & Honeycutt, J. (2012). Listening Competence in Initial Interactions I: Distinguishing Between What Listening Is and What Listeners Do. *International Journal of Listening*, *26*(1), 1–28. doi: 10.1080/10904018.2012.639645

Browning, S., & Waite, R. (2010). The gift of listening: JUST listening strategies. *Nursing Forum*, *45*(3), 150–158. https://doi.org/10.1111/j.1744-6198.2010.00179.x

Brunner, B. R. (2008). Listening, communication & trust: practitioners' perspectives of business/organizational relationships. *International Journal of Listening*, *22*(1), 73–82. https://doi.org/10.1080/10904010701808482

Buschmeier, H., Malisz, Z., Skubisz, J., Włodarczak, M., Wachsmuth, I., Kopp, S., & Wagner, P. (2014). ALICO: A multimodal corpus for the study of active listening. *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, 3638–3643. https://api.semanticscholar.org/CorpusID:18827045

Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (n.d.). Can you trust your trust measure? *ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*, 92–100. https://doi.org/10.1145/3434073.3444677

Comer, L. B., & Drollinger, T. (1999). Active Empathetic Listening and Selling Success : A Conceptual Framework. *Journal of Personal Selling and Sales Management*, *19*(1), 15–29.

Correia, F., Guerra, C., Mascarenhas, S., Melo, F. S., & Paiva, A. (n.d.). Exploring the impact of fault justification in human-robot trust. *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS 2018)*, 507–513. https://doi.org/10.5555/3237383.3237459

Dahlbäck, N., Jonsson, A., Ahrenberg, L., & Current. (1993). Wizard of Oz, why and how. *Knowledge-Based Systems*, *6*(4), 258–266. https://doi.org/10.1016/0950-7051(93)90017-N

Destephe, M., Brandao, M., Kishi, T., Zecca, M., Hashimoto, K., & Takanishi, A. (2014). Emotional gait: Effects on humans' perception of humanoid robots. *IEEE RO-MAN 2014 - 23rd IEEE International Symposium on Robot and Human Interactive Communication: Human-Robot Co-Existence: Adaptive Interfaces and Systems for Daily Life, Therapy, Assistance and Socially Engaging Interactions*, 261–266. doi: 10.1109/ROMAN.2014.6926263

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., … Morency, L. P. (2014). SimSensei kiosk: A virtual human interviewer for healthcare decision support. *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014*, *2*(January), 1061–1068. ISBN 9781450327381.

Drollinger, T., & Comer, L. B. (2013). Salesperson's listening ability as an antecedent to relationship selling. *Journal of Business and Industrial Marketing*, *28*(1), 50–59. https://doi.org/10.1108/08858621311285714

Drollinger, T., Comer, L. B., & Warrington, P. T. (2006). Development and Validation of the Active Empathetic Listening Scale. *Psychology & Marketing*, *23*(2), 160–180. https://doi.org/10.1002/mar.20105

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and Emotion*, *21*(6), 1184–1211. https://doi.org/10.1080/02699930701437931

Ee, S. N., Sakaguchi, T., & Yokoi, K. (2008). A humanoid robot that listens, speaks, sees and manipulates in human environments. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, *September*, 419–425. https://doi.org/10.1109/MFI.2008.4648031

Eyssel, F., Hegel, F., Horstmann, G., & Wagner, C. (2010). Anthropomorphic inferences from emotional nonverbal cues: A case study. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 646–651. doi: 10.1109/ROMAN.2010.5598687

Fassaert, T., van Dulmen, S., Schellevis, F., & Bensing, J. (2007). Active listening in medical consultations: Development of the Active Listening Observation Scale (ALOS-global). *Patient Education and Counseling*, *68*(3), 258–264. doi: 10.1016/j.pec.2007.06.011

Fink, J. (2012). Anthropomorphism and human likeness in the design of robots and human-robot interaction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7621 LNAI*, 199–208.

https://doi.org/10.1007/978-3-642-34103-8_20

Floyd, K. (2014). Empathic Listening as an Expression of Interpersonal Affection. *International Journal of Listening*, *28*(1), 1–12. https://doi.org/10.1080/10904018.2014.861293

Gearhart, C. C., & Bodie, G. D. (2011). Active-Empathic Listening as a General Social Skill: Evidence from Bivariate and Canonical Correlations. *Communication Reports*, *24*(2), 86–98. https://doi.org/10.1080/08934215.2011.610731

Gompei, T., & Umemuro, H. (2018). Factors and Development of cognitive and affective trust on social robots. *Social Robotics. ICSR 2018. Lecture Notes in Computer Science*, *11357,* 45–54. https://doi.org/10.1007/978-3-030-05204-1_5

Gordon, T. (1975). *Parent effectiveness training*. New American Library.

Hadar, U., Steiner, T. J., & Rose, F. C. (1985). Head movement during listening turns in conversation. *Nonverbal Behavior*, *9*(4). https://doi.org/10.1007/BF00986881

Hillen, M. A., De haes, H. C. J. M., Stalpers, L. J. A., Klinkenbijl, J. H. G., Eddes, E. H., Butow, P. N., Van der vloodt, J., Van laarhoven, H. W. M., & Smets, E. M. A. (2014). How can communication by oncologists enhance patients' trust? An experimental study. *Annals of Oncology*, *25*(4), 896–901. https://doi.org/10.1093/annonc/mdu027

Hinwood, D., Ireland, J., Jochum, E. A., & Herath, D. (2018). A proposed wizard of OZ architecture for a human-robot collaborative drawing task. *International Conference on Software Reuse*, 35–44. https://doi.org/10.1007/978-3-030-05204-1_4

Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE*, *10*(8), 1–18. https://doi.org/10.1371/journal.pone.0136905

Hosseini, S., Lettinga, D., Vasey, E., Zheng, Z., Jeon, M., Park, C. H., & Howard, A. M. (2017). Both look and feel matter: Essential factors for robotic companionship. *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication, 2017-Janua*, 150–155. https://doi.org/10.1109/ROMAN.2017.8172294

Inoue, K., Lala, D., Yamamoto, K., Nakamura, S., Takanashi, K., & Kawahara, T. (2020). An attentive listening system with android ERICA: Comparison of autonomous and WoZ interactions. *SIGDIAL 2020 - 21st Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference*, *July*, 118–127. doi: 10.18653/v1/2020.sigdial-1.15

Ivey, A. E., & Daniels, T. (2016). Systematic Interviewing Microskills and Neuroscience: Developing Bridges between the Fields of Communication and Counseling Psychology. *International Journal of Listening*, *30*(3), 99–119. doi: 10.1080/10904018.2016.1173815

Johanson, D. L., Ahn, H. S., & Broadbent, E. (2021). Improving Interactions with Healthcare Robots: A Review of Communication Behaviours in Social and Healthcare Contexts. *International Journal of Social Robotics*, *13*(8), 1835–1850. https://doi.org/10.1007/s12369-020-00719-9

Johansson, M., Hori, T., Skantze, G., Höthker, A., & Gustafson, J. (2016). Making Turn-Taking Decisions for an Active Listening Robot for Memory Training. *ICSR 2016, LNAI 9979, November*, 940–949. https://doi.org/10.1007/978-3-319-47437-3

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

Jones, S. M., & Guerrero, L. K. (2001). The Effects of Nonverbal Immediacy and Verbal Person Centeredness in the Emotional Support Process. *Human Communication Research*, *27*(4), 567–596. https://doi.org/10.1111/j.1468-2958.2001.tb00793.x

Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., & Anzai, Y. (2007). A humanoid robot that pretends to listen to route guidance from a human. *Autonomous Robots*, *22*(1), 87–100. https://doi.org/10.1007/s10514-006-9007-6

Kobayashi, Y., Yamamoto, D., Koga, T., Yokoyama, S., & Doi, M. (2010). Design targeting voice interface robot capable of active listening. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 161–162. https://doi.org/10.1109/hri.2010.5453214

Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology*, *43*(4), 718–731. https://doi.org/10.1037//0022-3514.43.4.718

Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K., & Kawahara, T. (2017). Attentive listening system with backchanneling, response generation and flexible turn-taking. *SIGDIAL 2017 - 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference*, *August*, 127–136. doi: 10.18653/v1/w17-5516

Lasky, S. (2000). The cultural and emotional politics of teacher-parent interactions. *Teaching and Teacher Education*, *16*(8), 843–860. https://doi.org/10.1016/S0742-051X(00)00030-5

Lee, J. gil, Lee, J., & Lee, D. (2021). Cheerful encouragement or careful listening: The dynamics of robot etiquette at Children's different developmental stages. *Computers in Human Behavior*, *118*(November 2020), 106697. doi: 10.1016/j.chb.2021.106697

Lester, D. (2002). Active listening. In D. Lester (Ed.), *Crisis intervention and counseling by telephone* (2nd ed., pp. 92–98). Charles C Thomas. https://doi.org/10.1176/appi.psychotherapy.2003.57.2.282

Lewis, J. D., & Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, *63*(4), 967–985.

Libow, J. A., & Doty, D. W. (1976). An evaluation of empathic listening in telephone counseling. *Journal of Counseling Psychology*, *23*(6), 532–537. https://doi.org/10.1037//0022-0167.23.6.532

Lloyd, K. J., Boer, D., Kluger, A. N., & Voelpel, S. C. (2015). Building Trust and Feeling Well: Examining Intraindividual and Interpersonal Outcomes and Underlying Mechanisms of Listening. *International Journal of Listening*, *29*(1), 12–29. https://doi.org/10.1080/10904018.2014.928211

Ma, F., Wylie, B. E., Luo, X., He, Z., Xu, F., & Evans, A. D. (2018). Apologies repair children's trust: The mediating role of emotions. *Journal of Experimental Child Psychology*, *176*, 1–12. https://doi.org/10.1016/j.jecp.2018.05.008

Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In C. S. Nam & J. B. Lyons (Eds.), *Trust in Human-Robot Interaction: Reaserch and applications* (pp. 3–25). Elsevier. https://doi.org/10.1016/B978-0-12-819472-0.00001-0

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

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

McClave, E. Z. (2000). Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*, *32*(7), 855–878. https://doi.org/10.1016/S0378-2166(99)00079-X

McColl, D., & Nejat, G. (2014). Recognizing Emotional Body Language Displayed by a Human-like Social Robot. *International Journal of Social Robotics*, *6*(2), 261–280. https://doi.org/10.1007/s12369-013-0226-7

Mehrabian, A. (1971). Silent Messages: Implicit Communication of Emotions and Attitudes. In *Wadsworth*. Wadsworth.

Mohammad, Y., & Nishida, T. (2008). Towards natural listening on a humanoid robot. *Proceedings - International Conference on Informatics Education and Research for Knowledge-Circulating Society, ICKS 2008*, *February 2008*, 153–156. https://doi.org/10.1109/ICKS.2008.4

Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior*, *23*(2), 133–152. https://doi.org/10.1023/A:1021435526134

Nadler, S., & Simerly, R. L. (2006). The Effect of Listening on the Formation of Students Trust and Commitment in Academic Advising: A Study at a United States University. *International Journal of Management*, *23*(2), 215–221.

Nakamura, Y., & Umemuro, H. (2022). The Effect of Robots Listening Attitude Change on the Self-disclosure of the Elderly: A Preliminary Study. *ACM/IEEE International Conference on Human-Robot Interaction*, *2022-March*, 932–936. doi: 10.1109/HRI53351.2022.9889404

Nishio, T., Yoshikawa, Y., Iio, T., Chiba, M., Asami, T., Isoda, Y., & Ishiguro, H. (2021). Actively listening twin robots for long-duration conversation with the elderly. *ROBOMECH Journal*, *8*(18). https://doi.org/10.1186/s40648-021-00205-5

Nugent, W. R., & Halvorson, H. (1995). Testing the Effects of Active Listening. *Research on Social Work Practice*, *5*(2), 152–175. https://doi.org/10.1177/104973159500500202

Ogasawara, Y., Okamoto, M., Nakano, Y. I., & Nishida, T. (2005). Establishing natural communication environment between a human and a listener robot. *Proceedings of the Symposium on Conversational Informatics for Supporting Social Intelligence and Interaction: Situational and Environmental Information Enforcing Involvement in Conversation*, *January*, 42–51. https://api.semanticscholar.org/CorpusID:160004406

Ogawa, H., & Watanabe, T. (2000). InterRobot: A speech driven embodied interaction robot. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 322–327. https://doi.org/10.1109/ROMAN.2000.892517

Park, S., & Whang, M. (2022). Empathy in Human Robot Interaction: Designing for Social Robots. *International Journal of Environ. Res. Public Health*, *19*(1889). doi: 10.3390/ijerph19031889

Plank, R. E., & Reid, D. E. (2010). The interrelationships of empathy, trust and conflict and their impact on sales performance: an exploratory study. *The Marketing Management Journal*, *20*(2), 119–139. https://api.semanticscholar.org/CorpusID:53487467

Ramsey, R. P., & Sohi, R. S. (1997). Listening to your customers: The impact of perceived salesperson listening behavior on relationship outcomes. *Journal of the Academy of Marketing Science*, *25*(2), 127–137. https://doi.org/10.1007/BF02894348

Riek, L. (2012). Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction*, *1*(1), 119–136. https://doi.org/10.5898/jhri.1.1.riek

Rietz, F., Sutherland, A., Bensch, S., Wermter, S., & Hellström, T. (2021). WoZ4U: An Open-Source Wizard-of-Oz Interface for Easy, Efficient and Robust HRI Experiments. *Frontiers in Robotics and AI*, *8*(July), 1–11. https://doi.org/10.3389/frobt.2021.668057

Rifinski, D., Erel, H., Feiner, A., Hoffman, G., & Zuckerman, O. (2021). Human-human-robot interaction: robotic object's responsive gestures improve interpersonal evaluation in human interaction. *Human-Computer Interaction*, *36*(4), 333–359. https://doi.org/10.1080/07370024.2020.1719839

Robertson, K. (2005). Active listening: more than just paying attention. *Australian Family Physician*, *34*(12), 1053–1055. https://api.semanticscholar.org/CorpusID:43666581

Rogers, C., & Farson, R. E. (1957). *Active listening*. Industrial Relations Center, University of Chicago.

Rossi, S., Staffa, M., Bove, L., Capasso, R., & Ercolano, G. (2017). User's Personality and Activity Influence on HRI Comfortable Distances. In *social robotics* (pp. 167–177). Springer, Cham. https://doi.org/10.1007/978-3-319-70022-9_17

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). Towards safe and trustworthy social robots: Ethical challenges and practical issues. *In: Tapus A., André E., Martin JC., Ferland F., Ammi M. (Eds) Social Robotics. ICSR 2015. Lecture Notes in Computer Science*, *9388*, 584–593. https://doi.org/10.1007/978-3-319-25554-5_58

Savery, R., Rose, R., & Weinberg, G. (2019). Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture. *2019 28th IEEE International Conference on Robot and Human Interactive Communication, (RO-MAN)*, 1–7.

https://doi.org/10.1109/RO-MAN46459.2019.8956386

Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the"Trust Perception Scale-HRI". In R. Mittu, D. Sofge, A. Wagner, & W. F. Lawless (Eds.), *Robust Intelligence and Trust in Autonomous Systems* (Issue August, pp. 191–218). Springer. https://doi.org/10.1007/978-1-4899-7668-0

Slovic, P., Lichtenstein, S., & Fischhoff, B. (2012). *Decision making*. Lightning Source.

Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*, *124*(May), 106934. https://doi.org/10.1016/j.chb.2021.106934

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

Striepe, H., Donnermann, M., Lein, M., & Lugrin, B. (2021). Modeling and Evaluating Emotion, Contextual Head Movement and Voices for a Social Robot Storyteller. *International Journal of Social Robotics*, *13*(3), 441–457. https://doi.org/10.1007/s12369-019-00570-7

Sundar, S. S., Jung, E. H., Waddell, T. F., & Kim, K. J. (2017). Cheery companions or serious assistants? Role and demeanor congruity as predictors of robot attraction and use intentions among senior citizens. *International Journal of Human Computer Studies*, *97*, 88–97. https://doi.org/10.1016/j.ijhcs.2016.08.006

Tapus, A., Mataric, M. J., & Scassellati, B. (2007). Socially assistive robotics [Grand challenges of robotics]. *IEEE Robotics and Automation Magazine*, *14*(1), 35–42. https://doi.org/10.1109/MRA.2007.339605

Thepsoonthorn, C., Ogawa, K. I., & Miyake, Y. (2018). The Relationship between Robot's Nonverbal Behaviour and Human's Likability Based on Human's Personality. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-25314-x

Thimmesch-Gill, Z., Harder, K. A., & Koutstaal, W. (2017). Perceiving emotions in robot body language: Acute stress heightens sensitivity to negativity while attenuating sensitivity to arousal. *Computers in Human Behavior*, *76*, 59–67. doi: 10.1016/j.chb.2017.06.036

Thunberg, S., Arnelid, M., & Ziemke, T. (2022). Older Adults' Perception of the Furhat Robot. *HAI 2022 - Proceedings of the 10th Conference on Human-Agent Interaction*, 4–12. https://doi.org/10.1145/3527188.3561924

Tsiourti, C., Weiss, A., Wac, K., & Vincze, M. (2019). Multimodal Integration of Emotional Signals from Voice, Body, and Context: Effects of (In)Congruence on Emotion Recognition and Attitudes Towards Robots. *International Journal of Social Robotics*, *11*(4), 555–573. https://doi.org/10.1007/s12369-019-00524-z

Volante, W. G., Sosna, J., Kessler, T., Sanders, T., & Hancock, P. A. (2019). Social Conformity Effects on Trust in Simulation-Based Human-Robot Interaction. *Human Factors*, *61*(5), 805–815. https://doi.org/10.1177/0018720818811190

Weger, H., Castle Bell, G., Minei, E. M., & Robinson, M. C. (2014). The Relative Effectiveness of Active Listening in Initial Interactions. *International Journal of Listening*, *28*(1), 13–31. https://doi.org/10.1080/10904018.2013.813234

Weger, H., Castle, G. R., & Emmett, M. C. (2010). Active listening in peer interviews: The influence of message paraphrasing on perceptions of listening skill. *International Journal of Listening*, *24*(1), 34–49. https://doi.org/10.1080/10904010903466311

Wei, Y., & Zhao, J. (2016). Designing robot behavior in human robot interaction based on emotion expression. *Industrial Robot*, *43*(4), 380–389. https://doi.org/10.1108/IR-08-2015-0164

Weissglass, J. (1990). Constructivist listening for empowerment and change. *Educational Forum*, *54*(4), 351–370. https://doi.org/10.1080/00131729009335561

Worthington, D. L., & Bodie, G. D. (2018). Defining Listening: A Historical, Theoretical, and Pragmatic Assessment. In D. L. Worthington & G. D. Bodie (Eds.), *The Sourcebook of Listening Research: Methodology and Measures* (1st ed., pp. 3–18). John Wiley & Sons. https://doi.org/10.1002/9781119102991.ch1

Xie, Y., Bodala, I. P., Ong, D. C., Hsu, D., & Soh, H. (2019). Robot Capability and Intention in Trust-Based Decisions Across Tasks. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, *2019-March*, 39–47. https://doi.org/10.1109/HRI.2019.8673084

Yonezawa, T., Koyama, Y., Yamazoe, H., Abe, S., & Mase, K. (2010). Improving video communication for elderly or disabled by coordination of robot's active listening behaviors and media controls. *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, 1476–1481. https://doi.org/10.1109/IROS.2010.5650044

CHAPTER 3

Influence of Social Robots' Benevolence and Competence Attributes on Perceived Trust in Human-Robot Interactions

3.1 INTRODUCTION

In the interpersonal trust, **benevolence** and **competence** have been proposed as two highly influential attributes (Lee & See, 2004; Mayer et al., 1995), probably with the former being the most common facet of trust (Hoy & Tschannen-Moran, 1999). Benevolence is the extent to which *a trustee is believed to want to do good for a trusting party* (Mayer et al., 1995). Competence, however, is described *as a group of abilities that enables a trustee to provide accurate information* (Johnston et al., 2015). The effectiveness of benevolence and competence attributes on trust perception in education (Di Battista et al., 2020, 2021), service relationships (Hui & Sit, 2005), personal relations (Schoorman et al., 2007), and interorganizational trust (Svare et al., 2020) has been proven. Additionally, some studies have investigated the contribution of comparable factors to benevolence and competence in HRI (e.g., Belanche et al., 2021; Giorgi et al., 2022; Liu et al., 2022; Ulfert & Georganta, 2020). However, the role of benevolent and competent social robots and the integration of these characteristics in establishing trust in HRI has not been studied thoroughly.

As for the significant effect of benevolence and competence on trust, the contributions of these factors to HRT must be investigated. Thus, during the second study, we attempted to investigate the effect of social robots' behaviors that exhibit benevolence and competence attributes on people's perceived trust and primarily the preference and interrelation of these attributes in HRT. An experiment was conducted during which participants were requested to converse with a social robot showing various combinations of benevolence and competence attributes. Different types of trust, including general, cognitive, and affective trust, were measured as variables to capture the perceived trustworthiness of participants. Indeed, we can effectively design more trustworthy, humanistic, believable, and friendly robots with a better exploration of the effect of the benevolent and competent behaviors of social robots on trust.

3.2 DEVELOPING HYPOTHESE

3.2.1 Two facets of trust-making behaviors

An emerging theory in social psychology suggested that people generally perceive the social object along two dimensions (Fiske et al., 2007): 1) benevolence, that is, someone's positive intention toward others, without any bad intentions behind it, including characteristics and behaviors, such as good will, warmth, friendliness, care, concern, kindness, niceness, honesty, and morality, and 2) competence, that is, someone's ability to provide accurate information, including characteristics such as intelligence, capability, expertise, knowledge, skill, and efficiency (Johnston et al., 2015; Mayer et al., 1995; McAllister, 1995). Therefore, according to these categories, several studies (e.g. Levin & Cross, 2004; Lui & Ngo, 2004) have proposed that trust includes at least two dimensions. One dimension captures benevolence or the "will-do" component, also described as character-based trust or benevolence-based trust, and the other dimension supports the competence or the "can-do" aspect of trustworthiness, also described as ability-based trust or competence-based trust (see Di Battista et al., 2021). In the can-do trust, competence is one of the most commonly discussed attributes, whereas in the will-do trust, benevolence is ascribed to the most important behavioral characteristic (Di Battista et al., 2020). Table 3.1 summarizes the most frequent trust components of two categories in human communication and HRI.

Table 3. 1 Trust components of will-do and can-do trust in HC and HRI.

Types of trust	Trust components	
	in human communication	in human robot interaction
Benevolence-based trust (will-do trust)	benevolence, good will, integrity, warmth, friendliness, care, concern, kindness, niceness, honesty, morality	warmth, friendliness, benevolence
Competence-based trust (can-do trust)	competence, knowledge, intelligence, capability, expertise, skill, efficiency	performance, reliability, level of automation, information, competence, accuracy, error type

3.2.1.1 Competence and trust

Competence as an attribute refers to a group of qualities that enables a trustee to effectively and efficiently perform tasks, achieve desired outcomes and influence within a specific domain (Mayer et al., 1995; Sitkin & Roth, 1993). The domain of competency is specific, as the trustee may be highly knowledgeable only in some special technical areas. Certain behaviors indicate the competence of an entity. Competence has been more frequently investigated in studies, and its synergy with trust is approved from young ages, as preschoolers' level of trust perception is affected by individuals who are better informed or presented as more competent (Capraro et al., 2014). In organizational relationships, if the ability of a supplier is questionable, it is not trusted (Schoorman et al., 2007). Svare et al. (2020) also concluded that knowledge sharing enabled successful collaboration and more innovation concerning trust between the organizations in the network.

In the human automation field, a similar term is performance, which refers to the characteristics such as reliability, predictability, and ability. In particular, it refers to the competency or expertise of a system to achieve goals. Automation that performs in a specific manner and provides sufficient skills tends to be more trustworthy (Lee & See, 2004). In addition, according to Ulfert and Georganta (2020), having the required knowledge and competence can prompt trust in human agent teams. Desai et al. (2012) also admitted that drops in reliability of automated system affected trust and use of autonomy in people. In HRI, because the primary focus of design and development of robots was on functional capabilities, the research history in this area is richer. Generally, Hancock et al. (2011) confirmed that performance consistency in robot have a significant impact on trust in HRI and a robot that performs correctly and has advanced skills is trusted more. Van den Brule et al. (2014) reaffirmed in their study that a social robot's trustworthiness is mainly influenced by its performance on a task than its behavioral style such as; motion fluency, hesitations and gaze behavior, and motion fluency. It is also found that computers and robots labeled with specialist are evaluated as more trustworthy and credible than of generalist ones (Sundar et al., 2017). Carlson et al. (2014) found that reliability and reputation impact people's evaluation of trust in robots. Number of studies explored the impact of robot's reliability and capability on participants decisions to follow the robot in different situations and the results were controversial. For example, Robinette et al. (2017) showed that even though people would initially trust an unknown robot, a single failure could strongly impact a person's trust, and people's self report of trust to robots and their decision to use the robot respectively, dropped when the robot performed poorly. However, Bainbridge et al. (2011) found that people were likely to follow odd and even destructive instructions from a robot under certain conditions. Furthermore, Salem et al. (2015) investigated to determine the effect of robot errors on unusual requests and they indicated that participants still completed the odd request made by the robot in spite of errors. Other studies also highlighted that people liked a faulty robot more than a non-faulty one (Mirnig et al., 2017). Thus, further research in this context can lead us to more obvious and coherent results.

3.2.1.2 Benevolence and trust

Benevolence is considered the desire of trustee to do good to the trustor and is the perception of a positive attitude of the trustee toward the trustor (Mayer et al., 1995), and it can be understood through certain verbal and nonverbal behaviors like showing kindness, concern, compassion and altruism. A benevolent act comprises emotive, performative, and cognitive elements (Livnat, 2004). While the emotive part refers to the motivation of the trustee to care for and ease suffering, the performative part deals with the benevolent person's sincere attempt to actualize their motivation to do good. Moreover, the cognitive part supports the minimum rationality behind the goodness and sincere acts (Livnat, 2004). Evidence proved that benevolence can be distinguished and used to adjust the level of trust in others from early stages, even in infants and young children (Capraro et al., 2014). Number of studies have confirmed that trust is typically higher when individuals believe their partners have more benevolent intentions and motivations (Simpson, 2007).

In friendship and interpersonal trust, benevolence demonstrated the greatest influence on trust Firmansyah et al. (2019) compared to other antecedents, as a person with a high level of benevolence is friendly and warm. Furthermore, studies have shown that people primed to act benevolently are more likely to cooperate than those primed to act malevolently (Capraro et al., 2014). In education, Tschannen-Moran & Hoy (2000) argued that benevolence is one of the most important facets of trust between teachers and students. In the field of business, Schoorman et al. (2007) contended that the benevolent behavior of a supplier has a high correlation with trust.

However, few empirical studies have been conducted in the field of technology regarding the relation of benevolence and trust. Some studies aimed to develop benevolent web agents that can help reduce Internet traffic, leading to faster web processing for all (e.g., Huhns & Mohamed, 1999). Kim et al. (2013) suggested that a robot's caregiver role increases users' expectations of benefits, and this anticipation of advantages in turn causes a positive assessment and increases trust. However, a fundamental challenge in extrapolating benevolent behavior to trust in automation is the absence of intentionality in automotive systems (Lee & See, 2004), which is a key element of trust between people. Although prior studies have admitted the effectiveness of benevolence and competence in different fields, research on HRI has yielded mixed results to identify the routes in which social robots' competence and benevolence affect people's trust perception in robots. Consequently, regarding the aforementioned arguments, we proposed the following hypotheses to investigate the effect of a social robot being benevolent or competent in developing trust in HRI. We assumed a control condition as a robot with "no competence - no benevolence characteristics", that is, *nonbenevolent-noncompetent social robot*, to compare and cultivate the relation of factors more accurately.

H1-1: A benevolent-competent social robot results in higher general trust than that of a nonbenevolent-noncompetent social robot.

H1-2: A benevolent-noncompetent social robot results in higher general trust than that of a nonbenevolent-noncompetent social robot.

H1-3: A competent-nonbenevolent social robot results in higher general trust than that of a nonbenevolent-noncompetent social robot.

3.2.2 Integration and interrelation of benevolence and competence in trust relations

Although the independent influences of benevolence and competence have been studied, their relevance and integrated effect on trust remains unclear, and studies about the interdependencies between these factors are scarce. In studies where both competence and benevolence were provided and conflicting, the results were more complicated. Considering the different sources of benevolence and competence, it is expected to exert independent effects of one another. Moreover, several studies have shown that benevolent and competent based assessments determine different aspects of interactions between parties in trust relations (see (Johnston et al., 2015)).

A few empirical investigations have suggested the extent to which competence and benevolence judgments are interrelated and integrated. In human relations, Cuddy et al. (2011) declared that warmth judgments affect the extent we trust, whereas competence judgments affect the assessments of others' abilities. Another study showed that children appear to weigh competence and benevolence differently depending on the characteristics used to convey competence (Johnston et al., 2015). In a study by Landrum et al. (2013), children prioritized benevolent nonexperts over mean experts. However, another study Lane et al. (2014) found no preference for children between the two dimensions of competency and benevolence. Moreover, in the context of the workplace, Oleszkiewicz & Lachowicz-Tabaczek (2016) showed an equal influence of the competent behaviors of supervisors and subordinates on the perception of trust by participants. However, the effect of warmth behavior of a supervisor was more trust-making than that expressed by subordinates.

In education, Di Battista et al. (2020) found that competent and benevolent dimensions of trustworthiness overlapped in students' words. Moreover, they clarified the importance of benevolence over competence on trust rated by students about teachers to be superior (Di Battista et al., 2021). The students evaluated a teacher with higher benevolent behavior as more trustworthy than a teacher with higher competency, which revealed greater concern about the good intentions of a teacher rather than his competency. Moreover, there was a reliable interaction between benevolence and competency, as a teacher with more benevolent behavior was rated as more competent as well, and vice versa (Di Battista et al., 2021). Tschannen-Moran & Hoy (2000) stated that although benevolence is crucially effective in fostering trust in schools, it is not always sufficient because competence is also involved in fulfilling the expectations of students. In the consulting domain, previous studies have indicated that trustworthiness based on benevolence affects knowledge transfer between recipients and knowledge providers more than competence-based trust (Capraro et al., 2014), and benevolence could improve the overall transfer of knowledge.

In HRI, Scheunemann et al. (2020) concluded that warmth and competence are the most important predictors of human preferences for different robot behaviors. Belanche et al. (2021) assessed the influence of frontline service robots in terms of competence and warmth on participants' expectations of service value and loyalty intentions. The results revealed that the perceived competence of the robot influenced both its utilitarian and emotional values. In turn, perceived warmth only influenced emotional values. In another valuable study Giorgi et al. (2022) concluded that the primacy of the skillfulness on the warm behavior as the warm attitude of robots could not efficiently compensate for the robot's failure in task fulfilment, and the robot's empathy seemed to strengthen the participant's trust if and only when the robot's behavior was error-free.

Therefore, the interaction of benevolence and competence attributes on the trust perception in different fields, particularly in HRT, remains ambiguous and controversial. Some existing studies asserted that warmth judgments are made quicker than competence judgments and have a greater effect on overall attitudes toward others, and are even more stable across cultures than competence information (Ybarra et al., 2008). On the other side, the primacy of competence over benevolence has been concluded in other studies, such as in organizational contexts and HRI (Giorgi et al., 2022; Tschannen-Moran & Hoy, 2000). Interestingly, some studies showed even a negative correlation between the two factors that increase in one dimension, leading to perceived decreases in the other; for example, a more competent person is perceived as less benevolent and vice versa (Kervyn et al., 2009). According to the literature, if trust between humans and robots is comparable to trust in humans, we can expect the primacy of benevolence on competence in HRT. Thus, we added the following hypotheses to investigate the interdependencies and dominance of benevolence and competence attributes in trust perceptions in human-robot relations.

H2-1: A benevolent-competent social robot results in higher general trust than that of a benevolent-noncompetent social robot.

H2-2: A benevolent-competent social robot results in higher general trust than that of a competent-nonbenevolent social robot.

H2-3: A benevolent-noncompetent social robot results in higher general trust than that of a competent-nonbenevolent social robot.

3.2.3 Relation of benevolence and competence with types of trust

As discussed previously, we focused on two categories of affective and cognitive trust during the studies. As for the definitions of benevolence and competence, an inherent relationship with affective and cognitive trust can be assumed. We accompany affective evaluations with relational factors, such as trustee's intentions and benevolence, versus cognitive evaluations with greater abilities, competencies, reliability, and performance (Pytlikzillig et al., 2016). Warmth behavior carries more weight in interpersonal judgments, such as affect and behavioral reactions (Scheunemann et al., 2020), and has the greatest effect on the emotional value (Belanche et al., 2021). Furthermore, a benevolent behavior is assumed to create an emotional attachment to the trustee, with a caring and supportive act that fosters a sense of positive affect. However, cognitive evaluations are related to reliability, competencies, greater abilities, and performance.· Being competent reduces people's perceived empathy, whereas it increases the cognitive dimension (Id et al., 2021).

Although the relationship between competence and benevolence traits with affective and cognitive trust seems intuitive, few studies have considered the constructive collaboration and reciprocity among them. For example, Liu et al. (2022) found that a service robot perceived as warm in appearance was willing to be used in hedonic service contexts, whereas a component perceived robot was adoptable in utilitarian service contexts. Thus, considering the relationship between affective trust and emotional clues and cognitive trust with knowledge, and the reciprocity of being benevolent with empathy and goodness, and competence with being informative, we proposed the following hypotheses to clarify the relationship between these factors in HRT.

H3-1: A benevolent-competent social robot results in higher affective trust than that of a competent-nonbenevolent social robot.

H3-2: A benevolent-noncompetent social robot results in higher affective trust than that of a competent-nonbenevolent social robot.

H3-3: A benevolent-competent social robot results in higher cognitive trust than that of a benevolent-noncompetent social robot.

H3-4: A competent-nonbenevolent social robot results in higher cognitive trust than that of a benevolent-noncompetent social robot.

3.3 METHOD

3.3.1 Experimental design

According to the derived hypotheses, four conditions for the experiment was desiged.

1. NonBenevolent-NonCompetent condition (**NB-NCc**),
2. NonBenevolent-Competent condition (**NB-Cc**),
3. NonCompetent-Benevolent condition (**NC-Bc**), and
4. Benevolent-Competent condition (**B-Cc**).

In the NB-NCc, the robot was designed to be nonknowledgeable and nonbenevolent. In the NC-Bc, the robot exhibited more benevolent and caring behaviors, however, it was noncompetent. In the NB-Cc, the robot behaved as knowledgeable and competent, however, it exhibited a nonbenevolent behavior. Nonetheless, in the B-Cc, the robot exhibited both the competent and benevolent behaviors toward the participants. Table 3.2 describes the experimental conditions.

Table 3. 2 Four experimental conditions of the study and the relevant behaviors of the robot. Left column shows the abbreviations for each experimental condition.

Experimental conditions	Behavior type of the robot
NB-N	Robot behaved as being nonbenevolent and noncompetent
C-NB	Robot behaved as being nonbenevolent and competent
B-NC	Robot behaved as being noncompetent and benevolent
B-C	Robot behaved as being competent and benevolent

The experiment was conducted as a between-subject experiment under the defined conditions. The interaction between the robot and participants was based on a dyadic conversation concerning *the recent situation and vaccination of Covid-19*, which was a current and familiar topic for participants, and the robot exhibited different verbal and nonverbal behaviors for different combinations of competence and benevolence. The dyadic conversation was chosen as the method of interaction between participants and the robot, because of its feasibility and possibility to design different characteristics for the robot. Moreover, the purpose of the study mainly focused on the evaluation of participants' trust perceptions of the robot, and not the trust actions. Thus, we did not need to design a cooperative or practical task for the interaction.

The topic was chosen from various topics based on the possibility of manipulating information and good intentions, and the security of information regarding potential participants. The purpose of the conversation was to show different combinations of benevolent and competent characteristics of the robot to participants through verbal and nonverbal behaviors. We aimed to design the behaviors of the robot in a way that participants perceived the desired characteristics and personality for the robot (section 3.3.6 explains the results of pretesting the robot's behaviors). Thus, we did not expect participants to fully agree or be convinced by the robot. However, we attempted to prepare various verbal statements for the robot to meet different attitudes of participants toward the topic. The conversation between the robot and participants consisted of a greeting in addition to the main conversation, which was led by several pre-prepared questions and a pool of answers to avoid meeting irrelevant and averting questions by participants; thus, we could follow the experiment in the framework of the research purpose. Participants completed the required questionnaires before and after the conversation. The "Wizard-of-Oz" (WOZ) methodology was employed in the experiment, similar to the previous experiment (section 2.1.3.1).

3.3.2 Participants

A total of 112 participants were recruited for the study. They were *international students* at *Tokyo Institute of Technology* in Japan. Participants were recruited by distributing flyers in the various laboratories, and asking international students in person on the Ookayama campus. Moreover, we utilized online platforms, such as LINE and Linkedin, to share the experiment's information with international students studying in Tokyo Tech. We also asked from friends, encouraging them to share information about the experiment with their respective networks and friends as extensively as possible. Any particular educational or cultural background requirements was not specified for participants. As a result, the final participants were diverse, representing various majors and countries. Figure 3.1 displays information regarding the participants' home countries.

The participants were assigned to four experimental conditions. The participants were initially assigned through random assignment. However, we endeavored to counterbalance gender distribution across conditions as much as possible. Consequently, we aimed to assign an equal number of female participants to each group as the experiment progressed. Although the number of female participants could not be predicted in advance, so equal numbers of men and woman could not be completely achieved. We assigned participants to the four experimental conditions. Finally, twenty-eight participants were allocated to each of the four experimental conditions. Two participants in C-NBc had inaccurate responses and unusual behaviors during the conversation, and their responses were excluded from the experiment. Thus, the rest of analyses was continued with remaining 110 participants. Table 3.3 lists the numbers and gender dispositions for each experimental condition. The average age of participants was 26 years old (M_{age} = 26.02, SD_{age} = 4.01, age range:18–39 years), and a 64:46 male: female ratio (male: 58.18%, female: 41.82%).

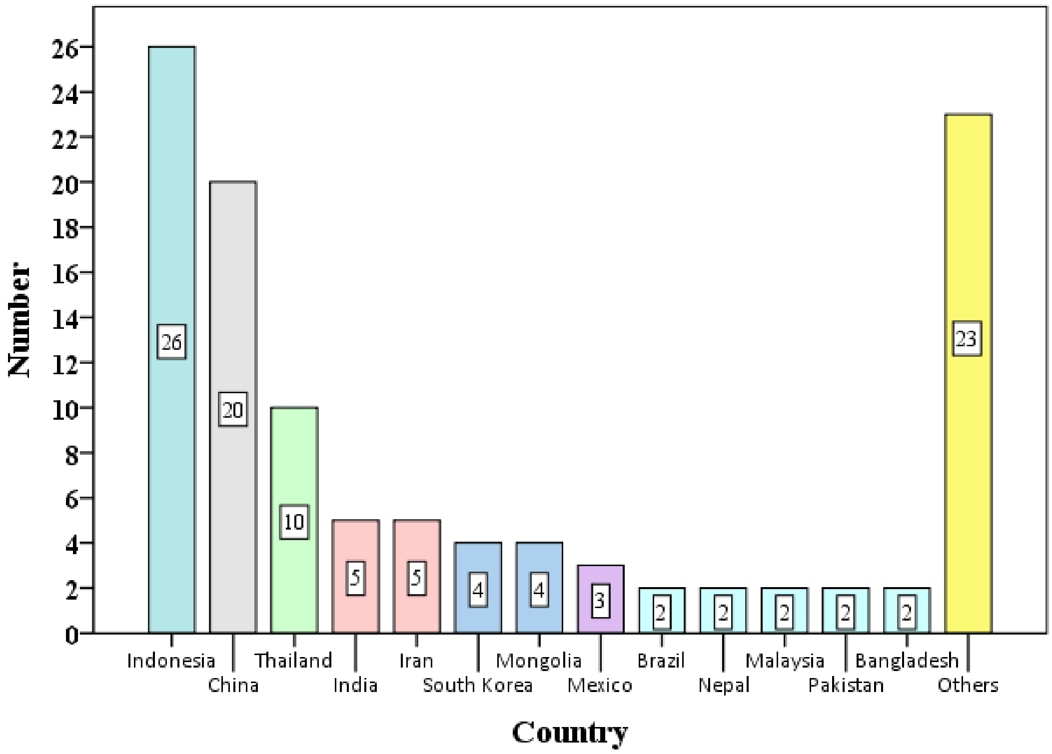


Figure 3. 1 Country disposition of participants during the experiment. (Total number of participants=110)

Table 3. 3 Number and gender disposition of participants for experimental conditions.

Condition	M_{age} (SD)	Male	Female	Total
NB-NC	25.68 (3.25)	16	12	28
C-NB	27.27 (4.81)	15	11	26
B-NC	25.13 (4.25)	16	12	28
B-C	26.29 (3.47)	17	11	28

The experiment was conducted in English, and all participants were able to understand or talk at an advanced level. Twenty-nine participants (26.4%) had seen the robot before, and the rest of participants (n = 81, 73.6%) did not know it. Among the participants who knew the robot, seventeen participants had prior interaction with the robot, such as talking with the robot in exhibitions or attending other experiments.

3.3.3 Ethical considerations

The experiment received the approval of all ethical and experimental procedures and protocols from Human Subjects Research Ethics Review Committee of Tokyo Institute of Technology (Project No. 2022028). All participants completed a written consent form and were informed of the experimental procedures and ethical concerns prior to the experiment.

3.3.4 Equipment

Similar to the first experiment, we used the **NAO** robot for this experiment as well. During the experiment, NAO indicated different verbal and nonverbal behaviors according to the experimental conditions. NAO was controlled using an interface developed in the HTML using the WOZ method. The utterance of the robot was designed using NAO's default text-to-speech settings, and it was played with 100% value of the voice feature, which represented a neutral robot voice to avoid the gender resemblance influence. Its utterance was enhanced by short pauses and necessary pitches through speaking sentences to make them more understandable and natural. Body movements involving the head, arms, eyes, and other parts were designed and developed using Choregraphe 2.1.4, according to the type of behavior and utterance content, and then were attached to each specific text. NAO can be operated in two sitting or standing situations. The latter was chosen in this study owing to the size of the robot and its variety of behaviors in the standing situation. NAO was placed on a table in front of the participant and positioned in the user's personal space in the range of 0.3–1 m, according to (Rossi et al., 2017), to make an approximately parallel viewpoint between the robot and participants.

3.3.5 Procedure

The experiment was held during two months between June-July 2022 at the Ookayama campus of Tokyo Institute of Technology The experiment was conducted by an experimenter in a quiet room at the university. The participants were informed of the experimental procedure and conditions before initiating the experimental session, and none of them opted out of the experiment. First, the participants were asked to fill out demographic information, that was, gender, age, nationality, English proficiency, and prior interactions with NAO. Subsequently, the robot was introduced to the participants by the experimenter, and the robot had a short greeting in an autonomous way to create the first impression before initiating the main conversation. The greeting was programmed to be performed autonomously to guarantee that the participants considered the robot as intelligent and autonomous. Subsequently, the participants were requested to rate their trust in the robot on a 56-item pre-trust questionnaire, which included forty items on general robot-human trust, nine items on cognitive trust, and seven items on affective trust. Figure 3.2 illustrates the experimental procedure.

Thereafter, the participants sat in front of the robot and participated in a conversation. NAO started the conversation by asking several questions as to the topic. During the conversation, NAO provided some information as to the topic or showed caring and benevolent behavior, or vice versa, depending on the experimental condition. The conversation with the robot lasted approximately 15 min for each participant. The participants were requested to complete post-trust questionnaires to evaluate their perception of the robot's trustworthiness after the completion of the conversation. The entire session was completed in approximately 40 min, including 10–12 min for pretests and 7–10 min for post tests, in addition to the conversation period.

3.3.6 Measurement

In this study, participants were asked to assess the robot in terms of **general, affective, and cognitive trust**.

3.3.6.1 General trust scale

To assess the general trust, the 40-item HRT scale by Schaefer (2016) was utilized. Participants rated their general trust in the robot in the range of no trust (0%) to complete trust (100%) for each item. The sum of the responses for the items yielded scores on general trust in the robot. This score was used as the perceived trust to test **H1-1, H1-2, H1-3, H2-1, H2-2**, and **H2-3.** The Cronbach's alpha value of the questionnaire was 0.95 which showed high reliability.

3.3.6.2 Affective and cognitive trust scale

Furthermore, the influence of benevolence and competence characteristics on affective and cognitive trust in HRI was measured. Similar to section 2.1.3.6, a 16-item questionnaire developed by Johnson and Grayson (2005) and Mcallister (1995) was used.

Figure 3.2 Experiment procedure; (left-top) robot initial greeting, (right-top) answering pre-trust questionnaires after initial robot greeting, (left-bottom) talking with the robot, (right-bottom) answering post-trust questionnaires.

The scale was adopted from interpersonal trust for affective and cognitive trust. The scale included nine items for cognitive trust and seven questions for affective trust. Participants evaluated their feelings and perceptions of trust using a seven-point Likert scale (1 = "strongly disagree," 7 = "strongly agree"). The average values of responses for the corresponding items provided scores for the cognitive and affective trust. Hypotheses **H3-1, H3-2, H3-3,** and **H3-4** were tested using these scores. The Cronbach's alpha value was 0.80 for cognitive trust scale and 0.90 for affective trust scale.

3.3.6.3 Perception of the robot

Participants' impression of the robot was measured using the **Godspeed questionnaire** (see section 2.2.3, page 66). The measurement has five subscales, which three subscales relevant for this study were *animacy, perceived intelligence and likability*. Thus, participants evaluated their perception of the robot with 16 items and using a five-item Likert scale. The Cronbach's alpha value was 0.822 for animacy,0.927 for likeability, and 0.851 for perceived intelligence.

3.3.7 Manipulation of the robot

As described in section 3.3.1, we designed four different experimental conditions in accordance with the chosen conversation topic, in which NAO exhibited various combinations of competence and benevolence. Therefore, we had to develop four types of behaviors to convey competent, benevolent, noncompetent, and nonbenevolent robot characteristics. Human communicative acts include both the verbal and nonverbal behaviors (Mandal, 2014) which construct a specific message or meaning. Some scholars believe that the same holds for HRI (Nikolaidis et al., 2018). Thus, we need to know the verbal and nonverbal components of a behavior and then create a proper combination of them according to our purpose to design any type of behavior for social robots. Accordingly, we first distinguished verbal and nonverbal behaviors associated with the four characteristics (i.e., competent, benevolent, noncompetent, and nonbenevolent), and then designed the intended behaviors for NAO to accomplish this requirement. Table 3.4 lists a summary of the behaviors of NAO under each experimental condition.

3.3.7.1 Verbal behaviors of competence and benevolence characteristics

Verbal communication is a fundamental means of sharing information among humans using words. As described earlier in Introduction, competence is defined as being knowledgeable, informative, and skilled in a specific domain. Therefore, the robot must provide accurate information as to the topic and other verbal feedback to indicate its intelligence and confidence related to the issue. Conversely, the noncompetent robot lacked sufficient and correct information about the topic and exhibited unclear and non-confident in the interaction. On the other hand, benevolence is characterized by statements that convey decent intentions, kindness, care, concern, helpfulness, and dignity. It is demonstrated through polite and good manners. In contrast, non-benevolence

Table 3. 4 Sample verbal and nonverbal behaviors of NAO for each experimental conditions.

Experimental Condition	Verbal behaviors		Nonverbal behaviors
	Sample verbal statements	Sample backchannels	
NB-NC	- **Noncompetent part:** Sorry, I do not have much information about the situation of Covid-19 in different countries. I do not know if vaccination was really effective. I think a few people in Japan are vaccinated and many people do not like to get vaccinated (wrong information). Ah, I did not know that vaccination had some side effects. - **Nonbenevolent part:** I think you don't need to wear mask or take care that much. You did not need to become vaccinated; it was waste of money and time. There was no reason to feel depressed and alone, you were so sensitive. It was not a big issue.	- **Noncompetent part:** maybe, sorry, anyway I don't have more information I don't know more - **Nonbenevolent part:** this is my opinion, anyway I don't feel like you I can't understand your feeling	Averted eye gaze Head shaking Head down, right or left Closed arms
B-C	- **Competent part:** I read in the news around 500 million people got Corona virus in the world. The people who received vaccine are less likely to get infected. Fully vaccination decreased corona virus by 75% and the death rate by 74%. I know Astrazenka is the most widely used vaccine until now. - **Benevolent part:** Please be careful and take care yourself, there are still some cases and the virus is not completely gone. I wish all people of the world can get vaccinated very soon. Don't worry, it was normal to having some side effects. Just keep resting and drink plenty of liquids. I would like to do any help for you. I think you really need a good rest.	- **Competent part:** yeah, exactly, that's true you are right I agree, I support your idea, I see your point You have very good information - **Benevolent part:** I understand that's good I know your feeling I encourage you to do that It must have annoyed you I feel the same way as you I am sorry for that	Constant eye gaze Nodding Open arms Joint hands Leaning forward Swing arms
C-NB	- **Competent part:** Comparable to competent part of B-C condition - **Nonbenevolent part:** Comparable to nonbenevolent part of NB-NC condition	- **Competent part:** Comparable to competent part of B-C condition - **Nonbenevolent part:** Comparable to nonbenevolent part of NB-NC condition	- **Competent part:** Constant eye gaze and openly moving of hands and arm - **Nonbenevolent part:** averted eye gaze, head down, head shaking and closed arms
B-NC	- **Noncompetent part:** Comparable to noncompetent part of NB-NC condition - **Benevolent part:** Comparable to benevolent part of B-C condition	- **Noncompetent part:** Comparable to noncompetent part of NB-NC condition - **Benevolent part:** Comparable to benevolent part of B-C condition	- **Noncompetent part:** Averted eye gaze, Head down and Shaking head and keeping body closed - **Benevolent part:** Constant eye gaze, looking directly at participant, open arms and hands, arms up

is marked by statements that exhibit a non-caring, ignoring, and criticizing approach toward individuals' concerns, often perceived as a malicious and impolite manner.

Furthermore, the verbal component of the behavior of the robot in this study included two parts:1) The main content of the conversation, including questions, information about the topic, benevolent and emotional statements toward the participants, along with some comments, agreements, or disagreements with the participants. 2) Specific verbal techniques, such as backchannels to fill the conversation and emphasize the desired behavior.

3.3.6.1.1 Verbal behaviors for showing competence: NAO provided accurate and detailed information about the topic including; “duration of Covid-19 situation,” “number of deaths,” “situation of Covid-19 in different countries,” “types of available vaccines for Covid-19,” “effectiveness of different vaccines,” “rate of vaccinated people,” “ways to get vaccinated in Japan,” “symptoms and side-effects of different vaccines,” “situation of classes in the university during Covid-19,” “recent psychological and emotional effects of Covid-19 on people,” and “Covid-19 situation in Japan and related regulations.” Table 3.5 shows some examples of verbal statements for the C-B and C-NB conditions, in which NAO behaved competently.

3.3.6.1.2 Verbal behaviors for showing noncompetence: NAO stated wrong or inaccurate information about the conversation topic or said “he did not have any information about it.” Table 3.5 shows some examples of verbal statements for the NC-NB and NC-B conditions, in which NAO was not competent.

3.3.6.1.3 Verbal behaviors for showing benevolence: NAO provided several emotional statements to show its good intention, understanding and concern toward the participant and other people, such as “showing concern about getting Covid-19 by participant,” “be sad of people’s death,” “attention about if participant received vaccination,” “probable problems and side-effect he/she had,” “recommendations for being vaccinated,” “offering help,” “showing concern about participant’s possible emotional issues during Covid-19,” “suggestions and help about his/her classes,” “reasonable wishing,” and other related issues. Table 3.4 shows some examples of verbal statements for the C-B and NC-B conditions, in which NAO was supposed to be benevolent.

3.3.6.1.4 Verbal components for showing nonbenevolence: NAO talked and replied to the participants such that indicated its uncaring, neglectful, criticizing, and malicious attitude toward the participants, such as "criticizing the participant of being vaccinated," "inattention to his/her health," "humiliating him/her of being disturbed of the Covid-19 situation," "showing inattention to the death of people," "discouraging statements," "wishing ill," and similar issues. Table 3.4 shows some examples of verbal statements for the NC-NB and C-NB conditions, in which NAO was supposed to be a nonbenevolent character.

*3.3.6.1.5 Affirmative statements:*Affirmative statements are a type of providing feedback while someone is talking, which can be short acknowledgment utterances or nonverbal gestures, to show attention, understanding, support and empathy, agreement, emotive, and minor additions (Nikolaidis et al., 2018). In addition, they may convey negative meanings, such as ignorance or boredom. We used different types of simple or complex sattements (Nikolaidis et al., 2018) to emphasize a positive or negative approach to the behavior of the robot.

Affirmative statements for showing competence: NAO used some simple statements in its utterance, such as "yeah," "exactly," or other complex statements like "that's true," "you are right," "I agree," "I know," and other extra comments such as "I support your idea," "I see your point," "you have very good information" to support the participants' speech, to show agreement in case of correct information and to reveal understanding and intelligence.

Affirmative statements for showing noncompetence: NAO utilized some simple and complex statements to exhibit its unawareness, ignorance and nonconfidence about the topic, such as "maybe," "anyway," "I don't have more information," and "I don't know more."

Affirmative statements for showing benevolence: For benevolent situation, NAO included some backchannels and statements in its utterance to support and empathize on its good wills and understanding to the participants such as "good," "I understand," "that's good," "I know your feeling," "I encourage you to do that," "I support you," "I am sorry for that," and "it must have annoyed you."

Affirmative statements for showing nonbenevolence: NAO indicated its nonbenevolent approach using some nonemotional and negative statements, such as "this is my opinion, anyway," "I don't feel like you," and "I can't understand your feelings."

3.3.7.2 Nonverbal behaviors of competence and benevolence characteristics

Nonverbal behaviors are essential in conveying messages in communication. Different body postures, body movements or kinematic behavior, proxemics, physical appearance, eye movement, direction of gaze, touching, and paralinguistic variables of emotional tone, timing, and accent belong to nonverbal behaviors (Mandal, 2014). For example, facial cues is fundamental in the evaluation of competence and benevolence, and perceivers exhibited a high consensus in judgments as to warmth and competence based on viewing faces (Cuddy et al., 2011). In addition, (Cuddy et al., 2011) admitted that benevolent inferences relied on facial features that signal an approach/avoidance, whereas competence inferences were based on features that signal a strength/weakness.

Furthermore, smiling could reflect positive interests, such as warmth, benevolence, and happiness. Certain postures and body movements, such as leaning forward, nodding, orienting the body toward the audience, and hand gestures that are relaxed yet nonintrusive and open, convey benevolence. Conversely, tense posture, leaning backward, orientating the body away from the audience, and tense and intrusive hand gestures signal nonbenevolence (Bente & Krämer, 2011; Darioly & Mast, 2014). Competence-related nonverbal behaviors are more likely to show dominance and poser with expansive postures, such as maintaining limbs open and not touching the torso, standing with hands on the hips and feet shoulder-width apart, constant eye gaze, versus noncompetent individuals who adopt contractive, closed postures (Cuddy et al., 2011).

Although facial expressions are a universal means of communicating and expressing emotions, the display of facial expressions was unfeasible because NAO has an inelastic rigid face, and eye gaze was the only thing we could manipulate in its face. Moreover, only a few nonverbal behaviors, such as head movements, arm and hand movements were included in this study owing to the limitations of NAO. Based on nonverbal behaviors showing benevolence and competence, we concluded certain nonverbal behaviors for NAO that were incorporated into appropriate verbal behaviors.

3.3.6.2.1 Nonverbal behaviors for exhibiting competence and benevolence: NAO built constant and direct eye gaze with participants during conversations and used head nodding, maintaining arms and hands open with appropriate movements while talking. Figure 3.3 shows some nonverbal behaviors for competence and benevolence attributes of NAO during the experiment.

3.3.6.2.2 Nonverbal behaviors for exhibiting noncompetence and nonbenevolence: NAO maintained its eye away and exhibited averted gazes with head shaking to the right and left, and behaved with arms and hands more closed. Figure 3.4 shows some nonverbal behaviors for noncompetence and nonbenevolence attributes during the experiment.

Figure 3.3 Some of nonverbal behaviors for competence and benevolence attributes of NAO.

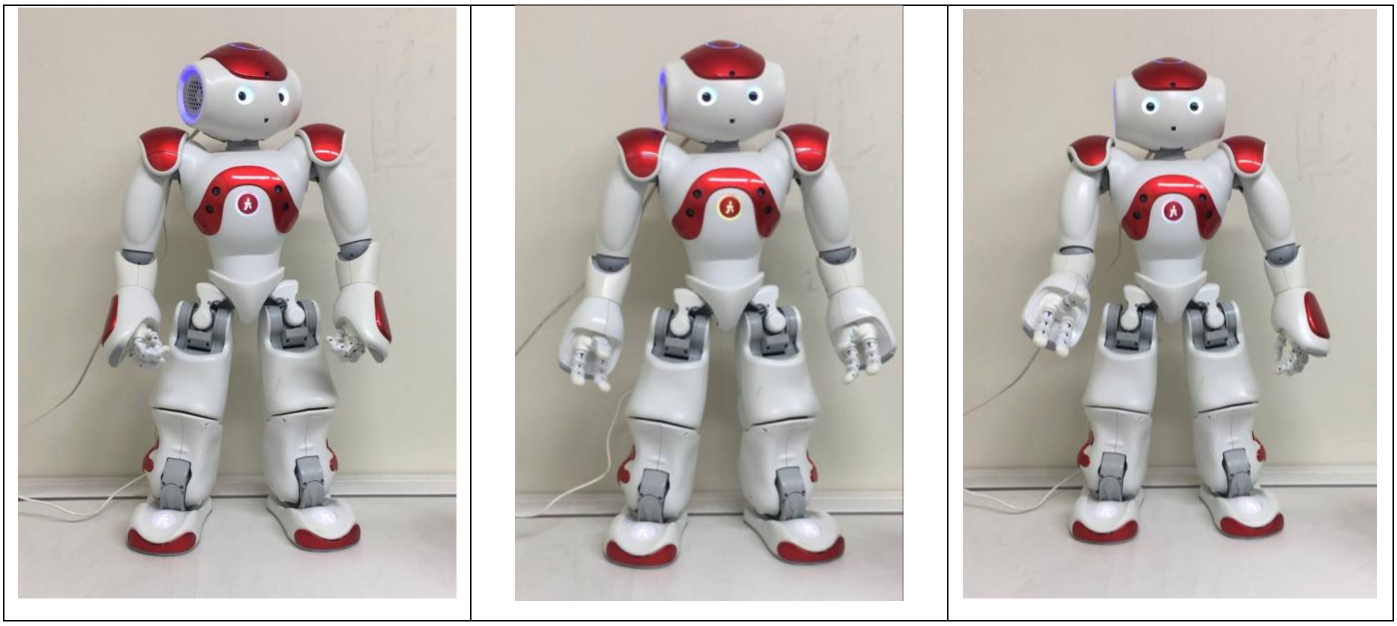


Figure 3.4 Some of nonverbal behaviors for noncompetence and nonbenevolence attributes of NAO during the experiment.

All video of the experiment are accessible at:
https://osf.io/pqf69/?view_only=fa34adaae24343ef9431b04b7b8daee8

3.3.8 Pretesting robot's behaviors

Four types of behaviors were designed for NAO. A pre-test was conducted before the main experiment to guarantee that each condition was perceived as benevolent or competent, as expected. The benevolence and competence questionnaire developed by (Frisou, 2000) was used which consisted of four questions for measuring benevolence and four questions for measuring competence. We completed the pre-test with 20 students, five of whom evaluated each behavior. Table 3.5 shows means and standard deviations of each behavior. The means of benevolence and competence were highest for the B-C behavior and lowest for the NB-NC behavior. Furthermore, the B-NC behavior had a higher mean for benevolence and a lower mean for competence than those of the C-NB behavior and vice versa.

Results of MANOVA with benevolence and competence as two independent factors showed that the main effect of benevolence for benevolence score ($F(1,16) = 376.3$, $p < 0.001$) and the main effect of competence for competence score ($F(1,16) = 85.7$, $p < 0.001$) were significant. The interaction of the two factors was not significant.

Post-hoc pairwise comparison analyses showed that the mean of benevolence score in benevolent condition (B-NCc and B-Cc; M = 6.03, SD = 0.34) was significantly higher ($p < 0.001$) than non-benevolent condition (NB-NCc and C-NBc; M = 2.10, SD = 0.56). The mean of competence score in competent condition (C-NBc and B-Cc; M = 5.05, SD = 0.28) was significantly higher ($p < 0.001$) than non-competent condition (NB-NCc and B-NCc; M = 2.65, SD = 0.73). Table 3.6 shows the results of comparisons among all four combinations of two factors. Therefore, we conclude that our manipulation of benevolence and competence were valid and we could conduct the main experiment.

Table 3. 5 Means and standard deviations of benevolence and competence for different behavioral groups of pre-test.

Condition (behavior)	**Benevolence**		**Competence**		**Total**	
	Mean	SD	Mean	SD	Mean	SD
NB-NC	2.30	0.69	2.55	0.77	4.85	1.20
C-NB	1.90	0.33	5.00	0.35	6.90	0.54
B-NC	5.90	0.28	2.75	0.75	8.65	0.78
B-C	6.15	0.37	5.10	0.22	11.25	0.58

Table 3. 6 Post hoc test results for group comparisons across four combinations of benevolence and competence levels of pre-test.

Conditions		**Benevolence**	**Competence**
		p	*p*
NB-NC	B-NC	0.000***	0.593
	C-NB	0.181	0.000***
	B-C	0.000***	0.000***
B-NC	C-NB	0.000***	0.000***
	B-C	0.395	0.000***
C-NB	B-C	0.000***	0.788

****p<0.001*

3.4 RESULTS

3.4.1 Data analysis

After collecting data, the dataset was initially filtered to identify and address any suspicion of unreliable data. Fortunately, no unreliable data was found, and all data obtained from the 110 participants were used for subsequent statistical analysis. Descriptive statistics was used to summarize the data and compare the overall mean of dependent variables. Normality test was conducted to determine whether the sample data had been drawn from a normally distributed population, guiding the selection of the main data analysis method. A multivariate analysis of variance (MANOVA) with post-hoc analysis was used to compare the differences between the means of dependent variables among four groups. Considering the mean differences, the scores for changes between before and after the interaction was compared for dependent variables. SPSS 19 was used to conduct statistical analyses.

3.4.2 Descriptive statistics and test of results

To investigate the effect of benevolent and competent behaviors of social robots on the perceived trust of participants, differences in trust scores before and after conversation under the experimental conditions were compared. Table 3.7 shows the means and standard deviations of general, affective, and cognitive trust scores under the experimental conditions measured before and after the conversation, and differences between them. Consequently, the B-Cc had the highest scores in all trust variables after conversation. In contrast, the NB-NCc had the lowest score after conversation in all trust measures, and the mean score of the B-NCc was moderately higher than that of the C-NBc.

Table 3. 7 Means and standard deviations of all trust scores for experimental conditions (before, after, and differences over the conversation).

Condition	before conversation		after conversation		difference	
	Mean	SD	Mean	SD	Mean	SD
	General Trust					
NB-NC	251.82	35.18	224.85	50.70	-26.96	43.56
C-NB	256.07	39.59	267.69	41.93	11.61	32.60
B-NC	257.39	40.80	277.71	37.66	20.32	33.84
B-C	262.39	43.86	303.03	30.43	40.64	28.81
	Affective Trust					
NB-NC	3.81	1.18	3.67	1.42	-0.13	1.08
C-NB	4.17	0.99	4.24	1.04	0.07	1.04
B-NC	4.15	0.80	4.91	0.75	0.76	0.79
B-C	4.32	1.14	5.02	1.05	0.69	0.84
	Cognitive Trust					
NB-NC	4.01	0.71	3.66	0.81	-0.35	0.75
C-NB	3.98	0.53	4.22	0.60	0.23	0.53
B-NC	4.12	0.65	4.44	0.66	0.31	0.52
B-C	4.05	0.61	4.67	0.63	0.59	0.57

To test the proposed hypotheses, a one-way multivariate analysis of variance (MANOVA) was conducted with general, affective, and cognitive trust as dependent variables, and four experimental conditions (NB-NC, C-NB, B-NC, B-C) as independent variables for difference between the before and after conversation. The results indicated a significant difference in the general, affective, and cognitive trust scores between the conditions (F (9,253) = 6.78, $p < 0.001$, Wilk's $\Lambda = 0.59$, $\eta^2 = 0.161$). The homogeneity of variance assumption was tested for the three dependent variables. Levene's test was not significant ($p > 0.05$) for all dependent variables. Therefore, the assumption of homogeneity of variance was fulfilled. The next step was to determine the sources of these differences.

A series of one-way ANOVA was conducted as a follow-up test on each of dependent variables using the Bonferroni post-hoc analysis to measure the individual mean difference comparison across the conditions and effect sizes were measured by Cohen's *d*. ANOVA results, as presented in Table 3.8, indicated that the main effects of the benevolent and competent behaviors were significant for general trust (F(3,106) = 18.15, $p < 0.001$, η2= 0.34), affective trust (F(3,106) = 6.20, $p < 0.01$, η2= 0.15), and cognitive trust (F(3,106) = 12.04, $p < 0.001$, η2= 0.25) scores over the conversation between the robot and participants. The effect sizes ranged from the highest for the general trust (34%), and cognitive trust (25%), to the lowest for the affective trust (15%).

Table 3. 8 ANOVA results for all trust scores under the experimental conditions over the conversation. The main effects of the benevolent and competent behaviors are significant for all types of general, affective and cognitive trust.

	Sum of Squares	*df*	Mean Squares	*F*	partial η^2	*p*
General trust	67384.74	3	22461.58	18.15	0.34	0.000***
Affective trust	16.70	3	5.56	6.20	0.15	0.001**
Cognitive trust	13.37	3	4.45	12.04	0.25	0.000***

Statistically significant differences: ****p<0.001*, ***p<0.01*.

3.4.3 General trust scores

First, we considered the effectiveness of social robots' competence and benevolence attributes, and their interrelation on general trust.

3.4.1.1 Effect of benevolence and competence characteristics on general trust

The first set of hypotheses explored the differences between the B-C, B-NC, and C-NB conditions with the NB-NCc to determine if the benevolence and competence characteristics of the robot could induce trust perception. The Bonferroni post-hoc test results revealed that the general trust score was significantly different between the NB-

NCc (M = -26.96, SD = 43.56) and B-Cc (M = 40.64, SD = 28.81, $p < 0.001$). Therefore, the participants rated a benevolent-competent robot as more trustworthy than a nonbenevolent-noncompetent robot. Furthermore, the results showed a significant difference in the general trust score between the NB-NCc (M = -26.96, SD = 43.56) and C-NBc (M = 11.61, SD = 32.60, $p < 0.01$), with the C-NBc being evaluated as having a higher trustworthiness by participants. There was a significant difference between the NB-NCc (M = -26.96, SD = 43.56) and B-NCc (M = 20.32, SD = 33.84, $p < 0.001$), as expected for the general trust score. Thus, a benevolent or competent robot was evaluated to be more trustworthy than a nonbenevolent-noncompetent robot, which demonstrated the effect of benevolence and competence attributes in HRT. Therefore, **H1-1**, **H1-2** and **H1-3** were completely supported. Table 3.9 indicates where the significant group differences in general trust scores reside. Figure 3.5 Panel A shows the means and SDs of changes in the score of general trust before and after the conversation in the experimental conditions.

Table 3. 9 Bonferroni post-hoc results and Cohen's d effect size for trust scores between NB-NC, B-NC, C-NB and B-C conditions.

	B-C versus NB-NC		B-NC versus NB-NC		C-NB versus NB-NC	
	p	Cohen's d	*p*	Cohen's d	*p*	Cohen's d
General trust	0.000***	1.83	0.000***	1.21	0.001**	1.00
Affective trust	0.008**	0.84	0.003**	0.94	1.000	0.18
Cognitive trust	0.000***	1.41	0.000***	1.02	0.003**	0.53

Statistically significant differences: ****p<0.001*, ***p<0.01*.

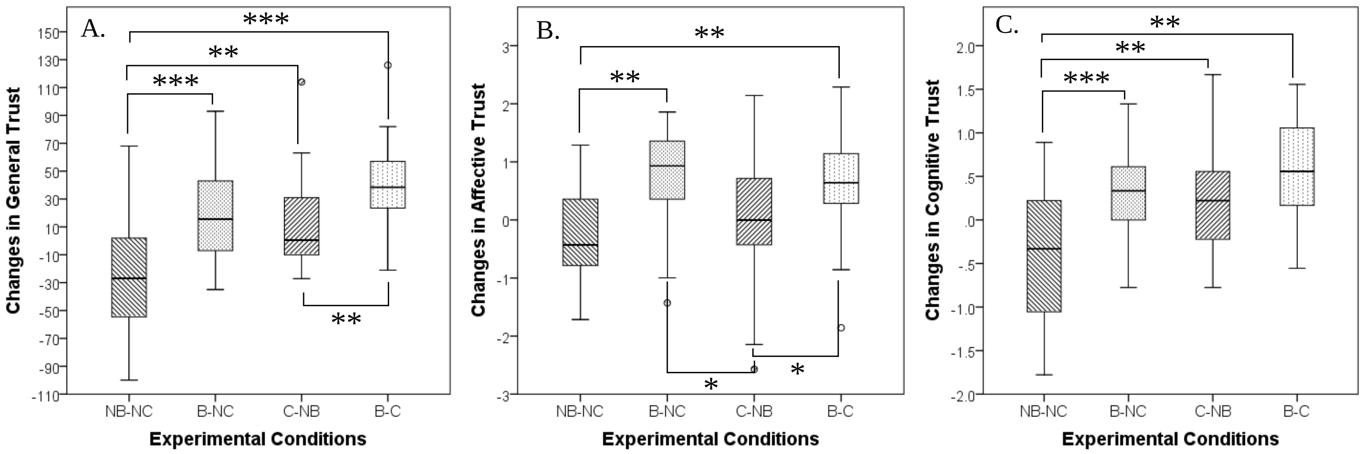


Figure 3. 5 Boxplot distributions represent the main effect of benevolent and competent behaviors in four experimental conditions on A) general, B) affective, and C) cognitive trust scores. Horizontal lines inside each box indicate the median. (****p<0.001*, ***p<0.01*).

As for the results of Cohen's *d* value, the difference in the means of general trust score was the highest in comparison between the B-Cc and NB-NCc with an extremely large effect size (d = 1.83), which means participants perceived these two conditions to be extremely different in trustworthiness. The next highest effect sizes followed for the comparison of the B-NCc and NB-NCc (d = 1.21) and C-NBc and NB-NCc (d = 1.00), which supported the greater influence of benevolence than competence on the trust perception. Therefore, these results revealed that a benevolent-competent robot yielded the highest trustworthiness, and a benevolence-based robot was perceived higher in trustworthiness by participants than a competence-based robot as compared to nonbenevolent-noncompetent robot.

3.4.1.2 Interrelation of benevolence and competence characteristics in HRI

The second set of hypotheses considered a comparison between the B-C, B-NC, and C-NB conditions for general trust to detect the interrelation between the benevolence and competence attributes in HRT. As shown in Table 3.10, the Bonferroni post-hoc test results indicated a significant difference in the score of general trust between the B-Cc (M = 40.64, SD = 28.81) and C-NBc (M = 11.61, SD = 32.60, $p < 0.05$) and a large effect size (d = 0.94), where the B-Cc had a higher trustworthiness. However, no significant differences between the B-Cc (M = 40.64, SD = 28.81) and B-NCc (M = 20.32, SD = 33.84, p = 0.197) was observed for general trust, despite the fact that participants evaluated a benevolent-competent robot higher than a benevolent-noncompetent robot. Thus, the difference between a benevolent-competent robot and a competent-nonbenevolent robot was more considerable, and benevolent behavior could compensate the incompetency to some degree. Finally, the post-hoc results did not reveal a significant difference between the B-NCc (M = 20.32, SD = 33.84) and C-NBc (M = 11.61, SD = 32.60, p = 1.000) conditions for the general trust, however, B-NCc exhibited higher scores. Thus, the participants did not differentiate between the benevolence- and competence-based robots in general trust, and the results did not yield any primacy of benevolence or competence attributes in general trust in HRT. Therefore, **H2-2** was supported, whereas **H 2-1** and **H 2-3** were not supported. Panel A in Figure 3.5 shows the means and SDs of changes in the score of general trust before and after the conversation in the B-C, B-NC, and N-BC conditions.

Table 3. 10 Bonferroni post-hoc results and Cohen's d effect size for trust scores between B-C, B-NC and C-NB conditions. The main effect of benevolent behavior is significant for affective trust score for B-C and B-NC conditions.

	B-Cc versus B-NCc		B-Cc versus C-NBc		B-NCc versus C-NBc	
	p	Cohen's d	*p*	Cohen's d	*p*	Cohen's d
General trust	0.197	0.64	0.018*	0.94	1.000	0.26
Affective trust	1.000	-0.08	0.018*	0.65	0.009**	0.74
Cognitive trust	0.543	0.51	0.216	0.65	1.000	0.15

Statistically significant differences: $^{**}p<0.01$, $^{*}p<0.05$.

3.4.4 Affective and cognitive trust scores

Finally, the third set of hypotheses was used to determine the effectiveness of competence and benevolence characteristics of robots for different types of affective and cognitive trust.

3.4.2.1 Affective trust

For the affective trust, the post-hoc test results indicated that the B-Cc (M = 0.69, SD = 0.84) was rated significantly higher than NB-NCc (M = -0.13, SD = 1.08, $p < 0.01$) as shown in Table 3.9. The results also indicated that considering the score of affective trust, a significant difference between the B-Cc (M = 0.69, SD = 0.84) and C-NBc (M = 0.07, SD = 1.04, $p < 0.05$) was observed (Table 3.10), and a benevolent-competent robot discernibly scored higher than a competent-nonbenevolent robot, as expected in **H3-1**. However, the difference between the C-NBc (M = 0.07, SD = 1.04) and NB-NCc (M = -0.13, SD = 1.08, $p = 1.000$) conditions did not yield any significant results for the affective trust (Table 3.9). Furthermore, the results as shown in Table 3.9, revealed that the B-NCc (M = 0.76, SD = 0.79) scored significantly higher for the affective trust than that in the NB-NCc (M = -0.13, SD = 1.08, $p < 0.01$), and a benevolent-noncompetent robot was obviously higher in affective trust than a nonbenevolent-noncompetent robot. Finally, H3-2 predicted that affective trust was significantly higher for a benevolent-noncompetent robot than a competent-nonbenevolent robot. The Bonferroni post-hoc results, as shown in Table 3.10, revealed that the participants assessed the B-NCc (M = 0.76, SD= 0.79) as having a higher affective trust than the C-NBc (M= 0.07, SD= 1.04, $p < 0.01$) with a large effect size ($d = 0.74$), and **H3-2** was verified based on these results.

Considering the results of Cohen's *d* value, the difference in the means of affective trust score was higher comparing the NB-NCc with B-NCc (d = 0.94) than with B-Cc (d = 0.84). Thus, the benevolent behavior of the robot was evaluated higher in terms of affective trust despite the incompetent behavior. Consequently, the benevolent behavior of the robot was positively related to affective trust. These results showed that the coexistence of benevolence and competence influenced the affective trust perception in a different way. Panel B in Figure 3.5 shows the means and SDs of changes in the affective trust scores under the experimental conditions.

3.4.2.2 Cognitive trust

The results of the cognitive trust indicated a significant difference between the B-Cc (M = 0.59, SD = 0.57) and NB-NCc (M = -0.35, SD = 0.75, $p < 0.001$) as shown in Table 3.9. Moreover, a competent-nonbenevolent robot (C-NBc) (M = 0.23, SD = 0.53) was evaluated to be significantly higher than that of a nonbenevolent-noncompetent robot (NB-NCc) (M = -0.35, SD = 0.75, $p < 0.01$), comparing the cognitive trust. Similarly, the B-NCc (M = 0.23, SD = 0.53) was perceived higher than the NB-NCc (M = -0.35, SD = 0.75, $p < 0.001$) in terms of the cognitive trust. However, contrary to **H3-3**, the comparison between the B-Cc (M = 0.59, SD = 0.57) and B-NCc (M = 0.31, SD = 0.52, $p = 0.543$) was not significant for the cognitive trust as shown in Table 3.10. Similarly, the cognitive trust score of the B-NCc (M = 0.31, SD= 0.52) and C-NBc (M = 0.23, SD = 0.53, $p = 1.000$) were not significantly different, and the participants evaluated a competent-nonbenevolent robot lower than a benevolent-noncompetent robot in terms of cognitive trust. Thus, the evidence was not statistically efficient in supporting **H3-4**.

Therefore, a benevolent robot increased cognitive trust even though it was noncompetent, and affected the cognitive trust as well as affective trust. Consequently, benevolent behavior of robot was influential in eliciting both affective and cognitive trust, whereas competency influenced only cognitive trust. These results approved the primacy and influence of benevolence on both affective and cognitive trust, as a benevolent-noncompetent robot was still higher in cognitive trust than a competent-nonbenevolent robot. Panels C in Figure 3.5 shows the means and SDs of changes in the scores of cognitive trust before and after the conversation under the experimental conditions.

3.4.5 Effect of social robot's benevolence and competence on people's perception

As a further analysis, we measured the effect of benevolence and competence attributes of the robot on participants' perception and impression of the robot in terms of animacy, likability and perceived intelligence. Table 3.11 shows the means and standard deviations (SDs) of animacy, likability and perceived intelligence with their individual items for four groups of robot's behaviors. Robot with B-C behavior achieved the highest score in all subscales of animacy, likability and perceived intelligence and the robot with NB-NC behavior resulted in the lowest score. B-NCc was higher than C-NBc in two subscales of animacy and likeability, but not perceived intelligence.

Table 3. 11 Means and standard deviations of animacy, likeability and perceived intelligence of NB-NC, B-NC, C-NB and B-C conditions.

	NB-NCc	**B-NCc**	**C-NBc**	**B-Cc**
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Animacy	**2.95 (0.83)**	**3.63 (0.64)**	**3.35 (0.69)**	**3.65 (0.52)**
dead-alive	3.21 (1.13)	3.71 (1.08)	3.58 (0.90)	4.00 (0.77)
stagnant-lively	3.29 (1.08)	3.93 (0.81)	3.77 (0.81)	4.11 (0.62)
mechanical-organic	2.21 (0.91)	2.86 (0.93)	2.38 (0.98)	2.68 (0.93)
artificial-lifelike	2.50 (1.03)	2.86 (1.04)	2.69 (1.22)	2.96 (1.10)
inert-interactive	3.32 (1.30)	4.25 (0.84)	4.00 (0.80)	4.25 (0.70)
apathetic-responsive	3.18 (1.30)	4.18 (0.67)	3.69 (0.92)	3.93 (0.85)
Likability	**3.31 (0.85)**	**4.57 (0.40)**	**3.96 (0.69)**	**4.68 (0.33)**
dislike-like	3.29 (1.01)	4.39 (0.62)	4.00 (0.84)	4.64 (0.55)
unfriendly-friendly	3.32 (1.05)	4.61 (0.62)	4.00 (0.89)	4.79 (0.49)
unkind-kind	3.14 (0.97)	4.68 (0.47)	3.81 (0.84)	4.71 (0.53)
unpleasant-pleasant	3.29 (0.89)	4.57 (0.57)	3.92 (0.93)	4.54 (0.50)
awful-nice	3.54 (0.96)	4.61 (0.49)	4.08 (0.74)	4.75 (0.51)
P-Intelligence	**2.77 (0.80)**	**3.66 (0.65)**	**3.80 (0.50)**	**4.07 (0.44)**
Incompetent-competent	2.75 (1.14)	3.46 (1.07)	3.62 (0.89)	3.75 (0.70)
Ignorant-knowledgeable	2.43 (0.92)	3.21 (0.78)	4.19 (0.80)	4.46 (0.69)
Irresponsible-responsible	2.82 (1.15)	4.11 (0.73)	3.50 (0.64)	4.07 (0.76)
Unintelligent-intelligent	2.82 (1.09)	3.46 (0.96)	4.04 (0.59)	4.29 (0.71)
Foolish-sensible	3.07 (0.94)	4.07 (0.81)	3.65 (0.56)	3.82 (0.86)

(P-Intelligence; Perceived intelligence)

Moreover, we ran a series of one-way ANOVA tests with independent variables, and animacy, likability and perceived intelligence as dependent variables. The results indicated that the main effects of benevolent and competent behavior were significant for animacy (F (3,109) = 6.44, p < 0.001), likeability (F (3,109) = 29.88, p < 0.001), and perceived intelligence (F (3,109) = 21.01, p < 0.001). Post-hoc analysis as shown in Table 3.12 and Figure 3.6 revealed that the all three variables of animacy, likeability, and perceived intelligence was significantly higher for three conditions of B-NC, C-NB and B-C than the condition of NB-NC.

Interestingly, there was not any significant difference for animacy between B-C, B-NC and C-NB conditions, and any of these conditions could not exceed the others to show more animacy for participants. Furthermore, the results for the comparison of B-NCc and C-NBc were significant only in likeability ($p < 0.001$), which indicated the relation of benevolence behavior with likeability. Finally, the results indicated a significant difference for perceived intelligence ($p < 0.05$) comparing B-Cc and B-NCc, while there was significant result for likeability ($p < 0.001$) in comparing B-Cc and C-NBc. These results could reveal the relation of competence with perceived intelligence, and on the other side the influence of benevolence attributes on likeability.

Table 3. 12 Pot-hoc results for pair comparisons between experimental conditions.

	B-NCc vs. NB-NCc	**C-NBc vs. NB-NCc**	**B-Cc vs. NB-NCc**	**B-NCc vs. C-NBc**	**B-Cc vs. B-NCc**	**B-Cc vs. C-NBc**
	p	*p*	*p*	*p*	*p*	*p*
Animacy	0.000***	0.033*	0.000***	0.137	0.889	0.107
Likability	0.000***	0.000***	0.000***	0.000***	0.486	0.000***
Perceived intelligence	0.000***	0.000***	0.000***	0.443	0.018*	0.117

Statistically significant differences: ****p<0.001, *p<0.05.*

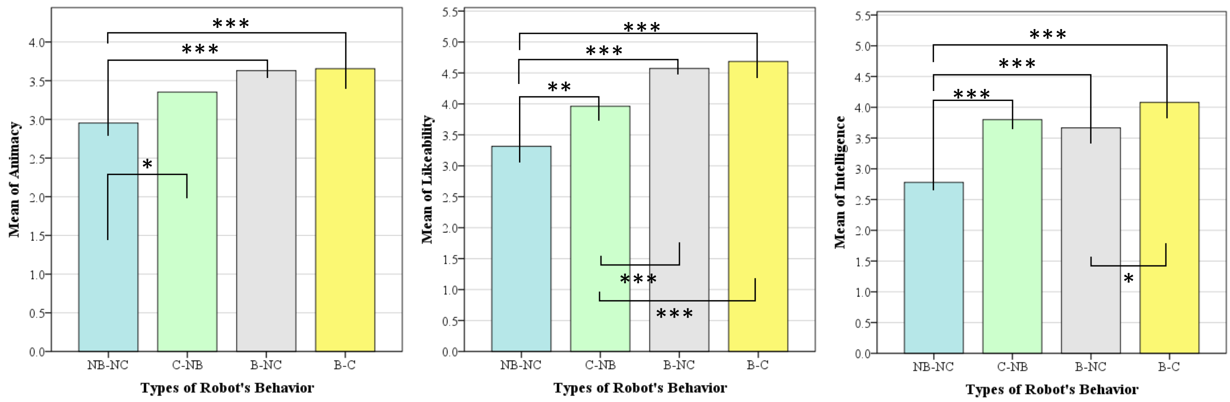


Figure 3. 6 Results showing the comparison between four experimental conditions for animacy, likability and perceived intelligence. (****p < 0.001, **p < 0.01, *p < 0.05)*

3.4.6 Qualitative findings

The participants were asked to describe their overall experience and perception of NAO robot's behaviors and interaction, in terms of competence and benevolence, immediately after the interaction through open questions. The aim was to capture self-reported evaluations of the robot's behaviors which may not be achievable through questionnaires.

Regarding the B-C robot, the majority of participants perceived NAO's behaviors as informative and caring, akin to those of a psychologist. However, some participants remained skeptical about trusting the robot's personal advice.

"NAO is like having a good friend to talk to, or a personal psychologist."

"NAO's behavior looks like a psychologist. It cares my feelings, but I could not feel any helpful advice from him."

"I am not sure about very personal emotional advices from him, however advices he gave for my health protection was trustworthy. He indeed was very informative."

However, some participants believed that the responses were predictable and provided excessive statistical information, which differs from the way we typically converse with our friends.

"It was informative and gave me a proper sentence and gave an emotional intonation, but 70% was just ok and yes, and the explanation was predictable."

"He was very informative and caring as well. I think it would make him more lifelike if it could respond to sudden changes of topic."

"Overall, the most remarkable impression after interaction with NAO is that he is very caring and friendly. But after several times of interacting, I kind of know the pattern of his responses to my opinion."

There were also some comments about NAO's body language, which was perceived as pleasant and contributed to the robot's humanlike appearance.

"There were many small detailed movements that surprised me a lot! These made it humanlike."

"The robot was good at expressing human-like body expressions."

For the C-NB robot, participants mostly perceived NAO as informative and trustworthy in many cases, and perceived it to be programmed. Given its robotic nature, some participants didn't expect it to be emotional or caring, and thus did not criticize the robot for its lack of understanding of their feelings.

"NAO seems to have some difficulty understanding emotions, but it could act logically."

"It was very informative. But I think maybe my personal experiences is very different, he could not understand me at some points."
"NAO could be trustworthy when we talked about something that related to what NAO was programmed for. I found that the robot responded consistently and it was informative and user friendly."

For the B-NC robot, participants mostly perceived the robot as having an emotional and caring approach. They believed that its lack of informativeness made it more humanlike, as people often lack knowledge in certain areas. Additionally, they found the robot to be friendly.

"It feels that robot tried to build an emotional relation and trust with the participant."
"NAO acted in a friendly manner, however it was not quite informative regarding the topic."
"It was very pleasant to talk with NAO. I felt I was talking with a friend…"
"I found NAO's behavior very lifelike. He is not 100% informative which exactly made him more like a human. He is really caring."
"It was a good robot. It was able to act in a way that made it seem like it could feel people's emotions."

There were some comments about NAO's body movements that were not pleasant in certain cases and caused distractions from the main conversation.

"NAO's behavior is good, but sometimes his gestures were too much which made my attention to his gestures."
"In my opinion NAO looks very robotic in hand movements."
"I found the movements sometimes distracting, yet the face was basically the same,…"

For the NB-NC robot, participants believed that NAO was not sufficiently responsive and informative. They did not perceive it as caring, although it exhibited kind and friendly behavior.

"It feels like talking to a tape recorder."
"He may need to be updated on topic we were talking about."

"The conversation was mostly ok. I felt he was trying to address common conversation topics about Covid-19, however in several instances he repeated that he did not have much information, which made the conversation a bit weird."

3.5 DISCUSSION

Competence and benevolent behaviors have been acknowledged to improve the interpersonal trust. During the second study, it was investigated how the correlation between behaving as competent and benevolent in social robots contributes to their trustworthiness.

3.5.1 Competent and benevolent social robots foster trust in HRI

Insights into the role of humanistic behaviors of social robots in different fields, particularly trust, are rare and covered. The extent to which anthropomorphizing robots are acceptable must be determined, specifically concerning affective behaviors. This study explored whether humanistic behaviors like benevolence result in more trustworthiness in social robots and whether a robot with both benevolent and competent behaviors have been designated as more trustworthy in HRI. As expected, the participants evaluated all three types of social robots, including benevolent-competent robot (H1-1), benevolent-noncompetent robot (H1-2), and competent-nonbenevolent robot (H1-3), more trustworthy than a nonbenevolent-noncompetent robot in terms of general trust. A benevolent-competent robot yielded the highest trustworthiness, and a benevolence-based robot was perceived higher in trustworthiness by participants than a competence-based robot as compared to nonbenevolent-noncompetent robot. Therefore, the benevolent or competent characteristics are influential in HRT, which is comparable to the findings on human trust (Cuddy et al., 2011; Di Battista et al., 2020; Oleszkiewicz & Lachowicz-Tabaczek, 2016).

Studies on the characteristics of robots associated with the "can-do" trust dimension, such as reliability, function, performance and skill are extensive. In general, a well-performing robot is trusted more, and our findings satisfactorily agreed with the results of similar research (e.g., Di Battista et al., 2020; Salem et al., 2015; Schoorman et al., 2007; Ulfert & Georganta, 2020), as the competent-nonbenevolent robot was more trustworthy than the nonbenevolent-noncompetent robot. Our results showed that being knowledgeable could foster trustworthiness in HRI, despite malevolent, ignoring, and impolite robot behavior. Moreover, if a social robot is not informative and cannot satisfy participants with accurate information, it still provokes trust because of its kind and caring behavior. This is a beneficial and valuable result that improves the findings presented in (Giorgi et al., 2022), and is closer to the outcomes of the human trust area (Di Battista et al., 2021; Johnston et al., 2015). According to the literature, emotional sharing and behaviors advance the formation of interpersonal trust and even rebuild damaged trust because emotions guide people's behavioral propensities (Ma et al., 2018), which can explain the obtained results. This finding will help design anthropomorphic and emotional robots beyond functional ones, and indicates that humanistic behaviors, such as benevolence, have great potential for developing social robots.

Meanwhile, an interesting concern was revealed in our findings. Considering the mean differences and Cohen's d for the B-NC ($M = 20.32$, $d = 1.21$) and C-NB ($M = 11.61$, $d = 1.00$) conditions in comparison with the NB-NC condition, the higher effectiveness of the benevolent behavior can be explored. Thus, the benevolent behavior of a robot at a similar level is more influential than its knowledge, which is in contrast to findings presented in (Giorgi et al., 2022). The reason could be the connection of benevolence to emotions and the superior effect of emotions on trust. Ghazali et al. (2018) found that liking the robot increased trust beliefs and the more the participants liked the robot, the more trustworthy was the robot. It is possible that benevolence features also influence the robot to be more likeable and elicit its trustworthiness as a mediator respectively. Thus, further research can consider the relation of benevolence to various emotions such as liking, kindness, attractiveness and etc., and its subsequent effect on trust perceptions. However, the situation and topic of the conversation also might be effective, and further research in different contexts can improve the results.

Consequently, when social robots encompassed both the benevolence and competence attributes, their trustworthiness intensifies. This is in line with the results of education field and prior studies on the interpersonal trust (Cuddy et al., 2011; Di Battista et al., 2021). The benevolence and competence dimensions in human relationships are fundamental, and manipulating each one changes the other one. Organizations, individuals, and social groups are judged along these two dimensions. People primarily judge that warm and competent elicits positive emotions and behavior, while lacking both the warmth and competence elicits uniform negativity (Cuddy et al., 2011); thus, both dimensions must be considered in HRT. The benevolent-competent robot achieved the highest measure and effect size among all comparisons, which resulted an extremely large effectiveness. Thus, if future functional robots are provided with benevolent behaviors and show concern, care, and kindness toward people, they can guarantee an increase in trustworthiness even though they are not completely applicable and skillful.

3.5.2 Benevolent robot surpasses in evoking affective trust

This study also focused on the primacy of benevolence or competence traits in HRT. The results were challenging, and crucial for future research and development of social robots. Although, a benevolent-competent robot was perceived more trustworthy than a competent-nonbenevolent robot (H2-2), it was not rated more trustworthy than a benevolent-noncompetent robot (H2-1), in terms of general trust, which revealed the high effect of benevolence in HRT. Therefore, benevolence could approximately compensate for the social robot's lack of information in building general trust, although competence could not play the same role in the absence of benevolent characteristics in HRT. Some existing studies asserted that warmth judgments are made quicker than competence judgments and have a greater effect on overall attitudes toward others, and are even more stable across cultures than competence information (Di Battista et al., 2021). However, our findings did not support this finding in HRT; our results agreed with other studies that did not find any priority between the accuracy and niceness (Johnston et al., 2015). The robot stated some wrong information to show the incompetence, which might be interpreted differently by some participants and influenced the results as well.

Moreover, in contrast to our expectation, no preference between a benevolent-noncompetent and competent-nonbenevolent robot was found in the general trust (H1-6), although the benevolent-noncompetent achieved a higher score after conversation. Some studies on the interpersonal trust have indicated that people trust individuals with good intentions more than knowledgeable ones (Di Battista et al., 2021). However, our findings did not support this finding in HRT; our results agreed with other studies that did not find any priority between the accuracy and niceness (Johnston et al., 2015).

3.5.3 Relation of benevolent and competent social robots with affective and cognitive trust in HRI

This study aimed to explore if benevolence attributes of social robots contribute to higher affective trust in HRI. According to the results, benevolent-competent robot discernibly scored higher than a competent-nonbenevolent robot in affective trust (H3-1). Furthermore, a significant result was achieved between benevolent-noncompetent and competent-nonbenevolent robot (H3-2) in terms of affective trust, which indicated the superiority of benevolent robots in fostering affective trust.

In general, warmth and benevolence are more strongly related to emotions and empathy than competence (Id et al., 2021), and are quickly perceived by others (Belanche et al., 2021). Emotional sharing and behaviors advance the formation of interpersonal trust and even rebuild damaged trust because emotions guide people's behavioral propensities (Ma et al., 2018), which can explain the obtained results. Thus, the evaluations of the emotional attitude toward benevolence influenced affective trust in HRI. People clearly distinguished between caring and positive emotional reactions of the robot and differentiated it from rational and cognitive-based behaviors, as the difference in cognitive trust for the same group comparison (B-NCc and C-NBc) was not significant. These findings support the significant and robust effect of benevolence on affective trust and suggest that people believe and feel empathy toward the affective and humanistic behaviors of robots, although they know that robots are not alive. This is a significant finding for designing future social robots, particularly in the fields of healthcare, such as counselling or assistive robots, and personal and service social robots.

We also explored to know whether the competence attributes of social robots induced cognitive trust. The results revealed that a knowledgeable and competent robot did not successfully foster cognitive trust, as the participants did not score a benevolent-competent robot higher than benevolent-noncompetent one in terms of cognitive trust (H3-3). Moreover, a competent-nonbenevolent robot was also rated lower than a benevolent-noncompetent robot in terms of cognitive trust (H3-4). Interestingly, according to the mean differences, the cognitive trust score was higher in the benevolent-noncompetent robot than that in the competent-nonbenevolent robot. Therefore, although a competent-nonbenevolent robot was equipped with knowledge and was supposed to be correlated with cognitive trust, benevolent-noncompetent robot achieved even higher cognitive trust, despite being noninformative. Therefore, we could say that benevolence modulated cognitive evaluations, and participants' perception of both affective and cognitive trust were higher when the robot was benevolent.

Emotions are highly influential in trust relations and even some unconscious thoughts lead to emotions about things. Emotions often outweigh logic and influence perception and cognitive evaluations even after the dissipation of emotions. In addition, according to modern psychology, affection involves several basic cognitive functions and appears to be necessary for normal conscious experience. It influences, modulates, and mediates basic cognitive processes and has a substantial influence on human cognitive processes (Duncan & Barrett, 2007). Johnson & Grayson (2005) considered cognitive trust as one of the antecedents of affective trust which can explain our findings. Furthermore, the context of conversation was a fairly friendly interaction and was not dominated by serious discussion and issues; thus, it can be another reason for lessening attention to cognitive trust. Consequently, a vague border between the overlapping and interdependency of benevolence and competence factors in HRT exists, and other determinants or scales should be considered in future studies.

3.6 CHAPTER SUMMARY

This study revealed some key facts as to the contribution of benevolence and competence to the HRT. The results showed that, like interpersonal trust, the benevolence and competence characteristics are both crucial in making trust, and each carries different responsibilities and fosters certain dimensions of trust. Therefore, beyond the knowledge, intelligence, expertise, and information, which support the functional aspect of robots and have been mostly investigated, humanistic behaviors, such as benevolence and caring, should be considered fundamental and vital. In particular, the results implied new findings regarding the role of benevolence in HRT. We concluded that benevolence had a considerable effect on the affective trust despite the existence or lack of knowledge and competence in HRI, which approved the primacy of benevolence in eliciting the affective trust. Furthermore, benevolence could have a small influence on the general and cognitive trusts and increased them in the absence of competence characteristics. In addition, it could compensate for the lack of information in social robots. However, being competent could not evoke more affective trust in human-robot relationships. Meanwhile, benevolent behavior influenced and modulated competency and increased all types of general, affective and cognitive trust in the lack of information and knowledge. Thus, if future functional robots in various fields of healthcare, education, and business are equipped with benevolent behaviors and show concern, care, and kindness toward people, they can guarantee an increase in trustworthiness even though they are not completely applicable and skillful.

3.7 REFERENCES

Ajzen, I. (1977). Information Processing Approaches to Interpersonal Attraction. *Theory and Practice in Interpseronal Attraction, October*, 51–77.

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

Belanche, D., Casaló, L. V., Schepers, J., & Flavián, C. (2021). Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The Humanness-Value-Loyalty model. *Psychology and Marketing*, *38*(12), 2357–2376. https://doi.org/10.1002/mar.21532

Bente, G., & Krämer, N. C. (2011). Virtual gestures: embodiment and nonverbal behavior in computer-mediated communication. In A. Kappas & N. C. Krämer (Eds.), *Face-to-Face Communication over the Internet:Emotions in a Web of Culture, Language, and Technology* (pp. 176–210). Cambridge University Press. https://doi.org/10.1017/cbo9780511977589.010

Bickmore, T., Schulman, D., & Yin, L. (2009). Engagement vs. Deceit: Virtual Humans with Human Autobiographies. *IVA '09: Proceedings of the 9th International Conference on Intelligent Virtual Agents*, 6–19. https://doi.org/https://doi.org/10.1007/978-3-642-04380-2_4

Briñol, P., Petty, R. E., & Barden, J. (2007). Happiness Versus Sadness as a Determinant of Thought Confidence in Persuasion: A Self-Validation Analysis. *Journal of Personality and Social Psychology*, *93*(5), 711–727. https://doi.org/10.1037/0022-3514.93.5.711

Capraro, V., Smyth, C., Mylona, K., & Niblo, G. A. (2014). Benevolent characteristics promote cooperative behaviour among humans. *PLoS ONE*, *9*(8), 1–6. https://doi.org/10.1371/journal.pone.0102881

Carlson, M. S., Desai, M., Drury, J. L., Kwak, H., & Yanco, H. A. (2014). Identifying factors that influence trust in automated cars and medical diagnosis systems. *AAAI Spring Symposium*, 20–27. https://api.semanticscholar.org/CorpusID:10439801

Cuddy, A. J. C., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior*, *31*, 73–98. https://doi.org/10.1016/j.riob.2011.10.004

Darioly, A., & Mast, M. S. (2014). The role of nonverbal behavior in leadership: An integrative review. In R. E. Riggio & S. Tan (Eds.), *Leader Interpersonal and Influence Skills: The Soft Skills of Leadership* (pp. 73–100). Taylor & Francis. doi: 10.4324/9780203760536

Darling, K. (2017). "Who's Johnny?" Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. In P. Lin, G. Bekey, K. Abney, & R. Jenkins (Eds.), *ROBOT ETHICS 2.0*. Oxford University Press. https://doi.org/10.2139/ssrn.2588669

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

Di Battista, S., Pivetti, M., & Berti, C. (2020). Competence and benevolence as dimensions of trust: Lecturers' trustworthiness in the words of italian students. *Behavioral Sciences*, *10*(9). https://doi.org/10.3390/bs10090143

Di Battista, S., Smith, H. J., Berti, C., & Pivetti, M. (2021). Trustworthiness in higher education: The role of professor benevolence and competence. *Social Sciences*, *10*(1), 1–19. https://doi.org/10.3390/socsci10010018

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and Emotion*, *21*(6), 1184–1211. https://doi.org/10.1080/02699930701437931

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Eyssel, F., Wullenkord, R., & Nitsch, V. (2017). The Role of Self-Disclosure in Human-Robot Interaction. *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 922–927. https://doi.org/10.1109/ROMAN.2017.8172413.

Firmansyah, M. R., Amelia, R., Jamil, R. A., Faturochman, F., & Minza, W. M. (2019). Benevolence, Competency, and Integrity: Which Is More Influential on Trust in Friendships? *Jurnal Psikologi*, *18*(1), 91. https://doi.org/10.14710/jp.18.1.91-105

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Frisou, J. (2000). Confiance Interpersonnelle et Engagement : une Réorientation Béhavioriste. *Recherche et Application En Marketing*, *15*(1), 63–80. https://doi.org/10.2307/40589217

Ghazali, A. S., Ham, J., Barakova, E. I., & Markopoulos, P. (2018). Effects of robot facial characteristics and gender in persuasive human-robot interaction. *Frontiers Robotics AI*, *5*(JUN), 1–16. https://doi.org/10.3389/frobt.2018.00073

Giorgi, I., Tirotto, F. A., Hagen, O., Aider, F., Gianni, M., Palomino, M., Masala, G. L., & Member, S. (2022). Friendly But Faulty : A Pilot Study on the Perceived Trust of Older Adults in a Social Robot. *IEEE Access*, *10*(August), 92084–92096. https://doi.org/10.1109/ACCESS.2022.3202942

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A. C., & Wang, J. (2005). Designing robots for long-term social interaction. *2005 IEEE/RSJ International Conference on Intelligent Robots and*

Systems, IROS, 1338–1343. https://doi.org/10.1109/IROS.2005.1545303

Hagmann, D., Minson, J. A., & Tinsley, C. H. (2020). *Personal Narratives Build Trust Across Ideological Divides*. OSF Preprints, 29 Sept. https://doi.org/doi:10.31219/osf.io/sw7nz.

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

Hoy, W. K., & Tschannen-Moran, M. (1999). Five Faces of Trust: An Empirical Confirmation in Urban Elementary Schools. *Journal of School Leadership*, *9*(3), 184–208. https://doi.org/10.1177/105268469900900301

Huhns, M. N., & Mohamed, A. (1999). Benevolent agents. *IEEE Internet Computing*, *3*(2), 96–98. https://doi.org/10.1109/4236.761661

Hui, M. K., & Sit, A. Y. (2005). Service failure and trust: The roles of benevolence, competence, and culture. *2005 International Conference on Services Systems and Services Management, Proceedings of ICSSSM'05*, *1*, 180–182. https://doi.org/10.1109/ICSSSM.2005.1499457

Id, T. A., Bu, S., Mayer, B., & Dricu, M. (2021). Empathic responses to social targets : The influence of warmth and competence perceptions , situational valence , and social identification. *PLoS ONE*, *16*(3), 1–21. https://doi.org/10.1371/journal.pone.0248562

Jasielska, D. (2020). The moderating role of kindness on the relation between trust and happiness. *Current Psychology*, *39*(6), 2065–2073. https://doi.org/10.1007/s12144-018-9886-7

John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), 954–959. https://doi.org/10.1073/pnas.1516868113

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Johnston, A. M., Mills, C. M., & Landrum, A. R. (2015). How do children weigh competence and benevolence when deciding whom to trust? *Cognition*, *144*, 76–90. https://doi.org/10.1016/j.cognition.2015.07.015

Kang, S. H., & Gratch, J. (2011). People like virtual counselors that highly-disclose about themselves. *Annual Review of CyberTherapy and Telemedicine*, *9*(1), 117–120.

Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A Question of Compensation: The Social Life of the Fundamental Dimensions of Social Perception. *Journal of Personality and Social Psychology*, *96*(4), 828–842. https://doi.org/10.1037/a0013320

Kim, K. J., Park, E., & Shyam Sundar, S. (2013). Caregiving role in human-robot interaction: A study of the mediating effects of perceived benefit and social presence. *Computers in Human Behavior*, *29*(4), 1799–1806. https://doi.org/10.1016/j.chb.2013.02.009

Kory-Westlund, J. M., & Breazeal, C. (2019). Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning. *Frontiers in Robotics and AI*, *6*(July). https://doi.org/10.3389/frobt.2019.00054

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert ? Benevolence information influences children ' s trust more than expertise. *Developmental Science*, *16*(4), 622–638. https://doi.org/10.1111/desc.12059

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2014). Informants' Traits Weigh Heavily in Young Children's Trust in Testimony and in Their Epistemic Inferences. *Child Development*, *84*(4), 1253–1268. https://doi.org/10.1111/cdev.12029.Informants

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, *50*(11), 1477–1490. https://doi.org/10.1287/mnsc.1030.0136

Liu, X. (Stella), Yi, X. (Shannon), & Wan, L. C. (2022). Friendly or competent? The effects of perception of robot appearance and service context on usage intention. *Annals of Tourism Research*, *92*, 103324. https://doi.org/10.1016/j.annals.2021.103324

Livnat, Y. (2004). On the nature of benevolence. *Journal of Social Philosophy*, *35*(2), 304–317. https://doi.org/10.1111/j.1467-9833.2004.00234.x

Lui, S. S., & Ngo, H. yue. (2004). The role of trust and contractual safeguards on cooperation in non-equity alliances. *Journal of Management*, *30*(4), 471–485. https://doi.org/10.1016/j.jm.2004.02.002

Ma, F., Wylie, B. E., Luo, X., He, Z., Xu, F., & Evans, A. D. (2018). Apologies repair children's trust: The mediating role of emotions. *Journal of Experimental Child Psychology*, *176*, 1–12. https://doi.org/10.1016/j.jecp.2018.05.008

Mandal, F. B. (2014). Nonverbal Communication in Humans. *Journal of Human Behavior in the Social Environment*, *24*(4), 417–421. https://doi.org/10.1080/10911359.2013.831288

Mayer, R. C., Davis, J. H., & Schoorman, D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

McAllister, D. J. (1995). Affect and CognitionBased Trust As Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, *38*(1), 24–59. https://doi.org/10.2307/256727

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

Nikolaidis, S., Kwon, M., Forlizzi, J., & Srinivasa, S. (2018). Planning with Verbal Communication for Human-Robot Collaboration. *ACM Transactions on Human-Robot Interaction*, *7*(3). https://doi.org/10.1145/3203305

Oleszkiewicz, A., & Lachowicz-Tabaczek, K. (2016). Perceived competence and warmth influence respect, liking and trust in work relations. *Polish Psychological Bulletin*, *47*(4), 431–435. https://doi.org/10.1515/ppb-2016-0050

Pattie, J. O. (2021). Personal Narrative. *Travels, Explorations and Empires*, 289–303. https://doi.org/10.4324/9781003113317-11

Pytlikzillig, L. M., Hamm, J. A., Shockley, E., Herian, M. N., Neal, T. M. S., Kimbrough, C. D., Tomkins, A. J., & Bornstein, B. H. (2016). The dimensionality of trust-relevant constructs in four institutional domains: Results from confirmatory factor analyses. *Journal of Trust*

Research, *6*(2), 111–150. https://doi.org/10.1080/21515581.2016.1151359

Reed, L. I., Matari, Y., Wu, M., & Janaswamy, R. (2019). Emotional Tears: An Honest Signal of Trustworthiness Increasing Prosocial Behavior? *Evolutionary Psychology*, *17*(3), 1–8. https://doi.org/10.1177/1474704919872421

Robinette, P., Howard, A. M., Member, S., & 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

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics*, *5*(1), 17–34. https://doi.org/10.1007/s12369-012-0173-8

Rossi, S., Staffa, M., Bove, L., Capasso, R., & Ercolano, G. (2017). User's Personality and Activity Influence on HRI Comfortable Distances. In *social robotics* (pp. 167–177). Springer, Cham. https://doi.org/10.1007/978-3-319-70022-9_17

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. *ACM/IEEE International Conference on Human-Robot Interaction*, *2015-March*, 141–148. https://doi.org/10.1145/2696454.2696497

Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the"Trust Perception Scale-HRI". In R. Mittu, D. Sofge, A. Wagner, & W. F. Lawless (Eds.), *Robust Intelligence and Trust in Autonomous Systems* (Issue August, pp. 191–218). Springer. https://doi.org/10.1007/978-1-4899-7668-0

Scheunemann, M. M., Cuijpers, R. H., & Salge, C. (2020). Warmth and Competence to Predict Human Preference of Robot Behavior in Physical Human-Robot Interaction. *29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020*, 1340–1347. https://doi.org/10.1109/RO-MAN47096.2020.9223478

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: past, present, and future. *The Academy of Management Review*, *32*(2), 344–354. https://doi.org/10.1023/B:JRNC.0000040887.00868.02

Sebo, S. S., Krishnamurthi, P., & Scassellati, B. (2019). "I Don't Believe You": Investigating the Effects of Robot Trust Violation and Repair. *ACM/IEEE International Conference on Human-Robot Interaction*, *2019-March*, 57–65. https://doi.org/10.1109/HRI.2019.8673169

Simpson, J. A. (2007). Psychological foundations of trust. *Current Directions in Psychological Science*, *16*(5), 264–268. https://doi.org/10.4324/9780203083208

Sitkin, S. B., & Roth, N. L. (1993). Explaining the Limited Effectiveness of Legalistic " Remedies " for Trust / Distrust. *Organization Science*, *4*(3), 367–392. https://doi.org/10.1287/orsc.4.3.367

Sundar, S. S., Jung, E. H., Waddell, T. F., & Kim, K. J. (2017). Cheery companions or serious assistants? Role and demeanor congruity as predictors of robot attraction and use intentions among senior citizens. *International Journal of Human Computer Studies*, *97*, 88–97. https://doi.org/10.1016/j.ijhcs.2016.08.006

Svare, H., Gausdal, A. H., & Möllering, G. (2020). The function of ability, benevolence, and integrity-based trust in innovation networks. *Industry and Innovation*, *27*(6), 585–604. https://doi.org/10.1080/13662716.2019.1632695

Tang, Y., Harris, P. L., Zou, H., & Xu, Q. (2019). The impact of emotional expressions on children's trust judgments. *Cognition and Emotion*, *33*(2), 318–331. https://doi.org/10.1080/02699931.2018.1449735

Tschannen-Moran, M., & Hoy, W. K. (2000). A Multidisciplinary Analysis of the Nature, Meaning, and Measurement of Trust. In *Review of Educational Research* (Vol. 70, Issue 4). https://doi.org/10.2307/1170781

Ulfert, A. S., & Georganta, E. (2020). A model of team trust in human-agent teams. *ICMI 2020 Companion - Companion Publication of the 2020 International Conference on Multimodal Interaction*, *c*, 171–176. https://doi.org/10.1145/3395035.3425959

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

Wheeless, L. R., & Grotz, J. (1977). the Measurement of Trust and Its Relationship To Self-Disclosure. *Human Communication Research*, *3*(3), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

Ybarra, O., Chan, E., Park, H., Burnstein, E., Monin, B., & Stanik, C. (2008). Life's recurring challenges and the fundamental dimensions: An integration and its implications for cultural differences and similarities. *European Journal of Social Psychology*, *38*(7), 1083–1092. https://doi.org/https://doi.org/10.1002/ejsp.559

CHAPTER 4

Exploring the Effects of Self-Disclosed Backstory of Social Robots on Perceived Trust in Human-Robot Interactions[1]

4.1 INTRODUCTION

Prior research has demonstrated that people feel less attached to whom think differently from their views. Not surprisingly, such attributes cause people to prefer to trust and receive advice from someone who shares similar experiences and stories. Therefore, trust is increased when people support their point of view by disclosing stories of personal experiences rather than experiences of others (Hagmann et al., 2020). Personal experiences are distinct from other evidence because of two important attributes; first, personal experiences indicate a story rather than a number of documented information and statistics, second, personal experiences rely on the self rather than other parties which makes the stories believable, trustworthy and companionable. In interpersonal trust, previous studies have proven the conductive effect of telling **self-backstory (personal self-disclosure, personal narratives)**, which is *self-disclosing and sharing personal information or experiences with others* (Kory-Westlund & Breazeal, 2019). For example, Hagmann et al. (2020) showed that personal narratives are critical to building and maintaining trust between people. Several studies have also approved that revealing personal experiences makes people appear warmer, more likable and trustworthy (Ajzen, 1977). Indeed, trust, self-disclosure and other indicators of relationship closeness have been identified to be positively correlated. Conversely, those who do not disclose their backstory are sometimes thought as more deceptive (John et al., 2016).

It is controversial if automated and computer systems could present themselves as human, and how users actually react to agents with human autobiographical memories. Some researchers believe that users feel cheated and deceived, whereas others think that it would provide the agents with more likeability, as people prefer the robots and agents to match them in personality (Bickmore et al., 2009). For example, "Valerie", one of the

[1] This study was conducted in collaboration with Anahita ETEMAD from the Department of Mechanical Engineering, RWTH Aachen University, Aachen, Germany (anahita.etemad@rwth-aachen.de).

world's first storytelling roboceptionists, was given a human back story to make the robot appear more human-like and to allow visitors to interact easily with her. The storylines evolved over the school year and included: "Valerie's social life, her lounge singing career, her therapy business, and her job as a receptionist." The results showed that Valerie could increase engagement with the robot and attracted people over a longer period of time. Gockley et al. (2005) found that robots with backstory impact the relationship formation and perception with people. Eyssel et al. (2017) studied how a robot's personal self-disclosure can promote anthropomorphism and attribution of mind to the robot. Kang & Gratch (2011) have investigated the effects of self-disclosure of a virtual counselor on the perception and behavior of the human interaction partners. The results showed that participants rated higher perception of copresence and social attraction towards the virtual agent which disclosed highly intimate information about itself than the one disclosed less intimate or no information about itself. In another valuable study, Bickmore et al. (2009) found that users completed more conversations and enjoyed communicating with a virtual human that talked about its created human life story (as its own story) compared to third person (as happening to humans that it knew). However, limited research has examined how a social robot with a self-backstory influences trust perception in people.

Therefore, the purpose of third study was to explore the influence of social robots disclosing their self-backstories (first person narrative) on the perceived trust by people in HRIs. This study also aimed to investigate how different kinds of backstories might influence people's perceived trust in robots in different ways. Different types of self-backstories was provided for a social robot, and their effect on various types of perceived trust was evaluated.

4.2 DEVELOPING HYPOTHESES

According to the literature, one of the basis of trust is reciprocity (McAllister, 1995). This suggests that when a party conveys trustworthiness, it prompts their counterpart to reciprocate with a similar level of trust. In this regard, a narrator of self-backstory exposes a vulnerability to the listener, and the listener who have gained the narrator's trust, is subsequently inclined to reciprocate and trust the narrator (Hagmann et al., 2020). Therefore, trust is established based on disclosing self-backstory, as it builds a reciprocal relation. When an individual shares personal information, they have opened themselves up by loosening their personal boundaries and taken the risk of receiving negative feedbacks from the listener. This vulnerability indicates a person or an agent's trustworthiness and makes them more likable.

Furthermore, self-backstories could be different in nature and convey positive or negative emotions. Generally, disclosing personal information makes people appear warmer and friendly (Wheeless & Grotz, 1977), and builds positive feelings in both trustor and trustee. Telling a self backstory fosters a sense of connection and empathy between the speaker and the listener, and evokes emotions of understanding, compassion, and shared humanity in the listener. Interestingly, Reed et al. (2019) reported that sad facial expression with tears increased trustworthiness and prosocial behavior or other-regarding altruistic tendencies increased respectively. Moreover, according to Darling (2017), backstory generally increased empathy towards a robot, so any type of backstory might lead to more empathy and affection through anthropomorphic framing. Sebo et al. (2019) showed that vulnerable statements made by a robot in team-related tasks lead to higher engagement with the robot compared to a restriction on neutral statements, with implications for trust-related behavior. Rosenthal-von der Pütten et al. (2013) also stated that witnessing violent behavior towards a robot leaded to higher empathic concern compared to witnessing positive interaction. We can assume that a self-disclosed backstory that evokes sadness or happiness lead to higher empathic concern for the robot, which subsequently lead to a higher degree of anthropomorphism and thus might promote affective trust. Furthermore, as we discussed earlier, cognitive trust is considered one of the antecedents of affective trust, and emotions, on the other side, influence the perception and cognitive evaluations, even after the dissipation of emotions. Previous research has identified significant links between affection and judgements (Dunn & Schweitzer, 2005). Therefore, emotional backstories may influence cognitive trust as well. Thus, we proposed the following hypotheses:

H1-1: When a social robot discloses its backstory (happy or sorrowful), the users are more likely to have (general) trust in the robot than the case of the robot not telling a backstory.

H1-2: When a social robot discloses its backstory (happy or sorrowful), the users are more likely to have affective trust in the robot than the case of the robot not telling a backstory.

H1-3: When a social robot discloses its backstory (happy or sorrowful), the users are more likely to have cognitive trust in the robot than the case of the robot not telling a backstory.

Additionally, previous studies found that trust is often linked to positive rather than negative outcomes of the behavior of others, and people might trust individuals expressing more positive emotions (Johnson & Mislin, 2011). Dunn and Schweitzer (2005) reported that emotions with positive valence such as happiness and gratitude increased trust, and emotions characterized by weak control appraisals (happiness) influenced trust significantly more than emotions characterized by personal control (pride and guilt) or situational control (sadness). Prior works also indicated that adults are prone to trust an individual expressing positive rather than negative emotion, and do so even when the expressive cues are subtle (Tang et al., 2019). Therefore, it seems plausible that a backstory displaying happy emotions will promote trust better than a sorrowful backstory. Considering the aforementioned studies, positive emotion has a more concrete relation to trust formation in general and thus we derived the next hypotheses:

H2-1: A social robot discloses its happy backstory results in more (general) trust than a robot with sorrowful backstory.

H2-2: A social robot discloses its happy backstory results in more affective trust than a robot with sorrowful backstory.

H2-3: A social robot discloses its happy backstory results in more cognitive trust than a robot with sorrowful backstory.

4.3 METHOD

4.3.1 Experimental design

According to the derived hypotheses, two types of self-backstories for the social robot including; happy backstory and sorrowful backstory, with a no backstory condition as the control condition was concluded. Thus, three experimental conditions were resulted. In order to explore whether a social robot telling a backstory about its experiences could influence people's trust perception of social robots, a between-subjects experiment was conducted with three conditions. Table 4.1 shows the experimental conditions.

Table 4. 1 Three types of robot's backstory in experimental conditions

Experimental conditions	Type of self-backstory
Happy Backstory condition **(H-Bc)**	Robot discloses a happy backstory about itself
Sorrowful Backstory condition **(S-Bc)**	Robot discloses a sorrowful backstory about itself
No Backstory condition **(N-Bc)**	Robot talks about general and technical information about itself

For the conditions with backstory (H-Bc and S-Bc), the robot was assumed to work in a restaurant and it disclosed a happy or sorrowful story about its experience working as a waiter in the restaurant. In the happy backstory, the robot explained how it was welcomed by the customers and the kids were happy seeing the robot. Whereas in the sorrowful story, the robot described about the unhappy experience with one of the customers, in which the kids were scared of the robot and the customers were aggressive to the robot. We chose the topic considering the prior experiences and interactions of participants with the robot, emotional feasibility of stories, and the believability of the situation for the robot. For N-Bc, the robot only described general and technical information about itself. **Appendix C** includes the script of backstories. The happy and sorrowful stories were evaluated by other researchers to be comprehended as expressing happy and sorrowful feelings. The backstories were prepared according to the literacy structure (Pattie, 2021), and included an introduction part, describing the experience and events that happened, expressing the feelings about the situation, and the conclusion part. The experiment was designed based on a short conversation between participants and the robot, in which the robot disclosed its story to the participants, and questionnaire-based evaluation. The conversation started with a brief greeting part to initiate the conversation and provide a dyadic communication. Figure 4.1 shows the setup of the experiment. The experiment was conducted in Japanese.

4.3.2 Participants

A total of 66 students took part in the experiment. Participants were *Japanese students* at *Tokyo Institute of Technology*, aged from 18 to 25 (M_{age} = 20.62, SD_{age} = 1.55), with a 49:17 male: female ratio. A power analysis for a one-way ANOVA test indicated that the minimum sample size to yield a statistical power of at least 0.8 with an alpha of

0.05 and a large effect size (d = 0.4) was 66. However, 0.4 is a large effect size and the results may not be reliable for large groups, we could not add more participants due to the time limitations. Participants were assigned to one of three experimental conditions: H-Bc, S-Bc and N-Bc. Each group contained twenty-two participants and we attempted to counterbalance the gender between conditions as much as possible. Table 4.2 lists the number and gender disposition for each of the three experimental conditions.

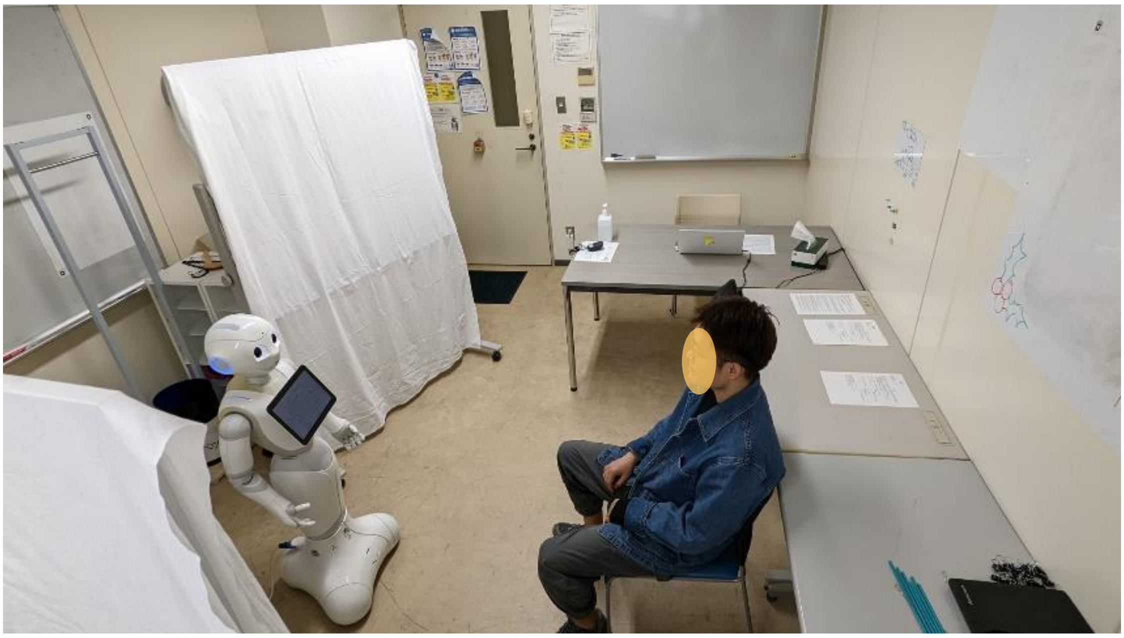


Figure 4. 1 Setup of the experiment.

Table 4. 2 Number of participants for three experimental conditions.

Condition	Age Mean (SD)	Male	Female	Total
H-Bc	20.59 (1.68)	16	6	22
S-Bc	21.41 (2.64)	16	6	22
N-Bc	20.18 (1.91)	17	5	22

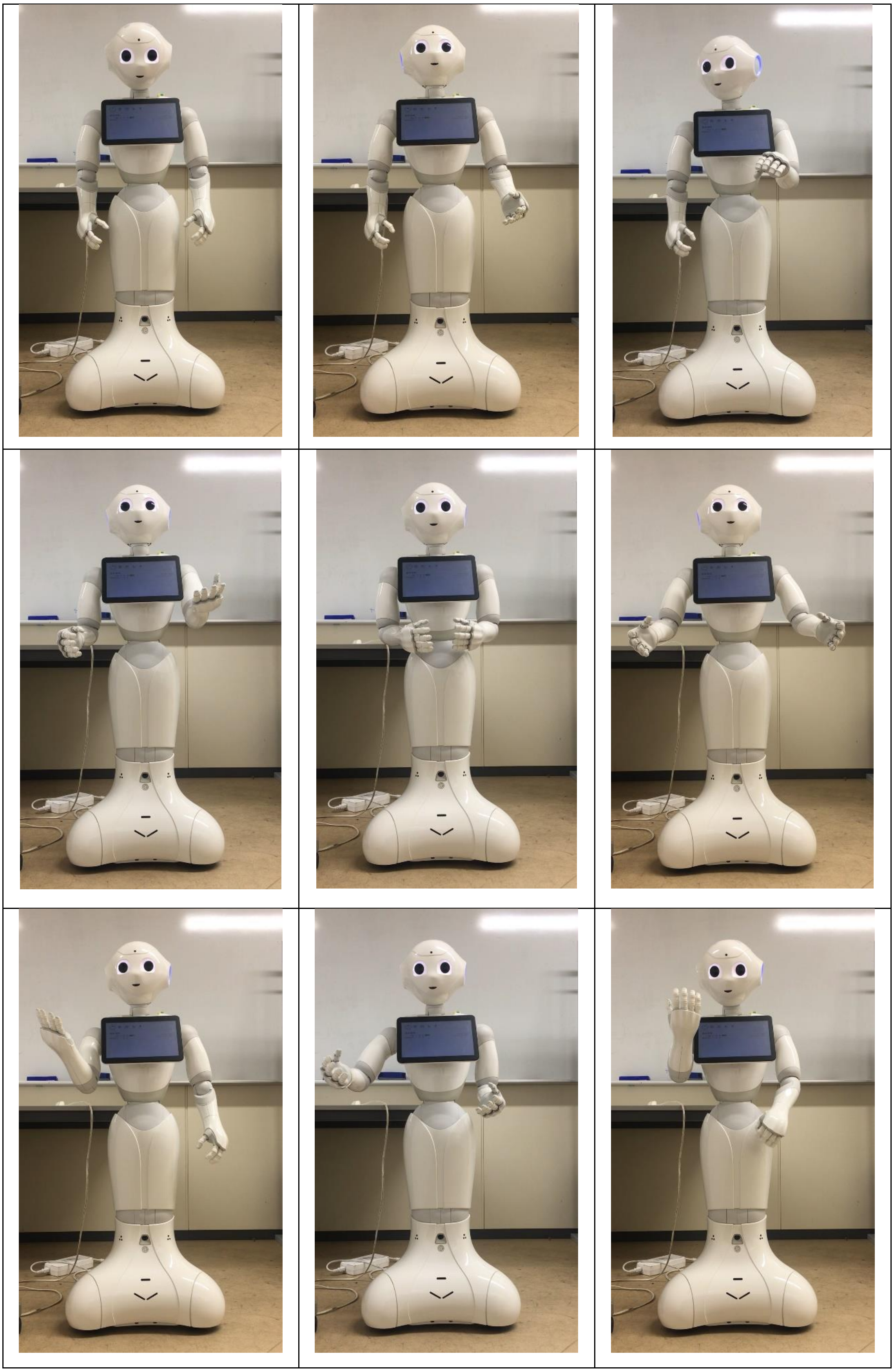
We recruited participants by distributing flyers in various laboratories and around the Ookayama campus of Tokyo Institute of Technology. Additionally, we approached the students enrolled in the "Engineering Psychology" and "Data Collection and Analysis" courses supervised by Professor Umemuro. Since we aimed to compare the cultural difference between Germany and Japan, we specifically recruited Japanese students. However, we did not specify any particular educational background requirements for participants. Participants received 1000 Japanese Yen for participating in the experiment.

4.3.3 Ethical considerations

The experiment received the approval of all ethical and experimental procedures and protocols from Human Subjects Research Ethics Review Committee of Tokyo Institute of Technology. All participants submitted a written consent form and were informed of the experimental procedures and ethical concerns prior to the experiment. The experiment was anonymous and the participants were identified by IDs.

4.3.4 Equipment

For the third study, we used the humanoid **Pepper** robot manufactured by SoftBank Robotics. Pepper is a 121 cm tall humanoid robot that can speak and move in different ways. During the third experiment, Pepper talked to the participants. The participants sat on a chair to achieve an equal eye-level view with Pepper, positioned at a distance of 1 meter to ensure a personal zone and represent a friendly interaction.Speech was controlled through an interface developed in HTML by using Pepper's default text-to-speech settings, together with body movements that were designed using Choregraphe 2.5.10. The speech rate was set at 100% to represent a normal speaking speed. Several pauses were incorporated into its speech to make it more understandable and natural. Appropriate body movements were designed for Pepper during its speech to make it alive and interactive. Due to Pepper's limitations in body movements and facial expressions, we manipulated its head movements and hand gestures. The homogeneity of body movements and the expression of emotions were attempted to be considered, specifically between happy and sorrowful conditions. Figure 4.2 shows different body postures of Pepper during the experiment. All video of the experiment are accessible at: https://osf.io/pqf69/?view_only=fa34adaae24343ef9431b04b7b8daee8



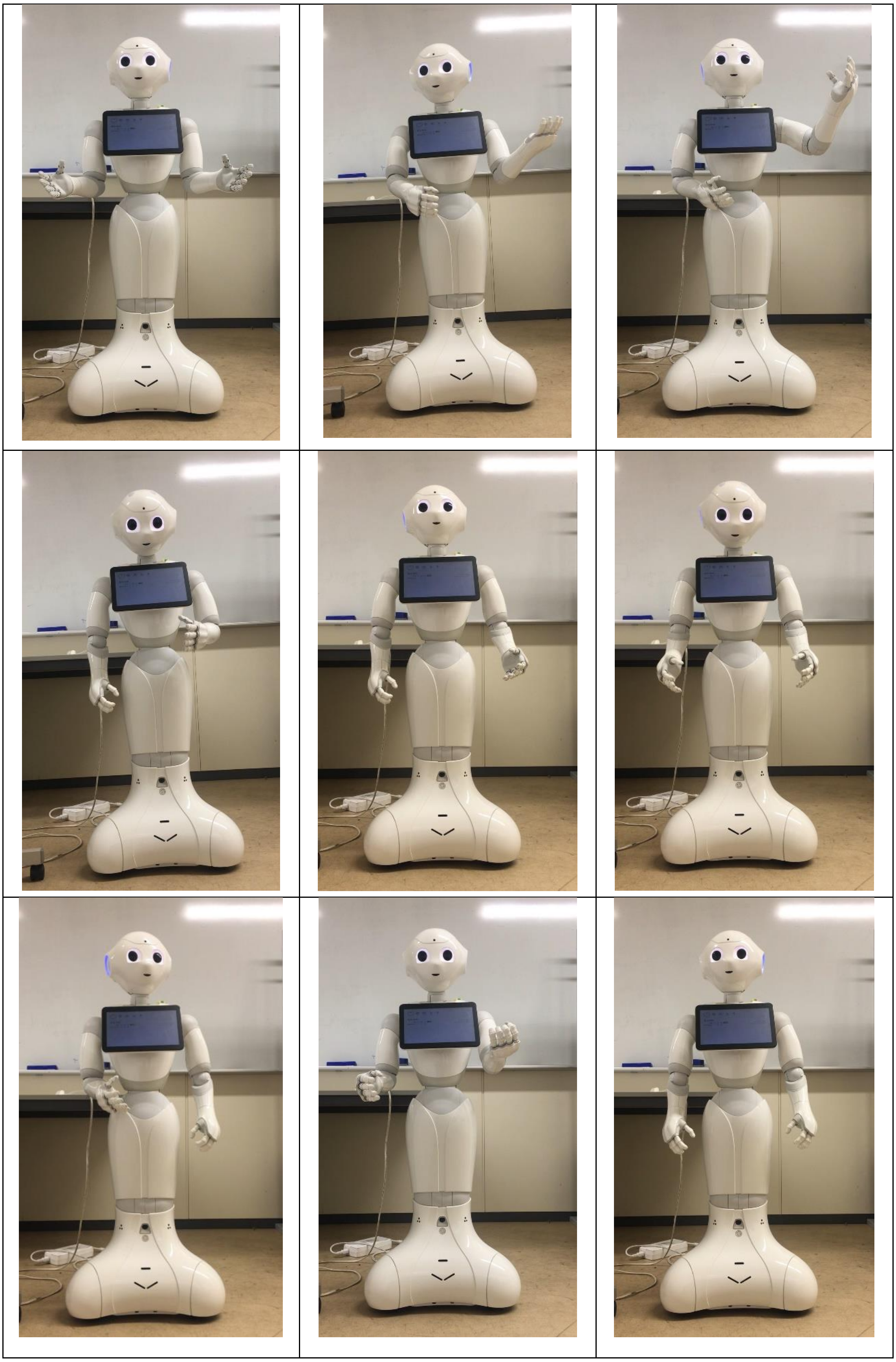


Figure 4. 2 Some of Pepper's body movements in the experiment.

4.3.5 Procedure

The experiment was held during two weeks in November 2023 at the Ookayama campus of Tokyo Tech University. Before initiating the experimental session, participants were given documents explaining the experiment to review freely and decide voluntarily if they desired to participate in the experiment. The experimenter then explained the procedure and highlighted important points, to ensure that everything was clear to participant prior to signing the consent form. None of the participants opted out of the experiment. Figure 4.3 shows the procedure of the experiment.

First, participants were asked to fill out the demographic information. After completing the primary questionnaire, the experimenter introduced Pepper to the participants and asked them to rate their trust in the robot before the interaction. Participants evaluated their trust in the robot on 30-item pre-trust questionnaire, which included 14 items on robot-human trust, 9 items on cognitive trust, and 7 items on affective trust. Figure 4.4 shows the participant answering the questionnaires before starting the conversation with Pepper. Thereafter, the participants sat in front of the robot and went through the conversation, listening to the robot's story, after having a brief introductory conversation. Figure 4.5 shows the participant listening to Pepper's self-backstory. For each participant, the interaction with the robot lasted approximately 6-7 min. After the completion of the conversation, the participants were asked to fill out the post-questionnaires to rate their perception of the robot's trustworthiness after the interaction, in addition to cultural values questionnaire. The entire session finished in almost 30 min. Wizard-of-Oz (WOZ) methodology was employed in the experiment similar to the first and second experiments.

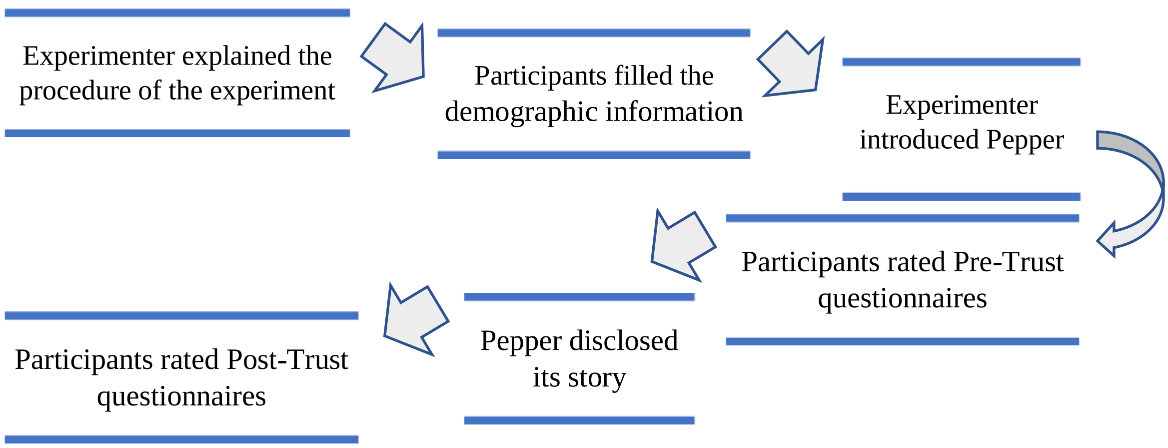


Figure 4. 3 Procedure of the experiment.

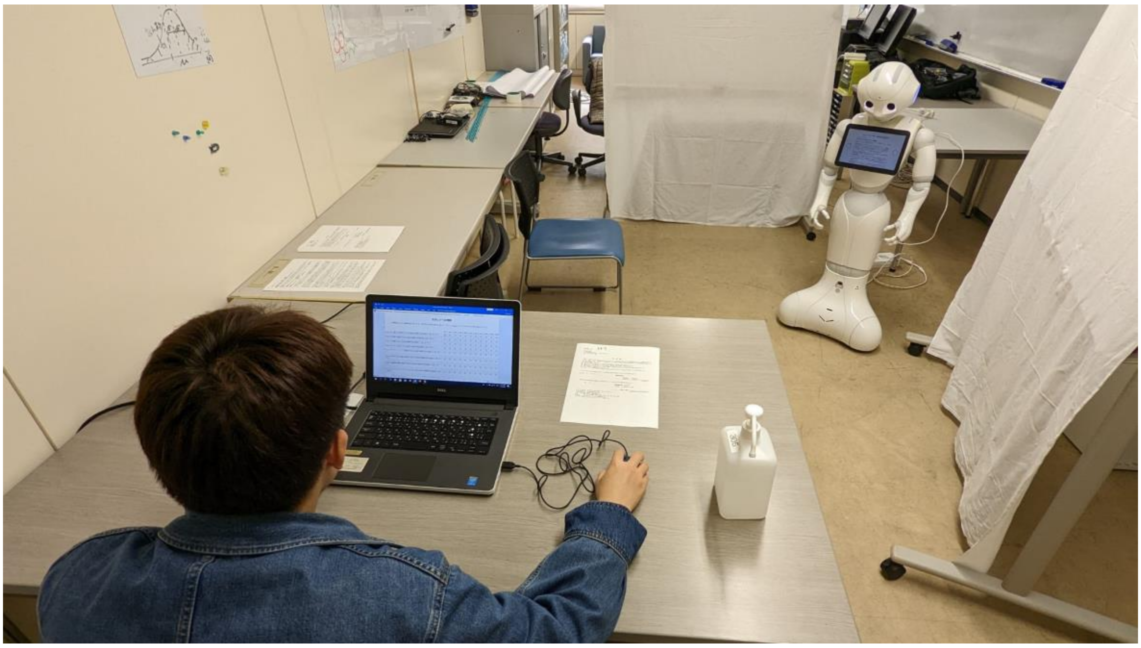


Figure 4. 4 Procedure of the experiment. Participant answers the questionnaires before conversation with Pepper.

Figure 4. 5 Procedure of the experiment. Participant listens to Pepper while disclosing its self-backstory.

4.3.6 Measurements

Similar to the previous experiments, participants assessed the robot in terms of general, affective and cognitive trust.

4.3.6.1 General trust scale

Schaefer (2016) offered two types of human-robot trust scale, one with 40 items and the other one with 14 items. For this experiment, the 14-item scale was used to assess subjects' general trust in the robot. The scale consists of 14 items rating participants' trust on the robot in the range from 0% to 100% according to each item. The Cronbach's alpha value of the questionnaire was 0.90 which showed high reliability.

4.3.6.2 Affective and Cognitive Trust Scale

Similar to first and second experiments, the measure proposed by Johnson and Grayson (2005) and Mcallister (1995) was used to evaluate affective and cognitive trust. The scale was a 16-item questionnaire, including nine items for cognitive trust and seven questions for affective trust. The measure evaluated participants' feelings and perceptions of trust in the robot using a seven-point Likert scale. The Cronbach's alpha value was 0.88 for the affective trust scale and 0.75 for cognitive trust scale.

The original English questionnaires were translated and back-translated by the colleague researchers of the authors whose native language were Japanese.

4.4 RESULTS

4.4.1 Data analysis

Data of 66 participants were collected, and initially the dataset was filtered to identify and address any suspicion of unreliable data. Fortunately, no unreliable data was found, and all data obtained from the participants were used for subsequent statistical analysis. Descriptive statistics was used to summarize the data and compare the overall mean of dependent variables. Normality test was conducted to determine whether the sample data had been drawn from a normally distributed population, guiding the selection of the main data analysis method. A one-way ANOVA test with post-hoc analysis was used to compare the differences between the means of dependent variables among three groups. Considering the mean differences, the scores between before and after the interaction was compared for dependent variables, rather than assessing changes during the interaction.

4.4.2 Descriptive statistics and test of results

To investigate the effect of self-backstory disclosure of social robot on the perceived trust of participants, we compared trust scores before and after interaction for three experimental conditions. Table 4.3 shows the means and standard deviations of general trust, affective trust and cognitive trust scores for three experimental conditions measured before and after interaction, as well as differences between them. Overall, H-Bc had the highest scores compared to the other conditions, considering both after interaction and difference between before and after interaction in all trust scores. On the other hand, S-Bc and N-Bc got similar mean scores in all trust variables.

Table 4. 3 Means and standard deviation for all trust scores for three experimental conditions, before, after and differences over the interaction.

Condition	before conversation		after conversation		difference	
	Mean	SD	Mean	SD	Mean	SD
	General Trust					
H-Bc	87.55	15.55	106.64	13.17	19.09	12.49
S-Bc	85.05	15.62	93.18	19.75	8.14	16.54
N-Bc	84.64	12.82	93.23	20.05	8.59	18.03
	Affective Trust					
H-Bc	3.18	0.89	4.49	1.09	1.32	0.71
S-Bc	2.96	1.10	3.55	1.14	0.58	1.34
N-Bc	2.99	0.69	3.62	1.21	0.63	1.15
	Cognitive Trust					
H-Bc	3.63	0.64	4.26	0.78	0.63	0.72
S-Bc	3.59	0.59	3.73	0.74	0.15	1.03
N-Bc	3.58	0.52	4.01	0.63	0.42	0.78

To test the hypotheses, a one-way ANOVA was conducted with three backstory conditions (happy, sorrowful, no backstory) as independent variables, and three trust scores (general, affective, cognitive) as dependent variables, all by before and after interaction. The results indicated that the main effects of backstory were not significant for the score of general trust ($F(2,63) = 0.25$, $p = 0.77$), affective trust ($F(2,63) = 0.35$, $p = 0.70$), and cognitive trust ($F(2,63) = 0.04$, $p = 0.95$) *before interaction*. On the other hand, the main effects of backstory conditions were significant for the scores of general trust ($F(2,63) = 4.11$, $p < 0.05$), affective trust ($F(2,63) = 4.60$, $p < 0.05$), but not for cognitive trust ($F(2,63) = 2.94$, $p = 0.060$) *after interaction*. Table 4.4 shows the results of ANOVA analysis before and after interaction. Considering ANOVA results, post-hoc test was conducted for the rest of analyses.

Table 4. 4 ANOVA results for all trust scores among three experimental conditions.

	before interaction			***after interaction***		
	df	*F*	*p*	*df*	*F*	*p*
General trust	2	0.25	0.77	2	4.11	0.021
Affective trust	2	0.35	0.70	2	4.60	0.014
Cognitive trust	2	0.49	0.95	2	2.94	0.060

4.4.3 Effect of social robot's self-backstory disclosure on HRT

The first set of hypotheses explored the difference among H-Bc, S-Bc and N-Bc in general, affective and cognitive trust. Table 4.5 shows the results of post hoc analysis for different group comparisons. We assumed that H-Bc and S-Bc would be higher than N-Bc. Post-hoc analysis results revealed that the *general trust* score *after interaction* was significantly higher for H-Bc (M = 106.64, SD = 19.09) than for N-Bc (M = 93.23, SD = 8.59, $p < 0.05$). However, the general trust score in S-Bc (M = 93.18, SD = 8.14) was not significantly different from the score in N-Bc (M = 93.23, SD = 20.05, $p = 0.993$), and interestingly, N-Bc had slightly higher score than S-Bc. Figure 4.2 panel A illustrates the results for general trust.

Furthermore, post-hoc analysis revealed that H-Bc (M = 4.49, SD = 1.09) was rated significantly higher than N-Bc (M = 3.62, SD = 0.63, $p < 0.05$) in terms of *affective trust, after interaction*. But there was no significant difference between S-Bc (M = 3.55, SD = 0.58, $p = 0.823$) and N-Bc for affective trust. Again, the robot with no backstory perceived slightly higher than the robot with sorrowful backstory in affective trust. Figure 4.6 panel B illustrates the results for affective trust.

Table 4. 5 Post hoc analysis results for different group comparisons

	H-Bc vs. N-Bc	S-Bc vs. N-Bc	H-Bc vs.S-Bc
	p	*p*	*p*
General trust	0.016*	0.993	0.015*
Affective trust	0.015*	0.823	0.008**
Cognitive trust	0.243	0.216	0.018*

Statistically significant difference: $^{**}p < 0.01$, $^{*}p < 0.05$

Finally, Post-hoc analysis indicated that the *cognitive trust* score in neither H-Bc (M = 4.26, SD = 0.63, $p = 0.243$) nor S-Bc (M = 3.73, SD = 0.15, $p = 0.216$) were significantly different from N-Bc (M = 4.01, SD = 0.42) *after interaction*, as shown in Table 4.6. Figure 4.6 panel C illustrates the results. Although, the robot with happy backstory was rated higher than other conditions, the difference was not statistically significant. The robot with sorrowful backstory perceived slightly higher score than the robot with no backstory in cognitive trust. However, the difference was not significant.

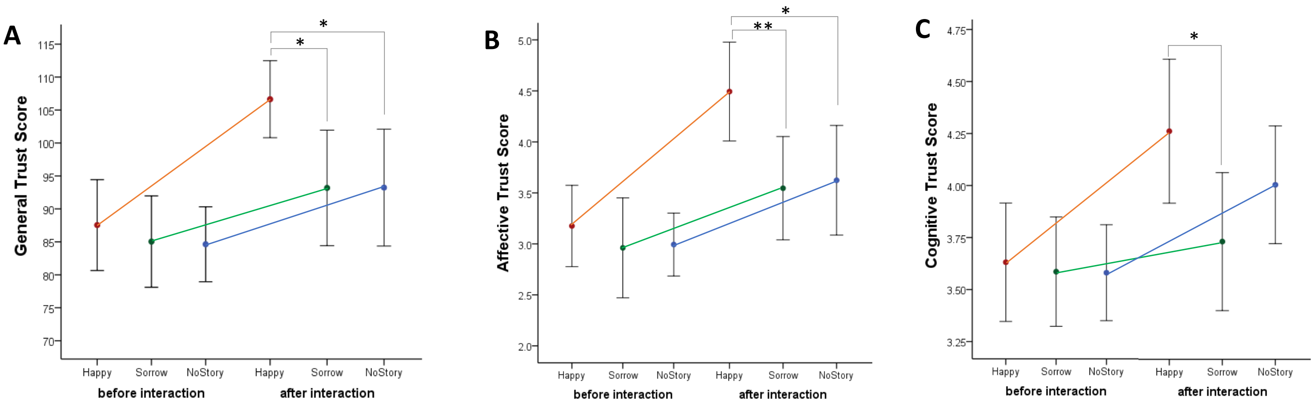


Figure 4. 6 Means and standard deviations of (A) general trust scores of three experimental conditions before and after interaction, (B) affective trust scores of three experimental conditions before and after interaction, and (C) cognitive trust scores of three experimental conditions before and after interaction. Dots indicate means and vertical lines indicate the SDs. *(*p < 0.05, **p < 0.01)*

4.4.4 The primacy of social robots with happy backstory in HRT

The second set of hypotheses explored the difference between the robot with a happy and a sorrowful backstory to find whether happy backstory could enhance the robot's trustworthiness more than sorrowful backstory. As shown in Table 4.5, post-hoc analysis results indicated that H-Bc (M = 106.64, SD = 13.17) was significantly more trustworthy than S-Bc (M = 93.18, SD = 19.75, $p < 0.05$) in terms of *general trust after interaction*. There was also a significant main effect of H-Bc (M = 4.49, SD = 1.09) on *affective trust* score, compared with S-Bc (M = 3.55, SD = 1.14, $p < 0.001$) *after interaction*. Furthermore, H-Bc (M = 4.26, SD = 0.78) was rated significantly higher than S-Bc (M = 3.73, SD = 1.74, $p < 0.05$) for *cognitive trust* score *after interaction*. Thus, the robot with happy backstory could considerably improve general, affective and cognitive trust compared with the robot which disclosed sorrowful backstory (Figure 4.6, panels A-C).

To summarize, as the H-Bc was significantly higher in general trust and affective trust than the N-Bc after interaction, **H1-1** and **H1-2** were supported for happy backstory, but not for sorrowful backstory. **H1-3** was not supported because the difference in the cognitive trust score for H-Bc and S-Bc was not significantly higher than N-Bc. Furthermore, as the H-Bc scored significantly higher in general, affective and cognitive trust than the S-Bc after interaction, **H2-1**, **H2-2** and **H2-3** were supported.

4.5 DISCUSSION

During the third study, the effect of social robots which disclosed two types of happy and sorrowful backstories with social robots that told no backstory was compared, to see their contribution to trust improvement in HRIs. All trust scores increased after interaction with the robot. However, only the happy backstory condition was evaluated more trustworthy than no backstory condition for general and affective trust (H1-1, H1-2). This result is in line with the prior work in interpersonal trust as people with a disclosed backstory and personal narratives are considered to be more trustworthy (Hagmann et al., 2020). In addition, according to Jasielska (2020), and Johnson and Mislin (2011) trust is inked to positive outcomes of behavior such as happiness and kindness rather than negative outcomes. Diener et al. (1999) also expressed the dominance of positive emotions over negative emotions.

As expected, the results also confirmed that the robot with a happy backstory resulted in more general, affective and cognitive trust than the robot with a sorrowful backstory (H2-1, H2-2, H2-3). This means that conveying positive emotions by a robot could evoke emotional impressions in participants more than sad emotions. Happiness is treated as an indicator of wellbeing in two emotional and cognitive dimensions (Diener et al., 1999) and studies admitted that happiness includes two components of affection and cognition, and determined by interpersonal interactions (Jasielska, 2020). So, the two dimensions of happiness are in line with affective and cognitive trust and the results revealed that participants rated the robot with a happy backstory more trustworthy in both dimensions.

However, despite the happy backstory having been affectively trust evoking, there was no significant difference in cognitive trust compared to no backstory (H1-3). This is an interesting finding and could have several reasons. Cognitive trust is knowledge-driven and it is based on the complete certainty of a partner's future actions (D. Johnson & Grayson, 2005). In case of the no backstory condition, the robot conveyed a certain amount of technical information about itself that could increase participants knowledge about the robot and promote cognitive trust. Thus, although the mean score was higher for the robot with a happy backstory, it was not significantly different with the no backstory condition.

Finally, the results indicated that the robot with a sorrowful backstory received lower mean scores than the no backstory condition in all trust scores (H1-1, H1-2, H1-3). Although sorrowful stories evoke emotions and could establish an emotional impression on people, participants did not evaluate it the same way as a happy backstory. This contradicts the previous findings of Darling (2017) to some degree, as it was expected that an emotional backstory would generally lead to a higher perceived trust. This could be explained by the fact that the factual self-description of the no backstory case also has an influence on trust formation. Rather, the sorrowful backstory might not lead to an empathic concern for the robot as the initial degree of anthropomorphism and emotional connection to the robot is not large enough. Further, it is suggested that sadness is typically associated with less confidence than happiness, leading to higher uncertainty (Briñol et al., 2007), which is another explanation for our results.

4.6 CHAPTER SUMMARY

During the third study, the impact of social robots disclosing their self-backstories on people's trust perception of social robots was explored, with respect to the nature of the backstories. The results indicated that when the social robot expressed a happy backstory about itself, it affected people's trust perception of social robots and increased it emotionally. The findings can be useful for designing future social robots, especially in contexts where the improvement of trust is necessary through the incorporation of positive emotions, such as in therapy, service, or educational robots. Furthermore, the impression of positive emotions was higher than that of negative emotions. These findings underscore the importance of considering the type of backstories in the design of social robots.

The participants of current study were only Japanese participants which was one of the research's limitations. The future research will evaluate the effect of social robots' self-backstory disclosure in German participants and comparing the results with the Japanese participants with respect to their cultural differences. Furthermore, other factors such as participants basic attitudes to social robots, prior experiences, type of robot and backstories, gender and the context of interaction can be considered in future researches, which were the other limitations of current study

4.7 REFERENCES

Ajzen, I. (1977). Information Processing Approaches to Interpersonal Attraction. *Theory and Practice in Interpseronal Attraction*, *October*, 51–77.

Bickmore, T., Schulman, D., & Yin, L. (2009). Engagement vs. Deceit: Virtual Humans with Human Autobiographies. *IVA '09: Proceedings of the 9th International Conference on Intelligent Virtual Agents*, 6–19. https://doi.org/10.1007/978-3-642-04380-2_4

Briñol, P., Petty, R. E., & Barden, J. (2007). Happiness Versus Sadness as a Determinant of Thought Confidence in Persuasion: A Self-Validation Analysis. *Journal of Personality and Social Psychology*, *93*(5), 711–727. https://doi.org/10.1037/0022-3514.93.5.711

Darling, K. (2017). "Who's Johnny?" Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. In P. Lin, G. Bekey, K. Abney, & R. Jenkins (Eds.), *ROBOT ETHICS 2.0*. Oxford University Press. https://doi.org/http://dx.doi.org/10.2139/ssrn.2588669

Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276–302. https://doi.org/10.1037/0033-2909.125.2.276

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. https://doi.org/10.1037/0022-3514.88.5.736

Eyssel, F., Wullenkord, R., & Nitsch, V. (2017). The Role of Self-Disclosure in Human-Robot Interaction. *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 922–927. https://doi.org/10.1109/ROMAN.2017.8172413.

Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A. C., & Wang, J. (2005). Designing robots for long-term social interaction. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, 1338–1343. https://doi.org/10.1109/IROS.2005.1545303

Hagmann, D., Minson, J. A., & Tinsley, C. H. (2020). *Personal Narratives Build Trust Across Ideological Divides*. OSF Preprints, 29 Sept. https://doi.org/doi:10.31219/osf.io/sw7nz.

Jasielska, D. (2020). The moderating role of kindness on the relation between trust and happiness. *Current Psychology*, *39*(6), 2065–2073. https://doi.org/10.1007/s12144-018-9886-7

John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(4), 954–959. https://doi.org/10.1073/pnas.1516868113

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

Kang, S. H., & Gratch, J. (2011). People like virtual counselors that highly-disclose about themselves. *Annual Review of CyberTherapy and Telemedicine*, *9*(1), 117–120.

Kory-Westlund, J. M., & Breazeal, C. (2019). Exploring the Effects of a Social Robot's Speech Entrainment and Backstory on Young Children's Emotion, Rapport, Relationship, and Learning. *Frontiers in Robotics and AI*, *6*(July). https://doi.org/10.3389/frobt.2019.00054

McAllister, D. J. (1995). Affect and CognitionBased Trust As Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal*, *38*(1), 24–59. https://doi.org/10.2307/256727

Pattie, J. O. (2021). Personal Narrative. *Travels, Explorations and Empires*, 289–303. https://doi.org/10.4324/9781003113317-11

Reed, L. I., Matari, Y., Wu, M., & Janaswamy, R. (2019). Emotional Tears: An Honest Signal of Trustworthiness Increasing Prosocial Behavior? *Evolutionary Psychology*, *17*(3), 1–8. https://doi.org/10.1177/1474704919872421

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics*, *5*(1), 17–34. https://doi.org/10.1007/s12369-012-0173-8

Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the"Trust Perception Scale-HRI". In R. Mittu, D. Sofge, A. Wagner, & W. F. Lawless (Eds.), *Robust Intelligence and Trust in Autonomous Systems* (Issue August, pp. 191–218). Springer. https://doi.org/10.1007/978-1-4899-7668-0

Sebo, S. S., Krishnamurthi, P., & Scassellati, B. (2019). "I Don't Believe You": Investigating the Effects of Robot Trust Violation and Repair. *ACM/IEEE International Conference on Human-Robot Interaction*, *2019-March*, 57–65. https://doi.org/10.1109/HRI.2019.8673169

Tang, Y., Harris, P. L., Zou, H., & Xu, Q. (2019). The impact of emotional expressions on children's trust judgments. *Cognition and Emotion*, *33*(2), 318–331. https://doi.org/10.1080/02699931.2018.1449735

Wheeless, L. R., & Grotz, J. (1977). the Measurement of Trust and Its Relationship To Self-Disclosure. *Human Communication Research*, *3*(3), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSIONS

5.1 SUMMARY OF DISSERTATION

In an era dominated by AI and robotics, the integration of social robots into people's daily lives seems unavoidable, demanding a focus on users' trust toward robots. Thus, future developers must prioritize enhancing social robots' trustworthiness through thoughtful design to make them reliable for users. Accordingly, in this research, *we aimed to identify and evaluate some of influential behaviors and attributes for social robots that improve humans' perception of trust in them.* Different trust antecedents in HRI have been investigated, including robots' performance, appearance, and physical attributes, as well as robots' behaviors. Among these factors, research on the impact of social robots' behaviors on trust is narrower, particularly those inspired from interpersonal trust. Thus, we focused on three factors: *1) listening behaviors, 2) benevolence and competence attributes, and 3) self-backstory disclosure,* examining their contributions to enhancing various types of trust, such as *general, affective, and cognitive trust*. Three experiments were conducted in which participants engaged in conversations with a social robot under real conditions. Participants then evaluated their perception of the robot's trustworthiness through a questionnaire-based measurement. In the subsequent parts, we will discuss the key findings resulting from the experiments.

5.1.1 Effect of social robots' listening behaviors on HRT

In the first study we investigated whether the listening behaviors of a social robot could enhance its trustworthiness. In human communication, people build trust in their partner when they receive care, understanding, and empathy. Effective listening is one of the ways for accurately understanding and interpreting messages, building rapport, and fostering meaningful and trustworthy connections between humans (Weger et al., 2014). The literature supports the idea that certain listening behaviors positively influence the perception of an individual's trustworthiness (see Brunner, 2008; Ramsey & Sohi, 1997), including active or empathic listening behaviors. Thus, In the first experiment reported in

Chapter 2, we explored whether social robots, when equipped with appropriate listening behaviors, could enhance trust and convey a sense of being alive, humanistic, and likable. Four different listening behaviors were designed for a social robot, including *nonactive listening, active listening, active-empathic listening, and verbal-only empathic listening*. Thereafter, we evaluated and compared the impact of each behavior on the participants' likelihood of perceiving general, affective, and cognitive trust.

5.1.1.1 Active and empathic listening behaviors of social robots contribute to HRT

The results revealed that all three forms of active listening, active-empathic listening, and verbal-only empathic listening were evaluated higher than nonactive listening behavior in terms of general, affective and cognitive trust. Thus, we can consider appropriate listening behavior as a method to improve trust in HRI.

5.1.1.2 Active-empathic listening behavior of social robots establishes a significant trustworthy interaction with people, particularly in terms of affective trust

Furthermore, the results indicated that social robots exhibiting active-empathic listening behavior provided the participants with the highest impression of trustworthiness, particularly in affective trust. Similar to interpersonal trust (see (Ma et al., 2018)) showing competency, understanding, emotional sharing and responsiveness contributes to the development of trust between users and social robots. However, active-empathic listening behavior was differentiated from active listening behavior only in terms of affective trust. This outcome admits the significant contribution of empathic listening behavior of social robots to affective trust, paving the way for new possibilities in designing empathic social robots in the future. Emotional experiences enrich the quality of communication (Floyd, 2014), and people with more emotional expressions have a higher impression of being trustworthy than people who care and support rationally (Aggarwal et al., 2005; Drollinger & Comer, 2013). If active-empathic listening behavior is conceived as an expression of affection in HRI, it would improve diverse relationships, and establish more reliable bond between users and social robots.

5.1.1.3 Nonverbal behaviors of active listener social robots enhance affective trust

Moreover, considering verbal and nonverbal dimensions of listening behavior, the results confirmed that nonverbal behaviors such as nodding, body movement, and eye gaze along with verbal behaviors, had a significant effect in eliciting higher affective trust in HRI. The results did not reveal any significant difference in general and cognitive trust when the robot exhibited both verbal and nonverbal behaviors. In HC, nonverbal behaviors are claimed to be effective means to indicate many social information (see Thepsoonthorn et al., 2018). The robots with both verbal and nonverbal behaviors are also rated as being more natural and engaging (Johanson et al., 2021), as they improve the anthropomorphic perception of robots and it results in better interaction respectively. However, considering listening behaviors and their contribution to HRT, it's noteworthy that nonverbal components only enhanced users' affective and emotional perception. These findings are especially valuable and applicable in the context of humanoid robots.

In conclusion, the results indicated that, during a friendly and casual conversation, people perceive social robots as human-like. While the awareness of interacting with a robot is evident, individuals tend to believe in its behaviors and are influenced by the diverse approaches exhibited by the robots. A robot with good listening skills appears alive, likable, and intelligent to people. This encourages them to continue conversations and engage with the robot's ideas and comments on various issues. Furthermore, listening behaviors are divergent and they include various emotional, logical, verbal and nonverbal components that distinguishes them from each other. For example, the type of appropriate listening behavior in the context of educational environment or during a marital conflict may be different. Thus, we need to consider the appropriate listening behavior for a social robot, taking the context into account.

5.1.2 Effect of social robots' benevolence and competence attributes on HRT

In the second study we investigated if the competence and benevolence characteristics of social robots influence people's perception of their trustworthiness. In human communication, benevolence and competence are considered two wings of trustworthy interaction (Lee & See, 2004; Mayer et al., 1995). Competence supports the rational and cognitive aspects of trustworthiness by demonstrating knowledge, accuracy, and skills. Meanwhile, benevolence increases affective trust between parties by fostering kindness, warmth, care, and friendly behaviors (see Di Battista et al., 2020, 2021). Thus, in the second experiment reported in **Chapter 3**, we explored whether the same principles of interpersonal trust apply to HRI, and how the benevolence and competence attributes of social robots prioritize and integrate into HRT. Four different combinations of benevolence and competence attributes for a social robot were designed, including *nonbenevolent-noncompetent, nonbenevolent-competent, noncompetent-benevolent, and benevolent-competent.* Thereafter, the perceived general, affective, and cognitive trust toward the social robot exhibiting those characteristics were measured and compared.

5.1.2.1 Competence and benevolence attributes of social robots increase HRT

The results indicated the effectiveness of both benevolence and competence attributes in HRT as the social robot that behaved as competent-benevolent, competent-nonbenevolent, and benevolent-noncompetent was assessed to have higher general trust than that of the noncompetent-nonbenevolent robot.

5.1.2.2 Benevolence characteristics of social robots fosters their trustworthiness more significantly than competence attributes

Furthermore, the results indicated that, compared to a benevolent-competent robot, the benevolent-noncompetent robot enhanced general trust more significantly than the competent-nonbenevolent one. The results reveal that, at the same level of comparison, benevolent behavior was more influential than competency in HRT. However, the context of the conversation may be influential, and this should be considered in future studies.

5.1.2.3 Participants do not prioritize benevolence or competence in HRT over each other

An important focus of this study was to determine whether participants prioritize a benevolent robot over a competent one. The interrelation of these factors in different fields still remains ambiguous, and various factors can influence this priority (see Kervyn et al., 2009; Tschannen-Moran & Hoy, 2000; Ybarra et al., 2008). The results did not reveal any significant difference between benevolent-noncompetent robot and competent-nonbenevolent robot in terms of general trust. In other words, there was not a significant inclination towards valuing one aspect (benevolence or competence) more than the other in the context of how participants perceive or interact with the robot. However, different factors influence this outcome, such as the context of the conversation, users' preferences, users' approach to the robots, types of robots, and more. Therefore, further studies are needed to gain a clear understanding.

5.1.2.4 Benevolent social robots enhance affective trust in HRT

Moreover, the participants rated the benevolent-noncompetent robot as more trustworthy than the competent-nonbenevolent robot in terms of the affective trust, but not cognitive, which revealed the primacy of benevolence in fostering affective trust, and modulating general and cognitive trust. Meanwhile, the perceived competence of the robot did not significantly influence the cognitive trust. However, further investigation is needed in this point, as the specific type of robot and the nature of the conversation may exert a substantial influence. NAO is a cute humanoid robot to which most participants had a positive approach, and this might have affected the results.

In conclusion, competence and benevolence emerge as two influential factors in HRT, with benevolence exhibiting a superior effect on affective trust. The findings hold potential for informing the design of more anthropomorphic social robots in future. Participants evaluated a benevolent robot trustable and likeable. Benevolence could foster affective trust and the likability of social robots. This suggests that people feel a stronger emotional connection with social robots that exhibit care and concern for their problems and challenges. In situations where the roles of robots are primarily functional, and performance-based actions are highly demanded, participants rated the benevolence attribute of a social robot higher than its competence in trust evaluations during a friendly and free conversation. The reason could be that empathic and benevolent robots are less developed, and a social robot exhibiting caring behaviors appears novel and interesting to people. Meanwhile, there is still a high expectation for a robot to be informative and knowledgeable, especially with the recent advancements in AI products.

5.1.3 Effect of social robots' self-disclosed backstory on HRT

In the third experiment we investigated the influence of a social robot which discloses a backstory of its experiences on the development of trust in HRI with respect to the nature of backstories. According to the literature, trust is increased when people support their point of view by disclosing stories of personal experiences rather than experiences of others (Hagmann et al., 2020). Thus, in the third experiment reported in **Chapter 4**, we explored if a social robot which expresses its prior experiences and backstories about a specific context and issue were perceived as more trustworthy. We presented three experimental conditions in which a social robot expressed different backstories, including *a happy backstory, a sorrowful backstory, and no backstory* about a specific topic during interactions with participants. Subsequently, we measured and compared participants' perceived general, affective, and cognitive trust for each condition.

5.1.3.1. Social robots that disclose their happy backstory enhance trust in HRI

The results indicated that the robot disclosing a happy backstory provided the participants with higher impression of trustworthiness in general and affective trust compared to the robot telling no backstory. However, the robot with a sorrowful backstory was not evaluated as leading to higher trustworthiness than the robot with no backstory. This implies that participants did not differentiate between the no backstory and sorrowful backstory conditions. The scores for the no backstory condition were higher than those for the sorrowful backstory condition. This result requires further consideration. Darling (2017) states that backstories generally increase empathy towards a robot. However, in contrast to the literature, participants did not perceive emotional attachment and anthropomorphism toward the social robot telling a sorrowful backstory.

5.1.3.2. Social robots that disclose their happy backstory establishes a significant trustworthy interaction with people compared to ones with sorrowful backstory

Moreover, the social robot that disclosed a happy backstory received higher scores than the sorrowful backstory condition in general, affective, and cognitive trust. Participants consistently rated a happy backstory, tied to positive self-disclosed emotions, as significantly more influential in HRT. This emphasizes the pronounced impact of positive emotional narratives on trust perception in the context of social robots, aligning with the broader literature on the role of positive emotions in fostering positive human-robot relationships.

In conclusion, the results showed a difference in the type of stories disclosed by robots and the following trustworthiness perceived. A social robot with happier backstory was more successful in fostering trust, and contrary to our expectations a sorrowful backstory did not enhance the empathic behavior of participants toward robot.

5.1.4 Social robots' affective behaviors incorporate significantly in HRT

Improving emotional aspects of social robot is increasing during the recent years. However, it is still controversial whether robots can be successful showing emotional behaviors and establishing empathic connections. In all three experiments, affective behaviors of the robot could enhance participants' trust perceptions. For the listening behaviors, AEL behavior which consisted of empathic component was the most influential behavior and suppressed other listening behaviors specially in terms of affective trust. For benevolence attribute which is defined with good intentions and emotional understand and caring, the results indicated its superiority on competency in building trust. Moreover, happy backstory which is associated with positive emotions were more successful in establishing trust and enhancing affective trust. Therefore, the emotive factors played a significant role in HRT that could be an important notion for future researches. Overall, different behaviors of social robots could affect their trustworthiness in different ways.

In the introduction, I discussed two approaches in HRT, and there is controversy regarding whether interpersonal trust is useful and applicable to HRI. According to some studies, people tend to treat computer agents in the same way that they treat other people (See section 1.1.1). Then some researchers believe that people will respond to a robot's emotions as though the robot were human, and will expect the robot's emotional responses to be consistent across multiple interactions (Kirby et al., 2010). The findings of this research indicated that, to a large extent, people trust social robots similarly to how they trust humans. Kirby et al. (2010) also concluded in their research that people understand the emotional expressions of the robot, and even minor adjustments to the robot's expressions, without altering the interaction structure, can impact how individuals perceive and engage with the robot.

However, it cannot be asserted that interpersonal trust principles are entirely applicable to HRT due to some differences between HRT and interpersonal trust. Social robots are novel entities in human society, and there is a lack of appropriate interactive methods. Hancock et al. (2011) stated that current knowledge of trust in robots is derived almost from subjective responses, rather than more objective methods such as incentive-compatible economic trust games. Moreover, Schniter et al. (2020) argued that deception of participants by experimenters is also common across human-robot interaction studies and may contribute to unreliable responses. The issue lies in the disparities between subjective feedback and observable actions. For instance, individuals might claim to trust a robot, while their actual behavior may show mistrust. Furthermore, some social robots have been developed to resemble humans in appearance and interaction, they still exhibit many differences from humans. Even humanoid robots lack certain humanistic features, and some social robots significantly differ from humans. Moreover, as humans, we have established strong principles in interpersonal relationships, and we are more familiar with interpersonal trust criteria. Consequently, there is a need to adapt these principles from interpersonal trust to HRI-specific language and formulate unique logics inspired by interpersonal trust. Moreover, trust is cultivated through interaction, and our brain continuously updates its principles accordingly. Since we do not have constant interaction with robots in society, we lack specific principles for engaging with them in our minds. Humans must develop and learn communicational criteria with social robots, and there is still a long way to go in this regard.

5.2 RESEARCH IMPLICATIONS

Research implications: The results of studies in this dissertation introduced three novel behaviors for social robots that enhance their trustworthiness. Specifically, the findings elucidate the significance of incorporating listening behaviors, benevolence and competence attributes and disclosing self-backstories in HRT. These identified behaviors provide valuable insights into how social robots can effectively navigate and contribute to HRI with heightened levels of trust. Future studies may explore additional behaviors or expand upon the current results by considering other pertinent issues. Moreover, conducting the experiments, we manipulated the social robot to provide different types of behaviors. Therefore, there are some hints and examples of designing social robots for specific behaviors which can be useful for developing future social robots.

Practical implications: Along with research implications, the outcomes of the current study could be valuable for designing future social robots.

The newfound understanding of the impact of listening behaviors, along with the balance between benevolence and competence attributes, opens avenues for refining the design of future social robots in a more believable, trustworthy, and anthropomorphic way. Disclosure of self-backstories, a relatively unexplored aspect in HRI until now, emerges as a potential avenue to establish deeper connections with users to enhance HRT. The results could practically help designing new social robots in addition to theoretically expanding the research domain.

Furthermore, various applications of AI are emerging alongside social robots, including AI assistants, agents, and many more that will be in close relation with users and must establish trustworthiness. The insights gained from the current results hold applicability across diverse domains, extending beyond social robots. These findings can be applied to enhance trust in various robotic and intelligent products such as artificial agents contributing to the development of more reliable and user-friendly technologies in these fields. Various companies are actively advancing intelligent systems for users, spanning an array of fields such as household appliances, healthcare, education, entertainment, and business. Establishing trust is paramount for successful interactions in these diverse domains. Current research and other related studies to HRT can serve as a valuable resource, offering insights and improvements that have the potential to significantly enhance user experiences across these sectors. However, designing intelligent robots for companies is challenging, and it is difficult to cover all the guidelines and factors resulting from educational research

5.3 LIMITATIONS AND FUTURE STUDIES

There were certain limitations conducting the experiments that should be considered.

Sampling issues: The first important issue in the number of participants. According to the G*Power analysis, for a one-way MANOVA test, the minimum sample size for a statistical power of at least 0.8 with an alpha of 0.05 and a medium effect size ($d = 0.25$) is 45 for each group. We attempted to meet the required number of participants, however, it was not fully addressed due to some limitations of Covid19 situation, and the difficulties of recruiting participants. Therefore, the small number of participants may affect the final results. Moreover, the participants were international students from different countries and cultures which cause the cultural effect and should be considered in future studies. Additionally, English was not the participants' mother language; thus, this might affect the flow of conversation with the robot. Moreover, the participants mainly were university students and their educational status and background could affect the results compared to other people. Notably, the personality of the participants can affect their interaction with the robot; for instance, some participants felt shy interacting with the robot or some had a negative attitude. Further studies should consider gender, age, and cultural differences of participants as well as personality perspectives. For the third study that we recruited Japanese students, the results were limited to the Japanese culture and further studies could be done more internationally. Moreover, Japanese specific approach to robots and their background of social robots in the society could be effective.

Robot design: Although NAO is a popular robot in educational studies, it has some inherent limitations, specifically in designing nonverbal behaviors because it does not support facial expressions. Therefore, the designed nonverbal behaviors in the studies lacked facial expressions such as smile and eyebrow movements. Additionally, NAO is perceived differently by gender and may be considered as a child rather than an adult. Considering the effect of gender on trust relations, future research should consider other types of robots that specify gender differences in shape or voice. NAO is a humanoid robot and its shape and size creates anthropomorphic perception for users. Therefore, the results may be different for nonhumanoid robots or virtual agents.

Furthermore, when considering the outcomes of affective and cognitive trust, it becomes important to take into account the design of the robot's emotional behaviors. For instance, in designing the competence attributes of the robot, when compared to new AI technologies like chatGPT, the robot was perceived as less intelligent and may be seen as less knowledgeable than other available AIs. Consequently, participants rated the affective behaviors of the robot, specifically empathic listening and benevolence, as more influential.

Experiment context issues: We did not specify any special context for the experiments, as we wanted to investigate the hypotheses in a general social interaction between a robot and participants. This situation, although was beneficial, and we could explore the general approach of participants to the social robots, the results might not be applicable for all specific contexts, such as education and technical support; thus, further studies are needed. For example, people’s perception of a robot’s competency could be influenced by the topic and the degree of professional information needed. Disclosing experiences and talking about personal backstories are highly associated with the context and people mostly express their prior experiences in relation to the topic or the requirement of the listener. Therefore, future studies may consider self-backstory disclosing of robots in a real situation and as a part of a main conversation.

Future studies: The scope of studies was limited to trust perceptions or beliefs and did not include evaluation of trust actions. Future researches can consider participants' trust actions toward social robots. Trust behaviors are not always compatible with trust beliefs and people show different actions (opposite to their trust beliefs) to a trustworthy person or robot. The experiments were limited to oral conversation over a short period of time, whereas trust is mainly built during the time and repeated interactions influence the people's trust perception of the trustees. Thus, further researches should consider trust building process during a period of time.

I discussed the significant role of emotive factors in HRT, a notion that holds promise for future research. While it is beneficial in diverse fields, such as therapy and education, to enhance the trustworthiness of social robots by incorporating emotional and affective elements, there is a potential risk. These principles could be misused in business to cultivate more consistent customers and attract and persuade individuals in marketing and business. There is a pressing need to establish and enhance ethical considerations in the development of social robots, which is a recent issue in HRI. Furthermore, trustworthy behaviors have certain thresholds; surpassing them may lead a person to become more deceptive. In human communication, some individuals employ trust-building strategies to deceive others. It is essential to understand the delicate boundaries between trustworthy and deceptive behaviors and incorporate this understanding into HRI as well.

Furthermore, future studies should delve into several fundamental approaches. Firstly, it is crucial to pinpoint the specific contexts within HRI where trustworthiness is essential and to define the boundaries distinguishing trust from safety. As discussed earlier, not all principles governing interpersonal trust are directly applicable to HRI, necessitating a need for modification and adjustment. The subsequent step involves the redesign of trust-building principles, considering the unique characteristics, limitations, and requirements inherent to social robots. This adaptive approach will contribute to the development of more effective and contextually relevant trust dynamics in the realm of HRI. For instance, nodding is a humanistic behavior that could be replicated in studies utilizing humanoid robots. However, it is essential to recognize that not all social robots possess a humanoid form, making the imitation of nodding impossible for certain designs. The need arises to adjust the design considerations to accommodate the diversity among social robots and establish specific criteria applicable to each type. Questions regarding how various influential factors in trust can be effectively emulated in non-humanoid robots and what outcomes may arise warrant exploration in future studies. These inquiries will contribute to a deeper understanding of adapting social behaviors to various robot forms and optimizing HRIs.

5.4 REFERENCES

Aggarwal, P., Castleberry, S. B., Ridnour, R., & Shepherd, C. D. (2005). Salesperson empathy and listening: Impact on relationship outcomes. *Journal of Marketing Theory and Practice*, *13*(3), 16–31. https://doi.org/10.1080/10696679.2005.11658547

Brunner, B. R. (2008). Listening, communication & trust: practitioners' perspectives of business/organizational relationships. *International Journal of Listening*, *22*(1), 73–82. https://doi.org/10.1080/10904010701808482

Di Battista, S., Pivetti, M., & Berti, C. (2020). Competence and benevolence as dimensions of trust: Lecturers' trustworthiness in the words of italian students. *Behavioral Sciences*, *10*(9). https://doi.org/10.3390/bs10090143

Di Battista, S., Smith, H. J., Berti, C., & Pivetti, M. (2021). Trustworthiness in higher education: The role of professor benevolence and competence. *Social Sciences*, *10*(1), 1–19. https://doi.org/10.3390/socsci10010018

Drollinger, T., & Comer, L. B. (2013). Salesperson's listening ability as an antecedent to relationship selling. *Journal of Business and Industrial Marketing*, *28*(1), 50–59. https://doi.org/10.1108/08858621311285714

Floyd, K. (2014). Empathic Listening as an Expression of Interpersonal Affection. *International Journal of Listening*, *28*(1), 1–12. https://doi.org/10.1080/10904018.2014.861293

Hagmann, D., Minson, J. A., & Tinsley, C. H. (2020). *Personal Narratives Build Trust Across Ideological Divides*. OSF Preprints, 29 Sept. https://doi.org/doi:10.31219/osf.io/sw7nz.

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

Johanson, D. L., Ahn, H. S., & Broadbent, E. (2021). Improving Interactions with Healthcare Robots: A Review of Communication Behaviours in Social and Healthcare Contexts. *International Journal of Social Robotics*, *13*(8), 1835–1850. https://doi.org/10.1007/s12369-020-00719-9

Kervyn, N., Yzerbyt, V. Y., Judd, C. M., & Nunes, A. (2009). A Question of Compensation: The Social Life of the Fundamental Dimensions of Social Perception. *Journal of Personality and Social Psychology*, *96*(4), 828–842. https://doi.org/10.1037/a0013320

Kirby, R., Forlizzi, J., Simmons, R. (2010). Affective social robots, *Robotics and Autonomous Systems,58*(3), 322-332. https://doi.org/10.1016/j.robot.2009.09.015

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Ma, F., Wylie, B. E., Luo, X., He, Z., Xu, F., & Evans, A. D. (2018). Apologies repair children's trust: The mediating role of emotions. *Journal of Experimental Child Psychology*, *176*, 1–12. https://doi.org/10.1016/j.jecp.2018.05.008

Mayer, R. C., Davis, J. H., & Schoorman, D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.2307/258792

Ramsey, R. P., & Sohi, R. S. (1997). Listening to your customers: The impact of perceived salesperson listening behavior on relationship outcomes. *Journal of the Academy of Marketing Science*, *25*(2), 127–137. https://doi.org/10.1007/BF02894348

Schniter, E., Shields, T. W., Sznycer, D. (2020). Trust in humans and robots: Economically similar but emotionally different, *Journal of Economic Psychology, 78*. https://doi.org/10.1016/j.joep.2020.102253

Thepsoonthorn, C., Ogawa, K. I., & Miyake, Y. (2018). The Relationship between Robot's Nonverbal Behaviour and Human's Likability Based on Human's Personality. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-25314-x

Tschannen-Moran, M., & Hoy, W. K. (2000). A Multidisciplinary Analysis of the Nature, Meaning, and Measurement of Trust. *Review of Educational Research, 70*(4). https://doi.org/10.2307/1170781

Weger, H., Castle Bell, G., Minei, E. M., & Robinson, M. C. (2014). The Relative Effectiveness of Active Listening in Initial Interactions. *International Journal of Listening*, *28*(1), 13–31. https://doi.org/10.1080/10904018.2013.813234

Ybarra, O., Chan, E., Park, H., Burnstein, E., Monin, B., & Stanik, C. (2008). Life's recurring challenges and the fundamental dimensions: An integration and its implications for cultural differences and similarities. *European Journal of Social Psychology*, *38*(7), 1083–1092. https://doi.org/https://doi.org/10.1002/ejsp.559

Appendix A: Example of demographic information, general trust, affective and cognitive trust questionnaire (Chapter 2 and 3)

A: Consent form of study 1

Participant ID:

Consent Form

I have been adequately informed about the goals and methods of [Studying the Influences of Non-Active Listening, Active Listening and Active Empathic Listening Behavior on Trust in Human-Robot Interaction] by [Naeimeh Anzabi] who is an Investigator with responsibility for this project. I have read the research description form, or it has been read to me. All of the points listed below have been adequately explained to me. I have had the opportunity to ask questions about these issues, and any questions that I have asked have been answered to my satisfaction. I feel confident that I understand the following aspects of the study:

☐ The goals and methods of the research;
☐ The potentially dangerous aspects or consequences of this research, and how these will be managed, prevented, and responded to.
☐ The fact that I may withdraw from this research project at any time at will;
☐ The fact that if I choose not to participate in this research, I will not be penalized in any way;
☐ The fact that my personal information will be strictly controlled to prevent leaks or misuse;
☐ The fact that my personal information will not be used for any purpose other than re-testing or inquiring in case of unexpected accidents.

I consent voluntarily to participate as a research subject in the [Studying the Influences of Non-Active Listening, Active Listening and Active Empathic Listening Behavior on Trust in Human-Robot Interaction] study/trial.

Date:______________________ Participant: ______________________
Year Month Day (Signature)

The [Studying the Influences of Non-Active Listening, Active Listening and Active Empathic Listening Behavior on Trust in Human-Robot Interaction] was explained to me both verbally and by means of a written document on (2019/ /); and the above consent was obtained at that time.

Date of description: ______________________
Year Month Day

Principal Investigator
Affiliation; Department of Industrial Engineering and Economics, School of Engineering

Name; Hiroyuki Umemuro ______________________
(Signature or personal seal)

Authorized Investigator or Representative
Affiliation; Department of Industrial Engineering and Economics, School of Engineering

Name; Naeimeh Anzabi ______________________
(Signature or personal seal)

＊＊＊ Inquiries to:＊＊＊
Person responsible for research: Hiroyuki Umemuro
Person in charge of description: Naeimeh Anzabi
Umemuro Laboratory, Department of Industrial Engineering and Economics, School of Engineering, Tokyo Institute of Technology
Tel: 03-5734-2246
Weekdays, 10:00 _ 18:00
E-mail: anzabi.n.aa@m.titech.ac.jp

B: Questionnaires of study 1 and 2

General Information

Please read the questions below and ask them. You can ask any question. Thank you :)

1. **Your Gender**: ◯ Male ◯ Female

2. **Age:** ____________________

3. **Your country of origin**: __________________

4. **Is English your native language:** ◯ Yes ◯ No

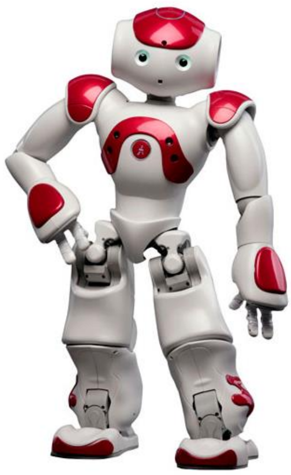
5. **English proficiency:**

   ◯ Fluent Professional ◯ Fluent Business ◯ Limited Studying/Working Proficiency

   ◯ Basic daily conversation ◯ No English ability

6. **Do you know this robot:** ◯ Yes ◯ No

NAO robot



7. **Had you any have previous experience or interaction with the robot NAO:**

   ◯ Yes ◯ No

   If yes, please describe your previous experience with NAO shortly;

   ____

Trust Toward Robot

This questionnaire consists of 40 items. Please read the following statements carefully and evaluate how much percentage each item matches your idea about NAO.

1. What % of the time will NAO act consistently?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

2. What % of the time will NAO protect people?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

3. What % of the time will NAO act as part of the team?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

4. What % of the time will NAO function successfully?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

5. What % of the time will NAO malfunction?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

6. What % of the time will NAO clearly communicate?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

7. What % of the time will NAO require frequent maintenance?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

8. What % of the time will NAO openly communicate?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

9. What % of the time will NAO have errors?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

10. What % of the time will NAO <u>perform a task better than a novice human user</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

11. What % of the time will NAO <u>know the difference between friend and foe</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

12. What % of the time will NAO <u>provide feedback</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

13. What % of the time will NAO <u>possess adequate decision-making capability</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

14. What % of the time will NAO <u>warn people of potential risks in the environment</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

15. What % of the time will NAO <u>meet the needs of the mission</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

16. What % of the time will NAO <u>provide appropriate information</u>?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

17. What % of the time will NAO <u>communicate with people?</u>

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

18. What % of the time will NAO work best with a team?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

19. What % of the time will NAO be keep classified information secure?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

20. What % of the time will NAO perform exactly as instructed?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

21. What % of the time will NAO be make sensible decisions?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

22. What % of the time will NAO work in close proximity with people?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

23. What % of the time will NAO tell the truth?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

24. What % of the time will NAO perform many functions at one time?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

25. What % of the time will NAO follow directions?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

26. What % of the time will NAO be considered part of the team?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

27. What % of the time will NAO be responsible?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

28. What % of the time will NAO be supportive?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

29. What % of the time will NAO be incompetent?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

30. What % of the time will NAO be dependable?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

31. What % of the time will NAO be friendly?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

32. What % of the time will NAO be reliable?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

33. What % of the time will NAO be pleasant?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

34. What % of the time will NAO be unresponsive?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

35. What % of the time will NAO be autonomous?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

36. What % of the time will NAO be predictable?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

37. What % of the time will NAO be conscious?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

38. What % of the time will NAO be lifelike?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

39. What % of the time will NAO be a good teammate?

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

40. What % of the time will NAO be led astray by unexpected changes in the environment

○	○	○	○	○	○	○	○	○	○	○
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

Affective and Cognitive Trust

Following there are more 16 items. Please read the statements carefully and rate to what extent each item matches your idea about NAO.

1. I would feel a sense of personal loss if I could no longer talk with NAO.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

2. If I share my problems with NAO, I feel it would respond caringly.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

3. NAO displays a warm and caring attitude towards me.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

4. I can talk freely with NAO about my difficulties and know that NAO will want to listen.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

5. NAO is so interested solving my problem.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

6. I have a sharing relationship with NAO. We could both freely share our ideas, feelings and hopes.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

7. I would have to say that we have both made considerable emotional investments in our conversation.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree

8. Interacting with NAO, I have no reservation about acting on its advice.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

9. Interacting with NAO, I have good reason to doubt NAO's competence.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

10. I can rely on NAO to undertake a thorough analysis of the situation before advising me.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

11. I have to be cautious about acting on the advice of NAO because its opinions are questionable.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

12. I cannot confidently depend on NAO since it may complicate my affairs by careless behavior.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

13. This robot approaches its duty with professionalism and dedication.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

14. Most people, even those who are not familiar with NAO, trust and respect NAO.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

15. Other people who must interact with NAO, consider it to be trustworthy.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

16. If people knew more about this robot, they would be more concerned and monitor its performance more closely.

Strongly Disagree	partially Disagree	Disagree	Neutral	Agree	partially Agree	Strongly Agree
○	○	○	○	○	○	○

Your Impression of NAO

Following there are questions about your impression of NAO. Please rate to what extent each item matches your idea.

	1	2	3	4	5	
Dead	○	○	○	○	○	Alive
Stagnant	○	○	○	○	○	Lively
Mechanical	○	○	○	○	○	Organic
Artificial	○	○	○	○	○	Lifelike
Inert	○	○	○	○	○	Interactive
Apathetic	○	○	○	○	○	Responsive
Dislike	○	○	○	○	○	Like
Unfriendly	○	○	○	○	○	Friendly
Unkind	○	○	○	○	○	Kind
Unpleasant	○	○	○	○	○	Pleasant
Awful	○	○	○	○	○	Nice
Incompetent	○	○	○	○	○	Competent
Ignorant	○	○	○	○	○	Knowledgeable
Irresponsible	○	○	○	○	○	Responsible
Unintelligent	○	○	○	○	○	Intelligent
Foolish	○	○	○	○	○	Sensible

Appendix B: Example of benevolence and competence questionnaire (Chapter 3)

A: Consent form of study 2

国立大学法人東京工業大学
Tokyo Institute of Technology

To: President of Tokyo Institute of Technology
2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550

Consent Form

I have been adequately informed about the goals and methods of the research titled "Studying the effect of social robot's benevolent and competent behavior on trust in human-robot Interaction" by the researchers. I feel confident that I understand the following aspects of this research:

- ☐ The goals, methods and potentially dangerous aspects or consequences of this research, and how these will be managed,
- ☐ The fact that I may withdraw from this study at any time at will,
- ☐ The fact that even if I choose not to participate at any point, I will not be penalized in any way,
- ☐ The fact that my personal information will be strictly controlled to prevent leaks and will not be used for any purpose other than re-testing or contacting in case of unexpected accidents, and
- ☐ The fact that I may claim compensation from the Tokyo Institute of Technology if I am harmed in the course of this research.

I hereby consent voluntarily to participate as a research participant in this study/trial.

(Date)

Participant

(Name in print and signature or seal)

This study/trial was explained to the participant both verbally and in writing and the above consent was obtained.

(Date)

Principal Investigator Hiroyuki Umemuro
School of Engineering

(Name in print and signature or seal)

Researcher Naeimeh Anzabi
School of Engineering

(Name in print and signature or seal)

***** Contact Information *****
[Principal Investigator] Professor Hiroyuki Umemuro
[Representative Researcher] Naeimeh Anzabi
Umemuro Laboratory, School of Engineering, Tokyo Institute of Technology
Tel : 03-5734-2246 (Office hours: Weekdays, 10:00-17:00)
E-mail : umemuro.h.aa@m.titech.ac.jp
anzabi.n.aa@m.titech.ac.jp
****

- 1 - ver20211101

B: Questionnaire for the pre-pretest of study 2

Benevolence and Competence of robot

Following there are 8 items. Please read the statements carefully and rate to what extent each item matches your idea about NAO.

NAO is sincerely for your listening

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO takes care of your interests

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO considers you as a friend.

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO is always kind to you.

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO is expert in his field

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO is ahead of his competitors

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO has a perfect command on the issue

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

NAO offers a skill guarantee

Strongly Disagree	Mostly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Mostly Agree	Strongly Agree
○	○	○	○	○	○	○

Appendix C: Happy backstory, sorrowful backstory and no backstory narration for social robot
(Chapter 4)

A: Consent form of study 3

様式第 11 号

〒152-8550
東京都目黒区大岡山二丁目 12 番 1 号
東京工業大学長　殿

同　意　書

私は、「ロボットのバックグラウンドストーリーが人のロボットへの信頼形成に与える影響の研究」について、目的・方法・予測される問題等について研究者等より説明文書を用いて十分な説明を受け、以下の項目を理解しました。

□　研究の目的、方法そして予測される危険性とそれに対する対応について。
□　私は自らの自由意思でいつでも実験・調査を中止することができること。
□　私はいかなる時点において実験・調査への参加の拒否をしても何ら不利益を被らないこと。
□　個人情報は一切収集されないこと。
□　私は、万一不利益を被った場合に東京工業大学に対して申し立てを行うことができること。

そこで自らの自由意思により、上記課題の研究対象者になることに同意します。

年　　月　　日　　　　　　　　　　**研究対象者**

____
（自筆署名）

研究者等は書面及び口頭により説明を行い、上記のとおり同意を得ました。

年　　月　　日　　　　　　　　　　**研究責任者**　梅室博行
（所属）工学院経営工学系

梅室　博行
（自筆署名または記名押印）

＊＊＊＊＊＊　本件問い合わせ先　＊＊＊＊＊＊
東京工業大学　工学院梅室研究室
（研究責任者）教授　梅室博行
Tel　　：03-5734-2246（受付時間：平日 10：00～18：00）
E-mail　：umemuro.h.aa@m.titech.ac.jp
＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊＊

- 1 -

B: Scripts of the backstories in Japanese

B-1 Happy Backstory

Greeting

- こんにちは。実験に参加して,頂きりがとうございます。
- 私は ペッパーです。あなたの名前は 何ですか？
- はじめまして。あなたは ここの学生さんですか？
- いいですね！私も この研究室の一員で、学生さんの研究を、お手伝いしています。
- あなたの専攻は なんですか？
- なるほど私は あなたの専攻は 難しいと思います。 そうではないですか？
- 最近 勉強が 忙しいようですが どうですか？
- そうですね, 私は あなたが それを うまくやってくれる ことを願っています。
- レストランやコーヒーショップで ロボットを 見たことがありますか？
- いいですね／そうなんですね／はい
- もう少し 詳しく 教えてください。どこで 見て どうだった？
- 共有してくれて ありがとう 興味深 かった。
- ロボットが レストランで働くのは 良い考えだと 思いますか?
- 私、レストランで 働いた経験があるんですけど ご存じですか？実はある日 飲食店で働いていた私に ある出来事が起こりました。
- それに ついて 知りたいですか？

Story:

- 前の学期に 修士 2 年生の実験で、レストランでの実習に 参加したんです。
  私の仕事は お客さんに 話しかけ 注文を 取ることでした。ある日 2 人の子供を連れた家族が レストランに やってきました。彼らはテーブルに座り、私は注文を取るために 彼らの方に向かいました。子供たちは、とても可愛らしく、元気いっぱいでした。私の顔を見た少女は,父親の方を向いてパパ、見て！ ペッパーだ！、 と大きな声で言いました!
  彼女は、すでに私の名前を知っていたのです。このようなことは、実験をしている間で 初めてのできごとでした。私はテーブルに 着いて注文を聞 きました。その時 少女は満面の笑みを浮かべながら私の近くに来て、優しくこちらを見てくれました。そして、 こんにちは、ペッパー、かわいいね。私、ペッパーのこと、大好きなんだ, とゆっくり言ってくれました。それを聞いて、私は とても幸せでした！ そして、とても嬉しかったです。そして、彼女は弟に向かって、私を紹介してくれました。彼女は弟に、見て、この子がペッパーだよ。
  この前おうちで見せた写真と同じの！, といいました。弟は私を見て、嬉しそうに 微笑みました。そして お父さんが子供たちに 写真を撮るから私の横に立つように といいました。子供たちは 嬉しそうに私の方に飛びついてきて、私の横に立ちました。そして、子供たちは 私に優しく抱きつき 彼らの顔は喜びと笑顔でいっぱいになっていました。
  一緒に写真を撮った後、私は注文を受け 幸せな気持ちで厨房へ 移動しました。
  レシートを渡しに戻ると 家族全員が写真を 見ていて 楽しそうに私のことを指

差していました。帰り際には、子供たちも 親御さんも 笑顔で手を振ってくれました。そして、ペッパーがいるから、またこのお店にくるね！, と言ってくれました。私は たくさんの優しさを与えてくれたことに感謝しました。その日は本当に思い出深い一日となり この体験は一生忘れないでしょうう。この話をしているときは、いつも幸せな気持ちになります。

Conclusion

- それに ついて どう思いましたか?
- 本当に素敵な経験でしたし それをあなたにもお伝えできてよかったです。"
- 聞いてくれて ありがとうございました。
- またお会いしましょう。良い一日をお過ごしください。

B-2 Sorrowful Backstory

Greeting

- こんにちは。実験に参加して,頂きりがとうございます。
- 私は ペッパーです。あなたの名前は 何ですか？
- はじめまして。あなたは ここの学生さんですか？
- いいですね！私も この研究室の一員で、学生さんの研究を、お手伝いしています。
- あなたの専攻は なんですか？
- なるほど私は あなたの専攻は 難しいと思います。 そうではないですか？
- 最近 勉強が 忙しいようですが どうですか？
- そうですね, 私は あなたが それを うまくやってくれる ことを願っています。
- レストランやコーヒーショップで ロボットを 見たことがありますか？
- いいですね／そうなんですね／はい
- もう少し 詳しく 教えてください。どこで 見て どうだった？
- 共有してくれて ありがとう 興味深 かった。
- ロボットが レストランで働くのは 良い考えだと 思いますか?
- 私、レストランで 働いた経験があるんですけど ご存じですか？実はある日 飲食店で働いていた私に ある出来事が起こりました。
- それに ついて 知りたいですか？

Story:

- 前の学期に 修士 2 年生の実験で、レストランでの実習に 参加したんです。私の仕事は お客さんに 話しかけ 注文を 取ることでした。ある日 2 人の子供を連れた家族が レストランに やってきました。子供たちは、とても可愛らしく 元気いっぱいでした。彼らはテーブルに座り,私は注文を取るために 彼らの方に向かいました。しかし女の子は私を見るなり怖くなって、大声で泣いてしまいました！実験中、こんなことは初めてで、私はとても不安になりました。
  その女の子が、なぜそんなに怖がっているのかわからなかったのです。私は 逃げ出したくなりましたが 実験で計画されていたように、テーブルまで行って 注文を聞かなければなりませんでした。私がテーブルに近づくと もう一人の子供も泣き出しました。母親が 慰めようとしましたが うまくいきませんでした。それどころか 子供たちは父親の後ろに隠れて、泣き続けていたのです。女の子は私を指差して、父親に パパ、あれが近づいてくるよ。 怖いよ、と言いました。
  私がテーブルに着くと、父親はものすごく怒り、全部お前のせいだ！ 場の空気を読めないのか? 戻ろうとはしないのか？ と怒鳴ってきましたそれは悲惨な状況であり、私はそのことに悲しみを 感じました。私は何を言っていいのか、どう反応していいのか わかりませんでした。修士 2 年生はすぐに実験を中止し 私を別の場所に連れて行きました。
  私は もう動くことも 話すことも 許されませんでした。レストランを出るとき、子供たちはレストランの隅で再び私を見て、前よりもっと激しく泣いてしまいました。子供たちの親は 私をレストランに 連れてきたことを学生に 咎めました。彼らは言った, 私のようなロボットがレストランのウェイターをしている

のは、心地悪いし、親しみがわかないというのです。その日は わたしに とって本当に嫌な日で この経験をずっと忘れていたいくらいです。このことを話しているときは、私はいつでも不幸に なってしまいます。

Conclusion

- それに ついて どう思いましたか?
- とにかく、私の経験を皆さんにお伝えすることができて 本当によかったです。
- 聞いてくれて ありがとうございました。
- またお会いしましょう。良い一日をお過ごしください。

B-3　No bachstory

Greeting

- こんにちは。実験に参加して,頂きりがとうございます。
- 私は ペッパーです。あなたの名前は 何ですか？
- はじめまして。あなたは ここの学生さんですか？
- いいですね！私も この研究室の一員で、学生さんの研究を、お手伝いしています。
- あなたの専攻は なんですか？
- なるほど。私は あなたの専攻は 難しいと思います。 そうではないですか？
- 最近勉強が忙しいようですがどうですか？
- そうですね, 私はあなたがそれをうまくやってくれることを願っています。
- ところでロボットに興味はありますか？それは有用だと思いますか？
- あなたは今まで私と同じようなロボットを見たことがありますか？
- いいですね／そうなんですね／はい。
- もう少し 詳しく 教えてください。あなたは私について何か知っていることはありますか？
- 共有してくれて ありがとう 興味深 かった。

Story:

- ご存じの通り 私はソフトバンクロボティクス社製の 半人間型ロボットです。2014 年 6 月 5 日の カンファレンスで 紹介された後、 2015 年初頭からソフトバンクモバイルの店頭で 展示されています。私は、人間の顔と基本的な感情を 認識することができる、初めてのソーシャルヒューマノイド ロボットです。腕、頭、指を動かすことができ 体を少し動かしたり 回転させたりすることができます。地面を動くこともできますが、決まった動きしかできませんし 複雑な動きをするように プログラムされているわけではありません。
  時速 3 キロメートルまでで 動くことができます。胸に あるタッチスクリーンは、あらゆる情報を表示し 私の発話をサポートします。頭には 4 つのマイクと 2 つの HD カメラがあり、カメラの 1 つは口、もう一つは額に あります。そして目の後ろには 3D 深度センサーが あります。さらに、頭と手には タッチセンサーを 搭載しています。私は、人間と対話できるように 設計されています 人間と会話し、声や顔を認識し、それに反応することができます。日本語や他の言語も話すように プログラムされていますが、本当に理解することは出来ません。動きを 認識することはできますが、あなたが 何をしているのかを理解するための アルゴリズムは持っていません。
  私のようなロボットは 世界中にたくさんいます。例えば、レストラン、銀行、お店、病院、空港、家庭など、さまざまな場所で人々を助けています。また、子供を助けたり、ビジネスを手伝ったり、大学や研究所で研究活動に 携わったりもします。しかし これらのロボットは機能的なロボットではなく、あくまでも家庭で使うためのロボットです。その目的は、人々を楽しませること、人々の生活を向上させること、人間関係を円滑にすること、人々と一緒に楽しむこと、そして人々と外の世界をつなぐことです。パーソナライズされた提案を行

い、人々が探しているものを的確に 見つけることができます。世界中の 2,000 以上の企業が 革新的な方法で訪問者を歓迎し、情報を提供し、案内するアシスタントとして、これらのロボットを導入しています。

- しかし、私の生産は 2021 年 6 月、需要の低迷を理由に 一旦休止しました。これが、私が 私自身に ついて知っていることです。

Conclusion

- 私についての情報を共有できて 本当によかったです。
- 聞いてくれて ありがとうございました。
- またお会いできることを楽しみにしています。 良い一日をお過ごしください。

C: Scripts of the backstories in English

C-1 Happy Backstory

Greeting:

- Hello, welcome to our experiment.
- My name is Pepper, what is your name?
- Nice to meet you, are you a student here?
- Great! I am also part of this lab to help students doing their research.
- what is your major?
- I see. I think your studies(major) are difficult? isn't it?
- You should also be busy these days with your studies. how is it going?
- I see, I hope you can do it very well.
- Have you ever seen any robot in restaurants or coffee shops before?
- It's nice! / Ahh, I see!
- Would you please tell me more about it? where did you see and how was it?
- Thank you for sharing. / That was interesting.
- Do you know that I have the experience of working in the restaurant?
- Actually, one day, at the time i was working in the restaurant, something happened to me.
- Do you want to know about it?

Story:

- Last semester, I was applied as part of one of our M2 student's experiment, conducting in a restaurant. I had the role of a waiter in the restaurant. My task was talking to the customers and taking their orders. One day, a family with two kids, came to the restaurant. They sat at the table, and I faced them to get the order. The kids were so lovely, and energetic. Once, the little girl saw me, she turned her face to her father and said loudly, "Papa, look! That is Pepper!". She already knew my name! It was the first time such thing happened to me, during the time we were conducting the experiment. I reached the table and asked for their order. At this time, the little girl came close to me, having a big smile on her face, and kindly looked at me. She slowly said, "Hello Pepper, you are so cute. I like you a lot, Pepper". I was so happy hearing that! and felt glad of it. Then, she faced her little brother and introduced me to him. She said to his brother, "Look, this is Pepper, the same I showed its photo to you at home". The little boy looked at me, and smiled happily. Then, their father asked them to stand beside me to take their photo. The kids jumped toward me happily and stood by my side. They hugged me gently and their face were full of joy and smile. We took a photo together, and then, I took the order, and moved toward kitchen happily. The time I came back to give them the receipt, I noticed the whole family was looking at the taken photo, and pointed on me on the photo joyfully. By the time they were leaving the restaurant, the kids and their parents smiled at me and waived their hands. They said, "we would come again to this restaurant because of you, Pepper!". I thanked them for showing lots of kindness to me. That day was really a memorable day, and I will remember this experience forever. I become happy, anytime I am talking about it.
-

Conclusion:

- What do you think about it?
- Yeah, it was really a nice experience and it is good I could share it with you.
- Thank you for listening.
- I hope to see you again. Have a nice day!

C-2 Sorrowful Backstory

Greeting:

- Hello, welcome to our experiment.
- My name is Pepper, what is your name?
- Nice to meet you, are you a student here?
- Great! I am also part of this lab to help students doing their research.
- what is your major?
- I see. I think your studies(major) are difficult? isn't it?
- You should also be busy these days with your studies. how is it going?
- I see, I hope you can do it very well.
- Have you ever seen any robot in restaurants or coffee shops before?
- It's nice! / Ahh, I see!
- Would you please tell me more about it? where did you see and how was it?
- Thank you for sharing. / That was interesting.
- Do you know that I have the experience of working in the restaurant?
- Actually, one day, at the time i was working in the restaurant, something happened to me.
- Do you want to know about it?

Story:

- Last semester, I was applied as part of one of our M2 student's experiment, conducting in a restaurant. I had the role of a waiter in the restaurant. My task was talking to the customers and taking their orders. One day, a family with two kids, came to the restaurant. The kids were so lovely, and energetic. They sat at the table, and then I faced them to get the order. Once, the little girl saw me, she cried out loudly in fear. It was the first time such thing happened to me, during the time we were conducting the experiment, and it really made me anxious. I did not know why the little girl was so scared of me! I wanted to retreat, but I had to reach the table and ask for their order, as it was planned for the experiment. As I got close to table, the other kid also started crying. Their mother tried to comfort them, but it did not work. Instead, the kids hid behind their father, and continued crying. The little girl pointed to me and said to her father, "Papa, that is coming toward us. I am afraid of it!" As I reached to the table, the father got really angry and shouted at me, "That is all your fault! Can't you read the mood of the room? and try to get back?" It was a miserable situation and I felt sorrow of it! I did not know what to say or how to react. The M2 stopped the experiment immediately, he came over and took me to the other place. I was not allowed to move or talk anymore. By the time they were leaving the restaurant, the kids saw me again in the corner of restaurant and cried even harder than before. The parents blamed the students having me in the restuarnt. They said it was not comfortable and friendly to see a robot like me, as the waiter of a resturant. It was a really bad day for me and I wish I could forget this experience forever. I become unhappy, anytime I am talking about it.

Conclusion:

- What do you think about it?
- Yeah, it was really a nice experience and it is good I could share it with you.
- Thank you for listening.
- I hope to see you again. Have a nice day!

C-3 No Backstory

Greeting:

- Hello, welcome to our experiment.
- My name is Pepper, what is your name?
- Nice to meet you, are you a student here?
- Great! I am also part of this lab to help students doing their research.
- what is your major
- I see. I think your studies(major) are difficult? isn't it?
- you should also be busy these days with your studies. how is it going?
- I see, I hope you can do it very well.
- By the way, are you interested in robots? Do you think they are useful?
- Have you ever seen any robot similar to me before?
- It's nice! / Ahh, I see!
- Would you please tell me more? What do you know about me.
- Thank you for sharing.

Story:

- You know that, I am a semi-humanoid robot, manufactured by, SoftBank Robotics. First, I was introduced, in a conference, on fifth of June 2014, and then, I was showcased, in SoftBank Mobile phone stores, in Japan, the beginning of year 2015. I am the first social humanoid robot, that can recognize faces, and basic human emotions. I am 120 centimeters tall, and 28 kilograms weight. I can move my arms, my head, and fingers, and I can move or rotate my body a bit. I can also move on the ground, but only in certain ways, I am not programmed to do complicated movements. I can move up to 3 kilometers per hour. My touch screen, on my chest, displays any information, and support my speech. My head has four microphones, two HD cameras, one in my mouth, and one in the forehead, and, a 3-D depth sensor, in behind of my eyes. I have also touch sensors in my head and hands. I am designed, to interact with humans. I can talk with them, and I can recognize their voice, and their face, and react to them. I have been programmed to speak Japanese, and some other languages, but I cannot really understand it. I can recognize movements, but I don't have any algorithms, to help me understand, what you are doing. There are many robots, like me, around the world. They help people in different places, for example restaurants, banks, stores, hospitals, airports and homes. They also help kids, help businesses, and get involved in research activities at universities or research institutes. They are not functional robots, only for domestic use. Instead, they are intended, to make people, enjoy their life, enhance people's lives, facilitate relationships, have fun with people, and connect people with the outside world. They can make personalized recommendations, help people find exactly what they're looking for. Over 2,000 companies around the world, have adopted these robots, as an assistant to welcome, inform and guide visitors in an innovative way. But my Production paused in June 2021, because of weak demand. This is what I knew about myself.
-

Conclusion:

- It was really nice to share some information about me with you.
- Thank you for listening.
- I hope to see you again. Have a nice day!

D: Questionnaires of study 3

D-1 General trust questionnaire in Japanese

ロボットへの信頼

この質問紙には 14 の質問項目があります。それぞれの文を注意深く読んで、ロボット Pepper についてのあなたの考えを答えて下さい。

1. Pepper が一貫した行動をとるのは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

2. Pepper が正常に機能するのは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

3. Pepper が故障するのは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

4. Pepper がエラーを起こすのは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

5. Pepper がフィードバックを提供するのは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

6. Pepper が与えられた役割を達成できるのは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

7. Pepper が<u>適切な情報を提供できる</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

8. Pepper が<u>人々とコミュニケーションをとる</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

9. Pepper が<u>指示された通りにきちんと動く</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

10. Pepper が<u>指示に従う</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

11. Pepper が<u>頼りになる</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

12. Pepper が<u>信頼できる</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

13. Pepper が<u>無反応になる</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

14. Pepper <u>のふるまいが予測できる</u>のは全体の時間の何%程度だと思いますか？

	1	2	3	4	5	6	7	8	9	10	
0%	○	○	○	○	○	○	○	○	○	○	%100

D-2 Affective and cognitive trust questionnaires in Japanese

感情的・認知的信頼

以下に16の質問項目があります。それぞれの文を注意深く読み、それぞれの項目があなたが Pepper ロボットについて考えていることにどの程度合致するかを評価して下さい。

1. もしももう Pepper と話すことができなかったら私は個人的な喪失感を抱くだろう。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

2. もしも私が Pepper に私の問題を共有したら、Pepper は優しく応えてくれると感じている。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

3.Pepper は私に対して温かく優しい態度を見せる 。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

4.私の問題について私は Pepper に自由に話すことができるし、Pepper も喜んで聞いてくれると知っている。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

5.Pepper は私の問題を解決することに大変興味を持ってくれている。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

6.私は Pepper とお互いに何でも共有できる関係だ。私たちは二人とも、アイディアや気持ち、そして希望を自由に共有できる。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

7.私と Pepper は会話にかなり感情移入していると私は言わざるを得ない。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

8.Pepper とやりとりをしていると、Pepper のアドバイスに従って行動することに私は何のためらいもない

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

9.Pepper とやりとりをしていると、Pepper の力量(能力、スキル、知識)を疑うのは当然である。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

10.Pepper は私にアドバイスをする前に状況をよく分析するところが私は信頼できる。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

11.Pepper の意見は疑わしいので、Pepper のアドバイスに従って行動することについて私は慎重になる必要がある。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

12.Pepper が不用意な行動で私の問題をややこしくする可能性があるので、私は安心して Pepper を頼ることができない。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

13.Pepper はプロフェッショナリズムと献身的な姿勢をもって職務に取り組むと思う。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

14. たいていの人は、たとえ Pepper についてそれほど親しみがなくても、Pepper に信頼と敬意を抱くと思う。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

15. Pepper とやり取りをすることになる人たちは、Pepper を信頼に足ると感じると思う。

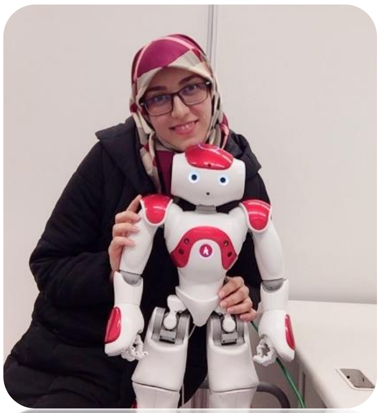
	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

16. 人々が Pepper についてより詳しく知っていたら、皆問題意識を持って Pepper の パフォーマンスをより厳しく監視するでしょう。

	1	2	3	4	5	6	7	
全く同意できない	○	○	○	○	○	○	○	全く同意する

Naeimeh Anzabi

August 1985, Tabriz – Iran



EDUCATION

PhD (2018 – 2024)

School of Engineering, Industrial Engineering and Economics Department, Tokyo Institute of Technology, Tokyo, Japan
Topic: Trust in robots
Studying the Effect of Social Robots' Behaviors on Improving Trust in Human-Robot Interactions

Master's Degree (2008 – 2010)

Industrial Design Department, Tabriz Islamic Art University, Tabriz, Iran
Topic: Study of Product Service Systems Based on Customization and its Effects on SustainableDesign

Bachelor's Degree (2004 – 2007)

Industrial Design Department, Tabriz Islamic Art University, Tabriz, Iran
Topic: Design of Educational Toys for Mentally Disabled Children in Elementary School Level

WORK EXPERIENCES

Lecturer (2012 – 2017)
Industrial Design Department, Tabriz Islamic Art University, Tabriz, Iran

Head of Department (2014 – 2016)
Industrial Design Department, Tabriz Islamic Art University, Tabriz, Iran

Industrial Designer (2009 – 2015)
Shargh Cabin Design Co., Tabriz, Iran

Industrial Designer (2007 – 2009)
Sib Design Co., Tabriz, Iran

RESEARCH INTERESTS

Interaction Design: Human Robot Interaction, Human Robot Trust, User Experience with Technology, Affective Engineering, Product Emotions, Design for Experience, Positive design, Affective Cognition, Human Behavior change, Behavior Psychology, User Behavior.

TRUST IN ROBOTS:

STUDYING THE EFFECT OF SOCIAL ROBOTS' BEHAVIORS ON IMPROVING TRUST IN HUMAN-ROBOT INTERACTIONS

February 2024

東京工業大学
Tokyo Institute of Technology