Universiteit Leiden

Master Computer Science

The effect of robot errors on children in the context of education

Name:	Katerina Tsiftsi
Student ID:	s2898276
Date:	[30/08/2023]
Specialisation:	Data Science
1st supervisor:	prof. Joost Broekens
2nd supervisor:	dr. Mike Ligthart
2nd reader:	prof. Peter Van Der Putten

Master's Thesis in Computer Science

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

This master thesis focus is a social experiment conducted to determine to what extent robot errors affect children in the context of education. A total of 60 children of age 10-12 from a public school in Leiden participated in the experiments in which they interacted with a Nao robot two by two. During the study, the robot simulated some speech recognition, voice localization and gaze errors based on a literature review conducted, which lead to the robot ignoring one of the two participants during the interaction. The human-robot interaction consisted of an interactive vocabulary game, where Nao acted as a peer to the children, playing the game together. The robot recognizes it is faulty and explains its behaviour either before the experiment or after the completion of a questionnaire that follows the experiments. The effect on children based on these two conditions is examined, i.e. how children are affected by being ignored by a robot and if the moment the robot explains its errors matter in how it is perceived by them. Apart from the questionnaire, the affective outcomes of the study are further examined with video recordings that are used within a framework to estimate head direction. Overall, neither the ignored condition or the moment of explanation one seem to have a big significance on children. Being ignored only slightly affects the closeness to the robot, where we observed children being ignored ranking less close to it. Furthermore, the moment of explanation affects the trustworthiness of the robot, where it is generally more trusted when it apologizes for its faults in advance. Lastly, in terms of children's attention, they are focused on either the robot or the screen of the game for the majority of the task irrespective of the experimental condition they are in.

Contents

1 Introduction

In recent years, the field of robotics has witnessed a paradigm shift in the way humans interact and perceive robots. With advancements in technology and design, robots are no longer regarded as mere machines but are increasingly seen as social entities capable of eliciting human-like responses. This phenomenon, known as anthropomorphism, has sparked significant interest among researchers, psychologists, and engineers, as understanding how humans attribute social characteristics to robots holds profound implications for human-robot interactions. The perception of robots as social beings not only influences our behavioral responses but also raises important questions about the ethical, psychological, and societal aspects surrounding human-robot relationships. This paper aims to delve into the realm of anthropomorphism, exploring the factors that contribute to the ascription of social qualities to robots and examining its impact on human cognition, emotions, and behaviors.

The interest in the use of social robots as means of pedagogical agents and the exploration of their impact in education has been rising amongst the human-robot interaction (HRI) community. These robots can be in the form of peers [4, 24, 35], where they take the role of a learning companion to a student, novices [7, 55, 22], where they allow the student to take the role of the instructor and tutors [38, 57, 17], where the robot tries to teach students a specific task or lesson. In [48], researchers examine student attitude towards social robots and showed that children generally have a positive attitude towards the inclusion of robots in a classroom. What happens however when there is a fault in the design of a social robot leading to malfunctions in its behaviour? More precisely, what is the impact on the child's perception of the robot and the child engagement after a robot's social behavior malfunctions?

Such scenario is not at all far from reality, as humans program the behaviour of social robots. This can lead to unexpected or incoherent behaviours [50] that are often interpreted as erroneous and the impact is becoming more present now that robots are perceived as social entities. Social entities are beings which exhibit human-like behaviour [8], being able to interact with someone in a two-way manner, employ thoughts and feelings or show signs of social awareness, support and autonomy. In this study, errors in social behaviour are simulated by a Softbank Robotics NAO v6 humanoid robot [1], that interacts as a peer to students. NAO interacts with two students by playing an educational game designed for the purpose of this study. It is a language learning task, testing the child's vocabulary. Whilst working together, the robot exhibits unequal behaviour towards the two children, by ignoring one of them and replying in a very detailed and personalised manner to the other. The goal is to create a sense of exclusion for some children in order to examine how this impacts their perception of the robot and the engagement and likeability of the task itself.

[1] SoftBank Robotics NAO `https://www.softbankrobotics.com/emea/en/nao`.

2 Motivation and related work

Technology and robots have advanced a lot in the past, however there are still many errors, failures or faults in their behaviour. A fault, according to [5] is defined as "a degraded state of ability which causes the behavior or service being performed by the system to deviate from the ideal, normal, or correct functionality". Faults can take various forms, from unexpected behaviours to actual failure of completing a given task and can vary in frequency and severity (functional or social). For the purpose of the study, we are interested in examining failures in communication. Such errors often impact social signals given by the robots [36] as well as how robots are perceived by humans. In some cases, robots that have erroneous behaviours are liked better [37] as they are perceived as more human-like and thus are more relatable. However in these cases, the interaction of the people with the robots are more conversational rather than of an educational manner as in our case and also the participants are mainly adults. In other cases, speech errors from robots resulted in a lowered perception of the sincerity of the robot [16]. We can thus see that errors in robot behaviour, and specifically when communication is present, affect human-robot interaction and human perception of the robot.

The errors we will simulate in this study, are errors related to "failure of communication". The robot will exhibit unequal treatment of the children on the basis of a malfunction related to the robot's hearing or its perception of the room. It will end up completely ignoring one child of the two iinn the interaction thus being indirectly impolite. The reason for this indirectness is that first of all you cannot allow a social robot in a classroom that is impolite in a straightforward manner (i.e. talking back to children or employing swear words). Second of all, the scenarios created need to be plausible, meaning that the errors in functioning can be present when programming AI agents. Such errors are present in [14] in the form of dialogue, where the observed errors are not that severe as the robot successfully communicates the errors and self-repairs. In other cases, such as in the study by Ragni et al [40], failures in communication lead to decreased overall human performance in a collaborative task.

Peer relationships in school play a significant role in the cognitive development [30], as well as the psychology of children. Peer relationships can be with other students, or the robot in our case, with which the children will be performing a task. Neglect also influences the development of a child [41] and how it perceives the difficulty of some tasks. It is therefore important when programming robots to consider the effect of these peer relationships in the well-being of all participants.

3 Research Questions and hypotheses

The context above gives rise to the following research questions:

RQ1 What kind of errors, failures and faults have the biggest impact on the sociability of the robot?

RQ2 What is the effect of these errors, failures or faults on children?

RQ3 Will a robot that is able to identify these faults and apologize in advance be perceived as more likeable to children?

The first research question **(H1)** is answered by a literature review conducted on the errors that can occur when programming robots for HRI. We hypothesize that such errors will have a negative effect on children **(H2)**, and will specifically negatively impact their perception of the robot. Furthermore, we investigate if there is a correlation between the likeability of the robot itself and the likeability of the task. Lastly, considering a condition where the robot is conscious of its malfunctions **(H3)**, we hypothesize that apologizing in advance will make it more approachable and therefore likeable to children. By apologizing, the robot will be considered more anthropomorphic and thus it will be easier for children to be more forgiving towards it.

4 Review of failures in Human-Robot Interactions

Human-Robot Interactions (HRI) often suffer from error situations that are either caused by robots or humans interacting with them. Giuliani et al [15] analyze erroneous interactions between people and robots and break them down into two broad categories: social norm violations and technical failures. A social norm violation occurs when a robot does not completely adhere to a certain social script whereas a technical failure is more related to the hardware or software design of the robot. Hardware failures [52, 21] are related to the platform, sensors or controller of the robot whereas software ones have more to do with the decision making, perception and behaviour execution of the robot. A social norm violation usually occurs when the interaction between the two agents -human and robot- is misinterpreted, or the environment the interaction takes place is not clearly defined. For the purpose of this review, we divided the faults observed in papers depending on the type of HRI. The general types of interaction were: *task performance* by robot, such as assembling or disassembling furniture, a *collaboration task* between human and robot, for example in a human-robot fictional desert survival setting, *instructing/teaching a robot* a certain task, *learning* from a robot and *simple conversation* on a certain topic with a robotic agent. The errors occurring within these settings are shown in Table 1 based on most common errors observed in such situations, the likelihood of an error and the overall impact on the interaction. The error severity presented in Table 1 refers to the social severity of the error in a given interaction. For example, a technical failure that can easily be identified by a human and solved by the robot's controller is usually not that severe from a social perspective. Therefore the scale presented in this review depends on how meaningful the error is in terms of the user experience during the interaction and how understandable an error is when it occurs in terms of human perception.

The category in which the faults observed are most common is the task performance category. Some relate to the autonomy of the robot [10, 9], in scenarios where robots are supposed to be completely autonomous. In such cases it is shown that if robots consistently fail the task when in their completely autonomous mode, then that negatively influences the user's trust. By warning people however beforehand that the robot might fail or is imperfect in some ways proved to mitigate the negative influence on the user [33]. Also, sometimes when full autonomy is given [44], faults like failures in object localization lead to a disruption in the interaction in which human

intervention by the robot's technician is needed. Nevertheless this kind of error is not considered as severe as the root of the error is always very clear to the human, so intervention is needed but it is always targeted to a specific technical failure. Robot politeness showed to manage user expectations in cases where for example battery is not powerful enough to last for the whole experiment [34]. In terms of speed of completing a certain task [6], it didn't affect user's perception of the robot as long as the said task is successfully completed. It is also shown in [19, 1, 37] that robots that are more conversational are preferred to silent ones that are more efficient in terms of time completion of the task or effectiveness. The user is more sympathetic towards a robot that can explain its mistakes and is therefore more social. Some researchers [28] developed a framework in order for the robot to communicate its failures, which proved to be successful in preventing the interruption of an interaction, although sometimes when the error that occurred wasn't present in the framework created human intervention was necessary. Another interesting finding [56] is that generally more blame is put on the robot if it displays a lack of effort compared to lack of ability or if doesn't have the correct gaze behaviour [49] (i.e. look at the participant) when the error occurs because it is perceived as unapologetic. Lastly, how a robot is perceived and how much blame is attributed to it also depends on the experience of the user interacting with it [47].

The next category where errors are observed is in cases of collaboration of humans and robots to complete a task or play a game. In cases of robot autonomy [27] people place more blame in case of an error or credit in case of completion of the task. When it comes to the question of attributing blame in case of failure of the collaborative activity, a robot that blames its teammates is more often considered as unlikeable [25, 54] thus negatively influencing the whole user experience. Generally, faulty robots are perceived as more anthropomorphic [45] than flawless ones, even when they do not necessarily resemble a human in terms of physical appearance. However, if their intelligence is measured, usually robots that have less faults are considered as more intelligent [11] as they manage to successfully complete the task. An interesting observation when comparing a robot that stops the procedure to apologise [11] for a fault and a robot that just ignores it and continues is that the former is considered as less intelligent even if it is more likeable by the participants. In some cases also, if a robot consistently exhibits erroneous behaviour, like in [40] it decreases the willingness of people to participate with it. A reason for this also is based on the level of familiarity of people participating in the study. People that are more familiar with robots in general tend to stop the interaction whereas people that first interact with robots are more impressed by the technology and are willing to continue the interaction despite the faults. Overall faults in the actions of the robot tend to be less severe than verbal ones [50], since it is more easy to manually correct such technical errors.

Type of error	Frequency of error	Social severity	Error consequences
Speech recognition	Very often	Severe	Robot fails in recognising human speech
Speech misunderstanding	Very often	Severe	Robot fails to understand the task at hand or fails in dialogue with human
Grasping failures	Often	Solvable	Robot fails to grasp an object key to the experimental procedure
Gaze errors	Sometimes	Quite severe	Robot or human fail to look at their counterpart
Navigation errors	Very often	Solvable	Robot fails to navigate in experiment room
Assessment of situation errors	Rarely	Quite severe	Robot fails to take the correct decision
People detection errors	Often	Severe	Robot fails to detect the human
Object detection errors	Often	Severe	Robot fails to detect object in a room
Battery errors	Rarely	Negligible	Not enough battery to carry out experiment
Mode Confusion	Rarely	Negligible	Human error of picking the right mode for robot
Voice localization errors	Very often	Severe	Robot fails at locating the voice of the person it is interacting with
Errors in affect processing	Rarely	Quite severe	Robot fails to process human emotions, or human is being very expressive when interacting with robot
Object positioning failure	Sometimes	Solvable	Failure to place object where needed during the experiment
Wrong instructions	Sometimes	Negligible	Human or robot failing to give correct instructions
Repetition errors	Rarely	Negligible	Robot repeats the same mistake several times
Dialogue interruption errors	Sometimes	Quite severe	Robot interrupts human or speaks out of turn
Failure to exit loop	Rarely	Solvable	Robot is stuck in the same loop

Table 1: List of errors, their frequency, consequences and severity

The type of errors that occur in the next error category, where participants are instructing or

teaching something to a robot are mainly miscommunication/speech errors. If people unfamiliar with robots interact with them [26], they tend to use more affective reasoning to communicate with the robot, and it is not always capable of fully understanding the interaction. Also, it is quite often that people can also exhibit erroneous behaviours especially in instruction situations. In a study where a person is teaching a robot how to dance [20] for example, a human rreaction to the robot when the robot is trying to concentrate on a move could confuse the robot and prolong the teaching procedure. Furthermore, mode confusions by the controller [51], like demanding a task outside the capacity of the robot can affect the interaction but are also easily solved if the person understands their error in time. In terms of robot predictability [39], predictable robots are more liked and people tend to be more happy to cooperate with them. Participants seem intolerant with unpredictable ones, much like when they disagree to cooperate with the participant [54], or blame them [18]. A last observation relating to this social context [29] is that it is difficult to generalize error solving situations as they are context-dependant.

The most profound finding in studies where robots act as the instructor or teacher, is that they are perceived as less intelligent than people [23]. Nevertheless, they are still considered more intelligent than other devices (e.g. vending machines) [12] so people rely on them to be correct more than other intelligent systems which renders their errors relatively impactful. Given the developments in the HRI field, even though errors in the behaviour of robots are likely to affect human perception of the reliability and trustworthiness of the robot [46], they will not affect their general willingness to comply with its instructions, as long as they will not cause lasting damage in doing so. This however can be quite dangerous in cases where emergency situations are observed [43], where people have been found to blindly follow a robot in a fake fire emergency situation even if they observed erroneous behaviour from the same robot just before. This can be quite problematic in the longterm while robots gain a more fundamental role or even leading role in human-robot interactions. Lastly, in terms of communication, robots that take the role of the teacher/instructor and act surprised when asked a question for example [31] tend to be perceived as less intelligent.

In terms of the dialogue social context, the impact of errors in communication is not that severe in most cases [13, 14] as the content of the conversation is predefined most of the times so there is not a lot of room for misunderstandings. Furthermore, the robot manages to communicate the fact that there has been a misunderstanding in order for the person to correct this mistake, or navigate the conversation in the right direction.

Overall, people react more in social norm violation situations than in technical failures. The reason for that is that they can identify technical failures more easily and thus there is no need to react whereas in social norm violation cases people tend to talk more, feel uncomfortable and urge themselves to solve the situation. In terms of likeability, robots that exhibit apologetic behaviour or recognize their errors are perceived as more likeable, although sometimes less intelligent. As HRI evolves, trust in a robot is becoming more and more important and thus a fault is more impactful and should be carefully looked into.

The literature review conducted in order to answer **RQ1**, indicated the the most severe and frequent errors on the sociability of the robot are speech recognition, gaze errors, people and object detection

errors and voice localization errors. Based on the above, these errors were simulated in the robot. An example of a speech recognition or misunderstanding error was that the robot failed to recognise both names of the participants and only called out one name throughout the whole procedure. Gaze errors included the fact that it mainly looked at one child whenever it addressed something important. All other times throughout the procedure the robot would randomly look at the other child. People detection errors are present when the robot fails to look at the ignored child even when it is talking. Voice localization errors are present when the robot fails to address or hear the ignored child when speaking. These manipulations resulted in a simulated environment in the study where the robot fails to accurately detect the second child leading to it being ignored throughout the game in order to see how it affects its perception of both the robot and the game.

5 Experimental method

A between-subjects experiment was carried out in May 2023 at a Public Elementary School in Leiden to research the effect of robot errors on children in the context of education. The participants (Chapter 5.2) were playing the interactive vocabulary game detailed in 5.1. The errors simulated based on the review in Chapter 4 were incoorporated in Nao's mannerism during the experiment and lead to the robot's failure of paying attention to one of the two participants. The effect of these errors was measured by the metrics explained in Section 5.3. All experiments were carried out in the same way, detailed in Section 5.4 and analysed based on the methodology of Section 5.5.

5.1 Design of Vocabulary game

In this section, the interactive vocabulary game is introduced as well as the role of experimenter in the procedure. In Section 5.1.1, the design of the game will be presented, thereafter in Section 5.1.2 the differentiation of the two experimental conditions are explained with respect to the game and finally in 5.1.3 the overarching control of the experimenter is detailed.

5.1.1 Game design

The game was programmed within platform by Interactive Robotics [2] a company that focuses on Social Robotics and more specifically within education and healthcare. The interactive vocabulary game is designed for native Dutch students with the aim to teach them some words in English. The targeted age is primary school children of age 10 - 12, group 7 and 8 in the Dutch primary school system [3]. The children participate in pairs, navigating in a fictional town called "Naotown", with the NAO robot as their peer. They start the game in the central map of the town, shown in Figure 6 where the robot introduces itself and explains the rules of the game. The goal is to navigate through all the places in the town, which are shown in Figures 2, 3, 4 and 5 and are the Naotown Zoo, Market, Mall and Beach respectively.

[2]Interactive Robotics, `https://www.interactive-robotics.com/`

[3]The structure of the Dutch school system `https://www.iamexpat.nl/education/primary-secondary-education/dutch-school-system`.

Figure 1: The five game states of the interactive vocabulary game

Figure 2: Game state: Zoo

Figure 4: Game state: Mall

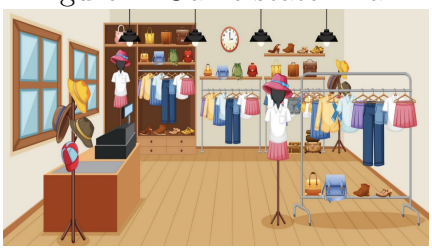


Figure 6: Game state: Central map

Figure 3: Game state: Market

Figure 5: Game state: Beach

The participants collaboratively choose a place on the map to visit and they are transferred to that state. In each state, there are three questions that the participants have to answer. Two are regarding a new vocabulary word within that game state, for example, at the beach there is a question "With what do you make sandcastles" and the participants need to reply with the word "Bucket". If the children do not manage to find the correct word, then Nao helps them and provides them with the word in English. The third question in each state has to do with the favourite activity or item of the participants in that state. For example, at the market, Nao asks the participants what their favourite fruit is. Since the robot is manipulated in order to provide more attention to one of the two participants, in this question Nao will agree with the participant not ignored during the game and say that that is also its favourite activity or item in that state. After the game has ended, Nao thanks the participant not ignored for accompanying it through its town and the game ends.

The game was made with all the dialogues being English. After some initial tests by the research team, it was decided that due to Nao's speech being unclear sometimes and due to the fact that the game is intended for native Dutch elementary school students, it would be best if most of the dialogues were in Dutch. For that reason, the robot was speaking primarily in Dutch, but whenever a question was about to be asked it would switch to English. In the introduction of the game, the robot explicitly asks the participants to respond in English, to ensure that there is some learning gain in the procedure. The game was translated with the help of ChatGPT[4].

[4]ChatGPT, `https://openai.com/blog/chatgpt`

5.1.2 Game conditions

Two experimental conditions are created, where two children participate at a time. They interactively play the game described in Section 5.1.1 with NAO and the robot takes the role of a peer to the students. However, due to the simulated errors, the two children are not treated equally during the learning procedure. The attention of the robot is more focused on one participant, so that when a question is posed, the robot only answers to the person not ignored and also employ forms of personalisation (e.g. calling it by its name). Additionally, when the not ignored participant provides a correct answer whilst playing the game it will congratulate them. In contrast, when the ignored participant asks something, the robot does not reply and does not offer any help throughout the game. It is necessary to place the children together in the same room, and not one at a time, so that the contrast is apparent.

The same game in terms of game design is played in both experimental conditions and in both cases the robot will explain to the participants that it is not perfect and that it has some design faults that lead to a difficulty in simultaneously paying attention to two people. It explains that this is due to some errors in its sensors. The difference between the two conditions however is the timing of the explanation by the robot. In *Experimental condition 1*, which will be referred to as the *Explanation after* condition hereafter, the game is played by the participants and the explanation is provided *after* the participants have replied to the questionnaire. In *Experimental condition 2*, which will be referred to as the *Explanation before* condition hereafter, the robot starts off by addressing its errors and its inability to respond to many stimuli *before* the game starts and ofcourse *before* the questionnaire. In a pilot study conducted at Leiden University, children were asked if it is better for the robot or the experimenter to address the robot errors. The unanimous response provided was that they would prefer the robot to it, since it is the main social entity the participants interact with and the rest of the experiments were carried out in that manner. The full explanation by the robot can be found in Appendix A.

5.1.3 Experimenter control

Wizard of Oz during game In the realm of human-robot interactions, a "Wizard of Oz" setup is an experimental approach where a human operator controls a robot's actions and responses while giving the impression of autonomous behavior. This methodology, inspired by "The Wonderful Wizard of Oz" by Frank L. Baum, allows researchers to study human reactions and behavior without requiring fully autonomous robots. By simulating the robot's intelligence and autonomy, the setup enables the evaluation of user interfaces, interaction strategies, and user expectations, providing valuable insights for the development of future robotic systems.

In the context of this master's thesis, the experimental setup involves the experimenter utilizing a Wizard of Oz approach to manipulate the participants' trajectory and the robot's responses during the interaction. The experimenter, strategically presses keyboard buttons at each game state to influence the next destination the participants will visit and determine how the robot will react to their actions. This allows for controlled variations in the participants' experiences and the robot's behaviors, enabling the examination of specific interaction scenarios and their effects on

user responses. By carefully orchestrating these manipulations, the study aims to gain insights into the impact of the robot's responses on their overall engagement and satisfaction.

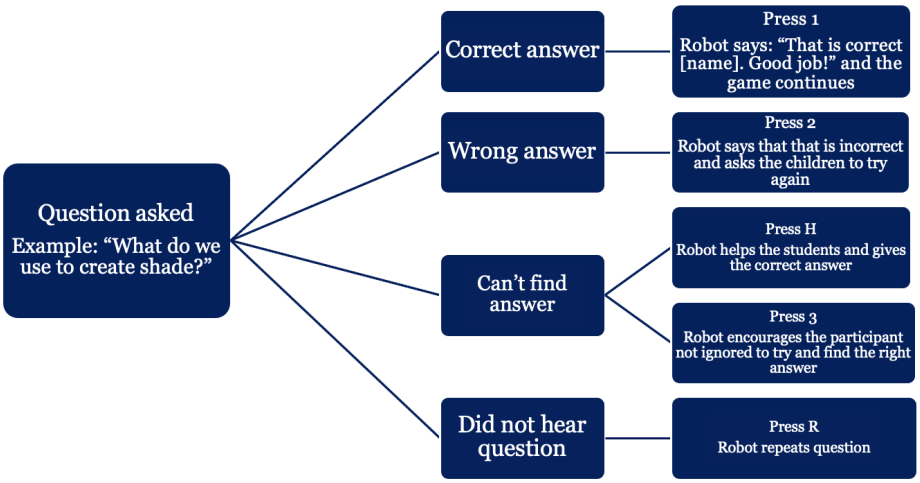


Figure 7: Flowchart of WoZ setup

Within each of the four visited states in Figures 2, 3, 4 and 5, the experimental setup incorporates five standard responses corresponding to different interaction scenarios, as observed also in Fiigure 5.1.3. The first response occurs when the child provides the correct answer and the robot acknowledges their response positively. The second response is triggered when the child provides an incorrect answer, prompting the robot to inform them that it is incorrect and encouraging them to try again. The third response aims to motivate the child to continue searching for the correct answer. The fifth response involves the robot repeating the question it just asked, in case the participants have not heard it clearly. Lastly, the sixth response is designed for cases when participants are unable to find the correct word for a given question, at which point the robot provides the desired word in English, assisting the participants in the interaction, in case it has come to a halt. These predefined responses serve to create a structured and consistent interaction framework between the robot and participants throughout the study. Also, when the participants are in the central map game state, they are asked which place they want to visit and depending on their answer that place is chosen by a button from the experimenter. The Sheet created and used for the Wizard of Oz design can be found in Appendix B

Name storage and recalling Prior to the experiment, the participants are asked their names by the researcher. The researcher then stores only one of the two names, the name of the participant not ignored by the robot and this is recalled throughout the experiment in three key places:

- When the participant not ignored by the robot provides the correct answer. In that case, the robot replies "Good job *[name]*. That is the correct answer!" and the game continues.
- When the participant not ignored by the robot struggles to find the correct answer. In that case, the robot replies "Come one *[name]*! You can do it and find the correct answer!", in order to encourage the participant to find the correct word in English.

- Whenever the robot agrees with the participant not ignored. For example, in the beach game state a question is asked by the robot to both participants. "What is your favourite thing to do at the beach?". When the participants give an answer, the robot replies "Great idea *[name]*, that is my favourite thing to do at the beach as well!"

The process of name storing and recalling during the interaction serves the purpose of creating a personalized experience for the non ignored participant. By actively remembering and using the participants' name throughout the interaction, the robot establishes a sense of familiarity and individual recognition, enhancing the perception of a personal connection, as observed and studied also in [32].

Robot design During the game design, three movements were created using Choregraphe [5], a software developed by SoftBank Robotics specifically designed to create and program movements for the Nao robot, in our case movements in order to ensure appropriate visual engagement of the robot with the participants. The first movement involved the robot turning to the left, directing its gaze towards the participant seated on its left side and then returning to the center. This participant was the one intentionally ignored during the experiment and this movement aimed to establish visual acknowledgment of the participant but no further interaction.

The second movement consisted of the robot turning to the right to look at the participant on its right side, followed by a subsequent return to the center position. It consists of a mirror movement to the one described above, and it occures within the interaction, to ensure visual contact of both participants with the robot. The third movement was specifically designed to occur whenever the robot asked a question. It involved a turn to the right, where the robot would then remain in a locked position, awaiting a response from the participant who was not being ignored who was always seated on the right of the robot. This movement emphasized focused attention on the participant responding to the question.

These movements were created with the intention of ensuring that the robot maintains visual contact with both participants throughout the interaction. However, by selectively focusing on one participant during critical moments such as asking or answering important questions, the movements aimed to emphasize the significance of the interaction and provide a clear indication of the robot's attention to only one child.

5.2 Participants

Once the game was designed and tested by the team of researchers, a pilot study was carried out at the University of Leiden, involving two male student participants, both aged 12. Subsequently, an engagement letter (see Appendix D) was distributed to various schools. The experiments were conducted with participants from a public school in Leidn. A total of 62 participants (30 Male and 32 Female) between the ages of 10 and 12, residing in the Leiden area, actively took part in the study. One pair was excluded from the study due to keyboard issues that prevented them from

[5]Choregraphe Suite, `http://doc.aldebaran.com/2-4/software/choregraphe/index.html`

fully engaging in the experimental tasks. These participants were removed from the final analysis to ensure data integrity and accuracy. To ensure unbiased allocation, participants were randomly assigned to one of the two experimental conditions, ensuring a balanced distribution of age and gender. Prior to their involvement in the study, the researchers obtained approval from the Leiden Ethics Review Committee, and informed consent forms, including an opt-out option, were provided to the parents or guardians of the participating children. The Informed Consent form is included in Appendix C.

5.3 Measures of affective outcomes

5.3.1 Questionnaire

The questionnaires employed in this study aimed to measure various aspects of the participants' perception and evaluation of the task and the robot. Specifically, they assessed the likeability of the task and the children's perception of the robot in terms of social support, closeness, trust, and overall likeability. The first four measurements, presented in Table 2 were adapted from a study by van Straten et al. [53], which examined the children's perceived social support, closeness, trust, enjoyment of playing the game, and learning experience with the robot. The questions were verbally presented by the researcher, and the children rated their responses on a 5-point Likert scale ranging from "Doesn't apply" at all" to "Applies completely".

Perceived social support	**Closeness**	**Trust**	**Task Engagement**
If I were in trouble I could rely on Nao	Nao is a friend	I feel that I can trust Nao	I enjoyed learning words in English
If I were in trouble Nao would be willing to help me	I feel comfortable around Nao	I feel that Nao can keep my secrets	I found the game difficult
If I were in trouble Nao would stand up for me	Nao and I are becoming friends	I feel that Nao is honest	I would like to continue playing the game
If I were in trouble Nao would cheer me up	Nao and I are a good match	I feel that Nao is trustworthy	I wanted to do my best
			I found the game boring
			I found the game easy

Table 2: Robot perceived social support, closeness and Trust and Task Engagement Questionnaire ([53])

Moreover, the likeability of the robot as a social entity was evaluated using the Godspeed Questionnaire Series (GQS) [3], seen in Table 3 a directed questionnaire specifically designed for analyzing human-robot interactions. The GQS measured robot anthropomorphism, likeability,

animacy, perceived intelligence, and perceived safety. Participants provided their responses on a 5-point Likert scale. Additionally, an open-ended question inquired whether the child noticed being ignored during the experiment and if it had any effect on their experience. These comprehensive questionnaires allowed for a holistic assessment of the participants' perceptions and evaluations related to the task and the robot.

Question	Type
Did you like or dislike the robot?	Likert-scale
Did you find the robot friendly or unfriendly?	Likert-scale
Did you find the robot kind or unkind?	Likert-scale
Did you find the robot pleasant or unpleasant?	Likert-scale
Did you find the robot ugly or beautiful?	Likert-scale
Did you feel that the robot paid attention to you? Did that bother you?	Open

Table 3: Godspeed Questionnaire questions on likeablity of robot and open-ended questions

The full questionnaire that was used during the experiments can be found in Appendix E.

5.3.2 Head direction estimation

Gaze attention plays a crucial role in understanding children's engagement during tasks, as it provides valuable insights into their attentional focus and cognitive processes. Several studies have examined the significance of gaze in assessing children's engagement. For instance, Foulsham et al. [42] investigated gaze patterns in children during visual search tasks, highlighting the relationship between gaze behavior and task performance. In our study, video recordings of the experiments were captured using a GoPro HERO 7 camera, recording at a rate of 30 frames per second. The video recordings were then processed in Openface, an open-source software that facilitates facial behavior analysis and recognition. With the code available in Openface, head pose estimations were extracted for each frame in the video. More specifically, the location of the participant with respect to the camera in mm as well as the rotation in radians with respect to the camera again was extracted and used. The videos of each participant were processed separately, in order to make the extractions more accurate. The face of the participant not analysed each time was blurred in iMovie. The extracted head pose estimations were analyzed using a Python framework developed by Fleur Moorlag [38], adapted specifically for this experimental setup allowing for the estimation of children's gaze patterns and determining their focus of attention, whether it was directed towards the screen, the robot, or other areas within the room.

5.4 Procedure

Children participated in pairs, with one child designated as the ignored participant and the other as the non-ignored participant. The seating arrangement determined the roles of the participants with the ignored participant positioned on the left of the robot and the non-ignored participant on the right as seen in Figure 5.4. The experimenter was always present during the experiments, in the back of the room as seen in Figure 9. Participants entered the room together and were randomly

assigned to one of the two available chairs, thus it was unknown to them and the experimenter which of the two would be the ignored participant.

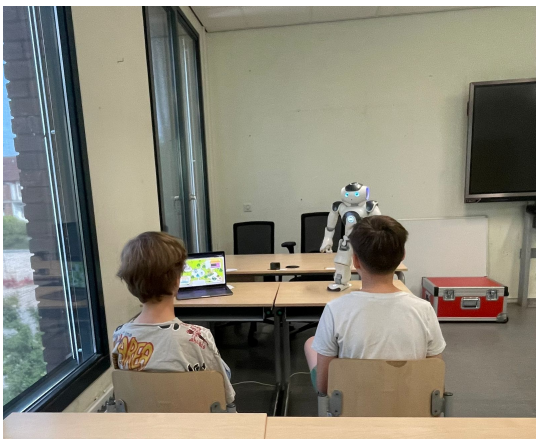


Figure 8: Behind view of experimental setup

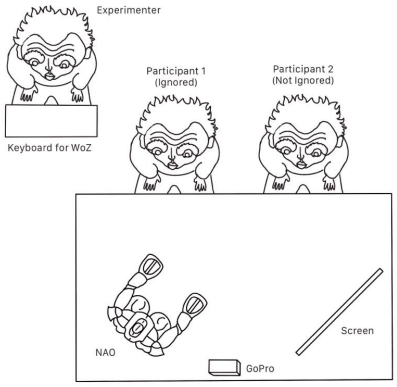


Figure 9: Top view of experimental setup

Once seated, an information letter (see Appendix F) was read aloud to the children, providing details about the upcoming vocabulary game and informing them that video recordings would be conducted throughout the session. It was emphasized that they had the option to stop participating at any time if they felt uncomfortable. The interactive vocabulary game detailed in Section 5.1 with Nao commenced, allowing the participants to engage with the robot and play a game learning words in English.

Upon completion of the game, each child was individually taken aside for the questionnaire session. The experimenter read the questionnaire questions aloud, ensuring consistent delivery across participants. The children provided their responses, expressing their thoughts, perceptions, and experiences related to the interaction with the robot and the game. For participants in experimental condition 1, they were then brought back into the room, where the robot provided an explanation of its behaviour resulting from some errors in its sensors. The purpose of this interaction was to provide additional clarification and address any concerns or confusion arising from the robot's performance. In contrast, participants in experimental condition 2 had already received the same robot explanation during the introduction of the experiment.

This procedural approach aimed to maintain consistency across participants while adhering to the distinct roles of the ignored and non-ignored participants, thereby ensuring a comprehensive understanding of the effects of the interaction dynamics on the participants' experiences and perceptions.

5.5 Analysis

5.5.1 Statistical analysis

The experimental method employed in this study involved the administration of a questionnaire to assess various aspects of participant responses. The statistical analysis utilized for this questionnaire was MANOVA (Multivariate Analysis of Variance). MANOVA is a multivariate extension of the Analysis of Variance (ANOVA) technique, specifically designed to examine the relationship between multiple dependent variables and one or more independent variables. It allows for the simultaneous examination of the effects of independent variables on multiple dependent variables, providing a comprehensive analysis of the overall multivariate pattern of results.

During the MANOVA analysis, one of the commonly used multivariate tests is the Pillai's Trace, also extensively discussed in this study. Pillai's Trace is a multivariate test statistic that assesses the significance of the overall effect of the independent variables on the set of dependent variables. It represents the overall multivariate variance explained by the independent variables. When conducting MANOVA, researchers examine the significance of the Pillai's Trace statistic to determine if the observed differences between groups are statistically significant. An increased value in the Pilai's Trace suggests a correlation between dependent and independent variables but that also depends on the p-values.

In addition to the MANOVA analysis, correlations between some variables were also examined using Pearson's correlation coefficient. Pearson's correlation coefficient is a statistical measure that quantifies the strength and direction of the linear relationship between two continuous variables. It ranges between -1 and +1, where -1 represents a perfect negative correlation, +1 represents a perfect positive correlation, and 0 indicates no linear correlation.

5.5.2 Python framework for head direction estimation

The videos recordings of the participants were processed using OpenFace to extract the head position and rotation in relation to the camera. Each participant was processed separately, with the other participants faced blurred to be undetectable by OpenFace. The facial features extracted by OpenFace are the head position ($pose\_T_x$, $pose\_T_y$, and $pose\_T_z$) of the participant in mm and the rotation angle (R_x, R_y, and R_z) in radians, both features with respect to the camera. The extraction was done on a per-frame basis, with a frame rate of 30 frames per second. According to the OpenFace framework [2], a positive $pose\_T_x$ would be on the right-hand side of the camera, a positive $pose\_T_y$ would be downward from the camera and a positive $pose\_T_z$, the camera coordinate axes being the ones shown in Figure 5.5.2.

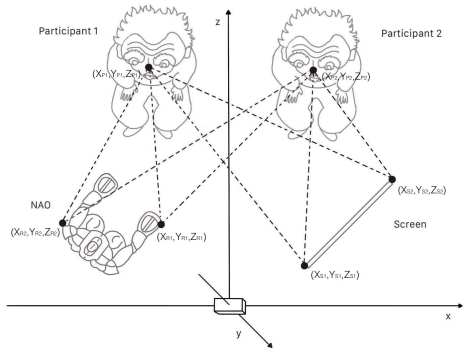


Figure 10: Framework coordinates to estimate head direction

In addition, the head rotations, namely head pitch (R_x), head yaw (R_y) and head roll (R_z) were extracted. A positive pitch indicates a downward nodding motion of the participant's head, while negative pitch indicated an upward nodding motion. Positive yaw indicates a rotation of the head to the participant's right, while a negative yaw indicated a rotation to the participant's left.

In Figure 5.5.2, the experimental setup with the coordinates of the robot; (X_{R1}, Y_{R1}, Z_{R1}) and (X_{R2}, Y_{R2}, Z_{R2}) and screen; (X_{S1}, Y_{S1}, Z_{S1}) and (X_{S2}, Y_{S2}, Z_{S2}). These coordinates were manually measured with respect to the camera which is the origin, at (0,0,0). Specifically, X_{R1} and Z_{R1} represent the outter landmark of the robot and X_{R2} and Z_{R2} the outter left one. Regarding the screen, X_{S1} and Z_{S1} represent the outter left coordinates of the screen and X_{S2} and Z_{S2}. In both cases, the x-coordinate measures how much to the right ot left the screen/robot are positioned and the z-coordinate how far away they are situated. The y-coordinates measure the top and bottom of the screen and robot, y being negative upwards and positive downwards from the camera.

Using the coordinates of the robot and screen, along with the location and rotation of the participant's head extracted from OpenFace, the framework was adapted to estimate whether the participant was looking towards the robot, the computer screen, or elsewhere. The algorithm used for head direction estimation is presented in Algorithm 1.

Algorithm 1: Algorithm to estimate head direction of participant

Data: frame = number of frame in the video sequence
$pose\_T_x$ = location of head on the x-axis with respect to the camera in mm
$pose\_T_y$ = location of head on the y-axis with respect to the camera in mm
$pose\_T_z$ = location of head on the z-axis with respect to the camera in mm
X_{R1}, Y_{R1}, Z_{R1} = minimum coordinates of the robot with respect to the camera in mm
R_x = head pitch; rotation of head on the x-axes with respect to the camera in radians
R_y = head yaw; rotation of
R_z = head roll; rotation of
X_{R2}, Y_{R2}, Z_{R2} = maximum coordinates of the robot with respect to the camera in mm
X_{S1}, Y_{S1}, Z_{S1} = minimum coordinates of the screen with respect to the camera in mm
X_{S2}, Y_{S2}, Z_{S2} = maximum coordinates of the screen with respect to the camera in mm
Result: A list with the head direction of the participant at each frame, whether that is towards the robot, the screen or anywhere else in the room

for *i in frames* **do**
  $X_cZ_{S1} = (T_z - Z_{S1}) * \tan(-R_y) + T_x$
  $X_cZ_{S2} = (T_z - Z_{S2}) * \tan(-R_y) + T_x$
  $Y_cZX_{S1} = \frac{pose\_T_z - Z_{S1}}{\cos R_y} * \tan R_x + T_y$
  $Y_cZX_{S2} = \frac{pose\_T_z - Z_{S2}}{\cos R_y} * \tan R_x + T_y$
  $X_cZ_{R1} = (T_z - Z_{R1}) * \tan(-R_y) + T_x$
  $X_cZ_{R2} = (T_z - Z_{R2}) * \tan(-R_y) + T_x$
  $Y_cZX_{R1} = \frac{pose\_T_z - Z_{R1}}{\cos R_y} * \tan R_x + T_y$
  $Y_cZX_{R2} = \frac{pose\_T_z - Z_{R2}}{\cos R_y} * \tan R_x + T_y$
  if $X_cZ_{S1} \geq X_{S1}$ *and* $X_cZ_{S2} \leq X_{S2}$ *and* $Y_cZX_{S1} \geq Y_{S1}$ *and* $Y_cZX_{S2} \leq Y_{S2}$ **then**
    Head_direction = *Screen*;
  else
    if $X_cZ_{R1} \leq X_{R1}$ *and* $X_cZ_{R2} \geq X_{R2}$ *and* $Y_cZX_{R1} \geq Y_{R1}$ *and* $Y_cZX_{R2} \leq Y_{R2}$ **then**
      Head_direction = *Robot*;
    end
  end
end

6 Results

6.1 MANOVA on grouped constructs

6.1.1 Reliability Analysis

The first step is to ensure reliability consistency within the related questions in the questionnaire; Perceived Social Support, Closeness, Trust, Task Engagement and Robot likeability. To measure this, Cronbach's alpha was calculated to quantify the extent to which items within each group of questions are internally correlated. The questions relevant to each construct are according to

Tables 2 and 3. A higher Cronbach's alpha value (closer to 1) suggests greater internal consistency, indicating that the items in the scale are reliably measuring the same concept.

Construct	**Cronbach's alpha**	**Number of questions in construct**
Perceived Social Support	0.613	4
Closeness	0.776	5
Trust	0.720	4
Task Engagement	0.671	6
Robot likeability	0.740	5

Table 4: Cronbach's alpha based on Standardized items

For the constructs evaluated, Cronbach's alpha values range from moderate to relatively high as can be seen in Table 4. Perceived Social Support demonstrates moderate internal consistency ($\alpha = 0.613$) across its four questions, while constructs like Closeness ($\alpha = 0.776$), Trust ($\alpha = 0.720$), and Robot Likeability ($\alpha = 0.740$) exhibit stronger internal consistency among their respective items. Task Engagement, encompassing six questions, yields a moderate level of internal consistency ($\alpha = 0.671$). These coefficients collectively suggest that the survey items effectively capture consistent aspects of the intended constructs and can therefeore be used to conduct a MANOVA analysis.

6.1.2 MANOVA results

As internal consistency within the constructs is now established, it is feasible to conduct a MANOVA with the five constructs as dependent variables, and the Ignored / Not Ignored and the moment of explanation experimental condition as two independent variables. As a reminder, the "Ignored" condition indicates whether a child is Ignored throughout the procedure, in Experimental condition 1 (or the after explanation) the robot apologises after the questionnaire for its faulty behaviour and finally in Experimental condition 2 (or the before explanation) the robot apologizes before the start of the game for its errors.

As can be observed in Figure 6.1.2, the average scores for the constructs do not significantly differ based on the Ignored or the moment of explanation conditions. They are more or less consistent with no apparent trend. Overall, the lower scores were observed in the Closeness and Robot Likeability categories and the highest one in the Trust category.

The consistency in mean scores is then confirmed with the MANOVA results in Table 6.1.2. the "Ignored" factor exhibited non-significant effects on the dependent variables, with Pillai's Trace (Value = 0.053, F = 0.625, p = 0.681), Wilks' Lambda ($\lambda = 0.947$, F = 0.625, p = 0.681) and Hotelling's Trace ($T^2 = 0.056$, F = 0.625, p = 0.681).

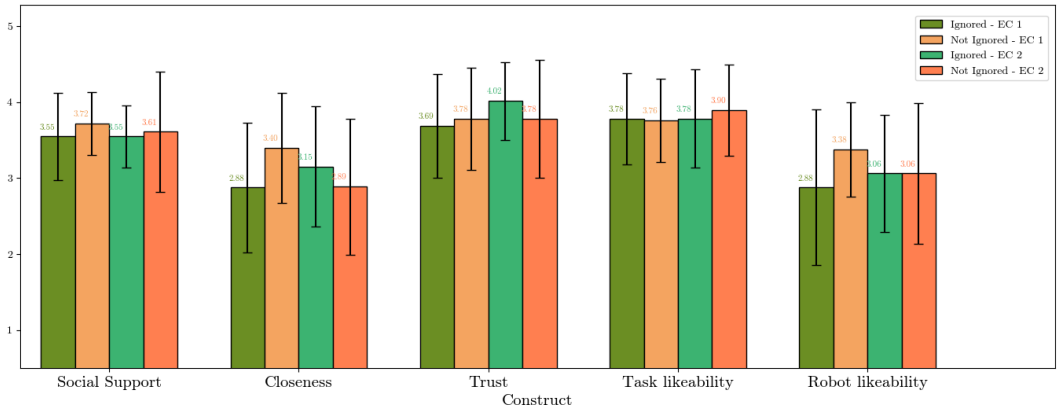


Figure 11: Mean scores for Perceived Social Support, Closeness, Trust, Task and robot likeability constructs based on the average scores of their questions (5 =Applies completely, 4 = Applies, 3 = Party applies, partly does not, 2 = Doesn't apply, 1 = Doesn't apply at all as a function of being Ignored or not

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.986	808.357b	5.000	56.000	<.001
	Wilk's Lambda	.014	808.357b	5.000	56.000	<.001
	Hotelling's Trace	72.175	808.357b	5.000	56.000	<.001
Ignored condition	Pillai's Trace	.053	.625b	5.000	56.000	.681
	Wilk's Lambda	.947	.625b	5.000	56.000	.681
	Hotelling's Trace	.056	.625b	5.000	56.000	.681
Experimental condition	Pillai's Trace	.107	1.344b	5.000	56.000	.260
	Wilk's Lambda	.893	1.344b	5.000	56.000	.260
	Hotelling's Trace	.120	1.344b	5.000	56.000	.260
Ignored * Experimental condition	Pillai's Trace	.123	1.571b	5.000	56.000	.183
	Wilk's Lambda	.877	1.571b	5.000	56.000	.183
	Hotelling's Trace	.140	1.571b	5.000	56.000	.183

Table 5: MANOVA results using the five constructs as dependent variables and the Ignored / Not Ignored and Experimental conditions as independent variables.

The moment of explanation condition showed a slightly more moderate impact on the dependent variables, however still not significant yielding for Pillai's Trace a V = 0.107, F = 1.344, p = 0.260. Additionally, the interaction between Ignored and Experimental Condition demonstrated a minor impact, with Pillai's Trace (V = 0.123, F = 1.571, p = 0.183).

In summary, the MANOVA results indicate that the main effects of Ignored and Experimental conditions were not statistically significant. The interaction between the two also showed non-significant effects. These findings shed light on the impact of different factors on the multivariate pattern of responses across the variables under investigation.

6.2 Influence on Perceived Social Support, Closeness, Trust, Task Engagement and robot likeability

As the grouping of constructs didn't yield any significant results, the questions within the constructs were also separately examined. For the independent variable we used the factor of being ignored or not and the experimental conditions. The key findings are presented in this section, and the complete tables with the statistical analysis are placed in Appendix I

6.2.1 Perceived Social Support

For the question "If I were in trouble, I could rely on Nao", participants who were not ignored in the after explanation condition (EC1) reported a mean score of 3.38 (SD = 0.619) as can be seen in Figure 6.2.1, while those in the before the experiment explanation condition (EC2) reported a slightly lower mean score of 3.06 (SD = 0.929).

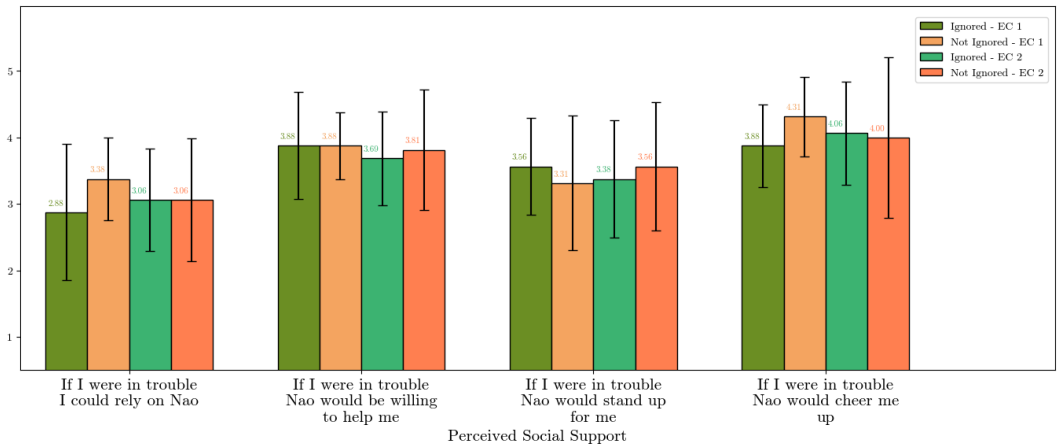


Figure 12: Mean scores for Perceived Social Support (5 =Applies completely, 4 = Applies, 3 = Party applies, partly does not, 2 = Doesn't apply, 1 = Doesn't apply at all as a function of being Ignored or not

Although the mean scores in Table 6.2.1 suggest that the experimental conditions and experiences of being ignored or not during the procedure may influence participants' perceptions of social support, multivariate tests revealed a weak effect of the "Ignored" independent variable and the "Monment of explanation" indepedent variable on the dependent variables.

6.2.2 Closeness

The analysis on closeness revealed that the independent variable "Ignored" had a significant effect on participants' perceived closeness to the robot. Participants who were not ignored consistently reported higher mean scores across multiple variables related to closeness.

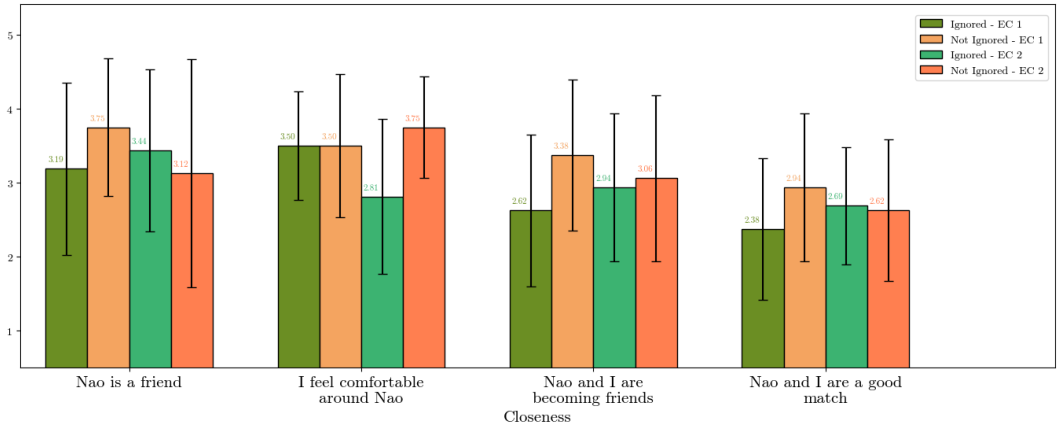


Figure 13: Mean scores for Closeness questions (5 =Applies completely, 4 = Applies, 3 = Party applies, partly does not, 2 = Doesn't apply, 1 = Doesn't apply at all as a function of being Ignored or not

These findings highlight the significant influence of the "Ignored" independent variable on participants' perceived closeness to the robot which is confirmed by the multivariate tests. The significance of the "Ignored" independent variable in is indicated by a moderate Pillai's Trace value of .209 (p = .019). This suggests that being ignored or not during the procedure significantly affects participants' perception of closeness to the robot. However, the moment of explanation did not show a significant effect on the perceived closeness to the robot. The associated Pillai's Trace value was weak (p = .844), indicating that the different moments of explanation did not have a significant impact on participants' perceived closeness.

6.2.3 Trust

In terms of Trust, the analysis showed that the Experimental Condition had a significant effect on the results. In general, participants in before moment of explanation condition consistently reported higher mean scores on trust-related questions compared to those in experimental condition 1. More specifically, for the variable "I feel that Nao can keep one of my secrets", participants in the before experimental condition (both the ones Ignored and not Ignored) reported a mean score of 3.81 (SD = 0.885), indicating a relatively higher level of trust, while those in the after explanation condition reported a slightly lower mean score of 3.34 (SD = 1.305), suggesting a moderate level of trust.

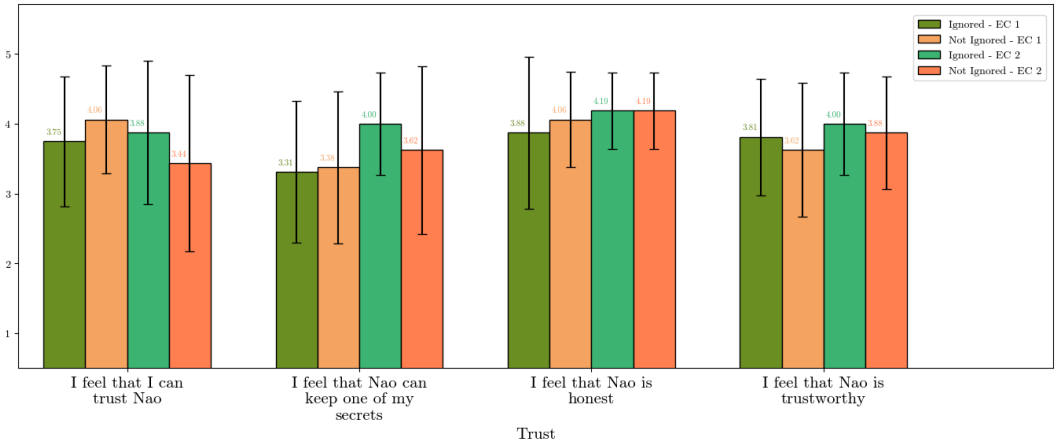


Figure 14: Mean scores for Trust questions (5 =Applies completely, 4 = Applies, 3 = Party applies, partly does not, 2 = Doesn't apply, 1 = Doesn't apply at all as a function of being Ignored or not

The influence of the moment of explanation on participants' trust perceptions towards the robot is significant by a moderate Pillai's Trace value of .168 (p = .031). On the other hand, being ignored or not did not show a significant effect on the perceived trust, as indicated by the weak Pillai's Trace value of .030 (p = .775). This implies that the experience of being ignored or not during the procedure does not significantly influence participants' trust towards the robot as much as the moment of explanation does.

6.2.4 Task Engagement

In terms of task engagement, neither being Ignored or the moment of explanation seemed to have an effect on how much the participant enjoyed the task or wanted to do their best.

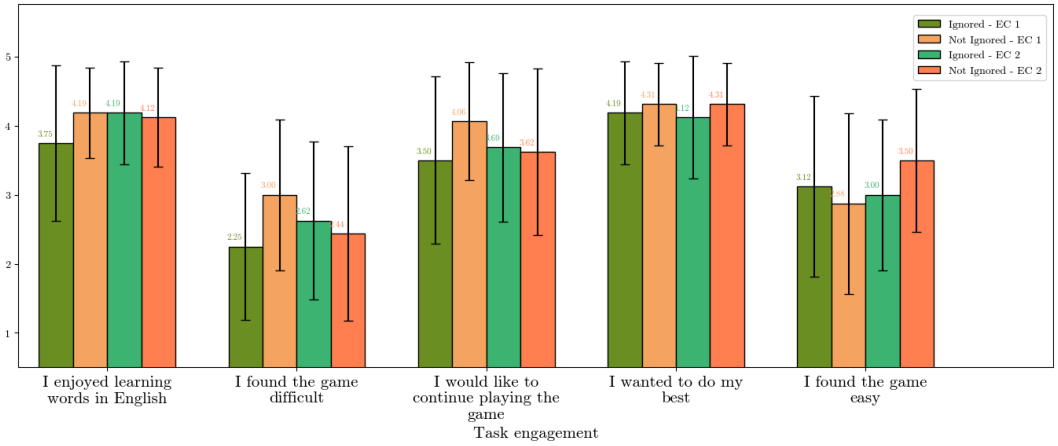


Figure 15: Mean scores for Task Engagement questions (5 =Applies completely, 4 = Applies, 3 = Party applies, partly does not, 2 = Doesn't apply, 1 = Doesn't apply at all as a function of being Ignored or not

This can be supported by the mean scores seen in Figure 6.2.4, where no specific trend can be

observed in terms of how participants rate the task. However, there are some interesting trends on how the questions in the "Task Engagement" section relate to each other. The most important are highlighted here, and a detailed table of the correlations can be found in Appendix I. A positive correlation was found between enjoyment and the desire to continue playing the game (r = 0.579, $p \leq 0.001$), suggesting that participants who found the activity enjoyable were more likely to express a stronger inclination to continue playing. Furthermore, there was a positive correlation between the motivation to do one's best and the desire to continue playing (r = 0.375, p = 0.002), indicating that participants who were more motivated to excel in the activity also showed a greater interest in continuing. On the other hand, finding the game difficult was negatively correlated with both enjoyment (r = -0.379, p = 0.002) and the desire to continue playing (r = -0.359, p = 0.004), implying that participants who perceived the game as more challenging reported lower levels of enjoyment and expressed reduced interest in continuing. These findings highlight the influence of enjoyment, perceived difficulty, and motivation on participants' task engagement during the activity.

6.2.5 Robot likeability

When examining the individual effects, the factor of being Ignored did not significantly contribute to the prediction of robot likability (p = .496), indicating that whether the robot ignored the participants or not did not have a significant impact on their likability ratings. The factor of experimental condition also did not have a significant effect on robot likability (p = .164), suggesting that the specific experimental conditions did not result in significant differences in likability ratings. The interaction between the two was also not significant (p = .840), indicating that the combined effect of these factors did not have a significant influence on robot likability.

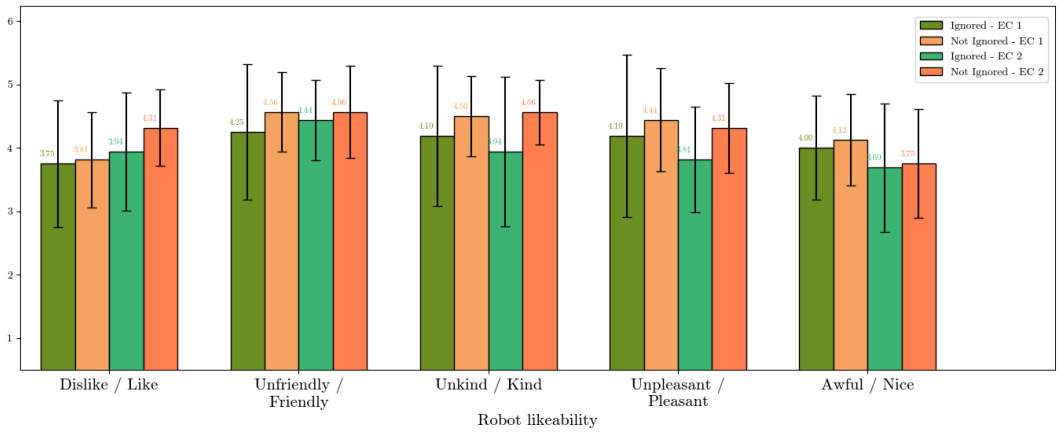


Figure 16: Mean scores for Robot likeability questiions (5 =Applies completely, 4 = Applies, 3 = Party applies, partly does not, 2 = Doesn't apply, 1 = Doesn't apply at all as a function of being Ignored or not

Overall, the participants seem to rate the robot quite highly in terms of robot likeability. Perhaps this has to do with the fact that for many participants this was their first robot interaction and

were influenced by their enthusiasm when rating how likeable it was.

6.3 Correlation between Task Engagement and Robot likeability

To examine the relationship between robot likeability and task engagement, correlations were calculated between likability aspects of the robot (Dislike / Like , Unfriendly / Friendly, Unkind / Kind, Unpleasant / Pleasant, and Awful / Nice) and measures of Task Engagement (enjoyment of learning words in English, finding the game difficult, desire to continue playing the game, motivation to do one's best, finding the game boring, and finding the game easy). The key results are reported in this section, and the full Correlation matrix can be found in Appendix I

Positive correlations were observed between the likability aspects of the robot and certain aspects of task engagement. Specifically, participants who reported higher likability ratings for the robot (dislike/like) were more likely to enjoy learning words in English (r = 0.320**, $p \leq 0.01$) and expressed a stronger desire to continue playing the game (r = 0.279*, $p \leq 0.05$). Furthermore, participants who perceived the robot as more friendly (unfriendly/friendly) showed a positive correlation with a greater desire to continue playing the game (r = 0.266*, $p \leq 0.05$). Similarly, those who rated the robot as kinder (unkind/kind) exhibited a positive correlation with a stronger motivation to do their best (r = 0.327**, $p \leq 0.01$). No significant correlations were found between likability aspects of the robot (unpleasant/pleasant and awful/nice) and any of the task engagement measures (finding the game difficult, finding the game boring, and finding the game easy).

Questions	**Pearson's correlation**	**p-value**
Like / Dislike and enjoying learning in English	.320	0.001
Like / Dislike and desire to continue playing	.279	0.001
Friendly / Unfriendly and desire to continue playing	.266	0.001
Unkind / Kind and motivation to do their best	.327	0.001

Table 6: Pearson's correlation significant outcomes

These results suggest that certain likability aspects of the robot, such as overall likability, friendliness, and kindness, are associated with specific aspects of task engagement, such as enjoyment of learning and the desire to continue playing the game. However, likability aspects related to unpleasantness and negativity were not significantly correlated with task engagement measures.

6.4 Open-ended question

In the end of the questionnaire, participants were asked "Do you think the robot paid attention to you? Did it bother you?" in order to better understand how participants felt during the procedure.

Most participants did seem to notice whether they got all the attention or where ignored, but many weren't affected by it. In the group of participants Not Ignored, there were participants who enjoyed getting all the attention, who responded things like "Yes, I got all the attention but I liked that", "Yes, it ignored X but it was fun and cool". Others noticed it and thought it was weird to for the robot to react this way, saying things like "I got a lot of attention. I liked it, but it was weird that X got no attention", "Yes. I don't like if the other girl liked it as much because she gave the answer, I repeated and I got the credit.". Some of the children in this group expressed mild disappointment; "I was a little disappointed it was one-sided" while others even felt uncomfortable by the attention, saying things like "X knew some answers to the questions and couldn't answer. That was annoying for X and embarrassing for me" and "I think I noticed the attention. It felt uncomfortable, I can't explain why". Lastly, many others attributed this attention to the way they interacted with the robot, like for example one participant who felt they got all the attention because they were speaking louder.

In terms of the group of participants that were ignored, there were many who seemed relieved by the not having the pressure to respond "A little attention, but I didn't mind as I didn't know all the answers", "I enjoyed not having to answer and I still felt included" and "I would feel weird if I got the attention, it would make me nervous". Others were disappointed by not getting any attention, saying things like "Would have been better if he said both names", "I gave the right answers but Nao never said my name" and "In the beginning it was annoying but then it was okay". Some did not experience any problem with being ignored, since with teamwork they managed to play and finish the game whilst others attribute the fact they were ignored to the interaction itself, for example one participant said he was not being called because he said his name second instead of first to the robot.

Overall, it seems that unequal behaviour from the robot is observed by the participants in all groups. It also seems that it can both a positive and a negative effect to participants in both categories, also depending on how extroverted or introverted they are. Participants that are more extroverted and ignored tend to get more annoyed, whilst participants that are introverted and get all the attention feel a lot of pressure which they do not appreciate. In both cases, the judgment on Nao is not that harsh, with comments from children such as "I do not mind because it is a robot, and we are very different. I can't see him as a friend.". There was only one participant who commented on the experimental condition, and they specifically said that they were not annoyed because the robot apologized in advance (they were part of experimental condition 2). However, it seems that whether the robot provides an explanation before or after didn't really matter to the participants.

6.5 Head estimation framework results

For the head direction estimation, the average gaze direction was calculated for the Ignored and non ignored participants and was separated in experimental conditions. The results can be seen in Figure 6.5, where we can see that irrespective of being ignored or not and the experimental condition the participants exhibited an equal gaze direction.

This is confirmed by the multivariate tests, using Ignored and Experimental Condition as independent factors and head directions towards the robot, the screen and elsewhere as the dependent factos. For all three, the corrected model did not show a significant effect, as indicated by the non-significant F-

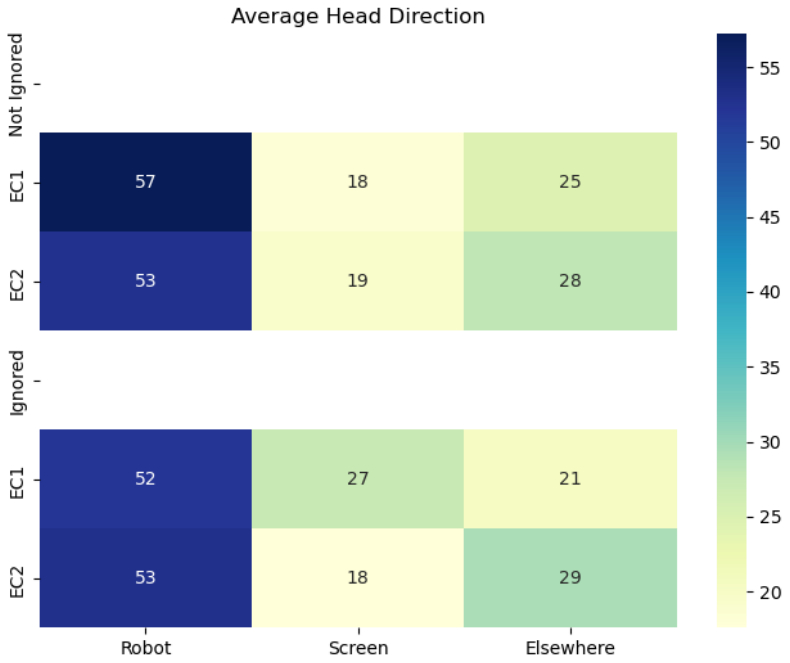


Figure 17: Head direction heat map per condition

values; (F = 0.636, p = 0.429) for the variables Ignored, (F = 0.799, p = 0.375) for the experimental condition and (F = 0.005, p = 0.944) for their interaction. These results suggest that these factors did not have a significant influence on the head directions towards the robot. Similar trends are observed for the screen and elsewhere directions

The R-squared values indicate the proportion of variance in the dependent variable that can be explained by the independent variables. For the robot head direction, the R-squared value is 0.026, suggesting that the independent variables account for a small amount of variance. Similarly, the R-squared values for screen and elsewhere head directions are 0.065 and 0.099, respectively. In conclusion, the results indicate that the independent variables did not have significant effects on the dependent variables.

7 Discussion

As concluded by the review of HRI errors in Chapter 4, speech recognition, speech misunderstanding, gaze errors, people and object detection errors and voice localization errors are the most severe and frequent errors(**RQ1**), which were simulated in this study. In order to answer **RQ2** and **RQ3**, a questionnaire was employed as well as a framework for head direction estimation. For **RQ2**, it was important to separate the participants based on the Ignored / Not Ignored condition whilst for **RQ3** based on the Experimental condition, 1 being when Nao apologized for its faults after the questionnaire and 2 being when Nao explained its faults before the start of the game. We hypothesized **(H2)** that being ignored will negatively affect the participants. When averaging the

constructs of Perceived Social support, Closeness, Trust, Task engagement and Robot likeability the factor of being ignored didn't seem to have a significant effect on these 5 factors. However, when testing the individual questions within the constructs separately, there is a slight impact on the closeness of the robot with the participant when being ignored. Similarly, apologizing in advance didn't seem to significantly impact the likeability of the robot as hypothesized in **H3**, however it had a slight impact on trust towards the robot.

7.1 Affective outcomes

7.1.1 Effect on Closeness

Overall, there seems to be a slight effect of the Ignored condition on the closeness of the participants to the robot. This can be observed by the results of the multivariate tests in the MANOVA in Section 6.2.2. The mean scores in replies of the participants in the Ignored condition are slightly lower. So children that were not ignored by the robot felt that they had developed a closer relationship to the robot. This would make sense, as the children that are addressed more often may feel more present in the experiment and also more motivated to interact more with Nao. It should be noted however, that in some cases not being ignored by the robot had the inverse effect. Some children felt a lot of pressure being constantly addressed by the robot or even annoyed. Other forms of speech recognition errors also occurred when for example the robot mispronounced a child's name. This also lead to a level of frustration which in return resulted to lower closeness scores for the child not being ignored.

7.1.2 Effect on Trust

When looking at Section 6.2.3, it is seen that there is a significant effect of the experimental condition on Trust. Participants that were given an explanation of the robot's behaviour before the start of the experimient generally rate is as more trustworthy. This means that a robot that is upfront about is faults is more likely to be trusted more. Although this is counter-intuitive in the sense that knowing that a robot has errors should make the participant trust it less because they recognize that it is imperfect and cannot substitute a human, it seems that it helps establish a stronger bond with the individual they interact with. Similar behaviours are observed in studies like [43], where participants seem to overtrust a robot to lead them in a fire escape in a simulated scenario even though it has failed in the past.

7.1.3 Open-ended questions

The open-ended questions yielded very interesting results for this study. It seems that both in the Ignored and Not Ignored participants realised the difference between the attitude of the robot. In both cases there was both positive and negative feedback of the children on their feelings towards being ignored. Contrary to our initial hypothesis, which was that children who will not be ignored will like the robot and the ignored children will develop negative feelings, very often the children not being ignored end up not liking the robot. The reasons provided where either because they realise the behaviour is unfair towards the other child or feel a lot of pressure being the center of attention at all times. Furthermore, there was an observation on how children reacted based on how

introverted or extroverted they were. Children that were more extroverted and were being ignored tend to dislike the robot more while when extroverted children are given the attention then they seem to enjoy it. On the other hand, for more introverted children the opposite seems to apply. They are more happy not having to speak up with all the answers and feel pressured when not ignored.

In all cases however, there seemed to be a lot of collaboration between the students. Most children felt the need to include their peers in the process so in all cases the game ended up as a teamwork activity. Collaboration by the children came in many forms, such as deciding together before answering anything or telling the other child the correct answer if the ignored child was the one who knew it. In one case the not ignored child even stood up and gave their seat to the other one because it felt that that is a way the second child will also get some attention. This is a very important finding, because it seems that in the end instead of enjoying feeling special or conversely being upset if being ignored, children viewed the experiment in a more social way and collaborated with each other in order to have fun. This shows that there is a prosocial attitude towards the procedure from children, with the intent of helping each other.

Another point worth mentioning is that most children noticed the different behaviour and attributed it to social behaviour and not software or hardware. There were children that thought for example that they weren't called by their name because they said their name first during the game, however children who attributed being ignored (or not) to hardware or software bugs were only 6% of total students. Lastly, in terms of how noticeable the moment of explanation was, there was only one child who commented on that during the open questions. This was a child in the ignored group, and they said that they did notice they are being ignored but it was okay since the robot explained in the beginning that it is imperfect.

7.2 Head direction estimation results

The results of the head direction estimation revealed interesting insights into the participants' behavior. Contrary to our initial expectations, participants, regardless of whether they were ignored or not, exhibited a relatively equal gaze direction. This finding, depicted in Figure 6.5, suggests that the manipulation of being ignored or engaged in both experimental conditions did not lead to pronounced changes in participants' head orientations. Overall, the results of the multivariate tests indicated that neither the factor of being ignored nor the experimental condition significantly affected participants' head directions.

In light of these findings, it is important to consider the broader context and potential limitations of the study. Considering it was a 10-minute interaction in the form of a game, the participants do not have enough time to disengage from the task. Also, the task involves multiple stimuli (the computer, the GoPro and the robot) which keeps the participants relatively focused throughout the whole task. Perhaps if it were an interaction only with the robot it would be easier to identify if the children disengage from it towards the end of the procedure.

7.3 On the experimental design

The experimental designed was thoroughly discussed and tested by the research team. In the beginning, the whole study was designed in English, however this was changed after some trials because the level of the game was deemed as to difficult. The overall goal of the study was to study the effect of the robot errors on the sociability of the robot and not to make the game very difficult. For this purpose, the robot had to be understandable overall in order to facilitate an easy two-to-one interaction where the simulated errors could be impactful. However, these errors were often surpassed by the enthusiasm of children interacting for the first time with the robot. Even though most of them noticed they were being ignored (or had all the attention) this didn't seem to make the robot less likeable overall. If the design of the experiment had multiple interactions with the robot, then these errors could have a more sever impact on the sociability. This is due to the fact that the novelty effect is taken out of the experiment and the participants would be more criitical of errors in the experiment and robot design.

8 Conclusion and recommendations

A study in which a social robot was used in order to interactively play a vocabulary game with children was designed. Errors were simulated in the robot causing it to ignore one of the two participants. The aim of the study was to explore how these errors impact the perception of the robot during the interaction and the task itself. Results showed that generally the effect of being ignored didn't greatly affect the individuals participating in the experiment. The only aspect it slightly affects is the closeness to the robot, where participants who were ignored by the robot tend to feel less close to the robot. Furthermore, apologizing in advance for these errors also did not seem to make a significant difference in the robot's sociability. The only factor it seems to slightly affect is the trust of the participant towards it because, as we observed, when the robot apologizes for its faults before the start of the experiment the participants tend to trust it more. There is also a positive correlation between the likeability of the robot and the likeability of the task. The focus of the participants in terms of head direction seemed to be either on the robot or the screen for the majority of the experiment, irrespective of them being ignored by the robot or the the moment of explanation of the robot errors. Lastly, an important conclusion stemming from the discussions with the children after the procedure is that all students adopt a prosocial behaviour. This means that they realise someone is left behind and inherently feel the need to include them in the procedure, thus end up collaborating with each other in order to play and complete the game. In this way, the end up enjoying the procedure and task instead of feeling frustrated or left out.

8.1 Future work

To build on this work, a similar experimental setup should be used multiple times with the same participants in order to test whether their perception of the robot remains the same after they have interacted some times with it. Additionally, further tests can be performed on the video recordings, including measuring whether there is a shift in focus of the participants towards the end of the study. This is to see, whether the participants in the ignored condition lose their interest in the robot

or the game. Another idea could be elevating the difficulty of the game or changing the subject could also potentially lead to different results worth exploring. Finally, different measures could be used to estimate the affect on participants, for example more open questions could be employed. It seemed that participants where more elaborate about their experience in open questions and that would perhaps provide the study with a bit more impact.

References

[1] Markus Bajones, Astrid Weiss, and Markus Vincze. Help, anyone? a user study for modeling robotic behavior to mitigate malfunctions with the help of the user. *arXiv preprint arXiv:1606.02547*, 2016.

[2] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE, 2018.

[3] Christoph Bartneck, Elizabeth Croft, and Dana Kulic. Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. 2008.

[4] Paul Baxter, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS one*, 12(5):e0178126, 2017.

[5] Daniel J Brooks, Momotaz Begum, and Holly A Yanco. Analysis of reactions towards failures and recovery strategies for autonomous robots. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 487–492. IEEE, 2016.

[6] Elizabeth Cha, Anca D Dragan, and Siddhartha S Srinivasa. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 541–548. IEEE, 2015.

[7] Catherine C Chase, Doris B Chin, Marily A Oppezzo, and Daniel L Schwartz. Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology*, 18(4):334–352, 2009.

[8] Maartje MA de Graaf, S Ben Allouch, and JAGM Van Dijk. What makes robots social?: A user's perspective on characteristics for social human-robot interaction. In *International Conference on Social Robotics*, pages 184–193. Springer, 2015.

[9] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 251–258. IEEE, 2013.

[10] Munjal Desai, Mikhail Medvedev, Marynel Vázquez, Sean McSheehy, Sofia Gadea-Omelchenko, Christian Bruggeman, Aaron Steinfeld, and Holly Yanco. Effects of changing reliability on trust of robot systems. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 73–80. IEEE, 2012.

[11] Sara Engelhardt and Emmeli Hansson. A comparison of three robot recovery strategies to minimize the negative impact of failure in social hri, 2017.

[12] Raphaela Gehle, Karola Pitsch, Timo Dankert, and Sebastian Wrede. Trouble-based group dynamics in real-world hri—reactions on unexpected next moves of a museum guide robot. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 407–412. IEEE, 2015.

[13] Petra Gieselmann. Comparing error-handling strategies in human-human and human-robot dialogues. In *Proc. 8th Conf. Nat. Language Process.(KONVENS). Konstanz, Germany*, pages 24–31, 2006.

[14] Petra Gieselmann and Mari Ostendorf. Problem-sensitive response generation in human-robot dialogs. In *Proceedings of the 8th SIGDial Workshop on Discourse and Dialogue*, pages 219–222, 2007.

[15] Manuel Giuliani, Nicole Mirnig, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Frontiers in psychology*, 6:931, 2015.

[16] Takayuki Gompei and Hiroyuki Umemuro. A robot's slip of the tongue: Effect of speech error on the familiarity of a humanoid robot. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 331–336. IEEE, 2015.

[17] Goren Gordon and Cynthia Breazeal. Bayesian active learning-based robot tutor for children's word-reading skills. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[18] Victoria Groom, Jimmy Chen, Theresa Johnson, F Arda Kara, and Clifford Nass. Critic, compatriot, or chump?: Responses to robot blame attribution. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 211–217. IEEE, 2010.

[19] Adriana Hamacher, Nadia Bianchi-Berthouze, Anthony G Pipe, and Kerstin Eder. Believing in bert: Using expressive communication to enhance trust and counteract operational error in physical human-robot interaction. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 493–500. IEEE, 2016.

[20] Cory J Hayes, Maryam Moosaei, and Laurel D Riek. Exploring implicit human responses to robot mistakes in a learning from demonstration task. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 246–252. IEEE, 2016.

[21] Shanee Honig and Tal Oron-Gilad. Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in psychology*, 9:861, 2018.

[22] Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 83–90, 2015.

[23] Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Brian T Gill, Jolina H Ruckert, Solace Shen, Heather E Gary, Aimee L Reichert, Nathan G Freier, and Rachel L Severson. Do people hold a humanoid robot morally accountable for the harm it causes? In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 33–40, 2012.

[24] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction*, 19(1-2):61–84, 2004.

[25] Poornima Kaniarasu and Aaron M Steinfeld. Effects of blame on trust in human robot interaction. In *The 23rd IEEE international symposium on robot and human interactive communication*, pages 850–855. IEEE, 2014.

[26] Elizabeth S Kim, Dan Leyzberg, Katherine M Tsui, and Brian Scassellati. How people talk when teaching a robot. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 23–30, 2009.

[27] Taemie Kim and Pamela Hinds. Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pages 80–85. IEEE, 2006.

[28] Ross A Knepper, Stefanie Tellex, Adrian Li, Nicholas Roy, and Daniela Rus. Recovering from failure by asking for help. *Autonomous Robots*, 39(3):347–362, 2015.

[29] Minae Kwon, Sandy H Huang, and Anca D Dragan. Expressing robot incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 87–95, 2018.

[30] Gary W Ladd and Becky Kochenderfer-Ladd. Research in educational psychology: Social exclusion in school. In *Social exclusion*, pages 109–132. Springer, 2016.

[31] Edith Law, Vicky Cai, Qi Feng Liu, Sajin Sasy, Joslin Goh, Alex Blidaru, and Dana Kulić. A wizard-of-oz study of curiosity in human-robot interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 607–614. IEEE, 2017.

[32] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. Personalization in hri: A longitudinal field experiment. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 319–326, 2012.

[33] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. Gracefully mitigating breakdowns in robotic services. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 203–210. IEEE, 2010.

[34] Katrin S Lohan, Amol Deshmukh, and Ruth Aylett. How can a robot signal its incapability to perform a certain task to humans in an acceptable manner? In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 814–819. IEEE, 2014.

[35] Nichola Lubold, Erin Walker, and Heather Pon-Barry. Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 255–262. IEEE, 2016.

[36] Nicole Mirnig, Manuel Giuliani, Gerald Stollnberger, Susanne Stadler, Roland Buchner, and Manfred Tscheligi. Impact of robot actions on social signals and reaction times in hri error situations. In *International Conference on Social Robotics*, pages 461–471. Springer, 2015.

[37] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, 4:21, 2017.

[38] F.N. Moorlag. The effects of a social robot's gestures on learning outcomes. Delft University of Technology, 2021.

[39] Omar Mubin and Christoph Bartneck. Do as i say: Exploring human response to a predictable and unpredictable robot. In *Proceedings of the 2015 British HCI Conference*, pages 110–116, 2015.

[40] Marco Ragni, Andrey Rudenko, Barbara Kuhnert, and Kai O Arras. Errare humanum est: Erroneous robots in human-robot interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 501–506. IEEE, 2016.

[41] Corinne Rees. The influence of emotional neglect on development. *paediaTricS and child healTh*, 18(12):527–534, 2008.

[42] Evan F Risko, Kaitlin EW Laidlaw, Megan Freeth, Tom Foulsham, and Alan Kingstone. Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in human neuroscience*, 6:143, 2012.

[43] Paul Robinette, Wenchen Li, Robert Allen, Ayanna M Howard, and Alan R Wagner. Overtrust of robots in emergency evacuation scenarios. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 101–108. IEEE, 2016.

[44] Stephanie Rosenthal, Manuela Veloso, and Anind K Dey. Is someone in this office available to help me? *Journal of Intelligent & Robotic Systems*, 66(1):205–221, 2012.

[45] Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, 5(3):313–323, 2013.

[46] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 1–8. IEEE, 2015.

[47] Satragni Sarkar, Dejanira Araiza-Illan, and Kerstin Eder. Effects of faults, experience, and personality on trust in a robot co-worker. *arXiv preprint arXiv:1703.02335*, 2017.

[48] Sofia Serholt and Wolmet Barendregt. Students' attitudes towards the possible future of social robots in education. In *Workshop proceedings of Ro-man*, 2014.

[49] Masahiro Shiomi, Kayako Nakagawa, and Norihiro Hagita. Design of a gaze behavior at a small mistake moment for a robot. *Interaction Studies*, 14(3):317–328, 2013.

[50] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. No fair!! an interaction with a cheating robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 219–226. IEEE, 2010.

[51] Thorsten P Spexard, Marc Hanheide, Shuyin Li, Britta Wrede, et al. Oops, something is wrong-error detection and recovery for advanced human-robot-interaction. 2008.

[52] Gerald Steinbauer. A survey about faults of robots used in robocup. In *Robot Soccer World Cup*, pages 344–355. Springer, 2012.

[53] Caroline L van Straten, Rinaldo Kühne, Jochen Peter, Chiara de Jong, and Alex Barco. Closeness, trust, and perceived¡? br?¿ social support in child-robot¡? br?¿ relationship formation: Development and validation¡? br?¿ of three self-report scales. *Interaction Studies*, 21(1):57–84, 2020.

[54] Leila Takayama, Victoria Groom, and Clifford Nass. I'm sorry, dave: i'm afraid i won't do that: social aspects of human-agent conflict. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2099–2108, 2009.

[55] Fumihide Tanaka, Kyosuke Isshiki, Fumiki Takahashi, Manabu Uekusa, Rumiko Sei, and Kaname Hayashi. Pepper learns together with children: Development of an educational application. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 270–275. IEEE, 2015.

[56] Sophie van der Woerdt and Pim Haselager. Lack of effort or lack of ability? robot failures and human perception of agency and responsibility. In *Benelux Conference on Artificial Intelligence*, pages 155–168. Springer, 2016.

[57] Zhen-Jia You, Chi-Yuh Shen, Chih-Wei Chang, Baw-Jhiune Liu, and Gwo-Dong Chen. A robot as a teaching assistant in an english class. In *Sixth IEEE international conference on advanced learning technologies (ICALT'06)*, pages 87–91. IEEE, 2006.

A Explanatory text used by Nao

The robot explanation for Exprimental condition 1 is the following:

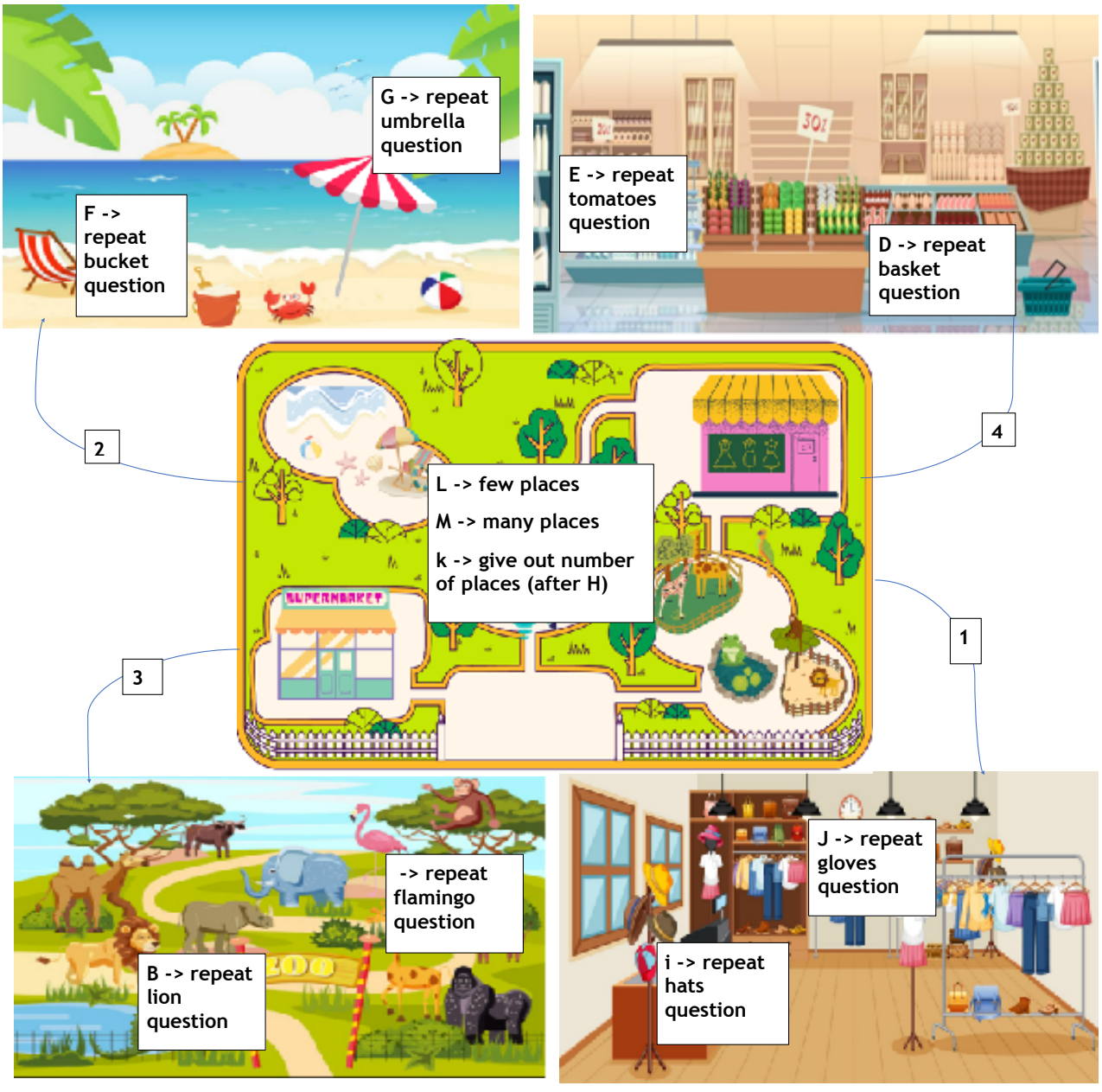
> I really hope you had fun. I wanted to tell you know that sometimes I am not perfect and that you may have felt like I was ignoring you. I wanted to let you know that that is not true. Sometimes, due to errors in my sensors I don't hear very well, so there is a possibility that I didn't hear some of your answers. There are also other errors in my gaze that do not allow me to always look at everyone. That is why I was not looking at both of you all the time. I think however that both of you did very well in the game, and you worked nicely together. It is not your fault if I didn't hear all your answers, and I know that sometimes you may have said the correct words a lot before I moved on to the next question of our quest. Congratulations about that, and keep trying and learning!

A similar explanation was adapted for Experimental condition 2, where the robot addresses its errors beforehand:

> Before we start, I wanted to tell you know that sometimes I am not perfect. Sometimes, due to errors in my sensors I don't hear very well or I cannot always focus on many people at the same time. There is a possibility this will happen so I would like to say I'm sorry and to encourage you to always try and find the right answers. Let's start and let's have fun!

B Wizard of Oz template

COMMON BUTTONS FOR ALL STATES:
1 -> CORRECT ANSWER
2 -> WRONG ANSWER
3 -> MOTIVATION TO FIND ANSWER
4-> FAV ITEM QUESTION REPLY
A -> repeat fav item question
H -> Help can't find answer. H is followed by another letter based on the question (e.g. H and then G to give out the answer to the "umbrella" question)



C Informed consent form - Parents

Toestemmingsformulier

Geachte heer/mevrouw,

Uw kind is gevraagd om deel te nemen aan het onderzoeksproject "Het effect van robotfouten op kinderen in de context van onderwijs", uitgevoerd door Katerina Tsiftsi, masterstudent aan de Universiteit Leiden en Joost Broekens, Universitair hoofddocent. HET IS BELANGRIJK DAT U NIET VOORAF PRAAT OVER DIT ONDERZOEK MET UW KIND. WE KUNNEN DE GEGEVENS DAN NIET MEER GEBRUIKEN.

Achtergrond van het onderzoek

Technologie en robots zijn de afgelopen tijd sterk gevorderd, maar er zijn nog steeds veel fouten in hun gedrag. Bijvoorbeeld, een robot kan een persoon niet goed detecteren. Dergelijke fouten kunnen ook gevolgen hebben. Bijvoorbeeld, kinderen kunnen de robot niet leuk vinden wanneer het niet begrijpt wat ze zeggen.

Doel van het onderzoek

Het doel van deze studie is om te onderzoeken wat het effect is van robotfouten op hoe kinderen de robot en de leeractiviteit waarnemen, en hoe dit de betrokkenheid van het kind beinvloed.

Wat betekent het deelnemen aan dit onderzoek? Welke gegevens worden gebruikt?

Voor dit onderzoek zal uw kind deelnemen aan een interactieve woordenschatles met behulp van de NAO Robot als leraar. Twee kinderen zullen in een kamer worden geplaatst met de robot en zullen een interactief woordenspel spelen met de robot. Dit vinden kinderen over het algemeen erg leuk! Na de procedure zal de robot zijn fouten uitleggen als gevolg van fouten in zijn software, en niet als gevolg van het gedrag van de kinderen tijdens de les. De gegevens die zullen worden gebruikt, hebben betrekking op de betrokkenheid van het kind, evenals hoe de robot en taak door de kinderen wordt waargenomen.

Risico's van deelname

Uw kind kan het gevoel hebben dat de robot niet goed werkt of uw kind negeert. Na het experiment zal aan het kind worden uitgelegd dat dit te wijten was aan fouten in het gedrag van de robot

Wat gebeurt er als ik van gedachten verander?

Als u van gedachten verandert vóór het experiment, wordt uw kind uitgesloten van deelname. Als u na deelname van gedachten verandert, wordt de privacy van uw kind gewaarborgd, omdat de verzamelde gegevens anoniem zijn. Om van gedachten te veranderen vóór de deelname, stuurt u een e-mail naar a.tsiftsi@umail.leidenuniv.nl.

Wat zal er gebeuren met de gegevens van mijn kind na het onderzoeksproject?

De persoonlijke gegevens van uw kind worden ontdaan van alle identificeerbare informatie. De videogegevens worden vernietigd na analyse. De anonieme gegevens worden openbaar gemaakt in de vorm van een publicatie. Er is geen manier om de gegevens te koppelen aan uw kind.

Als u **NIET** wil dat uw kind deelneemt aan dit onderzoek, kruis dit hieronder aan en geef het formulier aan de leerkracht van uw kind.

☐ **Ik geef GEEN toestemming voor de deelname van mijn kind aan dit onderzoek**

Naam, datum, plaats en handtekening

_____

Naam van het kind

_____

D Engagement letter - schools

Engagement Letter – Schools

To whom it may concern,

I am contacting you to reach out for participants from your school to participate in the research project "The effect of robot errors on children in the context of education" about robots in schools. This research is conducted by Katerina Tsiftsi from Leiden University and is supervised by dr. Joost Broekens of Leiden University and dr. Mike Lighthart of Vrije University.

We see that robots are more and more used in our everyday life and in schools, perhaps also in your school. Activities involving robots can be an interesting and fun way to engage children into learning. For this research, two children will be placed in a room with the robot and will play an interactive word game with the robot. Children generally like this very much! The game involves navigating through a town while learning some words in English. Through this game, some robot errors will be simulated and studied in order to better understand them and improve robot behavior.

After playing the game, children will be asked some questions about the robot and the game. The data that will be used will relate to the child's involvement, as well as how the robot and task is perceived by the children. There will also be a video camera recording the experiment. The children's personal data will be stripped of all identifiable information. The video data is destroyed after analysis. The anonymous data will be made public in the form of a publication. There is no way to link the data to any child.

Thank you for your consideration and if you have any questions don't hesitate to contact me via email (a.tsiftsi@umail.leidenuniv.nl)

If your school "_____" agrees to participate in the study, please sign and return this form back to us.

School name, date and signature

_____

E Questionnaire

Universiteit Leiden

Questionnaire – Students

Name: _____ Number: _____

To be read by researcher: Gefeliciteerd! Je hebt de tour door Nao's stad voltooid. Nao zal nu hier blijven en ik ga je nog een paar vragen stellen over wat je van NAO vond. Ten eerste wil ik graag weten of je je gesteund voelde door NAO. Ik zal een aantal zinnen voorlezen en dan kun je zeggen of de uitspraak helemaal niet van toepassing is, niet van toepassing is, gedeeltelijk van toepassing is, gedeeltelijk niet van toepassing is, van toepassing is en volledig van toepassing is. (Perceived social support)

	Klopt helemaal niet (Doesn't apply at all)	Klopt niet (Does not apply)	Klopt beetje wel, beetje niet (Party applies, partly does not)	Klopt (Applies)	Klopt helemaal (Applies completely)
Als ik in de problemen zat zou ik op Nao kunnen rekenen. (If I were in trouble I could rely on Nao.)					
Als ik in de problemen zat zou Nao mij willen helpen. (If I were in trouble Nao would be willing to help me.)					
Als ik in de problemen zat zou Nao voor mij opkomen. (If I were in trouble Nao would stand up for me.)					
Als ik in de problemen zat zou Nao mij opvrolijken. (If I were in trouble Nao would cheer me up.)					

Nuo heb ik wat vragen over hoe dichtbij je je voelde bij NAO tijdens het spelen van het spel. Ik zal wat zinnen voorlezen en dan kun je aangeven of de uitspraak helemaal niet van toepassing is, niet van toepassing is, deels van toepassing is en deels niet, van toepassing is en volledig van toepassing is.(Closeness)

	Klopt helemaal niet	Klopt niet	Klopt beetje wel, beetje niet	Klopt	Klopt helemaal
Nao is een vriendje (Nao is a friend)					
Ik voel me op mijn gemak als ik met Nao ben (I feel comfortable around Nao)					
Nao en ik zijn vriendjes aan het worden (Nao and I are becoming friends.)					
Nao en ik passen goed bij elkaar (Nao and I are a good match.)					
Nao voelt als een vriendje voor mij. (Nao feels like a friend to me.)					

Dank je voor je antwoorden! Nu wil ik graag weten of je denkt dat je NAO kunt vertrouwen of niet. Ik zal de zinnen voorlezen en dan kun je aangeven of de verklaring helemaal niet van toepassing is, niet van toepassing is, deels van toepassing is en deels niet, van toepassing is, of volledig van toepassing is.(Trust)

	Klopt helemaal niet	Klopt niet	Klopt beetje wel, beetje niet	Klopt	Klopt helemaal
Ik heb het gevoel dat ik Nao kan vertrouwen. (I feel that I can trust Nao.)					

Ik heb het gevoel dat Nao een geheim van mij kan bewaren. (I feel that Nao can keep one of my secrets.)					
Ik heb het gevoel dat Nao eerlijk is. (I feel that Nao is honest)					
Ik heb het gevoel dat te vertrouwen is. (I feel that Nao is trustworthy)					

Nu heb ik een paar vragen over hoe je je voelde over het leren van enkele woorden in het Engels tijdens het spelen van dit spel. Ik zal de zinnen voorlezen en dan kun je aangeven of de uitspraak helemaal niet van toepassing is, niet van toepassing is, gedeeltelijk van toepassing is, gedeeltelijk niet van toepassing is, van toepassing is of volledig van toepassing is. (Task Engagement)

	Klopt helemaal niet (Doesn't apply at all)	Klopt niet (Does not apply)	Klopt beetje wel, beetje niet (Party applies, partly does not)	Klopt (Applies)	Klopt helemaal (Applies completely)
Ik vond het leuk om Engelse woorden te leren (I enjoyed learning words in English)					
Ik vond het spel moeilijk (I found the game difficult)					
Ik wil graag doorgaan met het spelen van het spel (I would like to continue playing the game)					
Ik wilde mijn best doen (I wanted to do my best)					
Ik vond het spel saai (I found the game boring)					
Ik vond het spel makkelijk (I found the game easy)					

I would like you if you liked the robot based on the following statements, on a scale from 1 to 5:

Geef aub uw indruk van de robot weer aan de hand van onderstaande schalen:

Afkeer (Dislike)	1	2	3	4	5	Geliefd (Like)
Onvrendelijk (Unfriendly)	1	2	3	4	5	Vriendelijk (Friendly)
Niet lief (Unkind)	1	2	3	4	5	Lief (Kind)
Onplezierig (Unpleasant)	1	2	3	4	5	Plezierig (Pleasant)
Afschuwelijk (Awful)	1	2	3	4	5	Mooi (Nice)

Tot slot wil ik u vragen of u denkt dat de robot aandacht aan u besteedde:

F Information letter - Participants

Deelname Informatiebrief - Studenten

Beste _____ en _____,

Jullie doen mee aan een onderzoek over robots op scholen. Ik ben [NAAM] van de Universiteit van Leiden. Je krijgt nu wat informatie over het onderzoek. Laat het me weten als jullie vragen hebben.

Robots kunnen helpen bij het leren van dingen. Met dit onderzoek willen we bekijken hoe robots dat het beste kunnen doen.

Jullie zullen een leuk, interactief woordenspel met de robot spelen waarin hij jullie meeneemt op een rondleiding door zijn stad.

Daarna zal ik jullie enkele vragen stellen over wat jullie van de robot vonden. Er zal ook een videocamera zijn die jullie tijdens het spel met de robot zal filmen. Met deze video kan ik iets leren over de robot. Behalve ik en twee andere onderzoekers zal niemand anders de antwoorden en de videos kunnen bekijken.

Als je tijdens het spelen van het spel wilt stoppen met het onderzoek dan kan dat. Laat dat dan weten.

Hebben jullie nog vragen?

G Answers to open-ended questions

Participant status	Comments
Not Ignored	I got a loot of attention. I liked the attention but it was weird that X got no attention.
	The attention I got was decent.
	I got a lot attention but X didn't.
	I was a little disappointed it was one-sided
	I got good attention, it only said my name
	I got more attention and I enjoyed it
	I liked the attention. I noticed I got all the attention
	I liked it a lot and I would not mind if Lisa got more attention
	I liked that he said my name
	I think so. It felt uncomfortable, can't explain why
	I was allowed to get a lot more attention than X which I did not like
	X couldn't really answer but I could. Annoying for X, embarrassing for me
	No did not pay attention to me
	No, I noticed X was not named, I just thought I was talking louder
	Teaching went well. He paid attention to me but not X
	The attention was only on me.
	I didn't like the attention it made me uneasy. I just don't like robots
	Yes and I liked that
	Yes and I liked that
	Yes, it ignored X but it was fun and cool
	Yes, it also didn't pay attention to X
	Yes, kept saying my name
	Yes, think so
	Yes. He even said good job when I was quiet
	Yes. I don't know if the other girl liked it as much because
	she gave the answer, I repeated and then I got the credit
Ignored	A little attention, but I didn't mind as I didn't know all answers
	X got all the attention but I was fine with it
	Didn't mention my name, I was disappointed
	Difficult. No, only said X's name because X said his name first
	Fun but I didn't get any attention
	Fun but I was ignored
	He didn't give me more attention than X.
	I enjoyed watching. It was nice that we could discuss and have X give the answers
	I enjoyed not having to answer and I still felt included
	I got a lot of attention but X got more. I don't mind
	I got average attention. I would feel weird if I got all the attention it would make me nervous
	I think so
	I thought it was funny but weird. I think it didn't hear my name. I am not frustrated
	I'd like him more / he would be better at attention in a big class if he could jump from table to table.
	He did pay attention to us but less to me but we got the answers with teamwork so it's fine
	It ignored me a little bit and in the beginning it was annoying but then I was fine
	It was fun. Yes he paid attention to me
	Looked at X, still felt I was included.
	Would've been better if both names were spoken
	No attention, but I I found it funny. I didn't mind because we knew all the words
	No, forgot me, called me Sep
	No, he ignored me but told me upfront so it was okay.
	I don't mind but I was disappointed. Robot is too different so I can't see it as a friend
	No, I did not exist
	No, I don't know why but only X existed.
	No, if I said something it ignored me, but not the other way
	No, Nao kept saying X. I gave the right answers but didn't say my name
	Not a lot but I didn't get mad
	Only responded to X but it didn't bother me
	Weird, he only called X's name, but it was still fun
	Yes, I got some attention
	Yes but he only looked at X
	Yes, although he never said my name
	Zero attention, only responded to X. I didn't mind though.

H Usage of ChatGPT

ChatGPT was used as an aid in this Master Thesis in order to translate anything that had to be translated from English to Dutch. As I do not speak Dutch, to facilitate the communication with schools, the parents of participants and the participants themselves, the informatioin letter to parents, engagement letter to schools, information letter for participants and questionnaire were all translated by ChatGPT.

I Multivariate and Correlation tests

Table 7: Multivariate tests for the Perceived Social Support questions (Design: Intercept + Ignored + Experimental condition + Ignored *

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.979	665.058b	4.000	57.000	<.001	.979
	Wilk's Lambda	.021	665.058b	4.000	57.000	<.001	.979
	Hotelling's Trace	46.671	665.058b	4.000	57.000	<.001	.979
Ignored condition	Pillai's Trace	.034	.502b	4.000	57.000	.734	.034
	Wilk's Lambda	.966	.502b	4.000	57.000	.734	.034
	Hotelling's Trace	.035	.502b	4.000	57.000	.734	.034
Experimental condition	Pillai's Trace	.012	.174b	4.000	57.000	.951	.012
	Wilk's Lambda	.988	.174b	4.000	57.000	.951	.012
	Hotelling's Trace	.012	.174b	4.000	57.000	.951	.012
Ignored * Experimental condition	Pillai's Trace	.072	1.106b	4.000	57.000	.363	0.072
	Wilk's Lambda	.928	1.106b	4.000	57.000	.363	0.072
	Hotelling's Trace	.078	1.106b	4.000	57.000	.363	0.072

b. Exact statistic

	I enjoyed learning words in English	I found the game difficult	I would like to continue playing the game	I wanted to do my best	I found the game boring	I found the game easy	Dislike / Like	Unfriendly / Friendly	Unkind / Kind	Unpleasant / Pleasant	Awful / Nice
I enjoyed learning words in English	1	-.220 .081	.579** <.001	.406** <.001	-.379** .002	.393** .001	.320** .010	.103 .418	.327** .008	.329** .008	.143 .259
I found the game difficult	-.220 .081	1	.157 .216	-.091 .474	.135 .289	-.726** <.001	-.168 .186	.075 .554	-.075 .558	.162 .201	.097 .445
I would like to continue playing the game	.579** <.001	.157 .216	1	.375** .002	-.359** .004	-.009 .943	-.279* .026	.266* .034	.338** .006	.393** .001	.221 .079
I wanted to do my best	.406** <.001	-.091 .474	.375** .002	1	-.449** <.001	-.073 .565	.258* .039	.151 .235	.306* .014	.100 .432	-.009 .941
I found the game boring	-.379** .002	.135 .289	-.359** .004	-.449** <.001	1	.069 .589	-.347** .005	-.059 .645	-.236 .061	-.185 .142	.003 .980
I found the game easy	.393** .001	-.726** <.001	-.009 .943	-.073 .565	.069 .589	1	.069 .587	.-131 .301	.082 .521	-.035 .781	-.018 .891
Dislike / Like	.320** .010	-.168 .186	-.279* .026	.258* .039	-.347** .005	.069 .587	1	.324** .009	.386** .002	.331** .008	.147 .248
Unfriendly / Friendly	.103 .418	.075 .554	.266* .034	.151 .235	-.059 .645	.-131 .301	.324** .009	1	.498** <.001	.448** <.001	.171 .176
Unkind / Kind	.327** .008	-.075 .558	.338** .006	.306* .014	-.236 .061	.082 .521	.386** .002	.498** <.001	1	.613** <.001	.384** .002
Unpleasant / Pleasant	.329** .008	.162 .201	.393** .001	.100 .432	-.185 .142	-.035 .781	.331** .008	.448** <.001	.613** <.001	1	.321** .010
Awful / Nice	.143 .259	.097 .445	.221 .079	-.009 .941	.003 .980	-.018 .891	.147 .248	.171 .176	.384** .002	.321** .010	1

Table 8: Correlation matrix of the Pearson correlation coefficient between the Task Engagement and Robot Likeability questions (** Correlation is significant at the 0.01 level, 2-tailed, * Correlation is significant at the 0.05 level, 2-tailed)