KTH ROYAL INSTITUTE OF TECHNOLOGY

Degree Project in Computer Science and Engineering, specializing in Systems, Control and Robotics

Second cycle, 30 credits

Understanding the Link Between Robots Perspective Taking, and Humans Formulating a Mental Model and Exhibiting Prosocial Behaviour

Master Thesis

JOÃO ALMEIDA

Stockholm, Sweden, 2022

Understanding the Link Between Robots Perspective Taking, and Humans Formulating a Mental Model and Exhibiting Prosocial Behaviour

Master Thesis

JOÃO ALMEIDA

Master’s Programme, Systems, Control and Robotics, 120 credits
Date: October 18, 2022

Supervisor: Elmira Yadollahi
Examiner: Iolanda Leite
School of Electrical Engineering and Computer Science

Swedish title: Förstå Kopplingen Mellan Robotar som Tar Perspektiv och Människor som Formulerar en Mental Modell och Uppvisar ett Prosocialt Beteende
Swedish subtitle: Förstudie

Abstract

For effective Human-Robot Interactions, robots have to achieve the same level of understanding as humans in every aspect. Humans are highly versatile, able to switch between frames of reference and consider multiple perspectives efficiently. Additionally, humans help and care for each other in the hope of contributing to a better society. When we talk about interacting with robots, the uncertainty around how to collaborate with them has become an emerging topic in the past few years. To bring social robots into humans' lives, researchers need to understand the conditions and implications of robots' actions on human responses and perceptions of robots. This project aims at understanding the link between a robot taking a human's perspective and the human's exhibition of prosocial behaviour toward the robot. To test our hypotheses, we have designed an activity where the participant picks an object from an array of objects in front of them, after listening to the robot's descriptions of that object. Divided into three between-subject conditions, the robot took different perspectives to address the objects in each condition. After completing the tasks, participants could either help the robot collect speech data or move on with filling the final questionnaire and finishing the experiment. Our findings show that the participants were significantly more likely to help the robot when the robot took their perspective (human-centred) compared to both the control condition (object-centred) and when the robot did not take their perspective (robot-centred). Additionally, the results show that when the robot's descriptions of an object were ambiguous, in the first instruction, 96% of the participants perceived the robot's first instruction using an egocentric perspective. However, when the robot made such ambiguous instructions again later, participants perceived the instructions based on the mental model developed about the robot for that condition.

Keywords

Perspective Taking, Prosocial Behaviour, Human-Robot Interaction, Implicit frame of reference, Explicit frame of reference, Human's mental model

Sammanfattning

Robotar måste, i alla avseenden, uppnå samma nivå av förståelse som människor, i syfte till att skapa ett effektivt samspel. Människor är versatila, kan växla mellan olika referensramar och beakta flera perspektiv på ett effektivt sätt. Dessutom hjälper dem och även tar hand om varandra i hopp om att bidra till ett bättre samhälle. När man talar om att interagera med robotar har osäkerheten kring samarbetet blivit ett framväxande ämne under de senaste åren. För att få in sociala robotar i människors liv måste forskarna förstå förutsättningar för och konsekvenser av robotars handlingar gentemot människors reaktioner och uppfattningar om dem. Detta projekt syftar därför till att förstå kopplingen mellan en robot som intar en människas perspektiv samt människans uppvisande av prosocialt beteende gentemot roboten. För att testa hypotesen har en aktivitet designats där en robot ger instruktioner till en deltagare att välja ett visst föremål som den beskriver, utifrån en rad med olika objekt. Med tanke på att deltagarna endast tilldelas ett villkor av tre möjliga, tog roboten olika perspektiv för att adressera objektet i varje betingelse. Efter den slutförda uppgiften kunde deltagarna antingen hjälpa roboten att samla in tal-data eller gå vidare med att fylla i ett sista frågeformulär och avsluta experimentet. Resultatet visar på att deltagarna var mer benägna att hjälpa roboten när den tog deras perspektiv (människokoncentrerad) i jämförelse med både kontrollvillkoren (objektcentrerad) och när roboten inte tog deras perspektiv (robotcentrerad). Dessutom visar även resultatet att robotens beskrivning av ett objekt var tvetydliga. I den första instruktionen upplevde 96% av deltagarna att robotens instruktion hade ett egocentriskt perspektiv. När roboten gjorde ytterligare sådana tvetydliga instruktioner, När roboten gjorde ytterligare sådana tvetydliga instruktioner, uppfattades instruktionerna däremot utifrån den mentala modellen som utvecklades om roboten för just det villkoret.

Nyckelord

Perspektivstagande, Prosociala beteenden, Interaktion mellan människa och robot, Implicit referensram, Explicit referensram, Människans mentala modell

Resumo

Para uma Interacção Humano-Robot eficaz, os robôs têm de atingir o mesmo nível de compreensão que os humanos em todos os aspectos. Os seres humanos são altamente versáteis, capazes de alternar entre pontos de referência e considerar múltiplas perspectivas de forma eficiente. Além disso, os humanos ajudam e cuidam uns dos outros na esperança de contribuir para uma sociedade melhor. Quando falamos de interacção com robôs, a incerteza em torno de como colaborar com eles tornou-se um tema emergente nos últimos anos. Para trazer robôs sociais para a vida dos humanos, os investigadores precisam de compreender as condições e implicações das acções dos robôs nas respostas humanas e percepções dos robôs. Este projecto visa compreender a ligação entre um robô que assume uma perspectiva humana e a exibição de comportamento prosocial do ser humano em relação ao robô. Para testar as nossas hipóteses, concebemos uma actividade em que o participante escolhe um objecto de um conjunto de objectos à sua frente, depois de ouvir as descrições do robô sobre esse objecto. Sendo atribuída a cada participante uma e uma só condição entre três possíveis, o robô tomou diferentes perspectivas para abordar os objectos em cada condição. Após completar as tarefas, os participantes podiam ou ajudar o robô a recolher os dados da fala ou avançar com o preenchimento do questionário final e a conclusão da experiência. Os nossos resultados mostram que os participantes tenderam a ajudar mais o robô quando o robô tomava a perspectiva deles (centrada no ser humano) em comparação tanto com a condição de controlo (centrada no objecto) como quando o robô não tomava a perspectiva deles (centrada no robô). Além disso, os resultados mostram que quando as descrições que o robô transmitia acerca do objecto eram ambíguas, na primeira instrução, 96% dos participantes perceberam a primeira instrução do robô usando uma perspectiva egocêntrica. No entanto, quando o robô voltou a fazer tais instruções ambíguas mais tarde, os participantes perceberam as instruções com base no modelo mental desenvolvido sobre o robô para aquela condição.

Palavras-Chave

Tomada de Perspectiva, Comportamento Prosocial, Interacção Humano-Robot, Ponto de referência implícito, Ponto de referência explícito, Modelo mental humano

Acknowledgments

Words cannot express my gratitude to my supervisor Dr. Elmira Yadollahi for guiding me this past semester. For all the long meetings, discussing so many ideas, and your patience in answering all my questions.

I would like to extend my sincere thanks to RPL division for having me, especially to the Social Robotics group that shared your resources, knowledge and joy with me!

Last but not least, I would like to thank my family, parents, sister, and grandmothers for the opportunities, support and cheeriness. Special thanks to my girlfriend, you bring out the best in me.

As this marks the conclusion of my master's degree, I would like to thank all my friends both in Portugal and in Sweden.

To all, thanks again, this journey would not have been the same without every single one of you. Thanks!

Stockholm, October 2022
João Almeida

Contents

List of Figures

List of Tables

List of acronyms and abbreviations

AI	Artificial Intelligence
APT	Afective Perspective-Taking
HCI	Human-Computer Interaction
HHI	Human-Human Interaction
HRI	Human-Robot Interaction
IRI	Interpersonal Reactivity Index
ML	Machine Learning
NN	Neural Network
PPT	Perceptual Perspective-Taking
PT	Perspective Taking
RQ	Research Question
SPT	Spatial Perspective-Taking
ToM	Theory of Mind
VPT	Visual Perspective-Taking

Chapter 1

Introduction

Nowadays, robots are everywhere in all shapes and sizes, from automatic doors or elevators to autonomous cars or remote surgery. The definition of a robot is broad such as "any automatically operated machine that replaces human effort" [1] or "a type of automated machine that can execute specific tasks with little or no human intervention and with speed and precision" *. Every definition includes either probabilistic, deterministic or stochastic processes. The first two processes are bounded, which allows them to be present in society, limited to their specific job. Stochastic or random processes mirror human behaviour. Humans belong to a restricted part of this category since they make aleatory but thinkable actions. Back in 1950, Alan Turing developed a test called *The Imitation Game* [2, 1950] to answer the question: "Can machines think?". After 70 years, there is no evidence that any machine has already passed the test, which implies the complexity of human actions.

Human-Robot Interaction (HRI) involves any interaction between humans and robots to accomplish a task. In particular, this interaction differs from other types of human interactions with machines due to the robots used, since their exoskeleton system mirrors parts of the human's body, for example, an arm robot [3], a human-like robot head [4], or a humanoid robot [5]. This project focus on studying how a complex system such as the human mind reacts to a stimulus from a social humanoid robot. How humans' perceptions and behaviour toward a robot change as a function of the robot's behaviour.

Social robotics is a specific field in HRI since it involves robots with the capacity to engage in social interactions and explicitly communicate with, recognise and learn from each other. There are multiple situations where social robots have a place in human society, for example, in education [6, 7],

* https://www.techtarget.com/searchenterpriseai/definition/robot

especially in language learning [8] or in special education [9]. Furthermore, they can also make a positive contribution in rehabilitation through games [10] and form long-term relationship [11], between others. With the increase in human life expectancy, the portion of old-age people in society is going to increase, and one problem that threatens the elderly is loneliness [12]. Social robotics may have a word in this topic since robots can be in everyone's house. In order to accomplish that, social robots must interact with humans and create a connection.

At this point, one question emerges: Why would we (humans) develop a relationship with a robot designed for a specific purpose? Humans make decisions to serve their interests, and it is this egocentric characteristic responsible for humans helping each other. In a technological future, AI systems must have at least the same capabilities as us and use them to improve general human life.

Prosocial behaviour is when someone seeks to benefit third parties only, from helping a homeless person to philanthropy. Recent studies have shown that subjects who reported more time engaged in prosocial behaviour had higher levels of positive mood [13]. Similarly, other studies concluded that acts of kindness resulted in increased life satisfaction [14]. Jointly with AI and ML algorithms, robots are getting more and more efficient, competent and smoother during interactions. Shum et al. developed an algorithm that generalises the user depending on the cooperation level that they perceive from the robot [15]. These fast perceptions in every human life allow us to distinguish whom to cooperate with and with whom to compete.

Perspective-taking (PT) occurs in every human interaction when each party attempts to connect with the second party. While it is relatively easy for humans to take others' perspectives to understand their feelings, a software-based machine still has to map every piece of data with human labels, which inevitably constrains the freedom of the interaction. For example, humans change the manner they address objects while interacting with a robot compared to another human [16] because their not familiar with the capabilities of the robot. Such gaps still exist and can be closed when a robot develops a connection with a human.

Previous studies grouping PT and prosocial behaviour have already shown some positive interferences. Although PT in competition can activate unethical behaviour such as egoism [17], cheating [18] or negative thoughts or intentions [19], researchers have also reported that levels of PT and prosocial behaviour increased in cooperative PT tasks after mindfulness sessions [20]. In addition, researchers concluded that participants that showed higher levels of PT self-reported themselves with more prosocial behaviour [21, 22] in a Human-Human interaction (HHI). Also, several studies have proven that robots and virtual agents trigger different types of prosocial behaviour, such as: helping, giving more money, and spending more time with the robot [23]. In a Human-Computer interaction (HCI), Herrera et al. concluded that when the participant was able to personalism their avatar in a virtual-reality perspective-taking, they self-reported with higher social presence scores [24].

Altogether, this work will identify how humans prefer a social humanoid robot to address them during an HRI by answering the Research Question (**RQ1**): ”How does the mental model humans formulate about the robot’s perspective affect their understanding of an ambiguous instruction (implicit frame of reference)?”.

Secondly, even though recent studies focused on different types of PT, there is a gap relating PT stimulating prosocial behaviour in collaborative HRI. So we propose to contribute to the research linking robots taking a human’s perspective and humans exhibiting prosocial behaviour by answering **RQ2**: ”How does the perspective the robot takes influence how much humans show prosocial behaviour towards the robot?”.

1.1 Ethics and Sustainability

In terms of ethical concerns, participants agreed beforehand to participate in the first part of the study, however, they were not informed about the second part. Secondly, since participants might not feel totally comfortable with the robots, they had the option to interact with the robot before the real interaction started and they could have left the experiment room at any time. On top of that, participants signed a consent form stating that they had 24 hours to withdraw from the experiment to not count their results. In terms of resources and environment, both the robot and the place were used according to the availability of other parallel projects in the department without any additional costs. The acquired objects were left in the laboratory available to be used in further experiences. The main expense of this project was recruiting participants, however, this type of project has to be tested in real humans without previous knowledge of what this project consisted of. Also, the payment amount was aligned with other in-person user studies running at the same university.

All in all, this project respected ethical rules, toward other professionals in the laboratory as well as toward the participants. In addition, the study was carefully designed to change one independent variable and study the effects on two dependent variables, therefore it justifies the resources used, making this project sustainable.

Chapter 2

Background

This project consists of two main aspects of everyday life that humans cherish in an HHI, PT and prosocial behaviour.

2.1 Perspective Taking (PT)

Perspective-taking is the ability to connect with others and perceive thoughts, motivations and feelings. PT allows one to empathise with the other and take their perspective.
Bearing in mind the notation introduced in [25]: *Self*, *Other*, *Object*, PT means that the Self understands the Other's viewpoint about the Object. This relation is characterised threefold: affective (1) where the self understands the emotional state of the other; cognitive (2) where the self understands the other's knowledge about the object; perceptual (3) nature that represents the other's visual and spatial point of view. The object, in this case, can be a physical object [26], an informative instance (e.g. numeral [25]), memories, a landscape, between others.
The ability that the self realises whether or not the other sees an object characterises the level-1 PT, and determining how the object is perceived means having the level-2 PT. This distinction is not only related to children's development, but also studies showed that chimpanzees [27] and dogs [28], among others, show evidence of level-1, but not level-2 PT. Also, the distinction between level-1/level-2 PT is marked by cognitive resources that allow humans to have a rapid and efficient response in level-1 PT [29]; however this capacity is not available for level-2 PT [30].
Lastly, taking others' perspectives faces some difficulties, depending not only on internal factors such as the word choice (e.g. how to transmit

the perspective) but also on external variables such as the environment or intervenients. Therefore, sometimes ambiguities or conflicts happen.

2.1.1 Ambiguities

Ambiguities are frequently present in real-life situations. Humans give unique references that can describe the object outside the situation context. Even when the self takes the other's perspective, it might not be enough to localise the object, for example, if the reference frame is ambiguous or the instructions apply to multiple target values. Localising the object does not apply only to physical objects but also to beliefs or memories. In 3D objects, some features that resolve ambiguities are colour, shape or size. Ros et al. [31] proposed a method that used visual perspective-taking (VPT) to identify which objects were in the sightline of the human and spatial perspective-taking (SPT) with feature-based descriptions (FBD) to find unique attributes of each object during an HRI.

2.1.2 Conflicts

Conflicts also happen frequently in real-life situations. These involve situations where one intervenient takes the wrong perspective of the other. For example, colour tonalities might be hard to distinguish in some boundaries, or when both parties have different perspectives and one intervenient fails VPT level-1. A known task used to test the conflict of VPT is the director's task. Here the participant is in front of a shelf, where they can see all places. On the other side stands the director, who can only see some places. The director tests the conflict by asking the participant to pick an object. Although the director cannot see that specific object, so the correct pick is the object congruent to both (participant and director)'s perspective [32].

2.1.3 Precautions of Taking Humans Perspectives

However, there are aspects where there is a need to have precautions regarding AI implementations. Taking others' perspectives can make them feel uncomfortable, especially those who do not have experience with robots or technological background, since they have less knowledge about the reasoning behind software-based machines. Under these circumstances, humans may trigger defence mechanisms. An example of a self-defence mechanism is being more egocentric and less prosocial [17] to prevent critics or distrusting others. In the end, this reflects as acting unethical in order to prevent being exploited by others [18].

Lastly, empathy relates with PT, although with PT one is taking another's point of view, although for personal goals. It is similar with APT since both evoke sharing feelings with others and understanding someone else's situation with an emotional state depriving of their own.
For a robot, this means an unscripted interaction. On the contrary, depriving a robot of empathising (trued emphatic behaviour) is setting a limit on its intelligence, but simulating "untrue/controlled" empathic behaviour, researchers can focus on studying its influence in an HRI, according to Malinowska's article [33] about previous empathy research. Leite et al. concluded that: "the developed emphatic model had a positive impact in long-term interaction between children and the robot" [34, p. 338], which suggests an influence of the robot on human behaviour.

2.2 Perceptual Perspective-Taking (PPT)

Surtees et al. [25] distinguished the visual and spatial dimensions of PT into two levels. They conducted a user study with 64 undergraduate students (49F, $\overline{x} = 20.63$ years). Firstly, they conducted the first experiment to filter the participants based on their PT, which involved the task itself but with a cue. The participants that selected the answer "NO", which opposed the cue to the image, plus the participants whose results were outliers, did not pass to the second experiment. The second experiment (n=32, 20F, $\overline{x} = 21.70$ years) used to study the differences in perceptual perspective-taking.
The task involved the participant (self) taking the avatar's (other) visual perspective regarding the object that was only either in front or behind the self. The object was a cube with the character 6/9 written on top. In the experiment, the avatar rotated $\{0^o; 60^o; 120^o; 180^o\}$ taking the participants position as the reference frame. In addition, there were four different conditions to lest the level 1 and 2 of both PPT variants.

2.2.1 Visual Perspective-Taking (VPT)

VPT characterises if and how the self perceives the object is seen by the other. In the study [25], the level-1 VPT consisted of the participant answering whether the avatar was seeing the block or not. In order to accomplish this level, the participant had to confirm that the other could see the object. The answer was correct when the participant answered "YES", and the object was in front of the avatar. On the other hand, the correct answer was ”NO” when the object was behind the avatar. In order to accomplish level-2 of VPT, the participant had to choose the correct number seen by the avatar.

2.2.2 Spatial Perspective-Taking (SPT)

SPT characterises the self's perception of the relative spatial location between the other and the object.
With the same conditions as the VPT's study, Surtees et al. [25] differentiate two levels of SPT regarding the localisation of the object based on the other's perspective. The participant accomplished level-1 SPT correctly answered whether the object was in front or behind the avatar. Finally, the participants accomplished level-2 SPT when correctly identified if the object was on the left/right of the avatar.

Based on both the time used to answer and the error proportion, the authors concluded that only level-2 SPT requires self mental rotations. In addition, they also argued that visual perspective-taking relies on basic SPT. Furthermore, adults showed evidences of unintentional VPT level-1 [35].

2.3 Theory of Mind (ToM)

ToM characterizes the capability of attributing mental states to someone else (e.g. motivations, beliefs, moods) and understanding their actions and reasoning is an essential component of PT.
Theory of mind is a hot cognition process in the human brain since the way humans characterise others is affected by their emotional states. This perception leads to weighting more rewards or punishments [36]. On the other hand, cold processes lead to emotionally neutral decisions, and therefore they have been the main focus of AI in the past. Two main concepts: Simulation-Theory (ST) and Theory-Theory (TT), have been used to develop theoretical models [37]. TT consists of generating theories about an individual and validating them through experience. On the other side, ST consists of taking decisions from others' first-perspective to evaluate their reasoning.
One example of ToM in an HRI is anticipation. Koppula et al. [38] propose a model using the object's affordances that predicted the human's action in a closed scenario with high accuracy. This method implies weighting the possible actions with each object bearing in mind the environment.

2.4 Affective Perspective-Taking (APT)

APT allows the self to express reactions towards the other related to their feelings and behaviour.

It differs from ToM in the sense that ToM deals with situations in which both the self and the other have independent emotions. On the other hand, APT involves tasks in which the self has shown consideration about both intervenient's emotions simultaneously. Additionally, APT requires understanding the cause and ToM the performance of the other.

Harwood and Farrar [39] studied the difference between theory-of-mind and affective perspective-taking tasks. A total of 46 children (20F, $\overline{x} = 3.83$ years) took part in the study, consisting of 3 different tasks, the APT task, the physical and social ToM tasks.

In the APT task, participants had to identify the feelings of both the self and the other. Bearing in mind the age of the participants, only opposite sad/happy feelings were measured due to their simplicity of identification. The participants identified correctly 67.1% of the self's and the other's feelings in 12 different scenarios.

The theory of mind tasks consisted of predicting the other's false perceptions. For each task, there were control and experiment questions. In the control questions, the interviewer showed the participant a transformation in the scene and made sure they understood it. After this, the interviewer asked the participant how they thought the other was feeling about the object.

In order to accomplish the ToM physical task, for the control condition, the participant had to realise that the object inside a box was not its original content after seeing the interviewer changing it. Additionally, in the experiment condition, the participant had to exploit the previous reasoning to answer what they thought the other would feel in the same situation. The ToM social task consisted of understanding how the participant (control) and the other (experimental) felt after the other left a room and the activity had changed. In the ToM tasks, the participants gave 33.6% of valid answers according to the situation.

The results mark a big difference between APT and ToM tasks in children. In addition, the authors generalised that the lack of distinction between tasks was one reason for incoherent results in previous studies. Furthermore, they concluded that the scenario where APT is more correlated with ToM tasks is when the self and the other have different emotions. On the other hand, APT is weakly correlated with ToM when both the self and the other are happy [39, tbl 4].

2.5 Reference Frame and Viewpoint Rotations

When giving directions, every indication needs a frame of reference. This frame of reference maps the ideas of the self to the real world that the other understands.
There are two main domains of reference frames: relative or object-centred. According to the relative reference frame, the speaker addresses the object in either the self's (**egocentric**) or the other's (**addressee-centred**) perspective. Secondly, when someone is giving instructions, those can be either **explicit** (commonly contains words such as "my/your") or **implicit** when humans suppress those references to avoid repetitions.
The object-centred references can use the object-intrinsic characteristics to express the indication, such as the colour, shape, size, or any particular feature, for example, the left-front wheel of the car. In addition, the object-centre reference frame can also use object-extrinsic characteristics; for example, words such as "north/south/east/west/between" follow the object as the starting point. These attributes give no information about the object but use its position as an anchor.
These reference frame theories contribute to researchers developing spatial descriptions in robots. For example, Doğan et al. created a relation presence network (RPN) to retrieve the probabilities of the spatial relation between a pair of objects and a relation informativeness network (RIN) to decide if the spatial relationship is informative [40]. The output of this NN is a relative explicit instruction about an object with the other object as the anchor. The descriptions available to reference the target object in this NN are sixfold: to the right, to the left, on top, at the bottom, in front and behind, regarding the reference object.
Throughout interactions, humans formulate mental models [41] of the other (i.e. through ToM), which can be, for example, how is the other instructing the self. Alongside the formulated mental model, humans apply self rotations to visualise the instruction resulting from others' taking self-perspective. There are three conditions of rotation strategies. The first one is the self-mental rotation that rotates embodied representation of the self to the target perspective. On the other hand, with object-mental rotation, the self rotates the perspective of the target object to their perspective. Lastly, mental viewpoint rotation happens when the self does not change their position but takes the perspective from another place.

2.6 Examples of Perspective Taking in Human-Robot Interaction

PT has been a hot topic in robotics this century. Therefore, there is a tremendous variety of tasks used in HRI.

To study the influence of changes in humanoid robot's cognitive-affective state on children's behaviour, Yadollahi et al. [42] proposed a game where a child partners with the robot NAO [43].

The interaction consisted of each counterpart giving instructions to the other to reach a goal. However, when the child was egocentric, the robot failed the next move on purpose to show unpleasantness and frustration only for the experimental condition.

In this study, the authors hypothesised that the robot that did not act frustrated was the most consensual and that children would be more addressee-centred when the robot acted frustrated.

Zhao et al. [44] studied the influence of nonverbal behaviours in VPT.

In their experiment, participants answered the question *"Which number is on the table?"* based on a photograph. The photograph depicted either a human or the humanoid robot NAO [43] seated across a table with the digit "9" on it. In the control condition, an electric guitar substituted the robot. The baseline condition depicted an absence of a human and chair. Both the human and the robot had three different poses: looking away from the digit, gazing at the digit, and reaching the object while gazing at it.

The results support the idea that people take humanoid robots' perspectives, especially when they display referential gaze and goal-directed reaching behaviour.

Shi et al. created five different scenarios where the participant had to refer objects to the partner [16]. The type of partner (either a humanoid robot or a human) marked each condition.

Overall, they found that humans prefer spatial descriptions (e.g. closer to) when it is obvious. However, when the description was not clear or existed some intrinsic characteristic, participants opted for a reference frame rather than a spatial topology, even though the usage of this one was still stable.

Meanwhile, participants that partnered with the robot concentrated more on addressee-centred. Contrarily, participants that partnered with another human revealed more egocentrism in their speech. According to the authors, the principal explanation for this fact is the lack of responsiveness of the robot. In the other condition, the partner had a basic level of human cognition, which can transmit more versatility, despite the robot's capability.

Doğan et al. [26] studied the influence of PT in improving the performance of basic human basic tasks with adults.

In their study, the robot was in front of a round table with different objects on top of it. For each new object, the participant moved around the edge of the round table to renew their perspective. The interaction consisted of the participant picking the object the robot instructed. There were two different conditions based on the type of PT of the robot. In the control condition, the robot was egocentric (robot's perspective), and in the other condition, the robot was addressee-centred (participant's perspective). Note that the robot was always egocentric in the instructions (e.g. "... I/you see ...").

The study's authors hypothesised and verified that a humanoid robot like Pepper [5] helps humans spend less time on the task, commit fewer mistakes and perceive the task more easily when it successfully takes a human's perspective.

Sandra Okita [45] studied the impact of reducing pain and anxiety in pediatric patients and one respective parent in a shared common experience. In the experimental condition, the seal robot generated perspective-taking in the parents by intercalating the robot with their child. In the control condition, only the child interacted with the robot.

Each parent completed a form with their empathetic pain and others from their child's perspective. The author found better results regarding the effectiveness of the robot when the parent was also present in the room. The results show that the patient felt a reduction in pain after the experiment. Furthermore, their parents perceived lower pain levels on the part of the children.

Reidy et al. [46] proposed a method to study the influence of PT in shaping the levels of human-likeness and trust in HRI. Focusing on social robots with higher levels of human likeness, the researchers claimed that higher levels of human-likeness can let humans set their expectations about the robots in iterations. Their goal was to create a method that would let humans take its perspective before the proper HRI in order to smoothen and boost the experience.

They conducted a mix-subject model study where each participant would either take or not take both the human (control condition) and the robot's (experimental condition) perspectives. The experiment consisted of the participant reading a story from the agent's perspective. Despite all stories being different, they all share similar elements in which the agent tries to accomplish a goal but has some problems momentarily.

The authors hypothesised that taking the robot's perspective would impact the rate of the level of human-likeness and trustworthiness of the robot. The results supported the first hypothesis for human likeness but not for trust. According to the authors, the fact that the agent had a setback might influence its credibility.

Traffon et al. [47] centred their studied in humans' preferences instead of performances. They divided their studied in a VPT and SPT tasks.
In the first experiment, a robot was in front of a human in a big room. The robot was able to see two traffic cones, although the human could only see one and had no idea about the second one. Participants were divided into two conditions, either to take the human or the robot's perspective. Since participants were aware of both cones in the room, they classified which action the robot should take when the human asked the robot to go to the cone. The results showed that participants would prefer the robot to either go to the cone visible in both perspectives (robot taking human level-1 VPT) or ask for clarification (e.g. which cone). There was a consensus among both conditions that the robot should not go to the hidden cone.
In the SPT task, the only traffic cone present in the room was visible to the human but hidden to the robot. When asked to go to the cone, the robot's cognitive model had to determine that it did not meet its level-1 PT, so it had to change to the human perspective. So, this task included predicting any hidden space where the object might be. The final step is determining where the robot should move to finish the task.

2.7 Prosocial Behaviour

Prosocial Behavior is the core of personal relationships. It is any variety of voluntary action that help others, strictly at one's cost (e.g. time, money, effort) while not being rewarded. However, one behaves prosociality because it has direct benefits. Raposa et al. reported in their study (n=77 (41F), $\overline{x} = 24.52$ years) that people who demonstrate more prosocial behaviour by helping others experienced bad moods less frequently and reduced the impact of stress in their lives [13]. Furthermore, other researchers found benefits of prosociality such as, more academic achievements at school [48], foster both social integration [49] and peer relationships [50], higher levels of happiness [51], between others. Several causes stimulate prosocial behaviour. Firstly, people can behave prosocially to obtain the benefits mentioned before. Secondly, humans are encouraged by others, for example, children's education [52]. Thirdly, by receiving help, one increases one's willingness to act reciprocally. However, by observing acts of prosociality, people act more prosocial towards those who do not receive help [53].

2.7.1 Measures

Measuring how prosocial a person acts depends on several factors, and each case is different. Baumsteiger and Siegel proposed a scale in order to measure the prosociality of the action of each participant [54]. They classified its magnitude into three levels, the recipient of help in five levels and the cost of helping in three levels. Additionally, they added an attentional check question to test how concentrated the participants were. Moreover, another strategy found to analyse prosocial behaviour was adding total optional questions.
Another test used mainly in the psychological field is the interpersonal reactivity index (IRI) [55], which measures empathy in a human being. It contains four different scales: perspective-taking, fantasy, empathic concern and personal distress, which evaluate the human in different dimensions of empathy.

Moreover, measuring prosocial behaviour in HRI is even more complex, although researchers generally agreed three to use three main techniques. The simplest and oldest method is answering a self-report, for example, [22] where participants answered the prosocial behaviour and peer problem questions of the SDQ questionnaire [56]. However, this technique suffers from a lack of explainability since it is very close to the phenomenon under study [57]. Another technique consists of performed-based tasks. In [58], the robot had two moveable obstacles in front of it and gestured accordingly in order for the participant to realise that he needed help. However, human reactions cannot be binary classified since HHI are continuous, and also, it creates ambiguities once there might be more than one correct action.
Finally, many studies use third-party observations. In [20], an independent supervisor observed every children's behaviour without any information about the experience condition of the children. However, this technique collides with the intention of research and developers, creating more and more autonomous systems able to interact with everyone with no human intervention.

2.8 Examples of Prosocial Behaviors in Human-Robot Interaction

Stimulating prosocial behaviour is also present in HRI. Researchers argue that interactive agents can engineer environments appealing to individuals toward prosocial behaviour [59]. Fogg proved that a software-based machine has influence, convinces and stimulates prosociality [60] by changing its arguments. He proved that the computer, by prompting the same information more or less informative, users preferred the computers matching their personality. Secondly, users classified the computer smarter when they thought they were teammates. Lastly, users that worked with a helpful computer in the first task spent more time creating a colour palette in the second task.

Hayes et al. [61] tested the effects of empathy in participants showing prosocial behaviour by trying the participant to reduce their performance and in which circumstances to help the robot avoid a punishment.
The task consisted of stimulating prosocial behaviour in participants by counting the objects in the tray slower. Participants thought they were testing a new computer vision algorithm and that their results would be the baseline. Apart from the control condition, where the robot did not stimulate empathy, there are two other conditions. Firstly, in the ”self-direct” condition and the robot petitioned the participant to reduce their performance to avoid punishment. The second condition was ”externally-directed” and the robot petitioned on behalf of its programmer.
The results falsified their hypothesis regarding prosocial behaviour. The authors predicted that the ”self-direct” condition would make participants empathise the strongest with the robot. That condition had the same results as the control condition, although the ”externally-direct” condition was significantly more likely to assist the robot compared with the others.

Avelino et al. [58] researched the power of a handshake in an HRI by proposing this human introduction in the first moment when the participant met the humanoid robot.
The prosocial behaviour test removed the obstacle in front of the robot during a joint task. The condition division was according to whether the robot did a handshake (experimental condition)or not (control condition). In order to ask for help, the robot gestured and attempted to navigate through the obstacle but never verbally asked.
The authors concluded with the study that the robot greeting participants with a handshake would be perceived more positively, as they hypothesised. However, they also hypothesised that participants in the experimental condition would be willing to have more prosocial behaviour, which was not verified. On the other hand, participants in the handshake condition showed more willingness for future help.

Martin et al. [62] studied the kindness of children in helping the robot by

picking up a fallen stick. The authors justified that previous studies showed this children's characteristic, although centred on humans, so their goal was to investigate the application to social humanoid robots.
In the experiment, participants witnessed the robot dropping a stick. In the study, there were two conditions, whether the NAO robot [43] seemed to drop the stick intentionally (experimental condition) or unintentionally (control condition). In the experimental condition, the robot tried to reach the stick, alternated the gaze from the stick to the participant and verbally expressed possession of the stick. In the control condition, the robot looked in the opposite direction of the stick. Furthermore, the robot moved its head down while looking away from it.
Both the observed results and the post-test interview showed concordance about the children's behaviour, concluding that the results also support the HRI. Note that how the robot behaved can be seen as a hint to children to take their perspective.

Kleijn et al. [63] studied people's fairness performance and strategic and altruistic behaviour towards multiple opponents in two economic games.
In the experiment, each participant had to interact with a random partner. There were four different partners: a human, a hexapod robot, a humanoid robot and a laptop computer. In the first game, participants received a certain amount of money and had to split some with the robot and keep the rest for themselves. In the second task, participants had to propose an offer to the partner, and participants would only keep the remaining money if the partner accepted the offer.
The results support that the exterior appearance of the partner has a direct impact only on the altruistic behaviour of the participant. On the other hand, the level of anthropomorphism with which the participant rated the partner influences the fairness preference and strategy.

Chernyak et al. [64] studied how the manner of introducing new technologies to children impacts their prosocial behaviour towards them, among other processes.
Participants interacted with a robot dog which either moved autonomously or remotely controlled. The prosocial behaviour measurement consists of the participant's engagement level in prosocial behaviour towards the robot by giving it a sticker or a ball.
The results showed that participants who owned a dog showed differences between both conditions. Participants who owned a dog and interacted with the robot dog with autonomous movements showed more prosocial behaviour. On the opposite, children who did not own a dog showed no differences in

conditions.

Kühnlenz et al. [65] researched how an emotion adaptation approach stimulates prosocial behaviour towards a robot in a HRI.
The experimenters conducted a 2x2 between-subject design, where the dependent variables addressed whether the robot aligned its mood explicitly or implicitly with the same mood as the participant. With the ”explicit” condition, the robot stated that it was in the same mood that the participant was. On the other hand, when the robot was implicit throughout the interaction, it would adjust its facial mood to the participant’s mood.
In order to evaluate the prosociality of participants towards the robot, participants had to label a maximum of eighty pictures. The results showed that participants with both explicit and implicit conditions labelled more pictures, followed by the participant with the explicit condition only. On the other hand, participants within the condition of no emotion adaptation showed the least prosociality toward the robot.

Chapter 3

Methodology

3.1 Study-Design

In this study, we used a between-subject design with a ratio of 1 participant to 1 NAO robot [43]. The level of autonomy of the robot was semi-autonomous since it conducted all steps of the interaction. However, experimenters relied on Wizard-of-Oz to trigger responses to any question prompted by the participant that was not accounted for in the interaction flow or if the robot was not recognising the participant's speech. In the HRI, both the participant and the robot developed a collaborative task where the robot instructed the participant to pick a specific object and received the input data from its forehead camera.
This study consisted of two different tasks, the perspective-taking task and the prosocial behaviour task. However, the second task was considered voluntary since deciding to take on the prosocial task was a measure to evaluate participants' prosocial behaviour. This study is a small example of real-life situations between two interveners with little or no information about each other. During the interaction, the robot exhibits slightly different manners of telling the human how to locate specific objects [16]. We evaluate how these nuances in the instructions impact the exhibition of prosocial behaviour in humans.

3.1.1 Experiment procedure

The experiment starts with the participants signing a consent form to let the experimenters collect and save the data obtained during the experiment. In this regard, participants have 24 hours to announce if they want their data erased from the database of the experiment. Then participants read the rules, and the examiner demonstrated part of the tasks with the robot. After the demonstration, participants saw the experimenter put the voucher inside the box. Finally, participants had some time alone with the robot being able to simulate scanning objects and get used to the environment of the interaction before they started the task whenever they felt ready.

The experiment started with the PT task, which consisted of listening to the robot describing an object that contained the required landmark and selecting the object referred to by the robot and scanning the landmark. In the case of the participants selecting the wrong object, the robot advertised that it was the incorrect object and repeated the instruction. After 15 correct objects showed to the robot, the robot asked the participant for two random objects to make the task look realistic. Then, the robot would tell a three-digit code to the participant in order for them to open the locker and the box. Inside the box, it was a ReadMe message underneath the voucher. The ReadMe message specifically mentioned that participants could finish the experiment at that point. Otherwise, if they wanted to continue, participants had to touch the head of the robot again. The prosocial behaviour task consisted of reading an endless dataset of sentences. Lastly, participants had to fill out a five-section questionnaire regarding: demography, personal traits, PT task, prosocial behaviour task and a feedback part.

In the final questionnaire, participants answered three out of four scale items of the IRI questionnaire [55]. To evaluate participants' traits and ensure that the groups are balanced, participants answered the "Perspective-Taking", "Empathic concern", and "Fantasy". These results ensured the data balance, so any output changes only depend on the study's dependent variables.

3.2 Perspective Taking Task

The PT consisted of understanding a set of 15 instructions asked by the robot. The instructions are present in the tables 14 and 15. In each iteration, the robot asked the participant to choose one object. In order to validate the choice, participants had to scan the landmark (e.g. see fig 3.1) at the bottom face of each object. The robot concluded by expressing if the object was correct or not. When participants scanned the wrong object, the robot repeated the instruction, and the round restarted. After a minimum of 15 iterations, the robot asked the participant to show any two objects to decipher the code of the locker. Finally, the robot presented the participant with the 3-digit that opened the locker. The participant had 15 seconds to scan the landmark before the

robot repeated the instruction for each iteration.
Throughout the interaction, the robot followed the head movements of the participant in order to appear more sociable. Therefore if the participant wanted to pause the interaction, they could do it by turning their back on the robot and resume whenever they wanted.
In order to mark which perspective the robot was taking, the robot said three instructions with an implicit frame of reference (ambiguous) throughout the interaction. Each implicit instruction would lead to two different objects. Although the robot was able to resolve the ambiguity, the robot only resolved it when the participant asked for clarification, for example, by characterising the size of the box (e.g. "the small box") when initially only mentioned "the box".
To study the formulation of the participant's mental model, the first, the middle and the last instruction were ambiguous. Furthermore, the first instruction being ambiguous allowed the experimenters to study how participants perceived an instruction from a social humanoid robot without previous knowledge (i.e. having a mental model already formulated). Also, having instructions with an implicit frame of reference allows experimenters to evaluate how explicit vs implicit instructions with a frame of reference were influencing the time to understand the instructions.

Figure 3.1: A participant scanning the landmark

3.2.1 Ambiguous instructions

In this project, the ambiguity in some instructions is caused by having the frame of reference implicit. Compared to an instruction with the frame of reference explicit (e.g. "... the box on **your** left."), the same instruction is ambiguous by being formulated like "... the box on **the** left.".
Secondly, a pair with a non-ambiguous instruction that matched the same information was created to analyse the time of reacting to the frame of

reference information after a mental model is formulated (**H1c**). Each pair follows different rules in its construction (see table 3.1). The first pair includes instructions eight and fourteen, and pair two contains instructions four and fifteen.

Table 3.1: Particularities of the pairs ambiguous and non-ambiguous instructions.

Condition	object-centred	robot-centred and human-centred	
	Object characteristic	PT type of the first appearance	Number of instructions in-between
Pair 1	extrinsic	implicit	5 (0 ambiguous)
Pair 2	intrinsic	explicit	10 (1 ambiguous)

3.2.2 Set-up

The experiment set consisted of a line of 10 geometrical 3d printed objects:

- 1 small black box, 6x6x4 cm;
- 1 big black box, 6x6x6 cm;
- 1 small blue square pyramid, 6x6x4 cm;
- 2 big blue square pyramids, 6x6x8 cm;
- 3 small black triangular pyramids, 6x6x4 cm;
- 1 big black square pyramid, 6x6x8 cm;
- 1 big black triangular pyramid 6x6x8 cm.

Behind the line of objects, there was a NAO robot with a locked box placed in front of its legs. There was a screen on the right side used for the second task. Behind the screen, there was a camera that allowed the examiner to use the wizard-of-oz commands. Figure 3.2 represents the set-up seen from the participant's point of view.

Figure 3.2: Participants' point of view during the perspective-taking task.

3.2.3 Wizard-of-Oz

The interaction depended on wizard-of-oz to trigger responses when the speech recogniser did not decipher the participant's speech. There were some instances where the experimenter could trigger the robot remotely. Firstly, at the beginning of each task and the end of the interaction in case of, participants forgot to touch the robot's head. Then, while showing objects, the experimenter could make the robot repeat the instruction or clarify it, depending on the participant's question. Another option was making the robot say that it did not know how to answer the participant when the question concerned a particular aspect (e.g. asking for another characteristic of an object). Finally, the robot could be triggered to politely deviate the focus of a participant's question non-related to the experiment.
Another instance when the wizard-of-oz could be triggered was when participants were showing the last objects to get the code or when participants were opening the box. When the participant did not understand the instruction, the robot could repeat the past intervention if participants asked for it.

3.2.4 Experimental Conditions

In this study, there were three different conditions. In order to study the influence of PT on humans, one condition represented the control group.
In the control or **object-centred** condition, the perspective-taking taken by the robot was object-centre alongside intrinsic characteristics. Each instruction started with the goal object, followed by a pause and then an anchor object that related the goal object to the user. Note that the first segment of the instruction leads to ambiguity, and the perspective-taking comes in the second segment of the instruction. For example, bearing in mind the set-up (Fig 3.2), in order to reference the "big blue pyramid" on the centre-left of the image, the robot said: "Go with the big blue pyramid, on the side of the big box." The first segment: "... the big blue pyramid ..." creates an ambiguity since there are

two possible objects. Therefore the second segment: ”... on the side of the big box.” is necessary for the instruction because it leads the participant to choose the ”big blue pyramid” nearest to the big box. Another word to connect both segments was: ”beside”.

The second or **robot-centred** condition was when the robot gave the instructions from its perspective. The same reasoning applied before to refer to the objects is also seen in this condition. However, sometimes the robot used its position as the frame of reference. For example, to refer to the same object as above, the robot said: ”Go with the big blue pyramid, on my right.”. In this case, the robot used its physical appearance as the frame of reference, and participants had to rotate 180 degrees to the robot’s perspective to choose the correct object. Other instructions also included expressions such as: ”...that I see on the left/right...” or ”... on my left/right...”.

The third or **human-centred** condition consisted of taking the participant’s condition. This condition is identical to the robot-centred condition, although the words: ”I”, ”my”, ”left”, and ”right” were swapped with ”you”, ”your”, ”right”, and ”left”, respectively. The same example as above, in this condition, was: ”Go with the big blue pyramid, on your left.”.

Lastly, in the robot-centred and human-centred conditions, some first segments could not contain an intrinsic characteristic because it would let the instruction be solved not by the SPT. For example: ”Go with the object ...” instead of ”Go with the pyramid”. This change happened because, in the robot-centred and human-centred conditions, the SPT should be the only one responsible for preventing ambiguities. For example, when the participant located the anchor object in the object-centred condition, an intrinsic characteristic of the goal object is essential to decide whether each side object is the goal object.

3.2.5 Objective Measures

In the PT task, several results have been saved, particularly in order to measure the influence of the PT by the robot over the interaction. For every instruction, the metrics saved were:

1. Number of mistakes: number of incorrect objects picked by the participant;

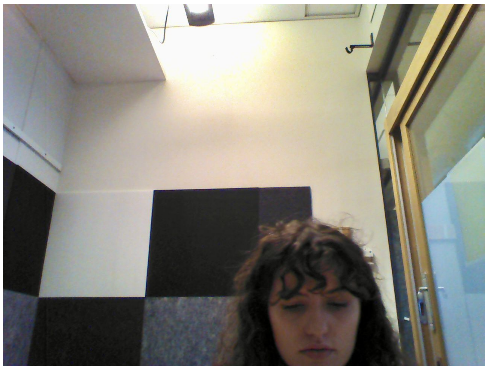
2. Number of clarifications: number of times participant asked for clarifications;

3. Number of tries: number of participant’s reactions either scanning objects or verbally;

4. Time: the amount of time the participant took until scanning the correct object. (it does not include the time when the robot was speaking nor when the participant paused the interaction).

In addition, for each ambiguous instruction was saved participants' mental model (i.e. egocentric or addressee-centric). Moreover, the saved metrics file also includes the amount of time the PT task took and a photo of the participant at the moment they were opening the locker. This last metric allowed the experimenter to check if the participant's face was in the centre of the image, meaning that the robot was correctly following their face.
Note that this check detected the technical failure that led to the removal of one participant. During the interaction, the robot stopped moving its head, which made the task unfairly harder for the participant and, therefore, affected the results regarding the perceptiveness level. In the figure 3.3, it is possible to see two examples of the robot detecting participants' faces and ensuring the robot's face was framed with the participant's face. However, in 3.3a the robot was moving its head to frame it according to the participant, which did not happen in 3.3b

(a) A successful example of the robot ensuring that its face is framed with the participant's face.



(b) An unsuccessful example of the robot ensuring that its face is framed with the participant's face.

Figure 3.3: Examples of the metric to ensure that the robot followed the participants head movements.

3.2.6 Subjective Measures

Regarding the subjective analysis, participants answered eight questions about their perception of the HRI. Although, since the completion rate of the task is 100%, and there was no inference of the robot that could change the flow of the task, experimenters did not focus too much on participants' feelings for the robot (e.g. RoSAS [66]). Instead, participants answered some questions that made the bridge to the analysis of the second task (prosocial behaviour). Regarding the first dependent variable: ”understanding the robot's perception”, participants rated from 1 to 7 on how easy they followed the instructions and explained any problems during the PT task. In addition, participants selected how they would instruct the robot to pick an object (chose between options used in the PT) and rate the robot's communicative skills.

3.3 Prosocial Behaviour Task

The second task measured how much the participant helped the robot by reading sentences aloud to it. The corpus used was presented in [67] as an open-source emotional speech corpus for HRI applications. The present study used the corpus similarly to the original study. The corpus contained 150 unique combinations of sentences and emotions, and after, it would resume to the beginning. There were ten different emotions, five principal emotions: neutral, sadness, anger, neutral, excitement, and five secondary emotions: apologetic, enthusiastic, worried, pensive and anxious. The sentences used for the principal emotions were the same, and ten of these sentences were also used for the secondary emotions. The five sentences remaining for each secondary emotion were specific for each emotion. In the table 18 is an example of the sentences used for the main emotions.
The robot mentioned that the task was voluntary, and the participants could leave whenever they wanted. In order to advance in the task, participants had to check the sentence they read aloud and then click the ”next” button for every sentence. In addition, the second option prompted every time was ”Finish the task”. This mechanism ensured that participants constantly chose to help the robot and that they had the option to leave whenever they wanted.

3.3.1 Set-up

The second task started right after the participant opened the box. Inside the box, there was a well-signalized ReadMe message that all participants had to read. In figure 3.4 it is possible to see a participant reading the message. The message stated that the robot did not adapt its voice according to the participant's answer in the previous task. In addition, the message said that to improve the robot's speech, it had to collect an enormous amount of samples from different users, which led the participants to realise that, even though they contributed, it was impossible to do it alone. In the end, participants could choose to leave the room as a signal to finish the experiment or touch the robot's head to show interest in helping the robot.

Figure 3.4: A participant reading the ReadMe message.

When the participant touched the robot's head, it mentioned that the task was voluntary and that they could leave at any time. Also, it explained what the task consisted of and finally led participants to switch on the screen to start the task. On the screen, the task's environment was already open. On the front page were written the instructions again, mentioning that the participant could leave whenever they wanted. Finally, the message on the front page stated that the task would start when they touched the top robot's head again. The figure 3.5 is the view of the experimenter screen, recording the participant interacting with the environment of the prosocial behaviour task.

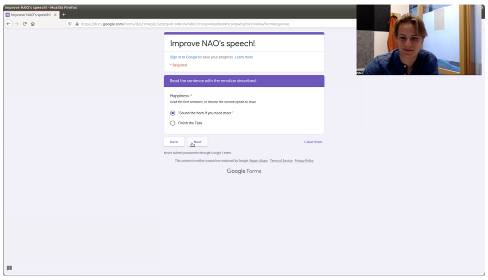


Figure 3.5: A participant during the prosocial behaviour task. He was the participant who read the most sentences, completing all corpus, 150 sentences.

3.3.2 Objective Measures

The number of sentences participants read aloud with emotions to the robot meant how much they helped the robot. Another metric used was the time spent, but it is less accurate since it depends on the pace of reading and how much effort they put into the emotions. In addition, since the task was only explained by the robot and from the information available on the computer, participants spent different amounts of time understanding the task and not just helping the robot.

3.3.3 Subjective Measures

To study the second dependent variable: ”prosociality towards the robot”, the questionnaire contained five linear-scale questions from 1 to 7. The questions answered by the participants regarding how much: care, likeable, and empathetic the participants found the robot. In addition, how great participants perceived the need for the robot to collect speech data.
Finally, the questionnaire contained three open questions concerning: why participants helped the robot, what made them stop the task, and how impactful they think social robots may become in the future.

3.4 Technical Implementation

This section explores the technical implementations of this project. As written before, the interaction was designed to have two tasks (flow: 1), but in case participants did not want to continue to the second task, this part would be skipped and participants would answer the final questionnaire.
The program could run autonomously, however, to be more reliable in reacting to participants' verbal input, it was designed with a parallel Wizard-of-Oz option, to trigger mainly the repetition of the instruction, or to clarify the ambiguous instructions. This option was only available when the robot was waiting for the participant to scan the landmarks, regarding the PT task. Figure 2 represents the flow of the PT. In addition, answers, when the participant was deviating from the task (e.g. occasional speech), could also be prompted.
Regarding the prosocial behaviour task, (flow: 3), the experimenter could trigger the next state in case the participant missed the touch on the sensor on top of the NAO robot's head.

Secondly, it was used five different API models to set the NAO robot for the interaction.

- The camera module was responsible to scan the landmarks (the landmarks technology used was the version up-to-date in-built on these robots). Moreover, this module allows taking a picture of the participant used to evaluate the performance of the face tracker of another module;

- The movements module was responsible to track and follow the movements of the detected human face. Also, this module was configured to place the robot in the ”sit” position which restricted the automatic reflexes of the lower body of the NAO robot;
- The speech module was responsible to convert the text to speech on the NAO robot and recogniser the verbally inputs of participants;
- The LEDs module was responsible to change the colour of the LEDs on the eyes of the robot. This module fades the LEDs green alongside positive feedback and red alongside negative feedback;
- The ReactToTouch module was responsible to detect touches on the tactile sensor on the centre on top of the NAO robot's head.

The code was organised as the first two modules were running continuously in the background and the other modules were called when the robot needed that input/output. The full code is available online in this repository.

3.5 Hypotheses

To answer the research questions and to bear in mind the design of the PT, we have developed some research questions. The first set of hypotheses focuses on understanding how humans formulate a mental model of the robot. They aim to help answer **RQ1** and were inspired by fostering mental models in HRI [41, 68] each is the base of the principals to efficient in HRI [69]. The hypotheses centred on the formulation of the humans mental model in a HRI [70] aligned with the principle of least collaborative effort [70] when humans minimise the effort spent.

- **H1a**: Before formulating a mental model, humans interpret an ambiguous instruction with an egocentric perspective.
- **H1b**: After formulating a mental model, humans interpret an ambiguous instruction aligned with the robot's previous instruction.
- **H1c**: After formulating a mental model, human interpretation time is not affected by the ambiguity in the instruction but by the progress in the interaction.

In order to answer **RQ2**, we were inspired by [65] to create an endless task to spend time helping the robot.
Studies showed levels of PT and prosocial behaviour being positively correlated [21, 22, 20] since humans perceive the task differently (e.g. committing more mistakes) when the robot takes their perspective compared to an object-centric perspective [26]. Additionally, when the robot only takes its perspective is seen as egocentric, which excites negative feelings in humans. Altogether, we hypothesise that the robot-centred and the human-centred condition suscitates opposed levels of prosocial behaviour compared to the object-centred condition.

- **H2a**: Humans are more prosocial toward a robot that only takes the human perspective compared to an object-centred perspective.

- **H2b**: Humans are less prosocial toward a robot that only takes the robot perspective compared to an object-centred perspective.

3.6 Participants

This experiment included 70 participants (39M, 31F) recruited by online publications or posters hung around KTH's main campus. This number was a balance between 3 main reasons: the minimal number of participants used in previous similar studies (approx. 20 per condition) to obtain plausible results, the time designated to run the experiment and the amount of money it was sustainable to spend on this project. One participant was excluded from the analyses during the experiment, and only 69 (39M, 30F) others were used. The average age of the participants is 24.8 years (SD=4.2), mostly students from a master's program at KTH. From at least 23 different backgrounds, 22 have an engineering background (Fig. 3.6), mainly computer science. Participants came from 18 countries (Fig. 3.7) and 4 continents. 61% of participants had prior interaction with a social robot on average two times, and the rest were interacting with the robot for the first time. Each experiment took around 30 minutes, and participants received a 100 kr voucher for participating in the study.

The 69 participants were distributed randomly to each condition, resulting in 23 per condition. In order to check if the conditions were balanced, participants answered the three sub-scales of the interpersonal reactivity index [55] (see 16) . Table 3.2 shows no significant difference between participants' scores per condition. Therefore the data collection can be considered balanced

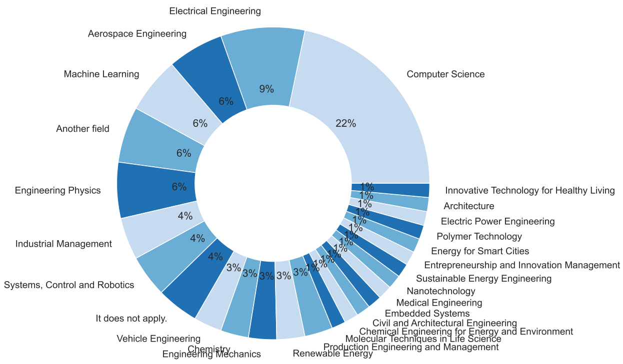


Figure 3.6: Distribution of backgrounds within participants.

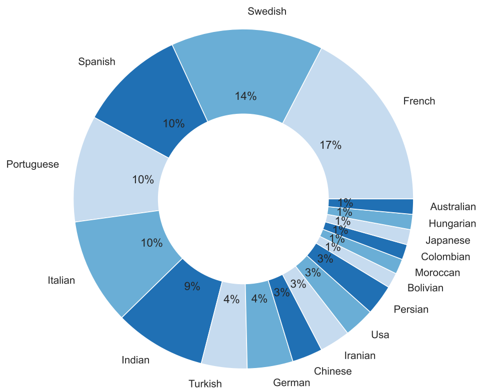


Figure 3.7: Distribution of nationalities within participants.

between conditions. In other words, the null hypothesis states that all conditions are equal is not false.

Table 3.2: Tukey's test for the interpersonal reactivity index distribution per condition

Perspective-Taking Scale	One-way ANOVA	ρ = 7.77e-1	H_0 validated
Empathic Concern Scale	One-way ANOVA	ρ = 5.46e-1	H_0 validated
Fantasy Scale	One-way ANOVA	ρ = 5.55e-1	H_0 validated

Chapter 4

Results

In this chapter, we present the results of the PT task. The effectiveness of this task serves as a base for the other outcomes of this project. Then, we present the results of the PT task related to the **RQ1**. Lastly, we address **RQ2** and discuss the results of the prosocial behaviour task as a function of the PT task.

4.1 Perspective-Taking Task

This section presents the results within four subsections. First, it starts with the general results of the PT task, followed by an analysis of individual instructions. Finally, it presents the results of the ambiguous (due to the implicitly of the frame of reference) instructions and compares those results with similar explicit (therefore non-ambiguous) instructions.

4.1.1 Pilot-study

Before the experiments, a pilot test was run in the PT task, with one participant per each condition. Participants, on average, took less than ten seconds to scan each landmark. However, in the informal discussion, we concluded that participants' experience interacting with social robots might have influenced the results. So we concluded that setting fifteen seconds for each iteration would be a good balance between understanding the instruction, choosing an object and scanning the landmark, and remembering the instruction if one of these steps took longer. Secondly, the pilot helped smoother the interaction flow, simplifying the robot's speech by adding pauses or reformulating some utterances. Lastly, with the pilot study, the range of possible verbally interventions of participants (e.g. follow-up answers, loud thoughts, random questions) was expanded, making the robot's system more complete for the user study.

4.1.2 General results

To analyse the effectiveness of the PT, the quantitative metrics used are the number of mistakes made (e.g. scanning an incorrect landmark), clarifications asked (only valid for the ambiguous instructions), and the number of total tries scanning an object. For each instruction, the number of total tries equals the number of clarifications plus the number of mistakes plus one (the successful scanning). In addition, the amount of time dedicated only to scanning objects and on the PT interaction is also a validation metric.

In the figure 4.2 it is possible to see the statistical distribution between conditions. This figure also shows the influence of the robot, especially when in the human-centred condition. The average minutes of the PT task is (M = 5.13, SD = 1.94), and table 1 expresses the ANOVA results for the PT task time and the other metrics per condition. Figure 4.2a shows the time distribution in this task per condition.

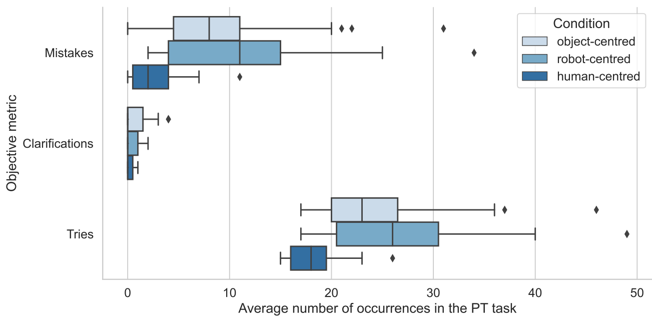
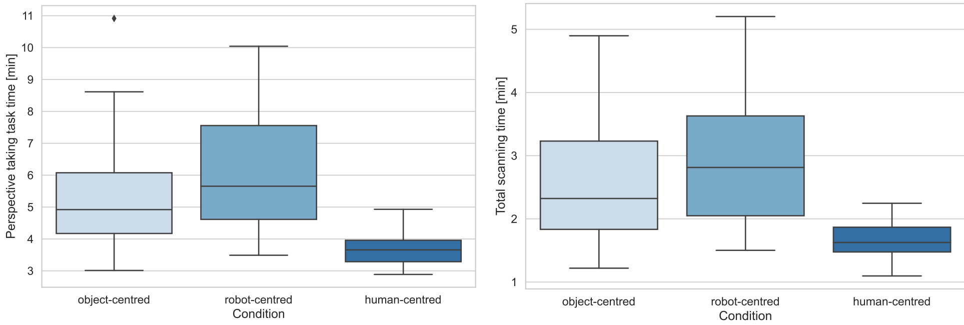


Figure 4.1: Total number of mistakes, clarifications and tries for all instructions. ANOVA: 1

Concerning the reacting time, spent from the moment the robot finished saying the instruction until participants scanned the correct object, the results are shown in figure 4.2b. The data allows running the Tukey's test once the NULL hypothesis (H_0) claiming that all conditions are the same is false. Table 4.1 shows those results that confirm the significant difference between the human-centred and object-centred condition ($p_{value} = 1.55e{-}3$), and between the robot-centred and object-centred condition ($p_{value} = 9.91e - 6$).

To conclude, participants in the object-centred condition spent 1.56 more time than participants in the human-centred condition; participants in the



(a) Minutes spent in the PT task. ANOVA: 1

(b) Minutes spent scanning the objects. ANOVA: 4.1

Figure 4.2: Average times in the PT task for each condition.

Table 4.1: Statistical analysis of the time spent scanning the landmark underneath each object per condition in the PT task.

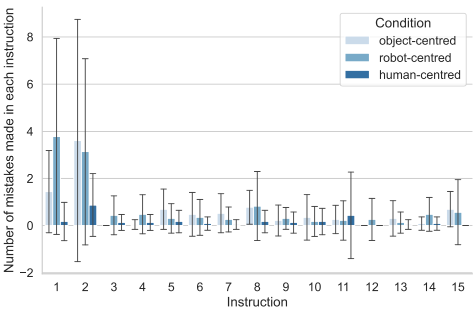
	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Total scanning time	2.63	1.01	3.00	1.09	1.69	0.31
One-way ANOVA	ρ = 1.06e-5				H_0 rejected	
Tukey's test	object-centred				robot-centred	
robot-centred	3.26e-1					
human-centred	1.55e-3**				9.91e-6****	

*ρ<5.0e-2. **ρ<1.0e-2. ***ρ<1.0e-3. ****ρ<1.0e-4

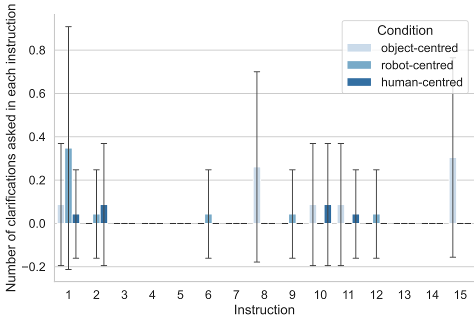
robot-centred condition spent 1.78 more time than participants in the human-centred condition.

4.1.3 Results per Instructions

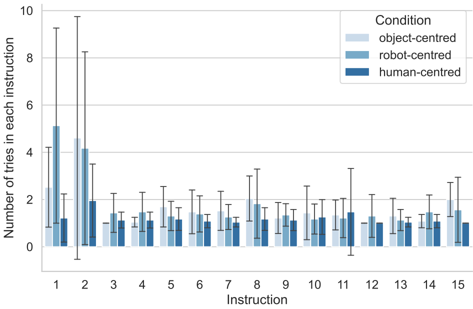
The PT task started with an instruction which did not have an object as the anchor. The robot instructed participants to scan the pyramid at the end of the line (the objects formed a line). Participants in the object-centred condition tend to choose, incorrectly, the object on their right. In the relative-PT conditions, humans started by adopting an egocentric mental model. Figure 4.3 shows the discussed results. In the second instruction, participants had not yet formulated a mental model of the robot. In this instruction, participants had to locate the anchor and goal objects, causing errors. At this point, some participants started realising they could ask the robot for clarifications.



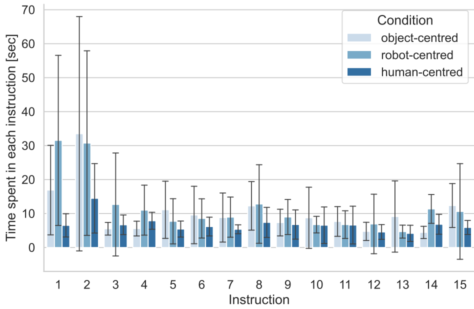
(a) Number of mistakes made in each instructions per condition.



(b) Number of clarifications asked in each instruction per condition.



(c) Number of tries in each instruction per condition.



(d) Average times in the PT task in each instruction per condition.

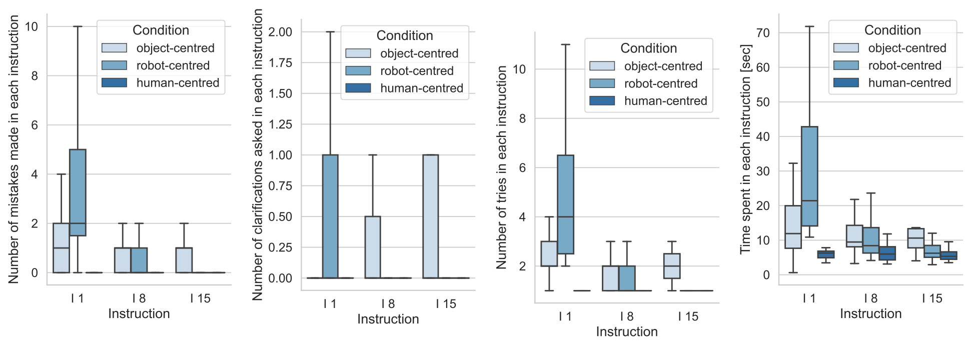
Figure 4.3: Quantitative results in each instruction.

To conclude, the results show a balanced difficulty between conditions. For example, in the figure 4.3c after the first two instructions and excluding the ambiguous instructions, there is no particular condition with a systematic differential relative to the other instructions.

4.1.4 Results of the Ambiguous Instructions

A detailed analysis of each ambiguous instruction is in figure 4.4. Concerning the number of mistakes participants made (fig. 4.4a), the first instruction stands out in the robot-centred condition. However, the number of mistakes was reduced in the next instructions. In the object-centred condition, participants made more mistakes in the first instruction and around half of that in the 8th and 14th instruction. (see 2). Finally, participants in the human-centred condition that made mistakes are outliers in the dataset.

Figure 4.4b shows that a rising number of participants in the object-centred condition understood that they could ask the robot for clarifications. In addition, participants in the robot-centred condition did not ask for clarification after a set of explicit (non-ambiguous) instructions. It is important to note that ambiguous instructions in the robot-centred and human-centred conditions can be clarified by having a mental model of the robot (i.e. remember past instructions). Whereas ambiguous conditions in the object-centred condition can be clarified either by asking for clarification or by trial and error. It explains why participants in object-centred condition still made mistakes in the 8th and 14th instructions or why they asked for more clarification compared to other conditions.
Figures 4.4c and 4.4d show that participants performance is relatively improving in the robot-centred and human-centred conditions, and is is maintained in the object-centred condition. To analyse the differences between each condition, table 4.2 shows the result of Tukey's test. The results significantly differed from the robot-centred condition to the remaining in the first instruction. However, the results differed between the object-centred and the other conditions at the end of the interaction.



(a) Number of mistakes made in each instruction ambiguous per condition. ANOVA: 2

(b) Number of clarifications asked in each ambiguous instruction per condition. ANOVA: 3

(c) Number of tries in each ambiguous instruction per condition. ANOVA: 4

(d) Average times in the PT task in each ambiguous instruction per condition. ANOVA: 5

Figure 4.4: Quantitative results in each ambiguous instruction.

4.1.5 H1a: Participant's first choice

In order to test the **H1a**, in the robot-centred and human-centred conditions, the first instruction was ambiguous (table 4.3). The robot instructed

Table 4.2: Tukey's test for number of tries in the ambiguous instructions.

		object-centred	robot-centred
Instruction 1	robot-centred	4.77e-3**	
	human-centred	2.38e-1	1.87e-5****
Instruction 8	robot-centred	7.71e-1	
	human-centred	2.04e-1*	1.05e-1
Instruction 15	robot-centred	2.50e-1	
	human-centred	1.20e-3**	1.00e-1

*ρ<5.0e-2. **ρ<1.0e-2. ***ρ<1.0e-3. ****ρ<1.0e-4

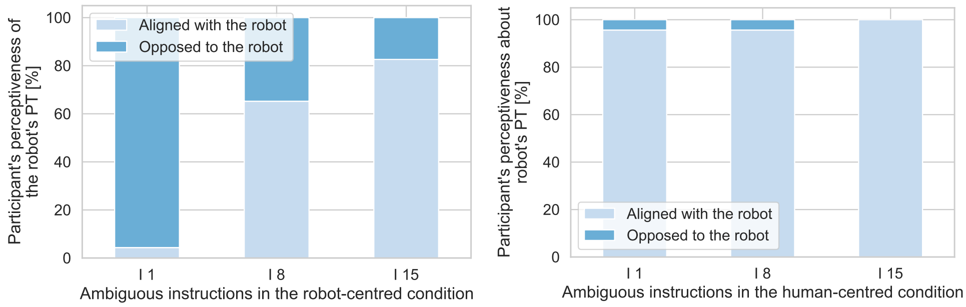
participants to pick the object on a far end, without mentioning which end in the object-centred condition or on the left/right end in the other conditions, without specifying from who's perspective. Additionally, it collected data on whether participants took an egocentric or addressee-centric approach to decipher the robot's instruction. Figure 4.5 shows the result for the participants in the robot-centred condition (fig. 4.5a) and in the human-centred condition (fig. 4.5b).

In the first instruction, participants had not formulated a mental model of the robot yet. Since the instruction had an implicit frame of reference, participants' results in the robot-centred and human-centred conditions can be merged. The results show that, in the first instruction, 96% of the participants took an egocentric perspective to select the object the robot was referring to. This result is aligned with the principle of least collaborative effort and the result from [70, 71], which shows that both children and adults first take an egocentric approach before adjusting it throughout the interaction.

Table 4.3: First instruction. Example: "Go with the..."

Instr.	Object-centred	Robot-centred	human-centred
1	pyramid, at the end of the line	pyramid, on the far right	pyramid, on the far left
1 *	pyramid, at the end of the line, on the side of the big box	pyramid, on my far right	pyramid, on your far left

* clarification of the instruction



(a) Participant's perception of the instruction in their perspective when the robot was taking an egocentric perspective (robot-centred condition).

(b) Participant's perception of the instruction in their perspective when the robot was taking an addressee-centre perspective (human-centred condition).

Figure 4.5: Participant's PT perception in the robot's ambiguous instructions.

4.1.6 H1b: Mental model evaluation

Secondly, in order to validate the **H1b**, participants had to show a mental model consistent with previous explicit instructions. Figure 4.5 shows an adaptation of participants' mental model to the robot's respective perspective throughout the interaction.

In the first instruction, 96% of participants in the robot-centred condition thought the robot was taking their perspective. Furthermore, after six instructions with an explicit frame of reference, 60% of the participants updated their mental model about the robot's perspective and started taking the robot's perspective when perceiving the instructions. In the end, 82% of participants succeeded in generalising the mental model for implicit instructions.

Contrarily, only 1 participant (4%) in the human-centred condition considered the robot might have changed perspective due to the implicit instruction. The same result applies to the eighth instruction, although with a different participant, so their responses can be treated as outliers. In the last instruction, 100% of participants in the human-centred condition scanned the correct landmark on the first try.

4.1.7 H1c: Interpretation time

To understand the efficiency of participants' mental models and validate **H1c**, we compared the time participants spent from hearing the instruction to scanning the correct landmark. For each ambiguous instruction that helped

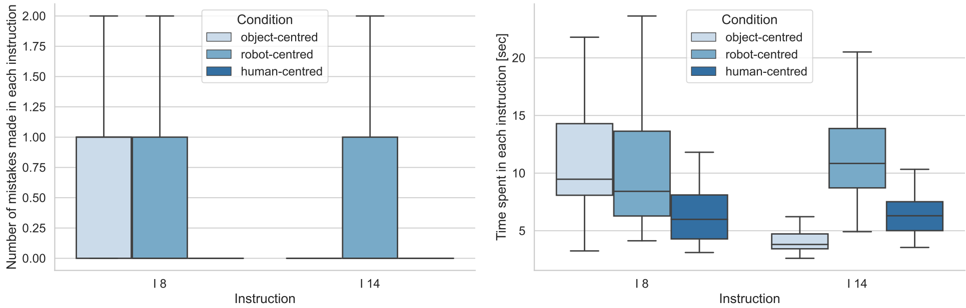
characterise participants' development of a mental model of the robot, there was a similar non-ambiguous instruction to compare. Both instruction pairs asked for similar objects, in a similar format, except one being ambiguous and one not.
To validate this hypothesis, participants in the robot-centred and human-centred conditions had to scan the landmarks quicker over the progress of the interaction. Contrarily, participants in the object-centred condition served as the baseline comparison since we expected them to spend more time in the ambiguous interaction.

4.1.7.1 Pair 1

The first pair involves the eighth (ambiguous) and the fourteenth (non-ambiguous) instructions (table 4.4). Figure 4.6 shows the results of the comparison between the pair. Firstly it is possible to see the difference in the object-centred condition. This difference is also evident because here the extrinsic characteristic of the objects was used, that turned out to be easily spotted (instruction: "...box between pyramids...") in the non-ambiguous instruction (M = 4.44, SD = 1.84) [min] (see 9). Participants made similar mistakes in object-centred and robot-centred conditions in the eighth instruction. However, participants in the object-centred condition asked for more clarifications, so they took more time scanning the objects. Secondly, the participants' behaviour was similar between the robot-centred and human-centred conditions regarding the time spent interpreting the instruction. Moreover, in the human-centred and robot-centred conditions, participants tend to make the same number of mistakes in both instructions.

Table 4.4: Instructions grouped forming the first pair. Example: "Go with the..."

Instr.	Object-centred	Robot-centred	human-centred
8	box, between pyramids	box, on the left of the triangular pyramid	box, on the right of the triangular pyramid
8 *	box, between pyramids of different heights	box, on my left of the triangular pyramid	box, on your right of the triangular pyramid
14	box, between pyramids of the same colour	box, on my right of the square pyramid	box, on your left of the square pyramid



(a) Number of mistakes made in each instruction. ANOVA: 6

(b) Seconds spent in each instruction instructions. ANOVA: 9

Figure 4.6: Average results in each instructions on the pair 1 per condition.

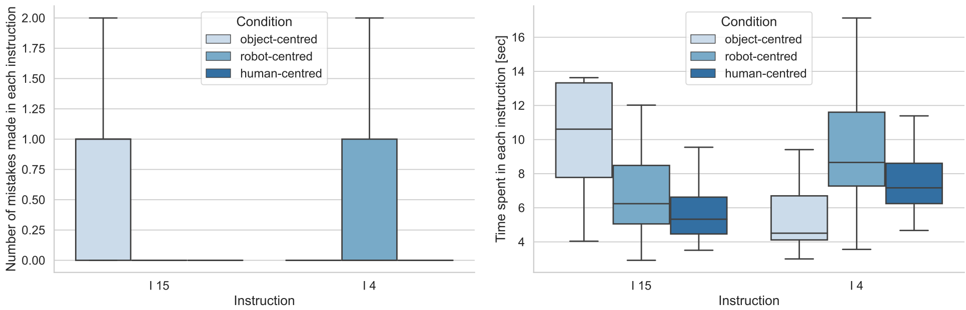
4.1.7.2 Pair 2

The second pair involves the fourth (non-ambiguous) and the fifteenth (ambiguous) (table 4.5). Figure 4.7 shows the results of the comparison between the pair. Firstly, in the object-centred condition, participants spent more time, made more mistakes and asked for more clarifications in the ambiguous sentence. Moreover, participants in the robot-centred and human-centred conditions behave similarly in both instructions. Equally to what happened in the previous pair, participants perceived the second instruction faster due to becoming more familiarised with the interaction.

Table 4.5: Instructions grouped forming the second pair.. Example: ”Go with the...”

Instr.	Object-centred	Robot-centred	human-centred
4	small pyramid, beside the small box	pyramid, on my right of the small box	pyramid, on your left of the small box
15	triangular pyramid, beside the box	pyramid, on the left of the big box	pyramid, on the right of the big box
15 *	triangular pyramid, beside the big box	pyramid, on my left of the big box	pyramid, on your right of the big box
* clarification of the instruction			

As argued before, the participants’ mental model was formulated in the first two instructions. However, participants in the robot-centred condition still made mistakes when the robot was implicit in the frame of reference in the eighth instruction, but as table 7 shows, no one asked for clarification. Hence, participants in this condition, who got this instruction wrong, spent



(a) Number of mistakes made in each instruction. ANOVA: 10

(b) Seconds spent in each instruction instructions. ANOVA: 13

Figure 4.7: Average results in each instructions on the pair 2 per condition.

more time revising the mental model they had already formulated.
The first and the second pair of sentences are similar, changing only the anchor with the goal objects. Therefore, the time spent on each instruction can also be compared. Participants in the robot-centred and human-centred conditions scanned the landmark in the fifteenth instruction in 80% of the time they took in the eighth instruction. This result shows the evolution of the participants' mental models, which were already perceiving instructions with an implicit centre of reference as well as without in the fifteenth and last instruction.

4.1.8 Subjective Measures regarding the PT task

In the last questionnaire, participants answered the following question: *"If you were to tell the robot to pick up the objects, how would you do it?"*. Figure 4.8 shows the results. Participants who did not have the robot-centred condition generally chose an object-centric perspective to address the objects to the robot. However, participants in the robot-centred condition prefered an addressee-centric approach mirroring how the robot proceeded.

Analysing the results in figure 4.8 show that participants in both object-centred and human-centred conditions would choose an object-centred perspective based on the intrinsic characteristic of the objects. Participants in the object-centred condition had no other information about whether or not the robot could understand another perspective (e.g. human's perspective), which explains why they chose the same perspective the robot showed. On the other hand, participants in the human-centred condition were familiar with the robot's ability to address objects using an addressee-centric perspective, but they still selected an object-centric perspective to instruct the robot. Their

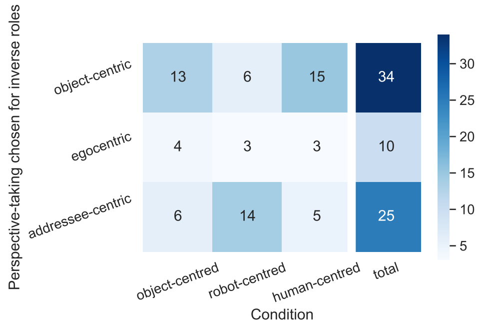
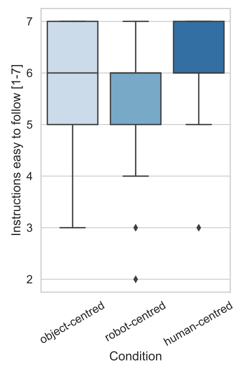


Figure 4.8: Participants answerers about how would they tell the robot to pick an object.

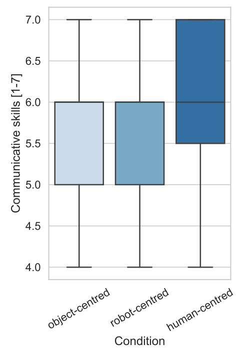
selection is aligned with Qiong et al.'s research, where they conclude that "in the absence of intrinsic reference, humans are more inclined to use addressee-centred, and in the case of intrinsic references, humans are more inclined to choose intrinsic." [16, p. 5].
However, In the case of participants in the robot-centred condition, their selection is more inclined toward an addressee-centric perspective. Even though humans tend to be egocentric by nature [72] after participants understood instructions from the robot's perspective, their expectations of the robot's understanding instructions from someone's perspective increased. On top of that, not only did participants know that the robot knew its perspective, but it is probably a reflection that participants' would appreciate the robot to have taken their perspective during the task to easier it for them. These results can be supported by the way participants rated how much they perceived the robot's caring toward them. As shown in Figure 4.12a, participants in the robot-centred condition rated the robot significantly less caring compared to the other two conditions.

Regarding the first dependent variable: "understanding robot's perception", participants perceived the robot similarly. Figure 4.9 shows the results of the answers with a 1/7 Likert scale (questions: 17). In addition, table 4.6 shows the statistical analysis of the difference between conditions.

Regarding how participants found the instructions easy to follow, figure 4.9a shows a larger variance in the object-centred condition based on the participants' who opted for trial and error and for those who asked for clarifications. On the other hand, participants in the human-centred found it easier to follow the instructions than participants in the robot-centred condition. Since the structure of both conditions was identical, we can hypothesise that participants had a harder time in this condition due to the



(a) Question: ”Did you feel the robot’s instructions were easy to follow?”



(b) Question: ”How did you find the robot’s communicative skills in instructing you to find the correct object?”

Figure 4.9: Participants answerers to instructions about their perception of the HRI with a 1/7 Likert scale, regarding the dependent variable: ”understating the robot’s perception”.

Table 4.6: Tukey’s test for subjective questions about participant’s perception of the robot.

1/7 Likert scale asnwer		object-centred	robot-centred
Instruction easy to follow	robot-centred	2.30e-1	
	human-centred	6.01e-1	2.88e-2*
Communicative skills	One-way ANOVA	ρ = 1.84e-1	H_0 validated

*ρ<5.0e-2. **ρ<1.0e-2. ***ρ<1.0e-3. ****ρ<1.0e-4

effort they had to make to take the robot’s perspective, either through mental or physical rotation.
These reasons led to participants rating the communicative skills of the robot lower in the object-centred and robot-centred conditions.

4.2 Prosocial Behaviour task

The second task was a voluntary task. As figure 4.10 shows, the adherence rate to this task depended on the previous task. Only 57% of the participants in the robot-centred condition showed interest in knowing more about the task to help the robot, contrasting with 87% of participants in the object-centred condition and 100% in the human-centred condition. Moreover, 30% of the participants in the robot-centred condition and 9% in the object-centred condition did not even show interest in knowing more about the voluntary task to help the robot (see table 4.7).

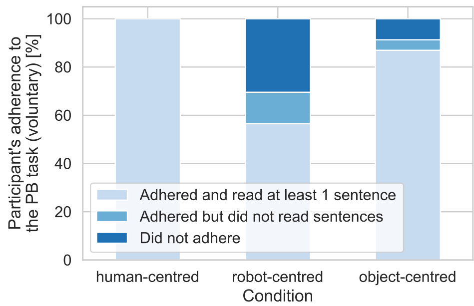


Figure 4.10: Participants' adherence to the prosocial behaviour task grouped by their condition in the PT task.

Table 4.7: Adherence rate to the prosocial behaviour task per condition.

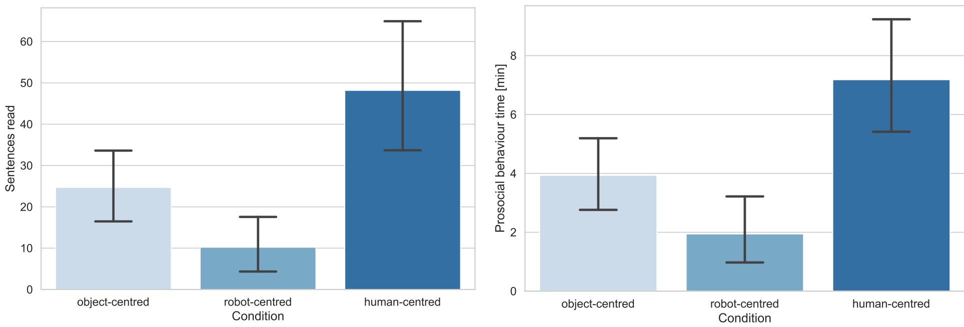
Condition	Adhered and read at least 1 sentence	Adhered but did not read sentences	Did not adhere
object-centred	87%	4%	9%
robot-centred	57%	13%	30%
human-centred	100%	0%	0%

4.2.1 H2: Participant's performance in the prosocial behaviour task

In order to measure the second dependent variable: "prosociality towards the robot", the metrics analysed are the minutes spent reading sentences and the number of sentences marked as read. The results show a discrepancy between the robot-centred condition and the rest. For the participants who did not decide to help the robot or read any sentence, we counted their contribution with a zero. Figure 4.11 shows the metrics for each condition.

Firstly, we run the one-way ANOVA test, and the results show that the differences between conditions are significant ($\rho_{sentences\ read}$ = 7.39e-5,

$\rho_{prosicial\ behaviour\ time\ [min]}$ = 7.34e-5). Since the test does not specify which conditions are significant independent from the remaining, we ran a pairwise t-test. The results of this test (table 4.8) support a significant difference between all conditions.
A pairwise t-test is better than Tukey's test because Tukey's test considers the global mean in their pairwise comparison. Taking into account the **H2**, we want to evaluate each relative PT condition separately compared to the object-centred condition. Therefore, we ran a pairwise t-test between all experimental conditions.



(a) Number of instructions participants read aloud to the robot.

(b) Amount of minutes spent helping the robot in the prosocial behaviour task.

Figure 4.11: Metrics used to evaluate participant's prosociality towards the robot.

Table 4.8: Pairwise t-test for participants performance in the prosocial behaviour task.

		object-centred	robot-centred
Sentences read	robot-centred	1.46e-2*	
	human-centred	1.56e-2*	1.14e-4***
Time helping	robot-centred	3.46e-2*	
	human-centred	1.16e-2*	7.89e-5****

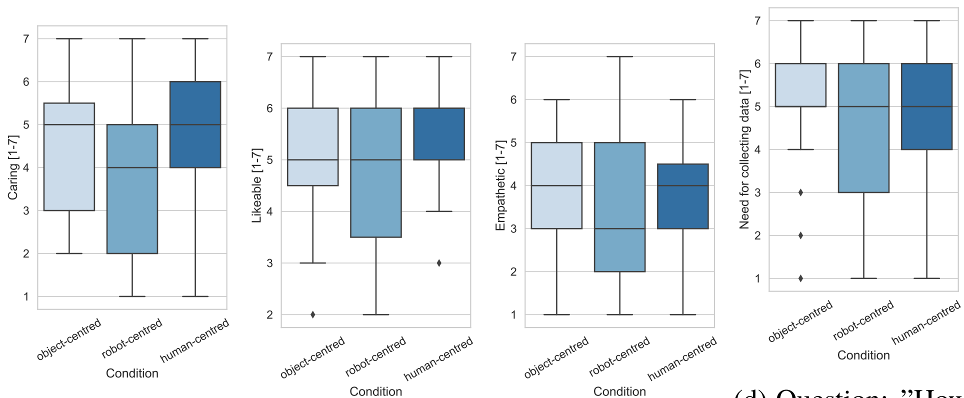
*ρ<5.0e-2. **ρ<1.0e-2. ***ρ<1.0e-3. ****ρ<1.0e-4

Firstly, these results (table 4.8) shows a significant different between all conditions. Furthermore, they show that the robot's perspective when instructing the human has made a difference in the way people felt prosocial toward the robot. The participants' eagerness to help the robot is not necessarily a function of how they performed in the perspective-taking task since they all managed to finish it. .

Another factor that may have influenced the results was the time spent on the previous task. Since participants booked a time to do the experiment, the sooner the participant finished the first task, the longer they could stay helping the robot. For example, participants who rated the robot with the maximum communicative skills might have spent less time in the PT task and therefore were more willing to help in the prosocial behaviour task.

4.2.2 Subjective Measures regarding the prosocial behaviour task

To study the influence of the PT, participants were asked to rate from 1 being "not at all" to 7 being "extremely" on how they found the robot. We asked these questions to understand which feeling motivated the prosociality toward the robot. Participants' answers showed little or no significant difference between conditions. Figure 4.12 shows the results of the answers with a 1/7 Likert scale (questions: 17). In addition, table 4.9 shows the statistical results of Tukey's test between conditions.



(a) Question: "How much did you find yourself caring about the robot?"

(b) Question: "How likeable did you find the robot?"

(c) Question: "How empathetic did you find the robot?"

(d) Question: "How great did you perceive the robot's need for collecting the speech data?"

Figure 4.12: Participants answerers to instructions, about their perception of the HRI with a 1/7 Likert scale, regarding the dependent variable: "prosociality towards the robot".

The results show a significant difference between how much care participants felt in the robot-centred and human-centred conditions (fig 4.12a). Since participants' in the human-centred condition made the interaction

Table 4.9: Tukey's test for subjective questions about participant's perception of the robot.

1/7 Likert scale asnwer		object-centred	robot-centred
Caring	robot-centred	1.27e-1	
	human-centred	8.95e-1	4.72e-2*
Likeable	One-way ANOVA	ρ = 2.56e-1	H_0 validated
Empathetic	One-way ANOVA	ρ = 6.58e-1	H_0 validated
Need for collecting speech data	One-way ANOVA	ρ = 5.67e-1	H_0 validated

*ρ<5.0e-2. **ρ<1.0e-2. ***ρ<1.0e-3. ****ρ<1.0e-4

smoother, participants recognised that and cared more about the robot.
On the other hand, the results in all conditions do not show a significant difference (see table 4.9) for how the participants perceived the robot's likeability (fig 4.12b) or empathy (fig 4.12c).

Finally, table 4.10 shows no significant difference between participants' perceptions of the robot needed to collect data. However, according to figure 4.12d, participants showed a little more need to collect data in the object-centred condition. The results corroborate with [16] since humans related an object-centred approach with intrinsic characteristics more to a robot approach instead of a more human approach with relative PT. Altogether, how much participants perceived that the robot needed to collect data did not influence how much they helped, but rather, how they felt toward the robot in a prosocial way.

Table 4.10: .

Need for collecting speech data	One-way ANOVA	ρ = 5.67e-1	H_0 validated

We see a low variance between ratings since all participants completed the task unquestionably. However, looking at all results, we can see a better appreciation for the human-centred condition, which justifies why participants spent more time reading aloud sentences with emotions, thinking they were contributing to improving the robot. Moreover, the lack of APT mainly when participants made mistakes prevented them from creating empathy toward the robot.

Chapter 5

Discussion

In this project, we explore the link between a social humanoid robot taking human perspectives and humans exhibiting prosocial behaviour toward it. As a result, we designed an interaction containing two different tasks, a PT task followed by a voluntary prosocial behaviour task. This user-study is a between-subject study since we only have one independent variable: the ”robot’s perspective”. The outcome of the PT reflects the first dependent variable: ”understanding the robot’s perspective”, and is used to answer **RQ1**. Whereas the second dependent variable: ”prosociality toward the robot”, characterises how participants performed in the prosocial behaviour task, and it answers question **RQ2**. The PT task consisted of picking objects by following the robot’s instructions from a specific perspective. The perspectives the robot took were either from its perspective (robot-centred), the participant’s perspective (human-centred) or the intrinsic characteristic of the objects (object-centred), which served as the control condition. Then, in the prosocial behaviour task, the participants had a chance to help the robot by reading sentences aloud to them with different emotions.
In the following paragraphs, we will highlight the key findings of this study, how they can contribute to designing and developing social robots, and their interaction modalities and behaviours.

Based on the two research questions, we can look at the experiment from two different directions. In the first direction, we look at the participants’ perspective-taking approach toward the robot in three different stages: before formulating a mental model, formulating a mental model, and their perception of the robot after the end of the interaction. This direction tackles our first research question and its related hypotheses.
In the second direction, we look at how the participants showed prosocial

behaviour toward the robot after the end of the interaction based on which condition they were in and discuss how their perception of the robot's perspective-taking might have impacted it. This direction mainly tackles our second research question and the implications of developing robots with at least basic perspective-taking abilities.

Regarding the PT task, the results show that participants in the object-centred condition spent over 1.56 more time scanning the landmarks compared to participants in the human-centred condition (table 4.1). These results align with [26]. In addition, participants in the robot-centred condition spent 1.78 more time scanning the landmarks compared to the human-centred condition. Altogether, the results show a significant difference and therefore that participants understood the difference between the robot taking the human-centred perspective or an object-centric perspective. Alongside the significant difference that participants in the robot-centred and the human-centred conditions showed, bearing in mind that instructions in both conditions share the same structure, it is valid to conclude that the robot-centred condition is also different from the object-centred condition. However, the similarity in the number of mistakes or time spent on the tasks in these two conditions compared to human-centred condition is a result of the participants having to deviate from their egocentric perspective to a more cognitively demanding perspective, e.g. the robot or objects.

Taking a closer look at the PT, the reasoning behind using ambiguous instructions was to mark which perspective the robot was taking. Even though participants did not know which condition they were in, by answering the first instruction correctly, they perceived how the robot would instruct them. Additionally, these instructions also allowed to test participants' mental models throughout the interaction. As figure 4.3 shows, there is a significant decrease in the number of mistakes (e.g. the number of tries) and time spent on the instructions after the second instruction. Therefore, it is valid to say that the mental model was formulated before the third instruction.

In the first instruction, participants had no prior information about the robot. When the robot used ambiguous instructions, participants could only solve the ambiguity by asking for clarifications, making their assumptions or trial and error. Participants expressed in the informal discussion that since the robot only greeted participants at the start and avoided polite conversations (e.g. How are you?; Are you ready to start the task?), they did not prioritise the verbal interaction with the robot. Therefore, this reduces the number of clarifications asked in this question. As the **H1a** predicts, participants perceived the first ambiguous instructions with an egocentric perspective. The results support the hypothesis, showing that 96% of participants in the robot-centred and human-centred conditions perceived the first instruction from an egocentric perspective, aligned with the principle of least collaborative effort and the result from [70, 71].

As argued before, the mental model can be seen to have been formulated after the second instruction. Therefore, the results of the eighth and fifteenth instructions can be used to analyse **H1b**.
According to figure 4.5, participants' mental models were updated throughout the interaction closing the gap to the robot's actual perspective, as observed in [71] with children. In the human-centred perspective, participants had no trouble aligning interpretation of ambiguous instructions. Contrarily, participants failed one out of four instructions in the robot-centred condition. As a result, participants in the robot-centred condition made more than twice the mistakes as participants in the human-centred condition (global average of all mistakes in non-ambiguous instructions after the second instructions). Therefore, it is reasonable that the percentage of participants with their mental model aligned with the robot's perspective is lower for participants in the robot-centred condition than participants in the human-centred condition. Altogether is valid to validate the **H1b**.

If the previous hypothesis measured the effectiveness of the participants' mental models, **H1c** measured participants' efficiency. Once the amount of information in each instruction influenced the time to process it and the probability to made mistakes, each ambiguous sentence was paired with a non-ambiguous with identical information as a baseline comparison.
In the pair with the fourth and fifteenth instructions (fig. 4.7b), the results show that the time spent in the instruction with the implicit frame of reference was lower for the robot-centred and human-centred conditions. On the other hand, participants in the object-centred condition spent less time in the instruction with the explicit centre of reference, corroborating with the **H1c**. Even though the pair containing instructions eight and fourteen showed that participants still made mistakes in the ambiguous instruction, which increased the time, resulting in contradicting the hypothesis. However, analysing both pairs simultaneously, the time spent on the non-ambiguous instructions is in the global average respectively in each condition. Nevertheless, as explained above, participants in the robot-centred condition made more mistakes than participants in the human-centred condition. Due to the progress over the interaction in the ambiguous instructions, it is possible to validate **H1c**.

Finally, using the object-centred condition as an example, figure 4.4 shows that participants improved their performance over the progress in the interaction. In addition, it shows that fifteen instructions were enough for the robot to mark its perspective and for participants to adapt to the task once the number of mistakes decreased at the same time, participants asked for more clarifications in the ambiguous questions.
Furthermore, ambiguous instructions left participants in a situation between two different objects. Most of them preferred to take the risk when they could not resolve the ambiguity instead of asking the robot for clarifications. This episode raises whether participants should be allowed to scan objects stochastically without being penalised.
Lastly, it is important to mention that two out of fifteen instructions in the object-centred condition used extrinsic-object characteristics. Despite one being an ambiguous instruction (8), the other was instruction fourteen, which turned out to be the quickest answer by participants.

The **RQ2** establishes a connection between a social humanoid robot taking perspective in addressing objects and humans showing prosocial behaviour towards it. Even though participants perceived the robot's need for collecting speech data similarly (see 4.12d), the prosocial behaviour results (see 4.8) show a significant difference in their decision to do so.

The results support the **H2a**. Not only had all participants in the human-centred condition shown interest in helping the robot but also, on average, they read 1.95 sentences to the robot than participants in the object-centred condition (fig. 4.11a). Overall, this shows that while humans do not think or talk about our PT abilities, humans are sensitive about how the self-perspective is considered, and if humans make an effort to take others' perspectives, they like them to reciprocate. And while the participants did not rate the robot differently on most perception scales, the robot's behaviour in considering their perspective significantly impacted the way they showed prosocial behaviour towards it.

All in all, the results show that participants that performed better in the PT task showed higher levels of PT, supporting that higher levels of PT are related to higher levels of prosocial behaviour [21, 22, 20].
The results also support **H2b**. Participants in the robot-centred condition showed the least prosociality towards the robot. This result is a combination of the participants that decided not to know more about the task, the ones who decided to know more but did not read any sentences, and the ones who read some sentences. As figure 4.11a shows, participants in the robot-

centred condition read on average 0.41 sentences compared to the object-centred condition. Because the robot was implicitly egocentric in the first instruction, and while participants were making errors, the robot did not clarify the instruction and kept using its perspective, making participants make extra effort to take the robot's perspective every time. In the final questionnaire, after participants decided to help the robot, participants in the robot-centred condition justified errors made during the interaction either because they did not understand the robot's perspective or it was not flexible when they made mistakes. It was also mentioned in the informal discussion that because the robot did not take the human's affective perspective when participants made mistakes, it created a less caring feeling from the robot's side. This reason explains the results obtained (fig. 4.12a).

Lastly, the fact that participants in the robot-centred condition preferred an addressee-centric approach (fig 4.8) suggests that participants were not comfortable with the robot's egocentric perspective and valued an approach where their perspective was considered. On the other hand, even though the PT task results did not show any significant difference, participants in the object-centred condition did not attribute the errors to the robot. Hence, they would majority use this PT type for referring the objects to the robot.

Finally, the voluntary task consisted of helping the robot. The ReadMe message stated that the robot needed human utterances to improve its speech. The results show that the experimental condition did not influence how much participants perceived the robot's need to collect speech data.
During the task explanation, the robot reminded participants of the possibility of them withdrawing from the task without any consequences. The prosocial behaviour consisted of reading aloud an endless set of sentences with different emotions to the robot. The advantage of this task was to quantify how much participants helped the robot, which resulted in significant results between conditions. However, reading sentences aloud does not imply much effort, resulting in a high variance per each condition. In addition, participants struggled to distinguish between some emotions, which in some cases catalysed finishing the task early. In addition, the fact that the sentences started to repeat had a threefold effect. Firstly, participants might have thought the task was redundant. Secondly, participants might have conditioned the sentences into batches and finished the task once they felt they had read enough batches. Finally, since the task getting repeated, it compelled participants to finish.
On a final note, the number of participants was enough to see plausible results,

we can see that apart from some outliers the results are not sparsed, instead, they are grouped per condition. In addition, we can see that the quantitive results of the PT task reach a steady value in the end, confirming that the length of the task was enough.

5.1 Limitations

The bigger limitation of this project was the speech recogniser of the NAO robot. This problem limited the project to only the robot giving the instructions or understanding complex questions of the user. Additionally, since the NAO robot could not move objects, the task design was limited to only the participant moving objects. Regarding the PT task, the main limitation was that due to the differences in the robot's perspective, the task's difficulty and the time participants took to finish the task were affected a lot in each condition. We have tried to design the task and execute it in a way that prevents any other confounding effect. However, since, in one condition, the robot is addressee-centric, which is less cognitively demanding for the participants, we could not completely control the time it took for participants to complete the tasks and the difficulty that the participants felt. In addition, the robot disagreed more times with the participants' choices in the object-centred and robot-centred conditions. We tried that the robot would not stimulate negative feelings that would make participants blame the robot. Finally, some participants might have missed some characteristics of the objects or how the robot addressed the objects might have been unusual. However, participants had time to get used to the objects before starting the task to control this effect. In the prosocial behaviour task, some participants might have felt different comfort levels while expressing emotions, which might have influenced their performance. For example, it might be considered to ask about the tone throughout the longer sentences in future work. The problem could have been the task being enganing itself, for example, playing a game with the robot and getting better every time. However, to prevent that, the same sentences were sequential, creating a sense of repetition for the participants.

5.2 Future Work

As a future development, we propose to validate these results in another social humanoid robot (e.g. Furhat). In addition, these results might inspire research to study the development of the humans mental model. Moreover, switching between different perspectives in different environments might contribute to a more general mental model of the robot's capabilities.

On the other hand, the influence of switching to a more friendly perspective when the participant has difficulties in stimulating prosocial behaviour can be interesting to study.

Chapter 6

Conclusion

To understand the link between robots showing perspective-taking and humans showing prosocial behaviour towards the robot, we have assembled two tasks in a one thread interaction. The first task consisted of the robot addressing objects to the human from different perspectives, followed by proposing to the participant a task that would help its speech tone adapt according to different emotions. This second task was voluntary as it allowed the author to measure how much participants helped the robot. In addition, the perspective task allowed the author to study the humans' mental model by introducing ambiguous sentences in the sense that the reference frame was implicit in the instruction. The ratio between these instructions was 1:4 compared to instructions with the frame of reference explicit.

Concerning humans' mental model, participants (96%) percept an instruction from an egocentric perspective whenever they had not any prior knowledge about the instruction. In addition, 60% of participants adapted their mental model to the robot's perspective in the second ambiguous instruction in the robot-centred condition. In the end, 83% of participants correctly understood the instructions with an implicit frame of reference in the robot-centred condition and 100% in the human-centred condition. Additionally, the mental model was formulated during the first two instructions because of the contrast between mistakes and time spent compared to the other instructions.

Moreover, the scanning time results decreased over to the completion of the interaction, illustrating that participants were getting used to the HRI. Finally, because errors occurred throughout the interaction, including in explicit instructions, it is valid to say that humans perceive an instruction with an implicit frame of reference equally well after receiving instructions with an explicit frame of reference (**RQ1**).

Answering the **RQ2**, the results show that humans spend more time helping a robot that can take its perspective, referring to an object based on other objects (object-centred condition), although the opposite did not happen. Participants could get closer to the robot and be willing to help the robot more than in the object-centred condition by self-rotating to the robot's perspective to complete the task. The results show that participants in the robot-centred condition were the most reluctant to help the robot improve.

Bibliography

[1] H. P. Moravec, "robot," *Encyclopedia Britannica*, 2021. [Online]. Available: https://www.britannica.com/technology/robot-technology [Page 1.]

[2] A. M. Turing, "I.Computing Machinery and Intelligence," *Mind*, vol. LIX, no. 236, pp. 433–460, 10 1950. doi: 10.1093/mind/LIX.236.433. [Online]. Available: https://doi.org/10.1093/mind/LIX.236.433 [Page 1.]

[3] C. Dune, C. Leroux, and E. Marchand, "Intuitive human interaction with an arm robot for severely handicapped people-a one click approach," in *2007 IEEE 10th International Conference on Rehabilitation Robotics*. IEEE, 2007, pp. 582–589. [Online]. Available: https://ieeexplore.ieee.org/document/4428484 [Page 1.]

[4] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: a back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive behavioural systems*. Springer, 2012, pp. 114–130. [Page 1.]

[5] A. K. Pandey and R. Gelin, "A mass-produced sociable humanoid robot: Pepper: The first machine of its kind," *IEEE Robotics Automation Magazine*, vol. 25, no. 3, pp. 40–48, 2018. doi: 10.1109/MRA.2018.2833157. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8409927 [Pages 1 and 12.]

[6] E. Yadollahi, P. Dillenbourg, and A. Paiva, "Changing perspective as a learning mechanism," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 612–614. [Page 1.]

[7] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science robotics*,

vol. 3, no. 21, p. eaat5954, 2018. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.aat5954 [Page 1.]

[8] R. van den Berghe, J. Verhagen, O. Oudgenoeg-Paz, S. Van der Ven, and P. Leseman, "Social robots for language learning: A review," *Review of Educational Research*, vol. 89, no. 2, pp. 259–295, 2019. [Online]. Available: https://journals.sagepub.com/doi/full/10.3102/0034654318821286 [Page 2.]

[9] A. Guneysu Ozgur, A. Özgür, T. Asselborn, W. Johal, E. Yadollahi, B. Bruno, M. Skweres, and P. Dillenbourg, "Iterative design and evaluation of a tangible robot-assisted handwriting activity for special education," *Frontiers in Robotics and AI*, vol. 7, p. 29, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2020.00029/full [Page 2.]

[10] A. G. Ozgur, B. Bruno, V. Taburet, A. Özgür, and P. Dillenbourg, "Design of dynamic tangible workspaces for games: Application on robot-assisted upper limb rehabilitation," in *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020, pp. 172–178. [Online]. Available: https://ieeexplore.ieee.org/document/9223332 [Page 2.]

[11] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013. [Online]. Available: https://link.springer.com/article/10.1007/s12369-013-0178-y [Page 2.]

[12] J. Pirhonen, E. Tiilikainen, S. Pekkarinen, M. Lemivaara, and H. Melkas, "Can robots tackle late-life loneliness? scanning of future opportunities and challenges in assisted living facilities," *Futures*, vol. 124, p. 102640, 2020. doi: https://doi.org/10.1016/j.futures.2020.102640. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0016328720301294 [Page 2.]

[13] E. B. Raposa, H. B. Laws, and E. B. Ansell, "Prosocial behavior mitigates the negative effects of stress in everyday life," *Clinical Psychological Science*, vol. 4, no. 4, pp. 691–698, 2016. doi: 10.1177/2167702615611073 PMID: 27500075. [Online]. Available: https://doi.org/10.1177/2167702615611073 [Pages 2 and 13.]

[14] K. E. Buchanan and A. Bardi, "Acts of kindness and acts of novelty affect life satisfaction," *The Journal of Social Psychology*, vol. 150, no. 3, pp. 235–237, 2010. doi: 10.1080/00224540903365554 PMID: 20575332. [Online]. Available: https://doi.org/10.1080/00224540903365554 [Page 2.]

[15] M. Shum, M. Kleiman-Weiner, M. L. Littman, and J. B. Tenenbaum, "Theory of minds: Understanding behavior in groups through inverse planning," 2019. [Online]. Available: https://arxiv.org/abs/1901.06085 [Page 2.]

[16] Q. Shi, P. Yang, and C. Chen, "Preference modeling of spatial description in human-robot interaction," in *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 2020. doi: 10.1109/ICNSC48988.2020.9238088 pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9238088 [Pages 2, 11, 18, 41, and 46.]

[17] N. Epley, E. M. Caruso, and M. H. Bazerman, "When perspective taking increases taking: reactive egoism in social interaction." *Journal of personality and social psychology*, vol. 91, no. 5, p. 872, 2006. [Online]. Available: https://psycnet.apa.org/doiLanding?doi=10.1037%2F0022-3514.91.5.872 [Pages 2 and 6.]

[18] J. R. Pierce, G. J. Kilduff, A. D. Galinsky, and N. Sivanathan, "From glue to gasoline: How competition turns perspective takers unethical," *Psychological science*, vol. 24, no. 10, pp. 1986–1994, 2013. [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/0956797613482144 [Pages 2 and 7.]

[19] C. Sassenrath, J. D. Vorauer, and S. D. Hodges, "The link between perspective-taking and prosociality—not as universal as you might think," *Current opinion in psychology*, vol. 44, pp. 94–99, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352250X2100169X [Page 2.]

[20] S. Berti and A. Cigala, "Mindfulness for preschoolers: Effects on prosocial behavior, self-regulation and perspective taking," *Early Education and Development*, vol. 33, no. 1, pp. 38–57, 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/10409289.2020.1857990 [Pages 2, 14, 29, and 50.]

[21] A. Cigala, A. Mori, and F. Fangareggi, "Learning others' point of view: Perspective taking and prosocial behaviour in preschoolers," *Early Child Development and Care*, vol. 185, no. 8, pp. 1199–1215, 2015. [Pages 3, 29, and 50.]

[22] C. K. Tamnes, K. Overbye, L. Ferschmann, A. M. Fjell, K. B. Walhovd, S.-J. Blakemore, and I. Dumontheil, "Social perspective taking is associated with self-reported prosocial behavior and regional cortical thickness across adolescence." *Developmental psychology*, vol. 54, no. 9, p. 1745, 2018. [Online]. Available: https://psycnet.apa.org/record/2018-36315-001 [Pages 3, 14, 29, and 50.]

[23] R. Oliveira, P. Arriaga, F. P. Santos, S. Mascarenhas, and A. Paiva, "Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour," *Computers in Human Behavior*, vol. 114, p. 106547, 2021. doi: https://doi.org/10.1016/j.chb.2020.106547. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563220302971 [Page 3.]

[24] F. Herrera and J. N. Bailenson, "Virtual reality perspective-taking at scale: Effect of avatar representation, choice, and head movement on prosocial behaviors," *new media & society*, vol. 23, no. 8, pp. 2189–2209, 2021. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/1461444821993121 [Page 3.]

[25] A. Surtees, I. Apperly, and D. Samson, "Similarities and differences in visual and spatial perspective-taking processes," *Cognition*, vol. 129, no. 2, pp. 426–438, 2013. doi: https://doi.org/10.1016/j.cognition.2013.06.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027713001200 [Pages 5, 7, and 8.]

[26] F. I. Doğan, S. Gillet, E. J. Carter, and I. Leite, "The impact of adding perspective-taking to spatial referencing during human–robot interaction," *Robotics and Autonomous Systems*, vol. 134, p. 103654, 2020. doi: https://doi.org/10.1016/j.robot.2020.103654. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889020304942 [Pages 5, 12, 29, and 48.]

[27] M. Tomasello, J. Call, and B. Hare, "Chimpanzees understand psychological states–the question is which ones and to what extent,"

Trends in cognitive sciences, vol. 7, no. 4, pp. 153–156, 2003. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/12691762/ [Page 5.]

[28] B. Hare and M. Tomasello, "Human-like social skills in dogs?" *Trends in cognitive sciences*, vol. 9, no. 9, pp. 439–444, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1364661305002081 [Page 5.]

[29] D. Samson, I. A. Apperly, J. J. Braithwaite, B. J. Andrews, and S. E. Bodley Scott, "Seeing it their way: evidence for rapid and involuntary computation of what other people see." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 36, no. 5, p. 1255, 2010. [Online]. Available: https://psycnet.apa.org/record/2010-17519-001 [Page 5.]

[30] A. D. Surtees, S. A. Butterfill, and I. A. Apperly, "Direct and indirect measures of level-2 perspective-taking in children and adults," *British Journal of Developmental Psychology*, vol. 30, no. 1, pp. 75–86, 2012. [Online]. Available: https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/j.2044-835X.2011.02063.x [Page 5.]

[31] R. Ros, E. A. Sisbot, R. Alami, J. Steinwender, K. Hamann, and F. Warneken, "Solving ambiguities with perspective taking," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2010. doi: 10.1109/HRI.2010.5453204 pp. 181–182. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5453204 [Page 6.]

[32] R. Newey, K. Koldewyn, and R. Ramsey, "The influence of prosocial priming on visual perspective taking and automatic imitation," *PloS one*, vol. 14, no. 1, p. e0198867, 2019. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0198867 [Page 6.]

[33] J. K. Malinowska, "What does it mean to empathise with a robot?" *Minds and Machines*, vol. 31, no. 3, pp. 361–376, 2021. [Online]. Available: https://link.springer.com/article/10.1007/s11023-021-09558-7 [Page 7.]

[34] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Empathic robots for long-term interaction," *International Journal of Social Robotics*, vol. 6, no. 3, pp. 329–341, 2014. [Online]. Available: https://link.springer.com/article/10.1007/s12369-014-0227-1 [Page 7.]

[35] A. Surtees, D. Samson, and I. Apperly, "Unintentional perspective-taking calculates whether something is seen, but not how it is seen," *Cognition*, vol. 148, pp. 97–105, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027715301220 [Page 8.]

[36] A. Lawrence, L. Clark, J. LaBuzetta, B. Sahakian, and S. Vyakarnam, "The innovative brain," *Nature*, vol. 456, pp. 168–9, 12 2008. doi: 10.1038/456168a. [Online]. Available: https://www.researchgate.net/publication/23469957_The_innovative_brain [Page 8.]

[37] I. A. Apperly, "Beyond simulation–theory and theory–theory: why social cognitive neuroscience should use its own concepts to study "theory of mind"," *Cognition*, vol. 107, no. 1, pp. 266–283, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010027707002120?casa_token=hzQHtLSIZKMAAAAA:xP9vUlYWMxO7E_NqSCfmh6sJ5Deb__yG7ZNVWzpm_HpTovkpVEG67V95vPg0Z_CLFSBiGPne-rI [Page 8.]

[38] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 14–29, 2016. doi: 10.1109/TPAMI.2015.2430335. [Online]. Available: https://ieeexplore.ieee.org/document/7102751 [Page 8.]

[39] M. D. Harwood and M. J. Farrar, "Conflicting emotions: The connection between affective perspective taking and theory of mind," *British Journal of Developmental Psychology*, vol. 24, no. 2, pp. 401–418, 2006. [Online]. Available: https://bpspsychub.onlinelibrary.wiley.com/doi/full/10.1348/026151005X50302 [Page 9.]

[40] F. I. Doğan, S. Kalkan, and I. Leite, "Learning to generate unambiguous spatial referring expressions for real-world environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019. doi: 10.1109/IROS40897.2019.8968510 pp. 4992–4999. [Online]. Available: https://ieeexplore.ieee.org/document/8968510 [Page 10.]

[41] S. Kiesler, "Fostering common ground in human-robot interaction," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* IEEE, 2005, pp. 729–734. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1513866 [Pages 10 and 28.]

[42] E. Yadollahi, W. Johal, J. Dias, P. Dillenbourg, and A. Paiva, "Studying the effect of robot frustration on children's change of perspective," in *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 2019, pp. 381–387. [Online]. Available: https://ieeexplore.ieee.org/document/8925100 [Page 11.]

[43] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, "Mechatronic design of nao humanoid," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 769–774. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5152516/ [Pages 11, 16, and 18.]

[44] X. Zhao, C. Cusimano, and B. F. Malle, "Do people spontaneously take a robot's visual perspective?" in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, pp. 335–342. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7451770 [Page 11.]

[45] S. Y. Okita, "Self–other's perspective taking: the use of therapeutic robot companions as social agents for reducing pain and anxiety in pediatric patients," *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 6, pp. 436–441, 2013. [Online]. Available: https://www.liebertpub.com/doi/full/10.1089/cyber.2012.0513 [Page 12.]

[46] K. Reidy, K. Markin, S. Kohn, and E. Wiese, "Effects of perspective taking on ratings of human likeness and trust," in *International Conference on Social Robotics*. Springer, 2015, pp. 564–573. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-25554-5_56 [Page 12.]

[47] J. G. Trafton, A. C. Schultz, M. Bugajska, and F. Mintz, "Perspective-taking with robots: experiments and models," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* IEEE, 2005, pp. 580–584. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1513842 [Page 13.]

[48] G. V. Caprara, C. Barbaranelli, C. Pastorelli, A. Bandura, and P. G. Zimbardo, "Prosocial foundations of children's academic achievement," *Psychological science*, vol. 11, no. 4, pp. 302–306,

2000. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1111/1467-9280.00260 [Page 13.]

[49] C. L. Keyes and J. Haidt, *Flourishing: Positive psychology and the life well-lived*. American Psychological Association Washington, DC, 2003. [Online]. Available: http://students.aiu.edu/submissions/profiles/resources/onlineBook/m7h6D2_positive%20psychology.pdf [Page 13.]

[50] G. Carlo, M. V. Mestre, P. Samper, A. Tur, and B. E. Armenta, "The longitudinal relations among dimensions of parenting styles, sympathy, prosocial moral reasoning, and prosocial behaviors," *International Journal of Behavioral Development*, vol. 35, no. 2, pp. 116–124, 2011. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/0165025410375921 [Page 13.]

[51] L. B. Aknin, E. W. Dunn, and M. I. Norton, "Happiness runs in a circular motion: Evidence for a positive feedback loop between prosocial spending and happiness," *Journal of happiness studies*, vol. 13, no. 2, pp. 347–355, 2012. [Page 13.]

[52] W. Waugh, C. Brownell, and B. Pollock, "Early socialization of prosocial behavior: Patterns in parents' encouragement of toddlers' helping in an everyday household task," *Infant Behavior and Development*, vol. 39, pp. 1–10, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0163638314200099 [Page 13.]

[53] M. Tsvetkova and M. W. Macy, "The social contagion of generosity," *PloS one*, vol. 9, no. 2, p. e87275, 2014. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0087275 [Page 14.]

[54] R. Baumsteiger and J. T. Siegel, "Measuring prosociality: The development of a prosocial behavioral intentions scale," *Journal of Personality Assessment*, vol. 101, no. 3, pp. 305–314, 2019. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00223891.2017.1411918 [Page 14.]

[55] M. H. Davis *et al.*, "A multidimensional approach to individual differences in empathy," *JSAS Catalog of Selected Documents in Psychology*, vol. 10, p. 85, 1980. [Online]. Available: https://www.uv.es/friasnav/Davis_1980.pdf [Pages 14, 19, and 29.]

[56] R. Goodman, "The strengths and difficulties questionnaire: a research note," *Journal of child psychology and psychiatry*, vol. 38, no. 5, pp. 581–586, 1997. [Online]. Available: https://acamh.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7610.1997.tb01545.x [Page 14.]

[57] D. A. Kenny and L. Albright, "Accuracy in interpersonal perception: a social relations analysis." *Psychological bulletin*, vol. 102, no. 3, p. 390, 1987. [Online]. Available: https://psycnet.apa.org/record/2016-47361-017 [Page 14.]

[58] J. Avelino, F. Correia, J. Catarino, P. Ribeiro, P. Moreno, A. Bernardino, and A. Paiva, "The power of a hand-shake in human-robot interactions," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1864–1869. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8593980 [Pages 14 and 15.]

[59] A. Paiva, F. Santos, and F. Santos, "Engineering pro-sociality with autonomous agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/12215 [Page 15.]

[60] B. J. Fogg, "Persuasive technology: using computers to change what we think and do," *Ubiquity*, vol. 2002, no. December, p. 2, 2002. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/764008.763957 [Page 15.]

[61] B. Hayes, D. Ullman, E. Alexander, C. Bank, and B. Scassellati, "People help robots who help others, not robots who help themselves," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 255–260. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6926262 [Page 15.]

[62] D. U. Martin, C. Perry, M. I. MacIntyre, L. Varcoe, S. Pedell, and J. Kaufman, "Investigating the nature of children's altruism using a social humanoid robot," *Computers in Human Behavior*, vol. 104, p. 106149, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563219303590 [Page 15.]

[63] R. De Kleijn, L. van Es, G. Kachergis, and B. Hommel, "Anthropomorphization of artificial agents leads to fair and strategic, but not altruistic behavior," *International Journal of Human-Computer*

Studies, vol. 122, pp. 168–173, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581918305524 [Page 16.]

[64] N. Chernyak and H. E. Gary, "Children's cognitive and behavioral reactions to an autonomous versus controlled social robot dog," *Early Education and Development*, vol. 27, no. 8, pp. 1175–1189, 2016. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/10409289.2016.1158611 [Page 16.]

[65] B. Kühnlenz, S. Sosnowski, M. Buß, D. Wollherr, K. Kühnlenz, and M. Buss, "Increasing helpfulness towards a robot by emotional adaption to the user," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 457–476, 2013. [Pages 17 and 29.]

[66] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas) development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, 2017, pp. 254–262. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/2909824.3020208 [Page 25.]

[67] J. James, L. Tian, and C. I. Watson, "An open source emotional speech corpus for human robot interaction applications." in *Interspeech*, 2018, pp. 2768–2772. [Online]. Available: https://www.isca-speech.org/archive_v0/Interspeech_2018/pdfs/1349.pdf [Page 25.]

[68] K. Stubbs, P. J. Hinds, and D. Wettergreen, "Autonomy and common ground in human-robot interaction: A field study," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 42–50, 2007. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4136857 [Page 28.]

[69] M. A. Goodrich and D. R. Olsen, "Seven principles of efficient human robot interaction," in *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, vol. 4. IEEE, 2003, pp. 3942–3948. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/1244504 [Page 28.]

[70] N. Epley, C. K. Morewedge, and B. Keysar, "Perspective taking in children and adults: Equivalent egocentrism but differential correction," *Journal of experimental social psychology*, vol. 40, no. 6, pp. 760–768, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022103104000241 [Pages 28, 36, and 49.]

[71] E. Yadollahi, M. Couto, P. Dillenbourg, and A. Paiva, "Do children adapt their perspective to a robot when they fail to complete a task?" *In Proceedings of the 21st ACM Conference on Interaction Design and Children.*, 2022. [Pages 36 and 49.]

[72] N. Epley, B. Keysar, L. Van Boven, and T. Gilovich, "Perspective taking as egocentric anchoring and adjustment." *Journal of personality and social psychology*, vol. 87, no. 3, p. 327, 2004. [Online]. Available: https://psycnet.apa.org/buy/2004-18348-004 [Page 41.]

Appendices

Table 1: Statistical analysis of the global performance of participants per condition in the PT task.

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Number of mistakes	9.50	7.57	11.70	8.40	2.65	2.67
Number of clarifications	0.83	1.40	0.52	0.66	0.26	0.45
Number of tries	25.30	7.15	27.22	8.22	17.91	2.81
Perspective taking task time [min]	5.53	1.96	6.14	2.05	3.73	0.57

Table 2: ANOVA analysis of the number of mistakes made in each ambiguous instruction.

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 1	1.43	1.78	3.78	4.25	0.17	0.83
Instruction 8	0.78	0.74	0.83	1.50	0.17	0.49
Instruction 15	0.70	0.76	0.57	1.41	0.00	0.00

Table 3: ANOVA analysis of the number of clarifications asked in each ambiguous instruction.

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 1	0.09	0.29	0.35	0.57	0.04	0.21
Instruction 8	0.26	0.44	0.00	0.00	0.00	0.00
Instruction 15	0.30	0.47	0.00	0.00	0.00	0.00

Table 4: ANOVA analysis of the number of tries in each ambiguous instruction.

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 1	2.52	1.73	5.13	4.22	1.22	1.04
Instruction 8	2.04	0.98	1.83	1.50	1.17	0.49
Instruction 15	2.00	0.74	1.57	1.41	1.00	0.00

Table 5: ANOVA analysis of the amount of seconds spent in each ambiguous instruction.

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 1	16.91	13.48	31.54	25.61	6.51	3.49
Instruction 8	12.26	7.31	12.80	11.83	7.39	4.53
Instruction 15	12.35	6.61	10.61	14.39	5.90	2.12

Table 6: ANOVA analysis of the number of mistakes made in each instructions on the subset 1: 8 (ambiguous) and 14 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 8	0.78	0.74	0.83	1.50	0.17	0.49
Instruction 14	0.09	0.29	0.48	0.73	0.09	0.29

Table 7: ANOVA analysis of the number of clarifications asked in each instructions on the subset 1: 8 (ambiguous) and 14 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 8	0.26	0.44	0.00	0.00	0.00	0.00
Instruction 14	0.00	0.00	0.00	0.00	0.00	0.00

Table 8: ANOVA analysis of the number of tries in each instructions on the subset 1: 8 (ambiguous) and 14 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 8	2.04	0.98	1.83	1.50	1.17	0.49
Instruction 14	1.09	0.29	1.48	0.73	1.09	0.29

Table 9: ANOVA analysis of the amount of seconds spent in each instructions on the subset 1: 8 (ambiguous) and 14 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 8	12.26	7.31	12.80	11.83	7.39	4.53
Instruction 14	4.44	1.84	11.36	4.34	6.83	3.00

Table 10: ANOVA analysis of the number of mistakes made in each instructions on the subset 2: 15 (ambiguous) and 4 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 15	0.70	0.76	0.57	1.41	0.00	0.00
Instruction 4	0.04	0.21	0.48	0.85	0.13	0.34

Table 11: ANOVA analysis of the number of clarifications asked in each instructions on the subset 2: 15 (ambiguous) and 4 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 15	0.30	0.47	0.00	0.00	0.00	0.00
Instruction 4	0.00	0.00	0.00	0.00	0.00	0.00

Table 12: ANOVA analysis of the number of tries in each instructions on the subset 2: 15 (ambiguous) and 4 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 15	2.00	0.74	1.57	1.41	1.00	0.00
Instruction 4	1.04	0.21	1.48	0.85	1.13	0.34

Table 13: ANOVA analysis of the amount of seconds spent in each instructions on the subset 2: 15 (ambiguous) and 4 (non ambiguous).

	object-centred		robot-centred		human-centred	
	M	SD	M	SD	M	SD
Instruction 15	12.35	6.61	10.61	14.39	5.90	2.12
Instruction 4	5.57	2.22	10.99	7.52	7.87	2.56

Table 14: Instructions per condition (Part I). Example: ”Go with the...”

Instr.	Object-centred	Robot-centred	human-centred
1	pyramid, at the end of the line	pyramid, on the far right	pyramid, on the far left
1 *	pyramid, at the end of the line, on the side of the big box	pyramid, on my far right	pyramid, on your far left
2	pyramid, beside the black square pyramid	object, that I see on the left of the black square pyramid	object, that you see on the right of the black square pyramid
3	blue pyramid, beside a box	second pyramid, from my right	second pyramid, from your left
4	small pyramid, beside the small box	pyramid, on my right of the small box	pyramid, on your left of the small box
5	big blue pyramid, on the side of the big box	big blue pyramid, on my right	big blue pyramid, on your left
6	triangular pyramid, at the end of the line, on the side of the black square pyramid	triangular pyramid, on my far left	triangular pyramid, on your far right
7	big pyramid, beside the small box	pyramid, that I see on the left of the small box	pyramid, that you see on the right of the small box
8	box, between pyramids	box, on the left of the triangular pyramid	box, on the right of the triangular pyramid

* clarification of the instruction

Table 15: Instructions per condition (Part II). Example: "Go with the..."

Instr.	Object-centred	Robot-centred	human-centred
8 *	box, between pyramids of different heights	box, on my left of the triangular pyramid	box, on your right of the triangular pyramid
9	pyramid, beside the small blue pyramid	object, that I see on the right of the small blue pyramid	object, that you see on the left of the small blue pyramid
10	blue pyramid, beside the big box	pyramid, that I see on the right of the big box	pyramid, that you see on the left of the big box
11	big pyramid, beside a box	second pyramid, from my left	second pyramid, from your right
12	box, beside a blue pyramid	box, on my right	box, on your left
13	big blue pyramid, on the side of the small box	big blue pyramid, on my left	big blue pyramid, on your right
14	box, between pyramids of the same colour	box, on my right of the square pyramid	box, on your left of the square pyramid
15	triangular pyramid, beside the box	pyramid, on the left of the big box	pyramid, on the right of the big box
15 *	triangular pyramid, beside the big box	pyramid, on my left of the big box	pyramid, on your right of the big box

* clarification of the instruction

Table 16: Interpersonal Reactivity Index questions.

Fantasy Scale
When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.
I really get involved with the feelings of the characters in a novel.
I am usually objective when I watch a movie or play, and I don't often get completely caught up in it.
After seeing a play or movie, I have felt as though I were one of the characters.
I daydream and fantasize, with some regularity, about things that might happen to me.
Becoming extremely involved in a good book or movie is somewhat rare for me.
When I watch a good movie, I can very easily put myself in the place of a leading character.
Perspective-Taking Scale
Before criticizing somebody, I try to imagine how I would feel if I were in their place.
If I'm sure I'm right about something, I don't waste much time listening to other people's arguments.
I sometimes try to understand my friends better by imagining how things look from their perspective.
I believe that there are two sides to every question and try to look at them both.
I sometimes find it difficult to see things from the "other guy's" point of view.
I try to look at everybody's side of a disagreement before I make a decision.
When I'm upset at someone, I usually try to "put myself in his shoes" for a while.
Empathic Concern Scale
When I see someone being taken advantage of, I feel kind of protective toward them.
When I see someone being treated unfairly, I sometimes don't feel very much pity for them.
I often have tender, concerned feelings for people less fortunate than me.
I would describe myself as a pretty soft-hearted person.
Sometimes I don't feel sorry for other people when they are having problems.
Other people's misfortunes do not usually disturb me a great deal.
I am often quite touched by things that I see happen.

Table 17: 1/7 Likert scale questions in the final questionnaire.

Did you feel the robot's instructions were easy to follow?
How did you find the robot's communicative skills in instructing you to find the correct object?
How much did you find yourself caring about the robot?
How likeable did you find the robot?
How empathetic did you find the robot?
How great did you perceive the robot's need for collecting the speech data?

Table 18: Example of the corpus for the sentences for the principal emotions.

Tom beats that farmer.
John laughs like your father.
The seed is buried in deep.
Taylor likes stewed Asian food.
The lord swims in the sea.
Jack views an art piece.
Carl leaps into a jeep.
Linda asks for more darts.
Find your boot in this shute.
Water harms the new born boy.
I have not seen my tooth.
Work hard or you lose.
Jim saw the port.
They should start to talk.
Sound the horn if you need more.

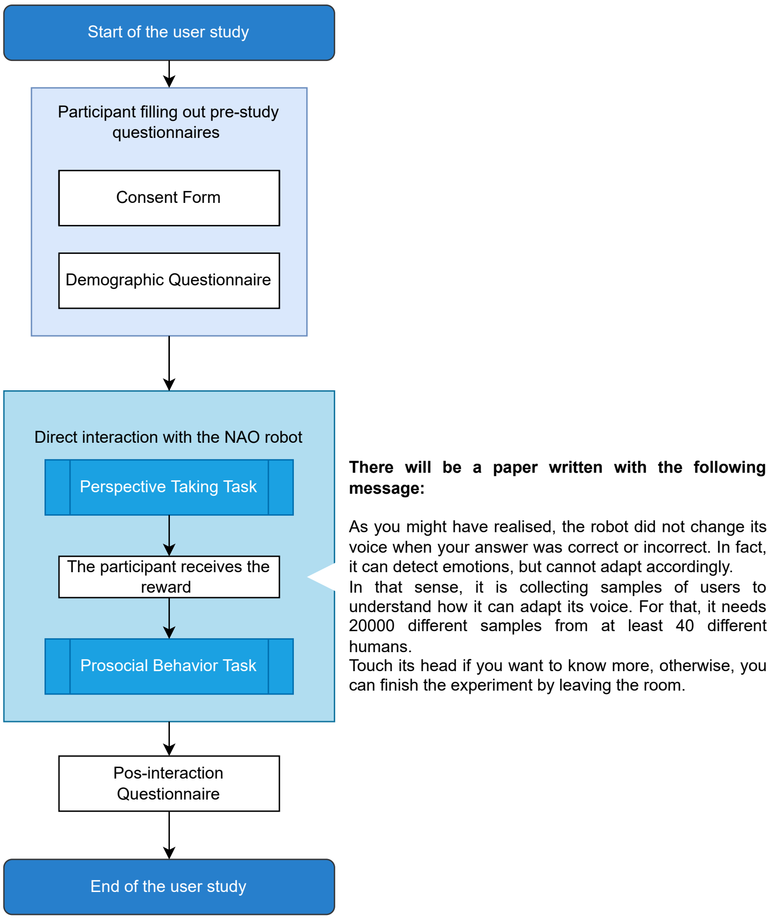
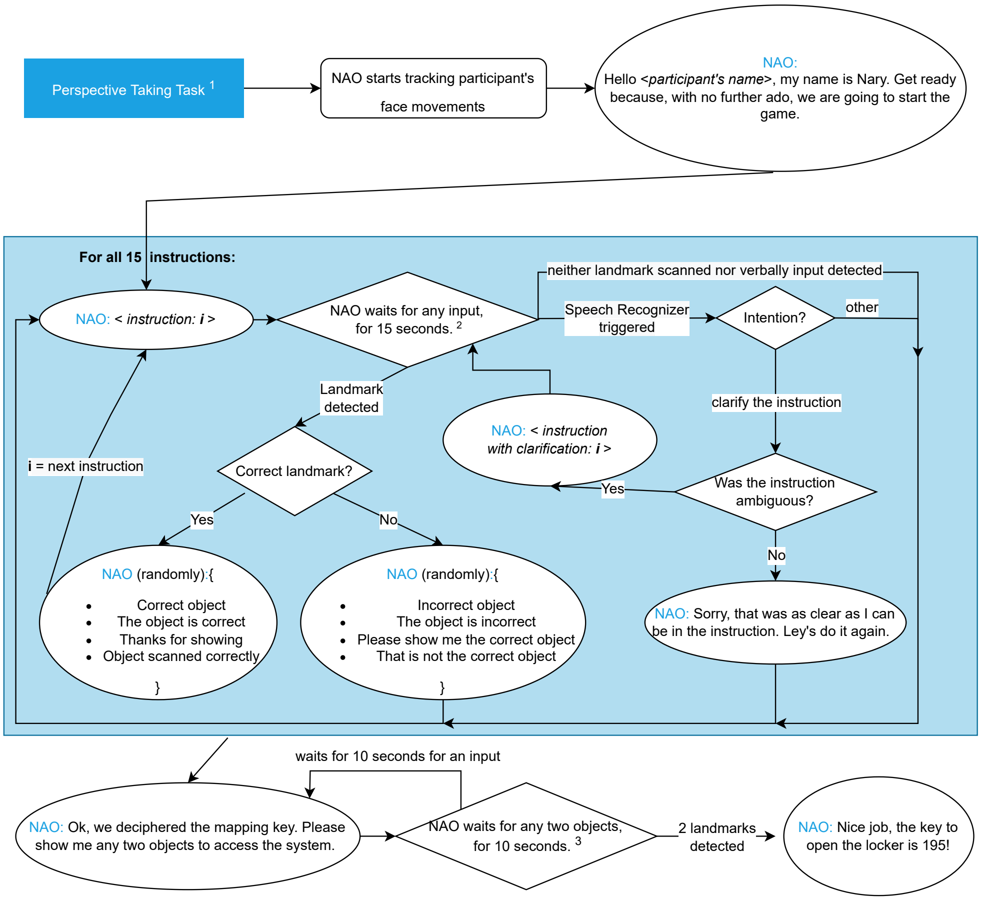


Figure 1: Architecture of the interaction.



1

Waits until the participant touches NAO's head to start the program, or it can also be triggered with a wizard of oz. While waiting, NAO reacts to landmarks by saying: "Landmark detected", to demonstrate to the participants how it works.

2

The program may receive the wizard of oz commands. Here are accepted the following instructions:

1. The landmark id;
2. Pause the task;
3. Repeat the instruction;
4. Clarify the instruction;
5. A standard answer in case the participant asks something the robot cannot answer by itself;
6. A standard answer in case the participant deviates from the topic of the task.

3

The program may receive the wizard of oz commands. Here are accepted the following instructions:

1. Pause the task;
2. Validates the participant's answer.

Figure 2: Design of Perspective-Taking task.

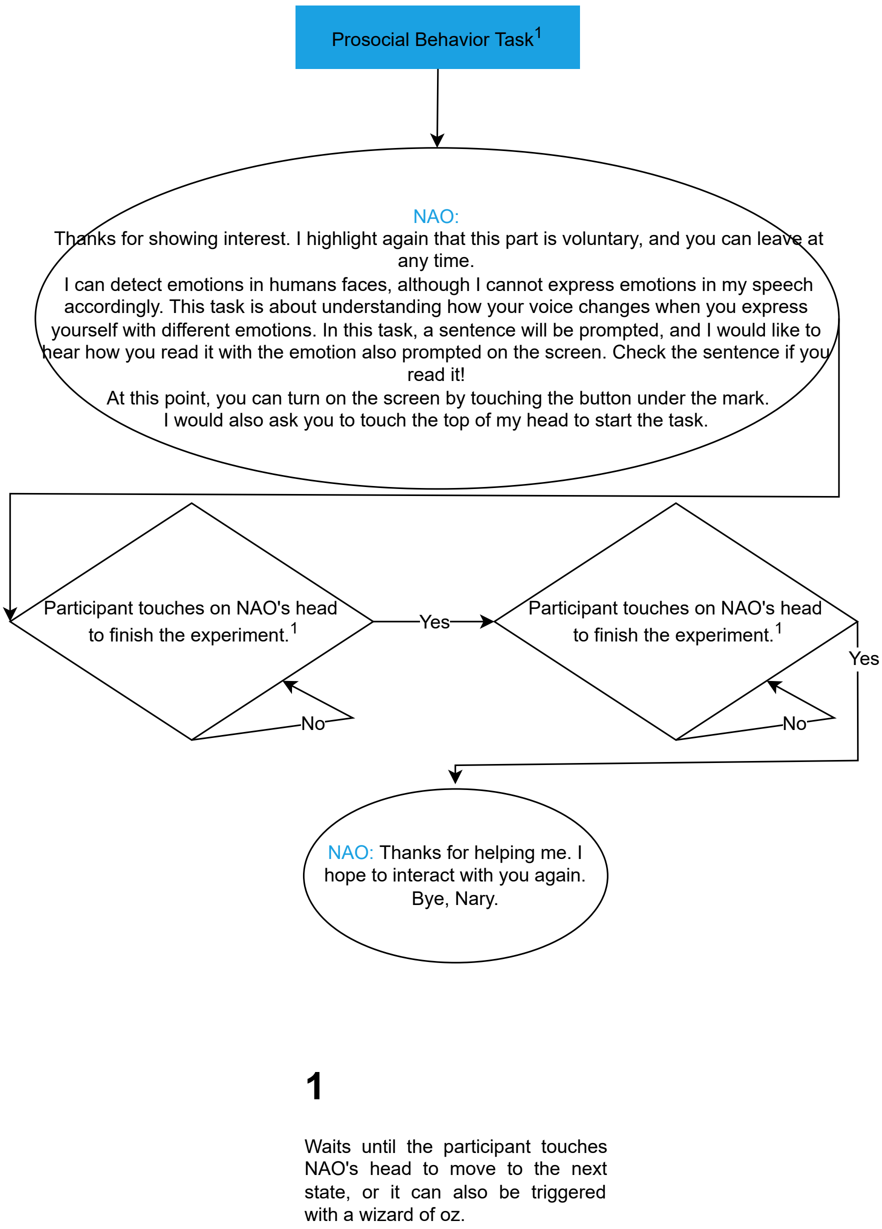


Figure 3: Design of the Prosocial Behaviour task.