EPFL

TÉCNICO LISBOA

Thèse n° 8352

Exploring Spatial Perspective Taking in Human-Robot Interaction

Présentée le 19 octobre 2021

à l'École Polytechnique Fédérale de Lausanne
Faculté informatique et communications
Laboratoire d'ergonomie éducative
Programme doctoral en robotique, contrôle et systèmes intelligents

et à l'Instituto Superior Técnico (IST) da Universidade de Lisboa
Doutoramento em Engenharia Informática e de Computadores
pour l'obtention du grade de Docteur ès Sciences

par

Elmira YADOLLAHI

Acceptée sur proposition du jury

Dr D. Gillet, président du jury
Prof. P. Dillenbourg, Prof. A. Paiva, directeurs de thèse
Dr V. Charisi, rapporteuse
Prof. K. Hindriks, rapporteur
Prof. F. Mondada, rapporteur

École
polytechnique
fédérale
de Lausanne

2021

In memory of my loving father,
who never saw the end of this adventure…

To my mother who taught me strength,
And to Thomas, with all my love…

Acknowledgements

This thesis would not have been completed without the support of many people spread around the world that extended their unconditional support and kindness despite the distance and time. First and foremost, I would like to thank my two wonderful supervisors Prof. Ana Paiva and Prof. Pierre Dillenbourg, who not just guided, inspired, encouraged, and advised me professionally but also supported me personally these past years. I started my PhD journey with Pierre in Switzerland where every meeting with him meant leaving the room with a newfound curiosity on the topics of discussion. The idea of working on perspective taking initiated from one of these meetings and it continues to be my passion for years to come. Working with Ana and moving to Portugal has been an exhilarating experience, where every meeting have made me more motivated and confident. Ana's unrelenting support throughout this process has made the experience more enjoyable and valuable. There is not enough words to express how grateful I am to have had her during my hardest times.

I would like to extend my gratitude to the members of the committee: Prof. Koen Hindriks, Dr. Vasiliki Charisi, Prof. Francesco Mondada and Dr. Denis Gillet, who kindly accepted to review this manuscript during the summer and provided me with valuable insights and feedback.

I am sincerely thankful to Prof. Wafa Johal for being one of the best mentors I have had. She is the first expert who taught me all the basics of human-robot interaction and a role model. I am grateful for our discussions and her invaluable support on every aspect of my work.

Special thanks to Dr. Marta Couto, for always being available to discuss my research, to check my statistical analyses and to just talk about life. It was an immense pleasure to learn from her not just as a colleague but also as a friend and I hope we can keep collaborating in the future.

To all my colleagues at CHILI lab, my home for the first half of my PhD, I am grateful for having you in my personal and professional life. From Skiminars to Christmas walks and all the wonderful discussions during lunchtime. To Mina Shirvani Boroujeni; for giving me invaluable advice, to Arzu Göneysu Özgür; a wonderful friend and collaborator, to Thibault Asselborn, never failing to make us laugh, to Lorenzo Lucignano, guiding me with all the Mafia tricks, to Khalil Mrini, a delight to talk, and to Ayberk Özgür, never failing to have a brilliant idea. I am also thankful to other members of the lab, Kevin Gonyop, Kshitij, Jennifer Olsen, Catharine Oertel, and Louis Faucon. Special thanks to Thanasis Hadzilacos, who kindly supported me during my candidacy exam and inspired me with great conversations. Most importantly, I am immensely grateful to Florence Colomb, who is a fantastic administrative staff and a true friend and supporter. To know that she is in the CHILI is to know that the lab is in great hands.

To my colleagues at GAIPS, my home for the rest of my PhD and during the pandemic, I could not have asked for a better lab in Portugal with such a dynamic and stimulating research environment. My special thanks goes to Maria José Ferreira for always helping me unconditionally and being a lovely friend and to Dr. Shruri Chandra for all the fantastic collaborations and fruitful discussions on research and child-robot interaction. To Miguel Vasco, a considerate friend, to Raquel Oliveira, a wonderful collaborator, to Mojgan Hashemian, a fantastic flatmate, to Dr. Filipa Correia, an inspiring scientist, to Ali Kordia, with tasty barbecues, and to Ricardo Rodrigues, always a delight to talk to. I am also thankful to Diogo Carvalho, Silvia Tulli, Diogo Rato, Dr. Samuel Mascarenhas, Carla Guerra, Kim Baraka, Raul Paradeda, Ramona Merhej, Rui Silva, and Miguel Faria. I would also thank Sandra Sá for being so supportive and making all the administrative processes as smooth as possible.

I am sincerely thankful to Prof. Francisco Santos for his continuous support and kindness, to Prof. Rui Prada for giving me an opportunity to learn from him in classes as a student and an assistant, and toz Prof. Carlos Martinho for his feedback and suggestions on my work. I am also grateful to all the administrative staff in both institutions, particularly Corinne Lebet and Sandra Roux at EPFL and the postgraduate office at IST, who have supported with the endless administrative work. Further, I am thankful to Fundação para a Ciência e a Tecnologia (FCT) PD/BD/135150/2017 for supporting my research.

My deepest gratitude goes to my friends who have supported me from distance, to Sony, Lym, and Pouya for being my extended family since south Korea and to Azadeh and Sareh for extending their kindness and advice and being there since our time in Iran.

To the memory of my father, Ghadir, I am truly blessed to be your daughter and never forget the impact you had by believing in me and trusting me. My toughest hours were seeing you battling cancer and still caring about us more than yourself and I want to dedicate this achievement to you. And to my mother, Omimeh, you are my true inspiration and role model, without your guidance, I would have never succeeded to be where I am today. Your everlasting love, dedication, and sacrifice have always given me the strength to keep pushing forward. To my brothers Ali and Mohamadreza, you are the greatest brothers I could have asked for. You have been always there for me when I needed support and sacrificed a lot during our hardest times, I am so proud of you, and believe that you both will have a bright future.

To my extended family, Jean-Claude, and Jung-Yea, I am deeply grateful for your love and guidance and to Marie and Konstantin, I am blessed to have had your support during my hardest time. I truly appreciate all of your encouragements and kindness in the last stages.

Last but not least, to my partner Thomas, my heartfelt and everlasting gratitude goes to you. Thank you for your unfailing love, exceptional understanding, and never-ending support through the most challenging stages of my life and for the exciting future we have ahead of us.

Lisbon, 26 July 2021 Elmira Yadollahi

Abstract

Humans often rely on their perspective taking skills to thrive within the world's complex relations and connections. An adequate understanding of others' spatial perspectives can increase the quality of the interaction, not only perceptually but also cognitively. This thesis is dedicated to exploring children and adults' spatial perspective taking abilities emerging from the interaction with embodied and virtual robots in different contexts. While most previous approaches were limited to particular circumstances and targeted adults, the proposed approach developed a cognitive model incorporated in agents to foster perspective taking abilities in different contexts. The developments of the model also detail the processes used by humans to infer spatial connections from other's viewpoints and it is adaptable to add or remove processes based on the context of interaction.

First, this thesis explores different interaction modalities and cognitive abilities through user studies with children. Each interaction outlines a set of components and processes required to develop a perspective taking model for robots and agents. The platform developed for the first study aims at evaluating the effect of a robot's non-verbal gestures, such as pointing, on children's joint attention during reading activity. The second study evaluates children's perspective adaptation to the robot in the context of collaborative activity. We expanded our observations to include children's first perspective choice, how they tried to accommodate the robot's perspective, and how they updated their mental model during the interaction. The third study explores children's spatial perspective taking abilities using game-based interaction and non-verbal channels.

Inspired by the psychological studies and the findings from the exploratory studies, the thesis proposes a cognitive model that uses automatic and cognitive controlled processes to generate behaviors and decisions for the robot. The model mainly focuses on processes linked with taking spatial perspectives and can be integrated into any agent architecture that deals with decision making and reasoning. The agent processes are then adapted to a new interaction scenario inspired by one of the exploratory studies. Finally, the thesis evaluates the model through a user study with adult participants and within a virtual platform, a drastic change from the exploratory study designs caused by the pandemic. The final studies are designed as two-player games with two virtual robots that interact with each other in two contexts of competition and cooperation. In the competitive version of the game, the robot guided by the human plays against the robot equipped with the perspective taking model, while in the cooperative version, the robots guide each other to win the game as a team. We performed two between-subject studies with more than 180 adults to evaluate the participants' perception of a robot with complete perspective taking abilities, compared to one with limited abilities. Participants were more influenced by the robot's perspective taking abilities in the cooperative game compared to the competitive one, which was reflected in their ratings of the robot's intelligence and game fun. Experimental results on the model evaluation can open up future possibilities for exploring links between the perspective developments in children through cooperation and competition. Furthermore, the model can be extended to study other perceptional, cognitive, and affective dimensions of perspective taking, such as prosocial behavior and transparency.

Keywords: Spatial Perspective Taking, Human-Robot Interaction, Child-Robot Interaction, Cognitive Modelling, Joint Attention, Reading, Gamified Interaction, Cooperation, Competition

Résumé

Les humains comptent souvent sur leur capacités à la prise de perspective pour évoluer au sein de relations et d'interactions complexes. Une comprehension adéquate des perspectives spatiales de l'autre peut décupler la qualité d'une interaction, non seulement au niveau de la perception mais également sur le plan cognitif. Cette thèse est dédiée à l'exploration des capacités de prise de perspective spatiale chez l'adulte et l'enfant émergeant d'interactions avec des robots physiques comme virtuels dans différents contextes. Alors que la plupart des précédentes approches étaient limités à des circonstances particulières est ciblaient les adultes, l'approche proposée a développé un modèle cognitif incorporé dans des agents afin de promouvoir une capacité à la prise de perspective dans plusieurs contextes. Le dévelopement du modèle précise également les processus utilisés par les humains pour déduire des connections spatiales du point de vue d'autres personnes et il est adaptable pour ajouter ou supprimer des processus basés sur le contexte d'interaction.

Dans un premier temps, cette thèse explore différentes modalités d'interaction et capacités cognitives au travers d'études utilisateur avec les enfants. Chaque interaction met en lumière un jeu de composants et de processus requis à l'élaboration d'un modèle de prise de perspective pour des robots et des agents. La plateforme dévelopée pour la première étude vise à évaluer l'effet d'une gestuelle non-verbale, telle que pointer du doigt, sur l'attention conjointe avec l'enfant dans un contexte de lecture. La deuxième étude évalue l'adaptation de perspective de l'enfant à celle du robot dans une activité collaborative. Nous avons élargis nos observations afin d'inclure le premier choix de perspective des enfants, la manière dont ils essayaient de s'accomoder à la perspective du robot, et comment ils mettaient à jour leurs représentations mentales lors de l'interaction. La dernière étude explore les capacités des enfants à la prise de perspective spatiale en utilisant des interaction basée sur le jeu et des canaux de communications non-verbaux.

Inspiré par les études psychologiques et les résultats des études exploratoires, la thèse propose un modèle cognitif utilisant des processus automatiques pour certains et pour d'autres sous controle congnitif pour générer des comportements et prises de décisions du robot. Le modèle se consacre principalement sur des processus liés à la prise de perspective spatiale et peut être intégré à n'importe quelle architecture d'agent affecté à des tâches incluant des prises de décision et du raisonnement. Les processus de l'agent son alors adaptés à un nouveau scénario d'interaction inspiré par une des études exploratoires.

Enfin, la thèse évalue le modèle au travers d'une étude utilisateur avec participants adultes sur une plateforme virtuelle, un changement drastique par rapport aux schéma des études exploratoires en raison de la pandémie. Les études finales sont conçues comme un jeu à deux joueurs avec deux robots virtuels interagissant dans uncontexte de compétition et de coopération. Dans la version compétitive du jeu, le robot dirigé par l'humain joue contre le robot équipé du modèle de prise de perspective, alors que dans la version coopérative, les robots se guident mutuellement pour gagner le jeu en équipe. Nous avons effectué deux études inter-sujet avec plus de 180 adultes pour évaluer la perception des participants d'un robot avec prise de perspective totale, comparée à la perception d'un robot avec des capacités limitées à la prise de perspective. Les partipants furent davantage influencés par les capacités du robot à la prise de perspective dans la version coopérative que dans la version compétitive, ce qui fut évalué dans leurs évaluation de l'intelligence du robot et de l'amusement qu'ils ont eu dans la pratique du jeu. Les résultats expérimentaux sur l'évalutation du modèle peuvent ouvrir d'autres possibilités pour explorer des liens entre les dévelopements de la perspective chez l'enfant au travers de la coopération et de la compétition. En outre, le modèle peut être étendu pour étudier d'autres dimensions perceptives, cognitives et affectives de la prise de perspective, telles que le comportement prosocial et la transparence.

Mots Clés : Prise de perspective spatiale, Interaction Homme-Robot, Interaction Enfant-Robot, Modelisation cognitive, Attention conjointe, Lecture, Ludification de l'interaction, Coopération, Compétition

Resumo

Os seres humanos confiam frequentemente nas suas capacidades de tomada de perspetiva para prosperarem nas complexas relações e conexões do mundo que os rodeia. Uma compreensão adequada da capacidade da perspetiva espacial dos outros pode aumentar a qualidade de uma interação, não apenas a nível preceptivo, mas também cognitivo. Esta tese é dedicada à exploração das capacidades de tomada de perspetiva espacial de crianças e adultos, que emergem da interação com robôs físicos e virtuais em diferentes contextos. Enquanto muitas abordagens anteriores se limitaram a circunstâncias particulares e populações adultas, a abordagem aqui apresentada, desenvolve um modelo cognitivo incorporado nos agentes para promover a tomada de perspetiva em diferentes contextos. O desenvolvimento do modelo detalha ainda os processos usados por humanos para inferirem conexões espaciais a partir do ponto de vista do outro, sendo ainda adaptável para que possam ser adicionados ou removidos, processos baseados no contexto da interação. Esta tese começa por explorar diferentes modalidades de interação e capacidades cognitivas, através de estudos com crianças. Cada interação descreve um conjunto de componentes e processos que são necessários ao desenvolvimento de um modelo cognitivo de tomada de perspetiva para agentes e robôs. A plataforma desenvolvida para o primeiro estudo, tem como objetivo avaliar o efeito de gestos comunicativos como o apontar, no comportamento de atenção conjunta das crianças durante uma atividade de leitura. O segundo estudo, avalia a adaptação da perspetiva das crianças ao robô, no contexto de uma atividade colaborativa. Neste estudo, expandimos a nossa observação para incluir a primeira escolha de perspetiva por parte das crianças, como estas tentam depois acomodar a perspetiva do robô e ainda como atualizam os seus modelos mentais durante a interação. O estudo final, explora a capacidade de tomada de perspetiva espacial, usando para isso um jogo e canais não verbais. Inspirada em estudos na área da psicologia e nos resultados dos estudos exploratórios, esta tese propõe um modelo cognitivo que usa processos cognitivos automáticos e controlados para gerar comportamentos e decisões para o robô. O modelo foca-se principalmente nos processos ligados à tomada de perspetiva espacial e pode ser integrado em qualquer arquitetura de agentes que lide com tomada de decisão e raciocínio. Os processos do agente foram posteriormente adaptados a um novo cenário de interação inspirado pelos estudos exploratórios. A tese avalia ainda o modelo, através de um estudo com utilizadores adultos e numa plataforma virtual. Uma mudança drástica em relação aos estudos exploratórios causada pela pandemia. Os estudos finais foram feitos com recurso a um jogo com dois robôs virtuais que interagem num contexto de competição e num contexto de cooperação. Na versão competitiva do jogo, o robô, guiado pelo humano, joga contra um robô equipado com o modelo de tomada de perspetiva. Na versão colaborativa, os robôs guiam-se um ao outro para ganharem o jogo como equipa. Realizámos dois estudos intersujeitos com mais de 180 adultos para avalia a perceção dos participantes em relação a um robô com um modelo completo de tomada de perspetiva comparado com um robô com capacidades limitadas de tomada de perspetiva. Os participantes foram mais influenciados pelo robô com capacidade de tomada de perspetiva no jogo cooperativo do que no jogo competitivo, algo que se refletiu nas avaliações que os participantes fizeram da inteligência do robô e do quão divertido foi o jogo. Os resultados empíricos da avaliação do modelo, abrem possibilidades que seja explorado o desenvolvimento de tomada de perspetiva em crianças através da cooperação e competição. Para além disto, o modelo pode ser expandido para estudar dimensões percetivas cognitivas e afetivas da tomada de perspetiva como por exemplo o comportamento pró-social e a transparência.

Palavras-Chave: Tomada de perspectiva espacial, Interação Humano-Robô, Interação Criança-Robô, Modelos Cognitivos, Atenção Conjunta, Leitura, Interação Gamificada, Cooperação, Competição

Contents

List of Figures

List of Tables

Acronyms

CogPeT	Cognitive Model of Perspective Taking. 103, 131, 135
CRI	Child-Robot Interaction. 4, 26
EC	Empathic Concern. 110, 173
EJA	Ensure Joint Attention. 34, 52
FS	Fantasy. 110, 173
HRI	Human-Robot Interaction. 3, 25, 26
IJA	Initiate Joint Attention. 21, 34, 52
IRI	Interpersonal Reactivity Index. 130
MOOC	Massive Open Online Course. 34
PD	Personal Distress. 110, 173
PT	Perspective Taking. 3, 13–15, 110, 131, 173, 176
RJA	Respond to Joint Attention. 21, 34, 52
RMI	Reading Miscue Inventory. 38
ROS	Robot Operating System. 36, 143–145
RWL	Reading While Listening. 33, 34, 53
SMA	Simplified Miscue Analysis. 38, 39
SPT	Spatial Perspective Taking. 7, 15–17
ToM	Theory of Mind. 4, 7, 15, 19, 20, 23
VPT	Visual Perspective Taking. 15–17, 23

Introduction Part I

1 Setting the Scientific Scene

We often rely on our cognitive skills to navigate the social world. To thrive within the world's complex relations and connections, we need to understand other people's minds and perspectives. This ability is what distinguishes us from other species. Depending on the type of interaction, different levels of comprehension are required to interact effectively; this can be as pivotal as perceiving someone's visual perspective e.g., what they see, or as complex as comprehending their mental state or belief system e.g., what they think. According to Bratman et al.'s model of human practical reasoning, our actions are directed by our goals, intentions, beliefs, and desires, which differs from others' (Bratman et al., 1988). Philosophers and psychologists define the ability to predict the actions of self and others as Theory of Mind (ToM) (Flavell, 2004). Additionally, to interact, cooperate, or compete with others, we need to have a grasp of their perspectives manifested in a fundamental skill called Perspective Taking (PT) (Piaget et al., 1960; Tomasello, 2010). In recent years different literature has pinned different definitions to perspective taking, all of which outline varying degrees of perception or understanding of another person's perspective (Surtees et al., 2013). This thesis is dedicated to exploring children and adults' spatial perspective taking emerging by interacting with embodied and virtual robots in different contexts.

Humans are inherently social beings able to carry out fluid and dynamic interactions with the ability to consider various aspects simultaneously (Boland Jr & Tenkasi, 1995; Clark, 1992; Clark & Marshall, 2002). Understanding our counterpart's perspective or taking it into consideration during interaction is one of our many efficacious abilities (Flavell et al., 1986; Galinsky & Moskowitz, 2000; Piaget et al., 1960). We tend to decide on the spur of the moment how to steer the interaction, and whether to consider our counterpart's perspective or not. Correspondingly, to enhance the quality of Human-Robot Interaction (HRI), one of the aspects worth consideration is perspective taking (Berlin et al., 2006; Torrey et al., 2009; Trafton et al., 2005; Zhao et al., 2016). Looking at perspective taking in human-robot interaction scenarios, various questions come to mind, and in recent years a growing number of studies have been trying to tackle these questions (Fischer, 2018; Robins et al., 2017; Warnier et al., 2012). While the majority of previous research in human-robot interaction looks at perspective taking dynamics with adults, more recent studies started looking at Child-Robot Interaction (CRI) scenarios with autistic children (Robins et al., 2017; L. Wood et al., 2017). Moreover, the introduction of robots into education and interaction with children has introduced new paradigms and modalities to the fields of education and human-robot. One of the crucial aspects of educational scenarios is maintaining mutual understanding between the child and the robot. To maintain such an understanding, it is inevitable for the child and the robot to develop a model of each other's minds and perspectives. To have robots with capabilities to carry out educational roles, play games, be peers in the activities of a classroom, and at the same time, support learning in different forms is a challenging task. To achieve that, we need to equip our robots with cognitive abilities that help them to become true learning companions. To endow the robots with cognitive abilities, we can either focus on the cognitive development, the interaction capabilities of the robot or develop both aspects simultaneously.

1.1 Research Statement

The research presented in this thesis has emerged from our desire to study and evaluate the cognitive abilities required by the robot to produce a socially and technically well-structured interaction with children in learning contexts. The initial step to achieve this goal was to determine which cognitive properties were required to be implemented in the robot. Essentially, we were interested in comprehending the dynamics between the robot and the child on a spatial perspective taking level and supplementing the robot with a framework that facilitates and strengthens its interaction capabilities in different contexts, particularly in educational settings. To better understand children's decision-making mechanism we got inspiration from studies in psychology that have studied the underlying mechanisms of perspective taking (Elekes et al., 2017; Newcombe, 1989; Todd et al., 2017; Vander Heyden et al., 2017). Additionally, we explored children's perspective taking abilities and their tendency to adapt their perspective to a robot through different interaction scenarios. We decided to start with spatial perspective taking as it lets us utilize the robot's embodiment and physical interaction and facilitate the assessment of children's perspective taking abilities through the activity. From a Theory of Mind (ToM) point of view, for children to master taking other's perspective, they need to master five levels of understanding informational states (Barnes-Holmes et al., 2004; Howlin et al., 1999). According to Barnes-Holmes and colleagues' model, visuospatial perspective taking contributes to the first two levels of theory of mind development. Our focus on spatial perspective taking can be considered as a starting point to develop a more comprehensive model that extends to other perspective dimensions and theory of mind levels. On the other hand, we paid special attention to the context of interaction as an important factor in steering the direction of our research. We designed activities and interaction scenarios that can contribute to children's development through educational and game-based learning. Ultimately, the components fundamental to developing our research are; spatial perspective taking, human-robot interaction, educational scenarios, and gamification.

1.2 Objectives

While thinking and conceptualizing the fundamental aspects of our research, we found many still unanswered questions that steered the direction of this thesis. To be able to address the questions within the scope of a doctoral thesis, we have formulated them into objectives. The objectives presented below are going to be translated into concrete research questions, investigated through studies with children and adults, and presented with details in each chapter. As shown in Figure 1.1, the objectives formulated in this thesis progress through time, from a data-driven approach to a model-driven one. Each objective has been investigated with user studies and described in detail in their related chapters. The main objectives of this thesis can be summarized as follows:

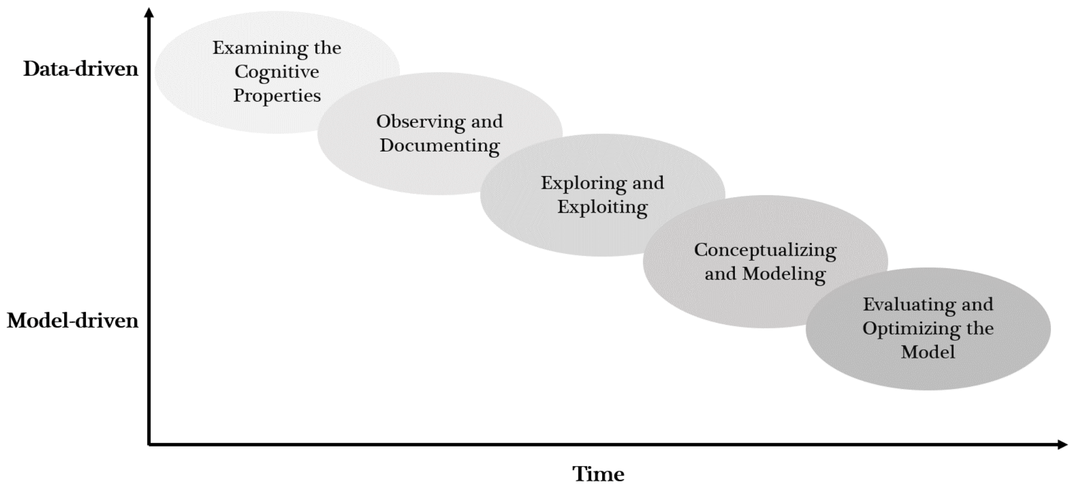


Figure 1.1 – The objectives of the thesis which guide the organization of the chapters and their associated studies.

- **Understanding the interaction and its cognitive needs.** The initial steps in defining and organizing the cognitive properties that enhance the interaction brought us to ask some fundamental questions about the basics of interaction. This led us to design a study to investigate joint attention as our first step into investigating perspective taking. The developed platform and the findings of the study on interaction modalities, spatial arrangement, and context of the task became the building block for the next steps.

- **Observing and documenting children's perspective taking behaviour.** Developmental psychology's account of how perspective taking abilities develop in children heavily relies on experimental observations and testing. We used the developmental models to design activities and interactions with robots for children. However, before incorporating a model inspired by developmental psychology, we decided to observe and document children's behaviour with the robot in the tasks we designed and developed.

- **Exploring and exploiting different modalities of perspective taking.** The observation and documentation part of the research brought valuable insights on how to develop a model for the robot, specifically by highlighting the features that needed more attention. Our initial studies heavily relied on the usage of natural language which is one of the weak spots of developing child-robot interaction scenarios. To explore the topic from different dimensions and provide more autonomous functionalities, we expanded the studies to a different mode of interaction and context. Our last study in the exploration phase focuses on game-based learning and non-verbal interaction.

- **Conceptualizing and developing a model for spatial perspective taking.** Inspired by the models in developmental psychology in conjunction with our studies, we conceptualized the components involved in an interaction with perspective taking. To investigate the processes that could contribute to the modeling, we used some of our findings from the exploratory research. Then, we developed the cognitive processes for agents engaged in such interactions. The proposed model can be integrated with any symbolic agent architecture.

- **Evaluating the model using human perception analyses.** To evaluate the model, we opted for designing a platform inspired by one of the exploratory studies. The platform included two virtual robots, one controlled by a human and the other was programmed using symbolic reasoning to play against or with a robot controlled by the human. Our method of evaluating the model performance was to document the participant's perception of the robot's abilities when it was equipped with the perspective taking model compared to when it had limited perspective taking functionality. To conclude the thesis, we provide details for how the mode can be adapted to other contexts.

1.3 Expected Contributions

We expect our work to contribute to the field of human-robot interaction in several aspects. Figure 1.2 shows an overview of the research carried out in this thesis visualized in three layers. From top to bottom the layers are; exploratory studies, model development, and model evaluation. The exploratory studies with embodied robots provide us with insights into the cognitive abilities of the child and robot. The model development layer allows us to formalize a perspective model for robots and agents. Model evaluation with virtual robots evaluates how adults perceive the robot equipped with the model that was developed in the previous section.

The first layer includes three studies exploring different interaction modalities and cognitive abilities for children when interacting with robots. Each user study can be looked at as a standalone project investigating the impact of robots on children's learning. Furthermore, we have developed different platforms incorporating NAO or Cozmo robot for each study. While each platform can be used to study different aspects of spatial perspective taking, it can

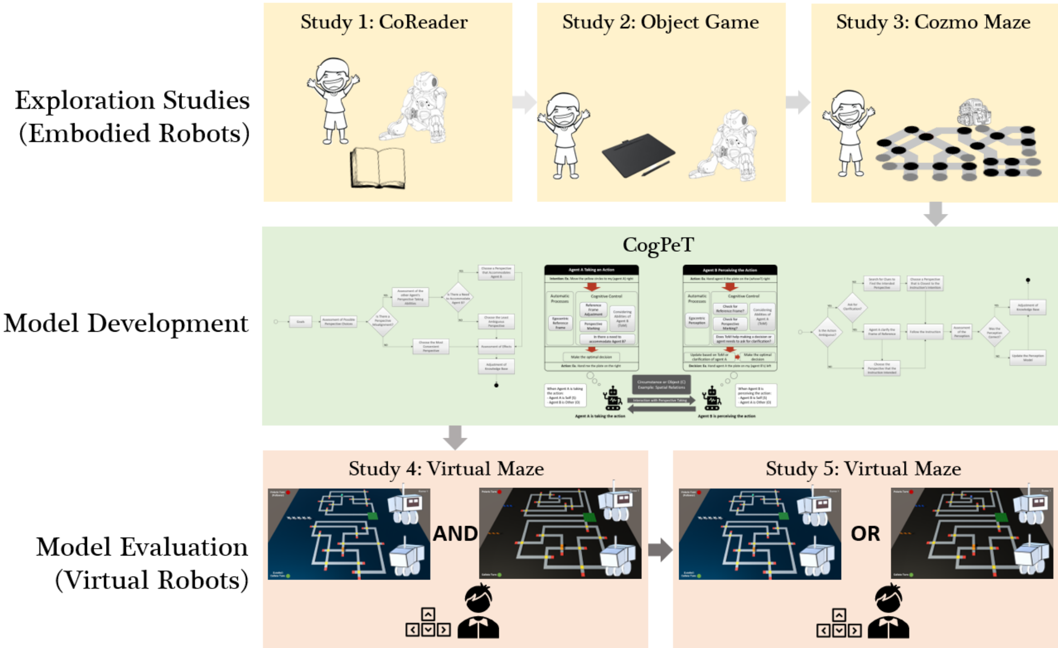


Figure 1.2 – Thesis overview.

also be expanded beyond the scope of this thesis to study other cognitive abilities required in child-robot interaction. The model development layer can provide valuable insights in developing a cognitive model of perspective taking for future research in the field.

The perspective taking model is developed with limited observations and functionalities to fit in the scope of a doctoral research, however, it has the potential to be expanded to a more comprehensive model not only to be used for Spatial Perspective Taking (SPT), but also in other perspective dimensions. For example, it can be used to study the effect of spatial perspective taking practices with competitive or cooperative priming on the participant's empathy or cognitive perspective taking abilities. The model can further contribute to the Theory of Mind (ToM) studies with robots and can be incorporated in their modelling and developments. In ToM models, visuospatial perspective taking is considered to be part of the early developmental levels. If the activities described in this thesis are expanded to other dimensions, together they can contribute to ToM studies with robot and furthermore to developing ToM models for robots. Additionally, the model can be improved by integrating transparency features and contribute to the field of transparency in robotics and AI. One of the approaches to develop a robot or agent with transparency is to equip them with the ability to understand their counterpart's perspective or theory of mind.

In the model evaluation layer, due to the restrictions imposed by the pandemic, we had to alter our developments to accommodate virtual interaction with adults. This shows the adaptability of the proposed model not only to embodied and virtual agents but also to adults and children. The developed virtual interaction used in the final two studies was inspired from the study with Cozmo robot presented in the exploration layer. The activity is developed with the capability of direct mapping to studies with embodied robots and children for possible future studies. This means, beyond the scope of this thesis, the studies can be carried out with children in both virtual and physical domains to further evaluate the model and children's learning. On the other hand, the platform used for the last two studies is designed for both cooperative and competitive interactions and it has shown promising results in training the adults spatial perspective taking abilities. The comparison between participants' performance and perception of the robots in the two contexts can inspire more research on the importance of perspective taking in human-robot cooperation. Taken together the results from that comparison raise questions that have the possibility to inspire new research.

1.4 Thesis Roadmap

This thesis is organized over the span of five Parts divided into nine Chapters with the addition of five Appendices as described below:

- **Part II: Theoretical Background**
  Provides the related background relevant to the development of the research presented in this thesis. This part is divided into two chapters with a focus on the two main contributing factors of the research; perspective taking and human-robot interaction.

  - **Chapter 2:** *Spatial Perspective Taking.*
    This chapter gives a detailed overview of perspective taking in social and developmental psychology. It describes all different dimensions and sub-dimensions and theories associated with perspective taking and in particular about the perceptual dimension that is the main focus of this chapter. Furthermore, it presents the state of the art research in robotics with respect to perspective taking and modeling it.
  - **Chapter 3:** *Human-Robot Interaction.*
    This chapter looks at different aspects of the research related to robots. It briefly describes the evolution of human-robot interaction, particularly child-robot interaction and includes applications that were covered in this thesis.

- **Part III: Exploring Perspective Taking in Interaction**
  We describe three user studies with children and robots developed to explore joint attention and perspective taking in the interaction. Each study is carried out using a platform developed specifically for the unique research questions in that study. All studies contribute to our understanding of different features and cognitive abilities that can improve child-robot interaction.

  - **Chapter 4:** *CoReader: An Exploratory Study.*
    The exploratory research initiates with a classic child-robot interaction study that evaluates children's joint attention in a reading activity with the NAO robot. The

study mainly evaluates the robot behaviour that can positively or negatively impact children's performance by impeding their joint attention during reading activity.

- **Chapter 5:** *Objects Game: A Behavioural Study*
  The second user study explores children's perspective adaptation when interacting with the NAO robot. The study mainly evaluates how children perceive the robot's perspective, build a mental model of the robot, adapt and update it according to the interaction.
- **Chapter 6:** *Cozmo Maze: An Exploratory Study*
  The final exploratory study is designed as a groundwork to develop interactions with non-verbal communication using the Cozmo robot. The interaction is primarily focused on gamifying perspective taking practices and it evaluates children learning and improvement during the interaction.

■ **Part IV: Modelling Perspective Taking and Evaluation**
We started conceptualizing the components and features that can contribute to perspective taking model. We later adapted the model to an activity inspired by the one used in Chapter 6 and carried out two studies to evaluate the model. The following two chapters detail the account of model developments and evaluation.

- **Chapter 7:** *CogPeT: A Cognitive Model of Perspective Taking*
  This chapter conceptualizes the components and features that can contribute to a perspective taking model. Then it describes how the model is inspired by the research in developmental psychology, including the components observed in the exploratory studies. The model particularly focuses on spatial perspective taking and its goal is to be more adaptable to the abilities of the human counterpart in child-robot interaction.
- **Chapter 8:** *Virtual Maze: An Evaluation Study*
  The model evaluation relies on documenting adults perception of the virtual robots while cooperating and/or competing with them. A group of adults interacted with a virtual robot with the perspective-taking model, while another group interacted with a limited perspective virtual robot. The chapter provides a full description of the platform development, participant's perception of the robot based on the context of the interaction, and robot's perspective taking abilities.

■ **Part V: Summary and Conclusions**

- **Chapter 9:** *Conclusions and Outlook*
  This chapter presents a summary of the thesis and draws conclusions from the findings of the research reported here. It highlights the research along with its limitations and discusses the potential future research that can be continued or adapted from the research carried out within this thesis.

■ **Appendices**

- **Appendix A** describes the robots and platforms used to implement the system.
- **Appendix B** provides details of the games and the tasks developed in the span of the thesis.
- **Appendix C** contains all the tests used in the user studies either adapted or modified from related studies.
- **Appendix D** presents all the questionnaires used within the user studies carried out as part of this thesis.
- **Appendix E** contains a list of all peer-reviewed publications that resulted from the thesis and a summary of how they contribute to the thesis.

This chapter provided the motivations for this research by introducing perspective taking and its implementation in human-robot interaction. It outlined the objective and potential contributions and presented an overview of the thesis chapter by chapter. The next two chapters will provide a detailed background on the two important components of this thesis, perspective taking and human-robot interaction.

Theoretical Background Part II

2 Spatial Perspective Taking

The theoretical background part has been divided into two chapters to cover the various research fields that contribute to this thesis. This chapter gives a detailed review of the theoretical and conceptual background on perspective taking in the fields of psychology and robotics. Section 1 is dedicated to looking at perspective taking in developmental psychology, including the definitions, different dimensions, and developmental stages. Section 2 covers the seminal and state of the art research in the field of robotics and human-robot interaction that revolves around perspective taking and cognitive modelling for perspective taking.

2.1 Perspective Taking in Developmental Psychology

The term perspective taking appears in a wide variety of fields, from developmental and cognitive psychology to social sciences and linguistics, from simple daily human interactions to preventing a nuclear war between two countries in a harrowing conflict (Galinsky et al., 2008). In general terms, Perspective Taking (PT) is the capacity to consider and understand the world from other viewpoints, with the perception ranging from acknowledging others having a different perspective to computing and perceiving the perspective of others. A breakthrough in the field happened through Piaget and Inhelder in 1956, which led to understanding human's ability to change perspective (Piaget & Inhelder, 1967). They used the three mountain tasks to determine children's visual perspective taking abilities and egocentrism. More research by developmental and comparative psychologists includes research on children and adults perspective taking development in its different dimensions and levels (Flavell et al., 1981; Masangkay et al., 1974; Michelon & Zacks, 2006; Surtees et al., 2013). In 1975, Kurdek and Rodgon proposed three dimensions associated with perspective taking: Perceptual, Cognitive, and Affective dimensions (Kurdek & Rodgon, 1975). In their work, they acknowledged how the literature associated with perspective taking has often failed to consider the multidimensional nature of this skill. Their mention of these dimensions date back to 1975, when Shantz also did a thorough review of the literature associated with these abilities (Kurdek & Rodgon, 1975; Shantz, 1975). In recent years the literature on perspective taking has spread on all different dimensions and sub-dimensions from understanding the developmental stages of infancy to adulthood (Flavell et al., 1981; Michelon & Zacks, 2006; Moll & Tomasello, 2006) to dissociating the underlying processes of different levels of perspective taking (Todd et al., 2017; Vander Heyden et al., 2017). A brief overview of these components, dimensions, levels and processes involving perspective taking helps to understand the importance of this skill in human communication and daily survival, which in turn leads to a better understanding of why incorporating such skill in agents and virtual robots should be the next step in developing these technologies.

Before going into the detail of all the dimensions and levels, first we start by describing the main components of a perspective taking task inspired by Surtees et al. According to Surtees et al. the three basic components involved in a PT tasks are (Surtees et al., 2013):

- a perspective-taker (Self), the person who judges, understands, or takes the other's perspective;
- a target's (Other) perspective that is being judged, understood, or taken (it is commonly referred to as another person, but it can also be a directional object, imagined self-perspective, or virtual or embodied entities such as agents or robots);
- an object or circumstance (Object) upon which the perspective is taken.

As described by Surtees et al., the aforementioned components involved in perspective taking elicit two levels of relation and representation (Rakoczy, 2012; Surtees et al., 2013). The first relation involves the relationship between the Self and Other; *Self-Other* relation in short, which answers the question "How does the Self represent the Other's perspective?". The second relation incorporates the relationship between the Other and the Object; *Other-Object* relation in short and it tries to answer the question: "What relationship between the Other and the Object is represented?". This representation only concerns the Self taking the perspective of the Other and does not require the Other to perceive anything. In accordance with these components and relations, we can say the nature of the *Object* and its relations with *Self* and *Other* correspond to the different dimensions and levels of perspective taking. The Object can be of perceptual nature, in which it represents the Other's visual or spatial point of view. It can be cognitive, which associates with the ability of the Self to assess the Other's knowledge in cognitive space. And it can be of affective nature which corresponds to the Self's understanding of the Other's emotional state.

2.1.1 Perceptual Dimension

There has been a wide range of research on what can be categorized as Perceptual Perspective Taking. As mentioned earlier perceptual perspective taking refers to the ability to perceive or imagine what other's see. This understanding can happen on two different sub-dimensions of visual and spatial perspective taking (Muto et al., 2019; Surtees et al., 2013). Visual Perspective Taking (VPT) consists of the Self's awareness of the Other's visual field of view and Spatial Perspective Taking (SPT) deals with the Self's spatial understanding of the Other's perspective or spatial relation with the Object. According to some models of perspective taking by Flavell et al., each sub-dimension can be explained with two levels (Flavell et al., 1981; Surtees et al., 2013). Moreover, Moll and Meltzoff proposes a model that includes a level-0 in addition to level-1 and level-2 (Moll & Meltzoff, 2011). Much of the perspective taking literature on the visuosaptial dimension does not differentiate between VPT and SPT (Kessler & Thomson, 2010; Michelon & Zacks, 2006). However, there are several research dedicated to understanding the acquisition and development of VPT in children and adults (Flavell, 1977; Flavell et al., 1992; Flavell et al., 1981; Moll & Tomasello, 2006). And more recent research is focused on understanding SPT, its acquisition, as well as the mechanisms and processes involved (Pickering et al., 2012; Surtees et al., 2012; Tversky & Hard, 2009).

Visual Perspective Taking (VPT)

The literature on VPT has been investigating its developments in children and non-human primates implicating that VPT is not a unitary ability (Call & Tomasello, 2011; Flavell, 2000; Masangkay et al., 1974). The distinctions between the two levels of VPT as defined by Flavell et al. correlate with the *what* and the *how* relations as proposed by Surtees et al. In level-1 VPT the Self simply needs to understand *what* Object is visually seen or accessible by the Other in the world. Research had shown that children as young as three years of age have been able to correctly acknowledge that adults have a different vision compared to them when they were presented with a card composed of a picture of a dog on one side and a cat on the other (Masangkay et al., 1974). On the other hand, level-2 VPT deals with the question of *how* Other sees the Object or the surrounding world, considering that based on different viewpoints the representation might differ. In this regard, the emergence and acquisition of level-2 VPT have been associated with the development of Theory of Mind (ToM) skills such as False Belief reasoning which emerges around 4 years of age (Flavell et al., 1983; Perner, 1991). Various PT tests are designed to evaluate level-2 VPT with the most well-known test being the "three-mountain task" developed by Piaget and Inhelder in 1967, in which the child is supposed to specify the viewpoint of a doll from different positions looking through a three-dimensional model with three mountains (Piaget & Inhelder, 1967). The mountains have different sizes and identifiers and the child selects the doll's viewpoint from a set of photographs. While this test was designed to evaluate children's egocentricity, it has been deemed to be too complicated for children and resulted in other researchers designing different tests that are more coherent with the acquired skills in different ages. For example for level-2 VPT, a simpler test with a more binary response called the "turtle task" was developed by Masangkay et al. in 1974. In this task, the child is supposed to specify if an adult seated in front of them sees a horizontally placed picture of a turtle in the "right-side up", similar to what the child sees, or "upside down" which is what they are really seeing (Masangkay et al., 1974). For example, if you replace the turtle picture with the number "9", the adult is supposed to see the number "6". The studies show that slightly more than 50% of children around 4.5 years of age and older were able to recognize that the adult had a different perspective, however, some still said adults had the same view as them. Several other studies replicating these tests or similar ones have shown that children below the age of 4-5 years old do not engage in level-2 VPT (Flavell et al., 1981).

In 2011, Moll and Meltzoff proposed a model of VPT that mainly overlapped with Flavell and colleagues model, however, their model differed on the ground of acknowledging joint attentional abilities as a prerequisite for the emergence of perspective taking. The joint attentional abilities of infants at around one year of age can be considered "level-0 perspective taking". At this stage, infants have no knowledge about other's perspectives, but they can have shared joint attention or engagement with others as evidenced by activities associated to joint attention such as gaze following, pointing, alternating gaze, and holding up and showing. After a year, around 2 years of age, children reach "level-1 perspective taking", which as explained before, includes the ability to perceive *what* others can see or not see in comparison to the child's viewpoint with the minimal visuospatial requirement. Moll and Meltzoff differentiate between the knowledge of what others see and what others are familiar with or not from past experience and call the latter "level-1 experiential perspective taking". Level-1 experiential perspective taking seems to develop significantly earlier than level-1 perspective taking and the ordering seems to be a consequence of particular challenges in understanding visual perspective compared to a more holistic view of the world and engagement with others. At "level-2 perspective taking" children understand *how* others see things in addition to *what* they see. This includes the understanding of the specific ways Objects are seen, constructed, or (re)presented. However, Moll and Meltzoff, proposed a two distinct sub-level for the acquisition of this skill which differs in having the ability to "confront". Children at around 3 years old reach the ability to recognize how the Other sees the Objects even if what they see differs from what the child sees; called level-2A, however, at this point, they are still unable to confront perspectives which entails the understanding of how, depending on viewpoints, one Object or event can be seen in different ways. Based on Theory of Mind studies, the ability to confront emerges at around 4.5 years of age, where children explicitly gain the knowledge about perspective in various domains and this is called level-2B (Wellman et al., 2001).

Spatial Perspective Taking (SPT)

Spatial perspective taking is the ability to understand the Other's spatial relation with the Object or the world. This dimension has received far less interest from researchers in social psychology, which might be a result of SPT not always implying mental content: meaning that by knowing that an Object is to your left is not dependent on how you represent the object or the relationship. Furthermore, some studies show there is a small distinction between adopting the spatial perspective of another person and an object with a front (Surtees et al., 2012; Surtees et al., 2013). One important component of spatial perspective taking is "Frames of Reference" which allows us to encode spatial information relative to Self/Other/Object. In the study of spatial cognition, frames of reference are considered to be an essential component that needs to be addressed across all modalities and disciplines (Levinson, 2003). We will look at this component in the next section with respect to its role in understanding SPT.

Looking at SPT from a developmental angle, it is evident that this ability has not shown a uniform developmental pattern. There is some evidence of earlier developmental timeline for notions of front and back (Bialystok & Codd, 1987; L. J. Harris & Strommen, 1972). By 3-4 years of age children consistently use the words “in front” and “behind”, however, it is much later when they exhibit the same consistent use for “to the right of” and “to the left of” (L. J. Harris, 1972; L. J. Harris & Strommen, 1972). Looking at Flavell and colleagues and Moll and Meltzoff’s developmental stages of VPT, Surtees et al. had proposed a 2 level developmental model for SPT (Flavell et al., 1981; Moll & Meltzoff, 2011; Surtees et al., 2013). The model is based on developmental delays of the abilities associated with each level and is meant to facilitate describing similarities and differences between visual and spatial perspective taking. Surtees et al. considers in front and behind judgement as “level-1 type” and to the left and to the right judgement as “level-2 type”. In both visual and spatial sub-dimensions, the level-1 perspective can be considered the early-developing and level-2 as the later-developing type. Surtees et al. have demonstrated that different processes are involved in the early-developing and later-developing perspective taking independent of whether the judgments are visual or spatial (Surtees et al., 2013). Furthermore, they have found that level-2 type judgements, which corresponds with knowing how things looked to someone else in VPT and left/right judgements in SPT, both required egocentric mental rotation. On the other hand, level-1 type judgements; what others see for visual and front/behind judgements for spatial perspective taking, did not involve mental rotation.

Frames of Reference

When perceiving and understanding the spatial relations with another person or object, one needs to adopt a frame of reference. The same adoption is required when one is producing expressions that describe the spatial relationships. For communication to occur the perceptual cues and the verbal cues that describe the spatial relationships should be mapped into a mental representation (Carlson-Radvansky & Irwin, 1993; Friederici & Levelt, 1990). Retrospectively, the frame of reference with respect to spatial positions can build bridges between perception and language. This is important to consider, as the definitions and distinctions in the literature include the linguistics approach to the topic. Levinson emphasizes that essentially the distinctions between frames of reference correlate to the distinction between their underlying coordinate systems and not the objects that invoke them (Levinson, 2003). To represent spatial representations among objects the literature consists of several classifications in different modalities. Levinson provides a brief overview and sketch of various distinctions of “frames of reference” across many disciplines such as philosophy, brain sciences, psychology, and linguistics (Levinson, 2003). Table 2.1 is adapted from Levinson and represents spatial frames of reference in different literature. We briefly discuss the distinctions here and use the ones that align with our line of research.

Table 2.1 – Spatial frames of reference with its distinctions in the literature adapted from (Levinson, 2003) with slight modifications.

'relative' vs. 'absolute'	*(philosophy, brain sciences, linguistics)*	- relative space is described by relations between objects, and absolute space is an abstract infinite void. - relative space is associated with egocentric coordinate systems, and absolute space with non-egocentric ones.
'egocentric' vs. 'allocentric'	*(developmental and behavioural psychology, brain sciences)*	- a coordinate system originating from the subjective body system (body-centered) of the Self versus a coordinate system originated anywhere else (environment-centered). - it can also translate to subjective (subject-centered) versus objective frames of reference.
'viewer-centred' vs. 'object-centred' or ' '2.5D sketch' vs. '3D models	*(vision theory, imagery debate in psychology)*	- distinction between the coordinate systems originating from the subjective body system of the Self versus a coordinate system originated anywhere else.
'orientation-bound vs. orientation-free'	*(visual perception, imagery debate in psychology)*	- absolute and relative frames of reference in language are both orientation-bound. - intrinsic frame is orientation-free.
'deictic' vs. 'intrinsic'	*(linguistics)*	- sometimes a third term **extrinsic** is opposed to include the contribution of gravity to the words like *above* or *on* - can be translated to speaker-centric versus non-speaker-centric. - centered on any of the speech participants versus not so centred. - ternary versus binary spatial relations.
'viewer-centred' vs. 'object-centred' vs. 'environment-centred'	*(psycholinguistics)*	- in a viewer-centered frame, objects are represented in a retinocentric, head-centric or body-centric coordinate system based on the perceiver's perspective of the world. - in an object-centred frame, objects are coded with respect to their intrinsic axes. - in an environment-centred frame, objects are represented with respect to salient features of the environment, such as gravity or prominent visual landmarks.

For representing the spatial relationships between humans and/or objects in the world, three distinct frames of references are commonly used: *viewer-centered frames*, *object-centered frames*, and *environment-centered frames* as shown in the last section of table 2.1 (Bloom, 1999; Levinson, 1996, 2003). This formulation of frames of reference is based on spatial perception and cognition and semantically it can be labelled with corresponding linguistic interpretations. As a result, *deictic*, *intrinsic*, and *extrinsic* are the other labels for viewer-centered, object-centered, and environment-centered, respectively (Bloom, 1999; Levinson, 1996, 2003).

In reference to the basic components of perspective taking tasks proposed by Surtees et al.; a perspective-taker (Self), a target's perspective (Other), and an object or circumstance (Object), we decided to use the frames of reference classification that fits the model. As used by Levinson, Surtees et al. and several other researchers, the frames of reference related to spatial perspective are:

- relative to ourselves: relative frames of reference;
- relative to another person/thing: the intrinsic frame of reference;
- relative to some non-varying degree in the environment: the absolute frame of reference.

In respect to the uses of frames of reference, there are several studies and evidence that show differences in the use of frames of reference in spatial perspective taking tasks involving another human in comparison to inanimate objects. Additionally, when taking the perspective of a person and an object, adults and children differ in the frames of references that they adopt (Surtees et al., 2012). There is also some evidence showing when adults make linguistic decisions they activate multiple frames of reference (Carlson-Radvansky & Irwin, 1993; Carlson-Radvansky & Jiang, 1998).

2.1.2 Cognitive Dimension

Cognitive perspective taking is considered to be the ability to imagine what other's experience (Taylor, 1988). This means the object or circumstance in the perspective taking task is "experience". This includes taking into account the other's perception and knowledge about a stimulus and potentially incorporating it with the self's experiences of similar stimulus (Hinnant & O'Brien, 2007). For many years the research on cognitive perspective taking has been intertwined with the research on Theory of Mind (ToM) (S. Baron-Cohen, 1997; S. E. Baron-Cohen et al., 2000; S. Baron-Cohen & Hammer, 1997). The research on this topic contributes to the design of intervention programs for populations with perspective taking deficits such as those diagnosed with autism (Barnes-Holmes et al., 2004). The link between cognitive perspective taking and prosocial behaviour has been studied by several researchers such as Eisenberg and colleagues and Farrant et al. (Eisenberg & Miller, 1987; Farrant et al., 2012). Hinnant and O'Brien have researched the link between cognitive and affective perspective taking in 5-year-old children. They found that the relation between affective perspective taking and empathy is moderated by cognitive perspective taking and gender has a role in the way this link works'(Hinnant & O'Brien, 2007). On the other hand, Eisenberg et al. studies with adults have found a positive correlation between the self-reported measure of cognitive perspective taking and empathy (Eisenberg et al., 1989).

2.1.3 Affective Dimension

Affective perspective taking is defined as the ability to understand what other's feel and emotionally experience (P. L. Harris et al., 1989). The definition of affective and cognitive perspective taking significantly differ since one involves understanding emotions whereas the other one involves understanding the other's knowledge and cognitive perspective. It is important to note that the terms empathy, affective perspective taking, and affective identification have been used interchangeably in the literature, (Kurdek & Rodgon, 1975; Shantz, 1975). *Empathy* comprises multiple psychological components and it refers to "the ability and tendency to share and understand other's internal states" (Zaki, 2017; Zaki & Ochsner, 2012). On the other hand, *perspective taking* is "the act of attributing mental states to others and reasoning about how situations relate to them" (Gopnik & Wellman, 1992). *Experience sharing* and *empathic concern* are also considered as other pieces of the puzzle contributing to prosocial behaviour. The research on the correlation between affective perspective taking and empathy and the rest of the aforementioned abilities are extensive and beyond the scope of this research. However, it is important to mention, some researchers have studied the possibility of using perspective taking exercises as a means to increase the measures of empathy or prosocial behaviour (Van Loon et al., 2018; Weisz & Zaki, 2017).

2.1.4 Theory of Mind and Perspective Taking

Considered as one of the foundational elements of social interaction, Theory of Mind (ToM) is defined as the ability to attribute mental states; beliefs, desire, to self and other (Barnes-Holmes et al., 2004; Frith & Frith, 2005; Premack & Woodruff, 1978; Wellman, 1992). Having Theory of Mind means that the individual is aware that other's mental states are different from their own. From a Theory of Mind point of view, for children to master taking other's perspectives, they need to master five levels of understanding informational states (Hadwin et al., 2015). The first three levels correspond to different levels of perceptual perspective taking and cognitive perspective taking. The last two levels involve true beliefs and predicting actions based on a person's knowledge and understanding false belief and predicting on the basis of false belief (Barnes-Holmes et al., 2004).

2.2 Perspective Taking in Robotics

The importance of perspective taking in composing effective interaction and collaboration can extend to scenarios with robots and agents. The prospect of humans and robots interacting with each other on different topics and levels calls for the need to study and investigate how to develop and incorporate this cognitive skill into robots and agents. An emerging body of study in human-robot interaction has discussed perspective taking in robotics, demonstrating that perspective-taking plays an important role in collaborative and learning scenarios with robots. Studying perspective taking development in robots requires understanding how humans perceive robots and their agency. It has been shown that the assumptions humans make about robots are similar to the assumptions they make about their human counterparts (Lee et al., 2005). For example, only showing certain nonverbal behaviours from the robot is enough for humans to attribute mental models to robots (Zhao et al., 2016). As a result, people tend to take the robot's perspective almost as much as they take other people's perspectives. Trying to answer Alan Turing's pivotal question "Can machines think?", Krach et al. investigate perspective taking with robots using fMRI, demonstrating that "the tendency to build a model of other's mind linearly increases with its perceived human-likeness" (Krach et al., 2008). Perspective taking in robotics was further studied for language understanding (Lemaignan et al., 2011; Roy et al., 2004; Steels & Loetzsch, 2009), imitation learning (Breazeal et al., 2006; M. Johnson & Demiris, 2007), mental state estimation (Devin & Alami, 2016; M. Johnson & Demiris, 2005), task understanding (Pandey et al., 2013; Sisbot et al., 2007), and facilitating human-robot communication (Berlin et al., 2006; Torrey et al., 2009).

In recent years, an emerging body of research has been dedicated to developing robots and agents with perspective-taking abilities and understanding how these abilities help improve human and robot interaction. One of the pioneering works on perspective taking by Trafton et al. shows how equipping the robot with visual perspective taking abilities can help to resolve ambiguous situations (Trafton et al., 2005). The ambiguous situation involved the robot having visual access to two objects where one was occluded from the human counterpart, a classic level-1 visual perspective taking situation. The study initially analyzed human-human interaction scenarios involving perceptual perspective-taking (Trafton et al., 2005). Then it proceeded to provide three important conceptual guidelines in building robotic systems in human-robot interaction. The authors evaluated their system in a collaborative interaction and with various frames of reference. Kennedy et al. studied level-2 perspective taking abilities in robot using "like-me" simulation (Kennedy et al., 2009). This includes the robot applying its own reasoning capabilities to the imagined situations. Another study by Ros et al. incorporates the object ontology into the resolution of ambiguity (Ros et al., 2010). All these studies show that perspective-taking plays an important role in collaborative and learning scenarios. In the next section, we also look at joint attention in robotics as a prerequisite of perspective taking based on the developmental psychology model proposed by Moll and Meltzoff.

2.2.1 Joint Attention in Robotics

Tomasello believes that to understand joint attention, both cognitive and social aspects of the process should be considered and studied (Tomasello et al., 1995). Joint attention should be perceived beyond simultaneously looking and orienting to an object or location. It should be expanded to a mutual awareness of the two parties' attentional state to the same subject and monitoring of the other's attention. Behavioural studies dedicated to joint attention in developmental stages focus on two kinds of behaviours displayed by infants: Respond to Joint Attention (RJA) and Initiate Joint Attention (IJA) (Mundy et al., 2007). Respond to Joint Attention (RJA) "refers to infants' ability to follow the direction of the gaze and gestures of others to share a common point of reference" (Mundy & Newell, 2007). On the other hand, Initiate Joint Attention (IJA) "involves infants' use of gestures and eye contact to direct others' attention to objects, to events, and to themselves" (Mundy & Newell, 2007). There have been several models and attempts to map the developmental models of joint attention of infants to robots. Most studies on the realization of joint attention have focused on responding to joint attention (RJA) abilities of the robots using two different approaches. In one approach the RJA ability was built as a model and programmed in the agent (Kozima & Yano, 2001; Thomaz et al., 2005). While in the other approach, a constructive model of joint attention has been developed, where the agent learns to respond to joint attention through multiple interactions with a model learning from the interactions (Carlson-Radvansky & Radvansky, 1996; Nagai et al., 2003). Other studies consider both responding and initiating joint attention (RJA and IJA) using eye gaze and pointing gestures (Baraka et al., 2019; Huang & Thomaz, 2010; Imai et al., 2003; Scassellati, 1999). Few studies also address a third skill for the robots with joint attention, ensuring joint attention (EJA) (Breazeal et al., 2005; Huang & Thomaz, 2010; Rich et al., 2010). Ensuring joint attention is “the ability to monitor another’s joint attention an ensure that the state of joint attention is reached” (Huang & Thomaz, 2010). There are few prerequisites to achieve joint attention abilities within the robots beyond the developmental aspect that can be learned or mapped from infants (Kaplan & Hafner, 2006). The agent should be able to develop *Attention Detection*, *Attention Manipulation*, *Social Coordination*, and *Intentional Stance* abilities. Kaplan and Hafner argue that a global developmental approach of joint attention is required in comprehending the links between each developmental stage and achieving autonomous capabilities and adaptive features (Kaplan & Hafner, 2006).

2.2.2 Computational and Cognitive Modelling of Perspective Taking

Several works have attempted to model perspective taking and other cognitive abilities in recent years. Hiatt et al. has proposed a perspective taking model using ACT-R/S (Hiatt et al., 2004). ACT-R/S is an extension of ACT-R model developed by Anderson et al., that implements a theory of spatial reasoning in ACT-R (Anderson et al., 1997; Harrison, Schunn, et al., 2003). The model proposed by Hiatt et al. is developed for a simple ’fetch’ task where the robot is listening to a speaker. The model presents the accounts for how the robot resolves the ambiguity in the speakers’ utterance to perform the fetching correctly. This work has later been extended and presented in more detail by Trafton et al. (Trafton et al., 2005). Ogata et al. proposed a model for developing processes of imitation sequences (Ogata et al., 2009). First, their model learns to map between self-motions and object movements. Then it is assumed that a teacher imitates the robot from four different viewpoints. The conversion modules from the robot’s egocentric perspective to the teachers’ imitation is then used for predicting the motion of others and the teacher. Nakajo and colleagues extends this work to be able to imitate known action from the viewpoints that have not been previously observed (Nakajo et al., 2015). To achieve this they represent the viewpoints and motion sequences separately. More recent work by Schrodt and colleagues introduces a similar model that uses an artificial neural network to learn how to correlate visual and proprioceptive data from motion sequences (Schrodt et al., 2015). Later they evaluate the model using recordings of full-body motion tracking. However, the input to their model is only available on a self-perspective basis and it is not sufficient for perspective taking. A similar approach is taken by Gentili and colleagues with a focus on imitation of what another person sees from an arbitrary viewpoint (Gentili et al., 2015). Recently Fischer and Demiris has proposed a computational model of perspective taking for iCub using forward and reverse engineering approaches to model and study perspective taking for their robotic system (Fischer, 2018; Fischer & Demiris, 2019).

2.2.3 Other Applications of Perspective Taking in Robotics

A more recent research direction tries to look at child-robot interaction scenarios either with Autistic or typically developing children. Cognitive processes for both VPT and ToM are reported to be impaired up to some degree in individuls with autism spectrum disorders (ASD) (Hamilton et al., 2009; Yirmiya et al., 1994). Studies show that robots can be used to teach and develop autistic children's visual perspective-taking skills (Robins et al., 2017; L. J. Wood et al., 2018, 2019; L. Wood et al., 2017). L. J. Wood et al. uses Kaspar robot's camera and a screen to show ASD children the robot's perspective e.g. what it sees through its eyes (L. J. Wood et al., 2019). This unique approach has the distinct advantage of using robots to practice and develop perspective taking abilities. Their study aimed at level-1 VPT and planned to increase the difficulty into level-2 over the spans of several sessions.

In another line of research, some studies have investigated the development of empathy in assistive robots (Tapus & Mataric, 2007). The study by Tapus and Mataric has proposed an model of empathy inspired by the work of Davis. In Davis's term, empathy is defined as " the capacity to take the role of the other, to adopt alternative perspectives vis a vis oneself and to understand the other's emotional reactions in consort with the context to the point of executing bodily movements resembling the other's" (Davis, 1983). Empathy is regarded as having both affective and cognitive processes, in which perspective taking is strongly associated to its cognitive component (Davis, 1983, 2018). More recent advances on designing robots for emphatic interaction with children includes the robots that recognize people and acknowledge their presence and simulate a theory of mind (Westlund et al., 2018; Williams et al., 2019). In developmental robotics, researchers are working on developing empathy in robots inspired by its developments in children as they train the robots to recognize emotions and differentiate between themselves and others (Asada, 2014; Lim & Okuno, 2015). There are also studies that focus on developing empathy and prosocial behaviour in children using robots. These studies express uncertainty about the ethical aspects of using robots to promote prosocial behaviour in children (Borenstein & Arkin, 2017).

2.3 Summary

This chapter presented some of the empirical and theoretical works that investigate perspective taking in developmental psychology and robotics. It was demonstrated that perspective taking is an ability that its different dimensions such as cognitive and social can be developed into robots and studied in human-robot interaction. The different dimensions of perspective taking have contributed to the various lines of research in human-robot interaction that focuses on studying and developing such skills in children and robots. As a result, the following chapter starts with introducing the field of human-robot interaction and some topics in education that have inspired the development of some of the user studies in the later chapters.

3 Human-Robot Interaction

This chapter is dedicated to presenting the research in human-robot interaction and in particular child-robot interaction that has inspired and contributed to this thesis. The chapter is organized as follows. Section 1 is dedicated to discussing child-robot interaction. The second section covers research in robotics dedicated to education and it details the applications of social robots on the topics studied in the thesis, such as reading and storytelling and games.

3.1 Child-Robot Interaction

The field of Human-Robot Interaction (HRI) is defined as "the study of humans, robots, and the ways they influence each other" (Fong et al., 2003). To better understand the field of human-robot interaction, one needs to gain a better understanding of both sides of the interaction; humans and robots. This makes the field of HRI inherently an interdisciplinary field that brings together fields such as psychology, cognitive science, linguistics, mathematics, computer sciences, engineering, human factors, and design. As described by Goodrich and Schultz the problem of HRI is "to understand and shape the interactions between one or more humans and one or more robots" (Goodrich & Schultz, 2008). Essentially, where there is a robot; even a fully autonomous robot, the interaction between the robot and human is present. This makes the designing technologies that improve the interaction an essential component of HRI. The fundamental goal of HRI is "to develop the principles and algorithms for robot systems that make them capable of direct, safe and effective interaction with humans" (Feil-Seifer & Matarić, 2009). It can be noted that some instances of the HRI field have been influenced by popular culture and particularly the work of prominent science-fiction writers. The three laws of robotics proposed by Asimov are considered as one of the benchmarks of HRI and ethics in robotics (Asimov, 2004). The three laws have provided the building block of the field of HRI and the theoretical implications revolving around designing robots and agents systems today (Wallach et al., 2008).

The recent advances in the field of Human-Robot Interaction and particularly social robotics have contributed to various domains such as education, elderly care, entertainment, cognitive modelling, and so on. In each of these domains, different types of robots with a wide range of functionalities are developed to interact with humans of all ages. One of the age groups that received particular attention from the field is children. As it is evident from developmental psychology, children are still in the developmental stages of their cognitive and social abilities (Belpaeme et al., 2013; Ros et al., 2011). Understanding the interaction modalities between the child and the robot has led to the creation of the field of Child-Robot Interaction (CRI) as a specific case of HRI. According to Ros et al., children's reaction to interaction with robots is stronger and more brisk compared to adults (Ros et al., 2011). Furthermore, children are less likely to use robots as tools which makes their interaction more social (Salter et al., 2008). This results in the need for researchers in the field of CRI to pay special attention to the social interaction when designing activities for robots with children (Salter et al., 2008). Additionally, according to the developmental paradigm of the socio-cultural theory by Vygotsky and colleagues children learn better in social contexts (Vygotsky et al., 1978). The work by Charisi et al. looks at a symbiotic co-development between the child and robot, where the robot interacts with the child and adapts its behaviour according to the child (Charisi et al., 2015). Their work is inspired by the 'constructivism' theory by Piaget et al. and the 'constructionism' theory by Papert, which have both enriched the field of CRI in understanding how children learn and grow (Ackermann, 2001; Papert, 2020; Piaget et al., 1960).

3.2 Robots in Education

In the conjunction of human-robot interaction and education, the role of the robots can vary from being a tool for learning certain subjects to agents that accompany the learner in learning environments. Looking closely, it is evident that robots can take on different roles to teach, assist, inspire and motivate the learner. Therefore, a robot can take on the role of a teacher or tutor, a learning companion, or a learner (Hood et al., 2015; Leite et al., 2013; Mubin et al., 2013). A study by Kanda et al. (Kanda et al., 2004) using robots as peer English tutors shows the different ways children interact with robots and can benefit from them. Long-term interaction studies by Tanaka et al. using QRIO (Tanaka et al., 2007) and Hyun et al. using iRobiQ (Hyun et al., 2010), show the level of social interactions between the robot and children. Through these experiments, they provide guidelines for using robots as tutors in the classrooms. In another study by Leyzberg et al. (Leyzberg et al., 2014) the effect of personalization of a tutor robot on the students' performance was evaluated. L2TOR by Belpaeme et al. (Belpaeme et al., 2015) is a recent project focusing on evaluating social robots for tutoring the second language to children in early childhood. L2TOR project tries to address current demands and define the pedagogy of robot-assisted tutoring. Learning a certain subject during the interaction with robots has also been studied (Brown & Howard, 2013; Hood et al., 2015). In Brown and Howard's study (Brown & Howard, 2013) they used a robot; Darwin, to assist children in solving math questions. Additionally, Hood et al.'s study (Hood et al., 2015; Lemaignan et al., 2016) introduced the CoWriter project focusing on writing activities using the robot Nao. CoWriter is a project that motivates children to help a robot with bad handwriting, and in the process, it engages children into practising their own handwriting in a more subtle yet productive way as they feel responsible to help the robot. This phenomenon is called the “Protégé effect” which emerges through learning by teaching paradigm.

The roles and cognitive abilities of robots in educational settings have been investigated and evaluated in different studies. Robots has been actively evaluated as learning companions (Lu et al., 2018), tutors (Belpaeme et al., 2015; Castellano et al., 2013; Gordon & Breazeal, 2015) and learners (Hood et al., 2015; Muldner et al., 2013; Yadollahi et al., 2018) in educational settings. Assigning the robots to any of these roles is subject to the learning objectives and the robot’s intelligence. As a result, different studies use different approaches in designing the robot, interaction, and educational content. For instance, some studies focus on formal *K12 education* content for child learners (Ahmed et al., 2018; Gordon & Breazeal, 2015; Hood et al., 2015; Schodde et al., 2017), some anchor around robot’s cognitive and artificial intelligence, and few base their development on *lifelong learning* scenarios (Parde & Nielsen, 2019). These studies all bring an understanding of how robots can be beneficial in educational settings and what developments are needed to reach that level of efficiency. A more recent research direction tries to look at child-robot interaction scenarios either with autistic or typically developing children. Studies show that robots can be used to teach and develop autistic children’s visual perspective-taking skills (L. J. Wood et al., 2019) or building their emotional intelligence skills. Furthermore, some studies try to evaluate how certain affective behaviours of the robot can influence children’s cognitive and emotional perspective taking skills.

3.2.1 Learning by Teaching

Previous research in education and cognitive sciences suggests that a powerful learning method is teaching others (Bargh & Schul, 1980). The potential of learning by teaching can also be deduced from methods such as peer-assisted tutoring, reciprocal teaching, small group interaction, and self-explanation each of which possesses a certain degree of teaching to others. There is also research demonstrating the positive effect of learning by teaching in computer-assisted learning environments (Biswas et al., 2005; Chase et al., 2009). The effect of using teachable agents during learning activities, on the students motivation and learning gain has been studied by Chase et al. (Chase et al., 2009). They observed students who taught to the teachable agents spent more time on learning behaviours and learned more than students who learned by themselves. Besides, they observed that the Protégé effect was particularly more helpful to low-achieving students. As a result, adoption of the Protégé effect in learning might bring an intrinsic level of motivation combined with a sense of responsibility for the robot, which can change the experience of learning.

3.2.2 Reading and Storytelling

Concerning reading activities, there have been multiple studies focusing on reading or storytelling activities with children aiming to make it more engaging (Rubegni & Landoni, 2016). Storytelling activities consist of telling a story by focusing on education or entertainment or creating a story based on available or given elements. These studies can be divided into three categories: 1) when the robot is reading or telling a story, 2) when the child is reading or telling the story, or 3) when someone else is reading to both of them. One method suggested to help children with reading difficulties is *Reading While Listening* (RWL). The effect of RWL is shown to be promising for poor readers in improving comprehension, word recognition, reading fluency, acquisition of word meaning, and learning to read (Carbo, 1978; Chomsky, 1976). One of the latest studies devising synchronized RWL is by Gerbier et al. (Gerbier et al., 2018). They evaluate the effect of visually highlighting the text as it is being read compared to simple text conditions. Their study shows that poor readers tend to benefit more from these visual cues. Another recent example of technology used to assist readers is PhonoBlocks, a Tangible User Interface (TUI) for children with dyslexia (Fan et al., 2016). In studies using robots, the work by Michaelis and Mutlu (Michaelis & Mutlu, 2017) proposes an in-home learning companion, in which the child is reading to a robot. They focus on how such a companion can help children expand their reading interests and abilities. *Family Story Play* is another reading platform that uses a digital companion for long-distance reading interaction between children and grandparents (Raffle et al., 2010). To facilitate the interaction, the platform uses a digital social agent in the form of Elmo to help enrich the interaction between the child and grandparents and strengthening dialogic reading. Storytelling delves into another aspect of interaction with a robot. For a robot to be a good storyteller being aware of its audience is critical. The study by Mutlu et al. shows how the robot's gaze behaviour during storytelling affects the audience's perception of the robot (Mutlu et al., 2006). Another system providing interactive storytelling is *AIBOStory* which aims to enrich remote communication experiences on the Internet using AIBO robots (Papadopoulos et al., 2013). In the study by Kory and Breazeal (Kory & Breazeal, 2014), a social robot interacts with children as a peer and plays a storytelling game with them. The interaction is designed to introduce children to new vocabularies in the context of storytelling games. It evaluates the effect of adapting the robot's language level to the child's learning, the complexities of the stories, and similarities to the robot's stories.

Studies by Pellegrini et al. and Bus et al. show that parent-child shared book reading can be beneficial for the children's emergent literacy (Bus et al., 1995; Pellegrini et al., 1985). However, various factors affect the interaction and its quality, such as the child's level of competence or the parent's level of adjustment to the child's needs and problems. These factors make it hard to distinguish between the effect of parent's behaviour on the child or child's behaviour on the parent (Pellegrini et al., 1985). MacNeil and de Ruiter have a typology distinguishing different types of gestures used during speech or interaction (de Ruirer, 2000; McNeill, 1992). Among the different gestures that accompany human speech, we are interested in deictic gestures that translate to pointing gestures during book reading. Justice et al. examine the effect of verbal and nonverbal references made by an adult during storybook reading on children's visual attention to the print(Justice et al., 2008). The nonverbal references in their case translate to pointing to the print by tracking the text while reading. Their study was conducted with preschool children, concluding that explicit referencing such as pointing increases the children's visual attention to the print.

3.3 Summary

This chapter presented the related work on the topics of human-robot interaction and education. It also highlighted the importance of child-robot interaction and designing robots and modalities for children particularly in the field of education. Most of the works presented here have inspired the design and the developments of the child-robot interaction studies presented in Chapters 4, 5, and 6. Further background is contained within the sections of the chapters where it seems necessary to provide the specific conceptualization and research that contributed to a specific design or decision.

Part III

Exploring Perspective Taking in Interaction

4 CoReader: An Exploratory Study

As a first step, we start with an exploratory study that provides the grounds for developing interactions between children and robots in the educational domain. Given our interest in understanding and modelling perspective taking in children and robots, we first opted for developing an activity that studies robot's deictic gestures e.g. pointing to the text, on children's joint attention during a reading activity. This chapter presents the design of a platform that aims to engage children in a Reading While Listening (RWL) activity with the robot reading the text and the child helping the robot when it makes mistakes. The idea is to use the learning by teaching paradigm to keep the child motivated and interested in the interaction. In the following sections, first, we describe our motivation, expected contributions, and general research questions. This is followed by the design of the task, platform, and interaction. Then, we describe the pilot and the main study carried out using the platform and present the results for each hypothesis. Finally, we detail the findings and limitations of the study and describe how the findings of this study contributes to the rest of this thesis.

4.1 Scope and Research Goals

4.1.1 Motivation and Contribution

Joint attention and perspective taking are two cognitive abilities with overlapping developmental pathways. Their developmental inquiries were mainly researched separately, resulting in most models failing to recognize joint attention as a foundational skill for perspective taking. Moll and Meltzoff argue that joint attention is a necessary step before perceiving other's perspective (Moll & Meltzoff, 2011). In their model of developmental stages of visual perspective taking, they introduce "sharing attention" as level-0 perspective taking. In our pursuit to study perspective taking, we developed this study as a starting point when we were formulating our research questions and before we discovered the required steps to develop a perspective taking model. The fundamental motivation behind designing and developing this study is twofold. First, to understand the basics of designing and developing interaction scenarios with robots in an educational domain. Second, to examine the cognitive properties required for such interaction and how the robot's behaviour can impact children.

While designing the robot behaviours that could help children in the reading activity, we thought about one of the important aspects of reading together; establishing joint attention. Tomasello et al. believe joint attention should be perceived beyond simultaneously looking and orienting to an object or location (Tomasello et al., 1995). It should be expanded to the two parties a mutual awareness of each other's attentional state to the same subject and monitoring the other's attention. This requires the robot to be able to Initiate Joint Attention (IJA) with the child, Respond to Joint Attention (RJA) cues from the child, and Ensure Joint Attention (EJA) during the interaction (Huang & Thomaz, 2010; Kaplan & Hafner, 2006; Mundy et al., 2007). Methods aiming at initiating joint attention embrace deictic gestures, such as pointing. Breazeal et al. suggest that implicit non-verbal communication positively impacts human-robot task performance (Breazeal et al., 2005). This inspired us to examine the effect of pointing gestures to the text during the Reading While Listening (RWL) activity. Consequently, we decided to explore the effect of the robot's gestures (e.g. pointing) on children's performance in the context of the reading activity with the robot. The main goal was to observe how much deictic gestures such as pointing to the text could help the child's Reading While Listening (RWL) proficiency. We expect that the effect of pointing on the child's reading, especially the ones with reading difficulties and poor readers can provide valuable information in designing reading companions for children. Understating such effects could be a building block of designing reading platforms and informative for knowing when and how the robot should use deictic gestures during reading or similar activities. Furthermore, exploring perspective taking in human-robot interaction can contribute to our future studies.

4.1.2 General Research Questions and Hypotheses

While the majority of previous works consider the robot as a learning companion, Kory and Breazeal found that a robot with lower abilities than a child can trigger teaching and mentoring behaviours from children (Kory & Breazeal, 2014). This behaviour is aligned with the studies of the CoWriter project and the teachable agents by Chase et al. (Chase et al., 2009; El Hamamsy et al., 2019; Hood et al., 2015). Our research aims to motivate and inspire children to practice more, challenge them, and improve their self-confidence particularly, for children with reading difficulties. In our study, the robot is designed to play the role of a learning companion, which encompasses the learning by teaching paradigm within the context of interaction. When the robot is reading to children, the idea of correcting the robot motivates children to read alongside the robot, and this creates a Reading While Listening (RWL) experience. While none of the previous works concerning reading and storytelling in robotics studied the effect of pointing with a robot, a study by Sharma et al. analyzed the effect of deictic gestures on video lectures from Massive Open Online Course (MOOC) (Sharma et al., 2016). Their study investigated if augmenting a video with deictic gestures or with the teacher's gaze could increase the students' learning gain. They obtained significant results for the gaze cues only, but the deictic modality also showed higher student performances. In this chapter, we use the learning by teaching paradigm to explore the effect of the robot's deictic gestures on the child's performance in a reading task. To evaluate the performance, we use the proportion of correctly identified robot's mistakes as a measure of child performance.

Our first research question aims at an overall understanding of how the robot's pointing gestures influence the co-learning interaction. Overall, we expect the robot's pointing gesture to positively affect children's performance in recognizing and correcting the mistakes. We also hypothesize that the robot's pointing gesture affects each type of mistake differently. The two hypotheses made to investigate this question are going to be discussed in more detail after presenting the activity and platform.

- **Research Question 1:** How does joint attention influences the co-learning interaction?
  Hypothesis 1: Children in the "with pointing" condition show a better performance in correcting mistakes than in the "without pointing" condition.
  Hypothesis 2: Children show a better performance in correcting Type-1 mistakes (illustrations related) compared to Type-2 and Type-3 (contextual and pronunciation related).

Apart from understanding the robot's gestures on children's joint attention, we also wanted to investigate whether children's reading level is a factor in how they perceive or are affected by the robot's gestures. The second research question focuses on understanding this factor. Concerning this question and aligned with the activity, we hypothesized that the robot's pointing might be more beneficial for children with lower reading levels than the higher ones.

- **Research Question 2:** Does pointing gestures affect children of different reading levels differently?
  Hypothesis 3: The robot's pointing gesture is more effective in helping children with lower reading levels than higher reading levels.

4.2 Development of the CoReader Platform

4.2.1 Robotic Platform

The CoReader platform, shown in Figure 4.1 supports the collaborative reading of stories between a child and a robot. The platform consists of a Nao robot, a paper book, and three tangible feedback buttons or three cards equipped with ARTags used in the earlier developments of the platform. During the interaction, the child and the robot are sitting beside each other facing a book at an angle between 30 to 50 degrees, which is considered to be a side-by-side F-formation (Kendon, 1990). The interaction is designed to be simple and understandable for children. The robot interacts with the child and reads the chosen book, making occasional

Figure 4.1 – CoReader platform; Left image shows the robot with the book and buttons; Right image shows the view from the robot's bottom camera looking at the book readable to robot through ARTags.

mistakes. To correct the mistakes, the child should use the cards or the buttons to inform the robot. In the end of the interaction, the child is asked to continue the rest of the book by reading it to the robot. The whole system is implemented using the Robot Operating System (ROS) (see Appendix A) and is fully autonomous. The use of *Augmented Reality Tags* ARTags package [I] allows for a flexible selection of books, since only few and simple modifications are required for each book e.g. sticking the visual tags, besides associating the book text to the corresponding visual tag. It is also low cost as neither complicated programming nor costly modification of the books is required.

To feedback the robot about its mistake, initially we designed feedback cards equipped with ARTags in three colors of red, green, and yellow. The child was supposed to wave the cards in front of the robot's camera and as soon as the robot recognized the tag it performs the action associated with the card. The card feedback system was used in the pilot experiment. To reduce the lags and delays with the robot recognizing the tags and facilitate the interaction, we upgraded the feedback system to using push buttons. The feedback buttons are placed in front of the child as shown in Figure 4.1 and inform the robot in real-time of the child's feedback. The buttons or the cards allow the following actions; stop reading due to making mistake (red), repeat a page (yellow) and give praise for successful completion of a page (green). In the rest of the chapter, when we refer to the feedback system, we are referring to the buttons, except in the pilot section where we used the cards.

The red button can be pressed at any moment when the robot is reading to signal a mistake. After the receiving the signal, the robot stops reading and it starts showing a sad reaction which includes using body gestures and change of LED colors of the eyes. The child tell the

[I] AR Track Alvar from ROS

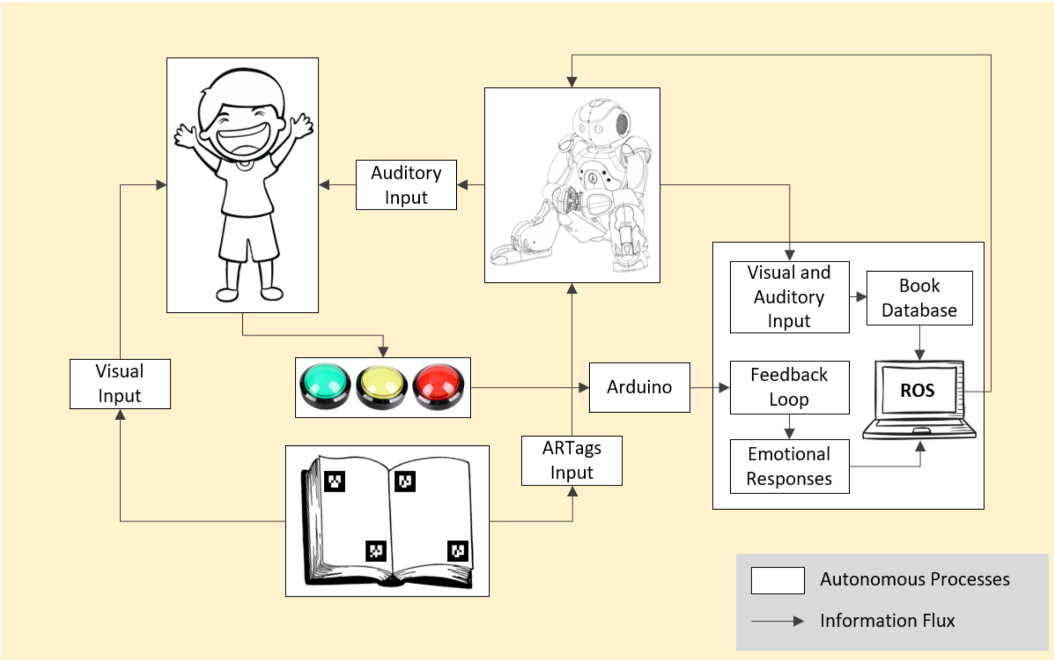


Figure 4.2 – Schematic of the study 1.

robot how to read the word correctly, then the robot re-reads the page and corrects its mistake. The yellow button, for repeating, can also be pressed at any time. After receiving a repeat signal, the robot stops reading and restarts from the beginning of the page. The green button can be used at the end of each page to inform and praise the robot of its success when it doesn't make any mistake. If the child accidentally presses the green button mid sentence, the robot continues with reading the book and ignores it until it reaches the end of the page. Pressing this button is followed by the robot's reaction with expressing happiness through body gestures and LED color patters.

Furthermore, the robot can express itself through a mixture of verbal and expressive actions that combine the robot's movement and dialogue acts. These allow the robot to react to making mistakes with a combination of sad and surprised behaviours and react to receiving praised by expressing happiness. Considering that the NAO robot cannot render facial expressions, to convey emotional responses to each reaction, we programmed the robot's eyes to imitate human emotions. For this purpose, we apply the LED patterns created and evaluated by D. O. Johnson et al. allowing to express six basic emotions (D. O. Johnson et al., 2013) . In this study, we only use *surprise* and *sadness* color patterns for reactions after making a mistake and *happiness* for reactions after being praised.

An overview of the interaction presented in this section is presented in Figure 4.2. The interaction as a whole is designed to be simple and understandable for children. At the beginning of each session, the robot introduces itself and expresses its desire to read the book, but it also explains that sometimes it needs help when it makes mistakes. Before starting the interaction, each child is informed of how the interaction flows, how to use the buttons or the cards, and their role in helping and correcting the robot in case of making mistakes. If the robot makes a mistake and the child does not recognize it, the robot keeps that state in memory and makes that mistake again if it is asked to repeat the page afterwards. For assessment purposes, the child is not informed about the mistakes that he/she didn't recognize, and the interaction will be carried out regardless of recognition of mistakes.

4.2.2 Mistakes Design

We created this scenario to engage children in reading with a robot while at the same time identifying the robot's mistakes. To do this, children need to read the story alongside the robot. To achieve this, the robot is designed to make mistakes while reading which gives the child the opportunity to correct the robot. As such, the design of the mistakes made by the robot was an essential part of the work. We considered that the robot should exhibit similar mistakes a child with reading difficulties makes. As such, we relied on the body of work on *Miscue Analyses* by Kenneth Goodman (K. S. Goodman, 1969, 1973) where a taxonomy to analyze the readers' deviations from the text (called miscues) was proposed. The deviation from the text is labeled as *miscue* and it is analyzed using 28 linguistic questions that investigate its compatibility with the text. The taxonomy was transformed into a diagnostic kit named Reading Miscue Inventory (RMI) by Yetta Goodman and Burke (Y. M. Goodman & Burke, 1972). This inventory includes the guide to mark, select and code the miscues for further analyses. While RMI provides a diagnostic tool for reading clinicians or special reading teachers, it also has several disadvantages regarding the time needed for administration, scoring and interpreting the result. According to the Simplified Miscue Analysis (SMA) by Cunningham miscues made by a reader can be analyzed through 4 questions (Cunningham, 1984). We used the first 3 questions from SMA to design our robot's mistakes. The questions from SMA examine the miscues based on similarity to the original wording, change in syntax, and change in meaning. These questions led us to define a property called the level of mismatch in designing the robot's mistakes. Furthermore, since we are working on books with images, we are also interested in adding a type of mistake defined as a mismatch with illustrations. This type of mistake was added to check the children's attention to the illustrations and the reading, and they are easier to recognize. As a result we designed 3 types of mistakes based on the different types of mismatch.

Type-1 mistakes are defined as any mismatch between the wording and the book's illustrations. An example of this type is when the robot reads *elephant* instead of *penguin* while there is an image of a penguin in the book.

Type-2 mistakes (question 3 from SMA) are contextual mistakes and correspond to a change in the meaning of the sentence after being replaced with another word. One example is saying *start* instead of *stop* in the text. Mistakes of Type-1 can also change the context at times, but

Table 4.1 – Different types of mistakes designed for the experiment each with an example sentence used in study 1. The last column shows the robots mistake.

Mistake Type	**Description**	**Correct Sentence**	**Mistake**
TYPE-1	Mismatch with illustrations	They saw a rainbow across the **sky**.	sand
TYPE-2	Mismatch with meaning	They **played** in the snow.	stayed
TYPE-3	Mismatch with pronunciation	Play in all the **colors**.	color
	or slight variation of the word	Kipper **picked** up his hat.	picking

since they are recognizable through images, the priority is to categorize them as Type-1.

Type-3 mistakes are defined based on the questions 1 and 2 from SMA and deal with pronunciation or syntax issues. These mistakes look like the original wording but have the wrong pronunciation of the word. One example is reading the past tense verbs such as *jumped* $/d\text{ʒ}\Lambda\ mpt/$ with wrong pronunciation such as *jumped* $/d\text{ʒ}\Lambda\ mpId/$. We also include mistakes that slightly modifies the original wording with changing the syntax to this type. One example is modifying the end of the verb from *-ed* to *-ing*, or similar modifications, or changing the original wording from singular to plural or vice-versa. We design these types of mistakes, because they can be recognized through pronunciation or understanding the syntactical changes. Table 4.1 shows each type of mistake with a sample that was used in the user study.

4.3 Pilot Study - Testing the Interaction with CoReader Platform

4.3.1 Experimental Design

The book used in the pilot (Figure 4.3) was not selected from the school curricula, however its difficulty was adjusted to children's reading level through confirming the book with the teachers. As mentioned before, children used cards equipped with ARTags to signal the robot to stop, start or repeat itself. Since the pilot uses the early developments of the platform children could only show the cards to the robot when the robot had finished reading. Furthermore, we also used speech recognition in this experiment using Google API, only to detect the child's correction of the robot's mistake. The study has two conditions of with and without pointing as shown in Table 4.2 and all the student read the same book called "Boo".

Table 4.2 – Between-subject experimental design with two conditions, "with pointing" and "without pointing".

Conditions	**Reading Session**
With pointing	Book: Boo
Without pointing	Book: Boo

Figure 4.3 – CoReader platform in the pilot.

4.3.2 Hypotheses

Our first hypothesis tries to consider the effect of robot's non-verbal cues, in this case pointing gestures on the child's performance. The second hypothesis considers different types of mistakes and tries to evaluate the effect of pointing on the child performance for each type.

- Ha: The robot's pointing gestures increase children's attention and performance in recognizing reading mistakes.

- Hb: The robot's pointing gestures have more positive effects for type-3 mistakes.

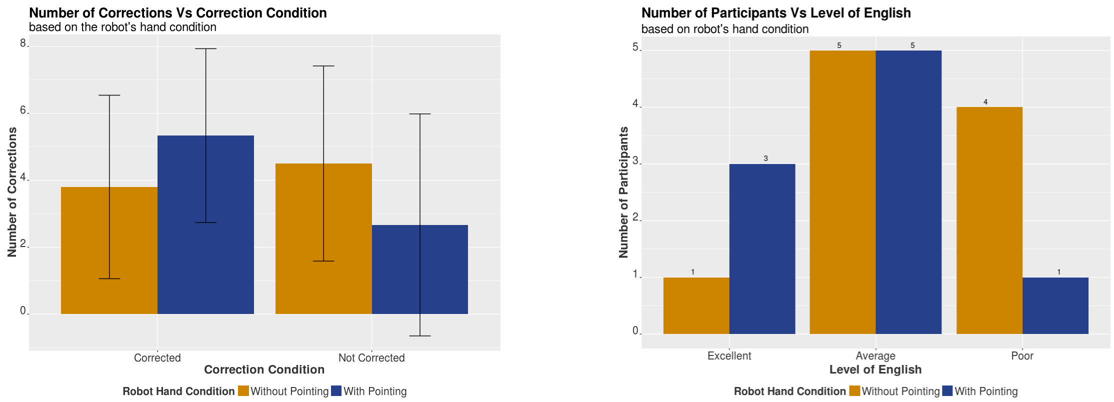
4.3.3 Participants

Twenty students in the 1st grade, drawn from two classes, participated in this study. The children were attending a bilingual school, with one-day English and one-day French curriculum. The level of speaking and writing in English among the subjects was varied, based on the observation and assessment by their teachers. One of the children was removed from the study due to the poor level of English to the point of not understanding the instructions. Among the 19 subjects, 11 were girls, and 8 were boys. Students were randomly assigned to two target groups, "with pointing" and "without pointing". In the first group the robot's hand was pointing to the text, and in the second one, it stayed motionless. The rest of the experimental conditions were the same for both groups except for robot's hand gesture. The with pointing group was composed of 9 students (5 F, 4 M) and the without pointing group was composed of 10 students (6 F, 4 M).

4.3.4 Results

Ha: Effect of Pointing Gesture on Attention

The first hypothesis focuses on the children's performance on the with pointing and without pointing conditions. Figure 4.4a shows the number of encounters that all children had with the robot categorized into corrected and not corrected encounters for both conditions. For the encounters in which the child corrected the robot; as shown in the graph, children in with pointing condition (M=5.33, SD=2.59) were more successful than without pointing condition (M=3.8, SD=2.7), however the t-test shows no significant difference; t=-1.251, p = 0.227. For encounters in which the child could not recognize the mistake and correct the robot, children in the pointing condition had fewer not corrected cases (M=2.66, SD=3.31) compared to without pointing condition (M=4.5, SD=2.91), yet again the difference was not significant; t=1.273, p = 0.220.



(a) The number of encounters based on the corrected and not corrected cases for both conditions

(b) Participants distributions in both conditions based on the reading assessment provided by teachers

Figure 4.4 – Ha. children's performance and distributions based on the pointing conditions.

Hb: Effect of Pointing Gestures on Different Mistakes

The second hypothesis focuses on the children's performance in each condition concerning the type of mistakes introduced in section 4.2.2. We want to show that the pointing will help children recognize type-3 mistakes more than other types of mistake. We expected the pointing to increase children's attention to the text. Unfortunately, due to some limitations in the design of the experiment, this hypothesis cannot be adequately analyzed. In the pilot, each page includes one or two mistakes, however, children could only give feedback at the end of each page. This has created a mismatch between the chances of finding the mistakes and evaluating children's performance. In the pages with two mistake, children had higher chances of recognizing the mistake. As a result, we do not present the analyses of this section and use what we learned to improve the interaction for the main study. Nevertheless, we will evaluate this hypothesis in the next experiment after changing the design and appropriate modification.

4.3.5 Findings from the Pilot

The pilot experiment provided us with valuable insights on how to redesign and improve the main experiment. To avoid having a bias in the distribution of children in the two conditions, we decided to change the experiment design from between-subject to within-subject and have an evaluation of children's reading level ahead of the experiment. Our second insight concerns the reading material. We realized it is better to carefully select the reading material after discussing it with the teachers and considering children's reading level and school curricula. The last thing we learned from the pilot was to change the feedback system from the card system to a real-time system. We have described both systems above. The reason for changing from the cards to the buttons was what we observed in the pilot. We noticed there were some delays and lags in recognizing the cards sometimes, especially if children did not hold it correctly in front of the robot's camera. Furthermore, waiting until the robot finish reading made the evaluation harder for us, and sometimes the students were becoming impatient. Furthermore, we decided to stop using the speech recognition to recognize children correction of the word and opt for automatic recognition of what they say. This means, the robot acts like it is listening but in reality it just pauses to give sometime to children to repeat the word.

4.4 Study 1 - The Effect of Deictic Gestures with CoReader

4.4.1 Reading Material

After selecting the school and the classroom for the main study, we had a discussion with the primary school teacher about selecting the appropriate reading material for children. The teacher proposed using books from the *Oxford Reading Tree* as it is one of the reading materials used in the classroom. Furthermore, the teacher informed us about children's differences in reading level even in one grade. In our target class, the teacher had already grouped children into two levels based on their reading abilities. To make sure that the reading material corresponded to children's reading level, we adapted the books to their level, hence, we selected two books from different stages of the Oxford Reading Tree. Based on the teacher's recommendation, we selected books one stage higher than the one children were currently reading in the class. The decision was made to ensure that the book is challenging enough for children and they have not read it before. We decided to call the reading levels as *Low level* and *High level*. For the *Low level* group, a book from stage 5[II] of the Oxford Reading Tree with 24 illustrated pages was selected . It was divided into two equal parts, first part was read on the first day of the experiment and the second part on the second day. On Day 1, the robot read 9 pages and upon reaching the 10^{th} page, it announced it was tired and asked the child to

[II] Oxford Reading Tree Stage 5: *Village in the Snow*; created by Roderick Hunt and Alex Brychta

Figure 4.5 – CoReader activity platform used in study 1.

continue reading until page 12^{th}. On the second day of the experiment, the robot welcomed the child back and showed excitement to know the end of the story. Like the previous session, the robot read 9 pages, and the child read the remaining 3 pages. The same procedure was carried out for the *High-level group*, except that they read stage 8^{th} book[III], with 32 pages. With the same principle, the robot reads 10 pages of the book and the child reads 6 on the first day and the same number on the second day until the end of the book. It is important to mention that, originally we were planning to use children's eye-tracking data for extra analyses and that is why we asked children to read. However, we do not report any eye-tracking related analyses as we were unable to use most of the eye-tracking data due to the loss in calibration and accuracy during the sessions.

4.4.2 Experimental Design

To understand the effect of pointing on the child's attention to the text, we designed an experiment with the CoReader platform that used pointing gestures as an independent variable and had two conditions. While reading the book, in one condition the robot looked and pointed at the text and in the other condition it only looked at the text. To avoid any grouping bias in the "with pointing" and "without pointing" conditions, the experiment was counterbalanced and had a within-subjects design. Each child went through two conditions on two different

[III]Oxford Reading Tree Stage 8: *The Rainbow Machine*; created by Roderick Hunt and Alex Brychta days. Apart from the pointing conditions, we also grouped children based on their reading level and assign them books accordingly. As explained earlier, each level has its own book that corresponds to the child's reading level. For the Low-level group, besides using an adapted book, the Nao robot's reading speed is decreased to 60% of its normal speed. On the other hand, for the High-level group, with more reading proficiency, the robot's reading speed is adjusted to 80% of its normal speed. As explained before, we designed three types of mistakes. On each day of the experiment, the robot makes 12 mistakes per book, which were carefully designed and randomly positioned throughout the text. The interaction started after the robot introduced itself and explained the procedure, then the experimenters made sure that the children understood the procedure. The duration of the experiment varied from 9 to 30 minutes in total, the effective time of the experiment was longer for some children due to some technical difficulties with the robot. After reading between 9 to 10 pages, the robot announced that he is tired and asked the child to continue reading the book. Compared to the pilot, in this study, children used buttons as presented in Figure 4.1 to inform the robot about its mistakes. Furthermore, considering the limitations of speech recognition in detecting the child's speech, in this study we avoided using speech recognition. Instead, the robot always acted like it understood the child in the first try.

In order to assess the impact of pointing on the performance, we measured how many corrective feedback the child gave to the robot. We defined the measure *Correction Percentage* to be the main measure of this experiment, assessing the percentage of corrections made by the children. However, the children could make correct and incorrect corrections. As such, a correction was considered *True Positive* if the robot made a mistake and the child corrected it. On the other hand, *True Negative* occurred when the robot did not make a mistake but the child considered it as a mistake. If the robot made a mistake but the child did not recognize the mistake it was considered as *False Positive*. For the analyses, when measuring the correction percentage, we only considered the *True Positive* and *False Positive* occurrences.

Table 4.3 – Plan of the experiment with a counterbalanced within-subject Design.

	Day 1		Day 2	
Level	**Book**	**Condition**	**Book**	**Condition**
Low	stage 5 - part 1	with pointing	stage 5 - part 2	without pointing
		without pointing		with pointing
High	stage 8 - part 1	with pointing	stage 8 - part 2	without pointing
		without pointing		with pointing

4.4.3 Hypotheses

Here, we briefly present the hypotheses for this study as described in section 4.1.2.

H1: Children in the "with pointing" condition show a better performance in correcting mistakes than in the “without pointing” condition.

H2: Children show a better performance in correcting Type-1 mistakes (illustrations related) compared to Type-2 and Type-3 (contextual and pronunciation related).

H3: The robot’s pointing gesture is more effective in helping children with lower reading levels than higher reading levels.

4.4.4 Participants

The study involved 22 typically developing children between the ages of 6 and 7 (11F, 11M) attending second grade in an international school in Switzerland, who participated in this within-subject experiment. Each child interacted with the robot during two sessions of around 30 minutes in two different days. The children were divided into two groups according to the reading level reported by their teacher. We refer to these two groups as *Low* and *High* level. We used the Oxford Reading Tree series to assign an appropriate book to each group based on their reading level (*Low*: stage 4 of Oxford Reading Tree and *High*: stage 7 of Oxford reading tree). According to the teacher, as soon as children in the Low level are proficient in their level they are promoted to the higher levels. Hence, as mentioned before we used books one stage higher than children’s reading levels, based on the teacher’s recommendation. The *Low* group consists of 9 children (5F, 4M) and *High* level of 13 children (5F, 8M).

4.4.5 Results

Validation Check for Mistake Types

As explained before, the mistakes design is one crucial part of this work. The mistakes are not just recognized by reading the text. Other aspects such as images, context, syntax, and pronunciation can also help the child to recognize them. While a combination of good reading proficiency, careful reading, and simultaneously listening to the robot should guarantee the detection of any mistakes, it is reasonable to consider additional means that can help the child to recognize the mistakes as an influential factor. And these additional methods were contributing factors in designing and categorizing the mistakes. In Figure 4.6, we display children’s success in correcting each type of mistake regardless of their book level and robot pointing condition. The figure shows that Type-1 mistakes have the highest correction percentage and Type-3 have the lowest. The correction analysis reveals that our mistakes design has been in accordance with our design implications. Pearson’s Chi-squared test shows no significant difference between correction of Type-1 and Type-2 mistakes ($\chi^2 = 2.2037, df = 1, p-value = 0.1377$), as well as Type-2 and Type-3 ($\chi^2 = 3.273, df = 1, p-value = 0.070$). However, the difference between recognizing Type-1 and Type-3 is significant with ($\chi^2 = 11.604, df = 1, p-value < 0.001$).

Regarding the Type-1 mistakes, we can say, since they were recognizable from the illustrations, most children were successful in recognizing them. This is aligned with our expectations regarding this type of mistake. On the other hand, Type-2 mistakes had a lower correction percentage compared to Type-1 mistakes, considering that they mismatched with the meaning. We expected to find the mismatch with images to be easier to detect than mismatch with meaning, and the results seem to be in line with this assumption. Type-3 mistakes are a bit more complex, as these mistakes have either slight or no deviation from original wording and their differentiation comes in either a change in pronunciation or the ending of the word. Thus, recognizing them requires having good reading skills, and good knowledge of words' pronunciation. As we expected, this type of mistake was the hardest to recognize by the children, who exhibited their lowest performance. In conclusion, the results seem to agree with our design assumptions regarding the mistakes levels of difficulty. We will discuss more on the interaction of mistakes' types with other experimental conditions in the next section.

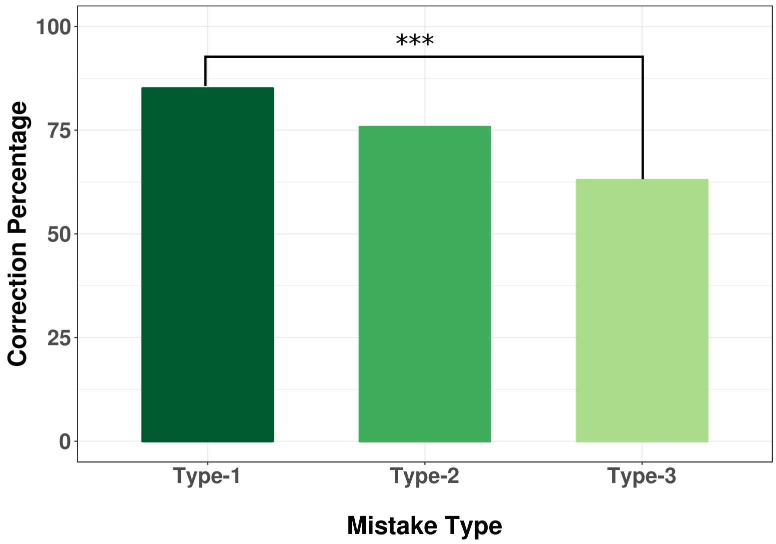


Figure 4.6 – Validation check for the types of mistake. $***p < 0.001$.

Participant's Statistics

On the first day of the experiment, 22 children participated in the study. However, on the second day, 2 children from the Low-level group and 2 children from the High-level group were absent. Furthermore, 2 children from the 18 participants who repeated the experiment were removed due to loss of data logs. The final results are obtained from 16 participants. Considering the within-subject design of the experiment, we checked if there was any effect between the first and second day of the experiment. The analyses shows no significant difference between days ($\chi^2 = 0.5537, df = 1, p - value = 0.456$).

H1: Effect of Pointing on Correction Percentage

Figure 4.7 shows the correction percentage according to the pointing conditions. Pearson's Chi-squared test with Yates' continuity correction is used to evaluate the difference between the two conditions. This test shows no significant difference between pointing and no pointing conditions ($\chi^2 = 0.68952, df = 1, p - value = 0.406$). Considering that our main hypothesis investigates if pointing improves children's performance in reading and recognizing the mistakes, no significant difference is observed, hence our first hypothesis is rejected.

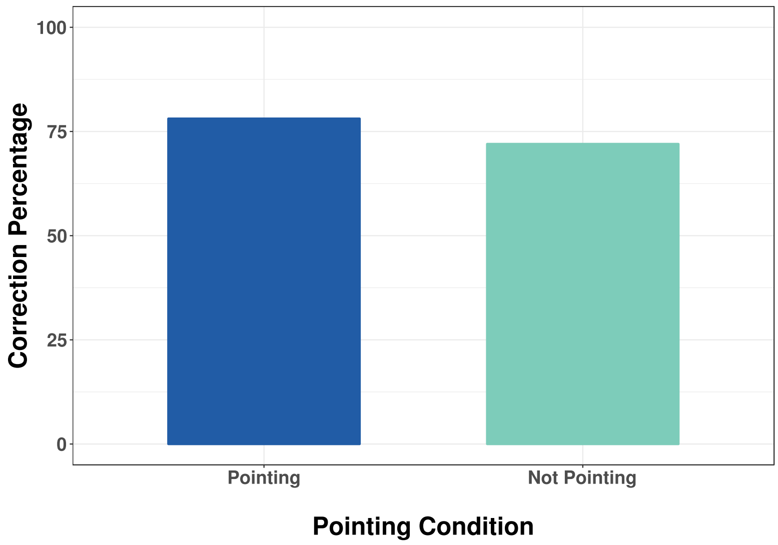


Figure 4.7 – H1. Correction percentage versus pointing conditions.

H2: Interaction Between Pointing and Mistake Types

Figure 4.8 demonstrates the correction percentage for each type of mistake for both pointing conditions. In particular, the interaction between mistake types and pointing conditions is presented here. We can observe that children in the pointing condition show much higher performances in correcting Type-1 mistakes. This difference in performance for Type-1 mistakes is significant based on Pearson's Chi-squared test ($\chi^2 = 11.389, df = 1, p - value = 0.0007388$). This result proves our second hypothesis that pointing gesture is most effective on the mistakes recognizable through illustrations. The differences for Type-2 and Type-3 mistakes between two pointing conditions are not significant with following test results ($\chi^2 = 0.19996, df = 1, p - value = 0.6548$) and ($\chi^2 = 0.53361, df = 1, p - value = 0.4651$) respectively. Nevertheless, we still explore the interaction between pointing conditions and mistake type in more detail in the next section when we also group the result based on the children's reading level.

H3: Interaction Between Pointing and Reading Level

In this section, we divide the results based on the children's reading level. We have already discussed the interaction between pointing conditions and mistake types for all of the children in Figure 4.8. In Figure 4.9, we present the interaction between the pointing conditions and mistake types separated according to High and Low-levels. Figure 4.9a, for High-level group,

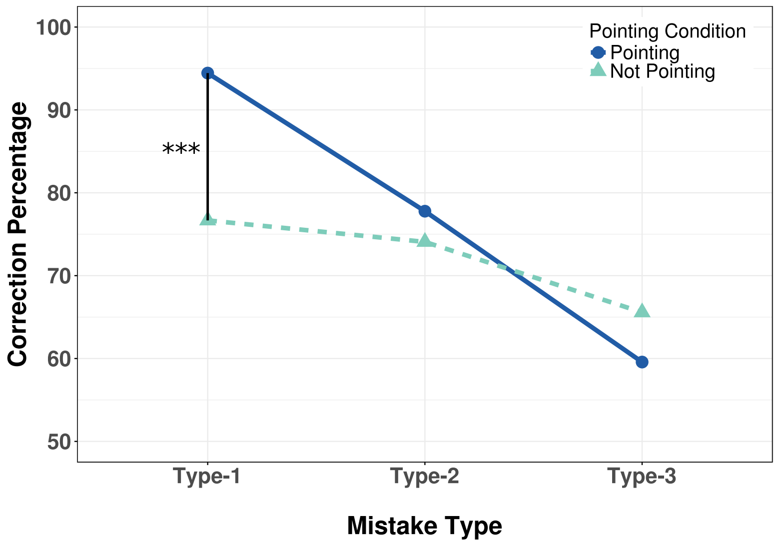
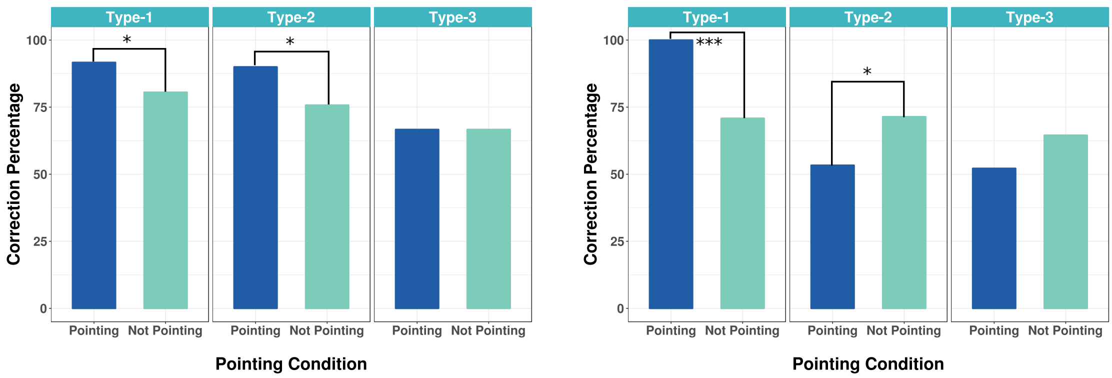


Figure 4.8 – H2. Interaction between pointing conditions and mistake types. $***p < 0.001$.

shows that children in the pointing condition are more successful in correcting the robot when it makes Type-1 and Type-2 mistakes. Pearson's Chi-squared test shows a significance of $*p < 0.05$ for both Type-1 and Type-2 mistakes with $(\chi^2 = 4.2741, df = 1, p-value = 0.0387)$ and $(\chi^2 = 6.1791, df = 1, p-value = 0.01293)$ respectively. The success in correcting the Type-1 mistakes corresponds to our second hypothesis. And for Type-2 mistakes, we can say that pointing could have helped the High-level children to be more concentrated on the story and context. There is no significant difference in correcting Type-3 mistakes between the two pointing conditions $(\chi^2 = 0, df = 1, p-value = 1)$. We can deduce that children who are proficient in reading benefit more from the pointing condition in correcting Type-1 and Type-2 mistakes.



(a) High level group. $*p < 0.05$

(b) Low level group. $*p < 0.05$, $***p < 0.001$

Figure 4.9 – H3. Correction percentage for each type of mistake, arranged by pointing conditions; for High-level in the left figure and Low-level in the right figure.

Figure 4.9b shows the performance of the Low level group. We observed that this group was differently affected by the pointing conditions. They showed a better performance for the Type-1 mistakes and lower performance for the Type-2 and Type-3 mistakes, in the pointing condition. In the pointing condition, children were significantly better in recognizing the Type-1 mistakes ($\chi - squared = 31.845, df = 1, p - value = 1.67e-08$). This is aligned with the result of our second hypothesis. However, surprisingly, pointing gestures had a negative effect on the recognition of Type-2 mistakes. The difference between the two conditions for Type-2 mistakes is significant with $*p < 0.05$ using Chi-squared test ($\chi^2 = 6.2267, df = 1, p - value = 0.01258$). As it will be discussed later, the justification can be that children in the Low-level group may have been distracted by the pointing gestures. Moreover, similar to the previous results, pointing doesn't have a significant effect on children finding Type-3 mistakes ($\chi^2 = 2.6466, df = 1, p - value = 0.1038$).

We can conclude that children at both levels can benefit from pointing gestures when having to recognize mistakes that are a mismatch with the image. However, pointing seems to distract children in the Low-level group from comprehending the text. And as a result, in the pointing condition, they achieved a lower performance in recognizing mistakes when there was a mismatch with the meaning of the text.

4.5 Discussions

The main goal of this study is to evaluate the effect of pointing gestures by a robot in a reading activity with a child peer. The result from our user study shows that overall the pointing has some significant, yet diverse, effects on the children's reading, and particularly on their capability to detect and correct the robot's mistakes. Our findings are aligned with the results from Justice et al. (Justice et al., 2008), that explicit referencing to the print such as pointing increases the children visual attention to the print. However, we notice that the pointing affects mistake recognition differently based on the type of mistake and based on the children's proficiency in reading. We observed that pointing has a significant effect in recognizing the mistakes that are a mismatch between text and images, for both reading levels. The effect on this type of mistake is more significant for early readers compared to children who are more proficient in reading. On the other hand, for the mistakes that are a mismatch between text and meaning, pointing has a different effect on children in Low-level compared to High-level. While pointing helps children with High-level proficiency to recognize the mistakes related to the meaning, it has a significantly negative effect on Low-level children. Presumably, while the robot pointing gesture brings the child's attention to the images, it may also distract them from comprehending the context, thus leading to a negative result. This can also be confirmed from the observations made, that some children in the pointing condition were actually looking at the robot's hand and were curious about it, rather than following the story and its reading. Subsequently, pointing does not have a significant overall positive effect on recognizing mistakes that are a mismatch with pronunciation or syntax.

Teachers' Input

One important aspect of this type of studies is the teachers' feedback. In general, the teachers were really cooperative and interested in the project. They were willing to let the experimenter attend a class to become more familiarized with the classroom and the methods they used. When we explained the general idea and the concept of the robot reading to the child, they were very positive and considered that challenging children with one level higher book was indeed a good idea. Regarding the development and design of the mistakes, they provided us with relevant input about the common mistakes that children usually make during reading. Their suggestions were incorporated into the the experiment design. The teachers also suggested adding more modes of interaction between the child and the robot to enrich the whole experience. The idea comes from the fact that when children have a problem reading a word, especially in the Low-level group, they start deciphering the word letter by letter and by sounding each letter. Usually, when a child starts deciphering, they either recognize the word and read it successfully or they get help from the teacher to complete the word. Some new types of interaction that could help this process would be very useful. In fact, the proposed idea was to incorporate the deciphering mode into the robot and let the child be the one who would help the robot to read the word.

Children's Input

Basing our study on the learning by teaching paradigm, we figured informing children of their role during the experiment was a fundamental component of this experiment. Children were aware that they would be helping and correcting the robot and they were pretty excited about the interaction, especially the idea of being the robot's teacher. Children were selected randomly from the class and had a one-to-one interaction with the robot. They were as excited about working with the robot in the second session as they were in the first one. After the end of the second session, nearly half of the students were asking when there would be the next session. Some children were suggesting which book they wanted to read in the next session, with statements such as *"Can we have another book, the other time? [experimenter: Yeah, sure what book do you like?] another adventure, I like adventures, about Gazelles, I love Gazelles, that's my favorite animal"*. Some children had a long interaction with the robot before they left the room, about seeing it in the next session. *"Are you gonna have a third test? [Experimenter:not now] When will we have the third test, when everybody has the second test? [to the robot] See you in the third test"*. Some of the interesting reactions from the children were about the robot's progress in reading and their concern that the book was too hard for the robot. Occasionally they were suggesting how the robot should practice reading by saying *"He need to read easier books"* or *"He get better if he reads 10 books [experimenter: In a day?] aaah... in a week"*.

Regarding the mistake types, children were sometimes amused by the Type-1 mistakes and for them, it was interesting that the robot was making such mistakes, yet they were very understanding about it. In addition, while attending a reading session with the class, the experimenter had observed children with lower reading proficiency make similar types of mistakes. As mentioned earlier, some mistakes were suggested by their teacher and inspired by the common mistakes that the children make during reading. For these mistakes, some children were quick to figure out the association as they probably heard them before. For example, when the robot said *saw* instead of *was* there were cases of children who understood the robot is reading the word backward or mentioned some of their peers make similar mistakes.

4.5.1 Limitations

Our study has its own set of limitations, that should be considered in the interpretation of our findings. We explain these limitations here and how we try to overcome them, in our future development section.

Limited number of participants
We found the participants of this user study by first having pilots with the first and third graders, in order to improve our design and find the right audience. As a result, we targeted one classroom in the second grade, to have a homogeneous level of children, considering different classes have different learning structures and speeds, and we wanted to have a comparable group of students. However, we are aware that a larger number of participants gives more validity to the results.

Limited number of reading sessions
One of the goals of our platform is to provide a sustainable interaction. It is important to increase the number of sessions to observe the children's behaviour over the sessions. We are aware that children's performances, in the first few sessions, can be affected by the novelty effect of using a robot. Due to limitations imposed by other factors and school curriculum, we were only able to have two sessions per child in this user study.

Fixed mistakes
In the current experiment, the mistakes were designed by the experimenter following our design guidelines. The goal was to test our design structure for creating and implementing the mistakes into the system in order to evaluate and improve it for future interactions.

Scripted interaction
The current interaction was autonomous and scripted, which can work for a limited number of sessions. But, by increasing the number of sessions, we need to make the robot's behaviour more diverse and flexible. The idea can be to expand the range of the robot's behaviours in conjugation with the new modes of interaction.

4.5.2 Future Developments

This platform is designed with the idea of providing a reading companion for children. It is supposed to be complementary to their existing practices and accompanies them in early stages of learning, especially for children with reading difficulties. Based on the user study and our goals, there are four main development points that we are willing to make.

Developing Adaptive Pointing

Our observations and results from the user study show that continuous pointing gesture has diverse effects on the children's attention to reading. While it can be constructive in some aspects, it can also be distractive in some other aspects. As a result, we decided to design an adaptive pointing system based on the positive and negative effects of pointing. According to our observations, sometimes children in the not pointing group were getting lost not knowing which page the robot was reading. As a result, they either asked the robot to repeat or considered it as a mistake.

Increasing Modes of Interaction

Considering that correcting the robot's mistake is just one mode of interaction, we have lots of possibilities to design other modes of interaction in this context. These new modes can be inspired by the children's reading habits, such as when they decipher the text to read a new word, or when they ask someone else to read a word for them. We have already designed the version when the robot sometimes asks the child to read a word, but we haven't included it in our current user-study due to issues with speech recognition for children. Another mode of interaction would be when the robot starts deciphering the word and hesitates to read it, in order for the child to help the robot to pronounce the word correctly. These new modes of interaction may need their own deictic gestures, which can be more informative and different from the continuous pointing gesture we tested in this study.

Adapting the Mistakes to the Child's Level

One important aspect of this interaction is the children's perception of being the teacher. Therefore, when the child corrects the robot, it seems essential that the robot shows some improvements. In the current study, as a design feature, the robot doesn't make any mistake on the last page in each session, to give the impression of improvement to the child. We would like to make two main improvements to the system directed to this aspect. First, to adapt the mistakes types and difficulty to the child's level of reading. This can be achieved by analyzing the child's reading style using eye-tracking in an initial reading or by the type of mistakes he/she makes during reading. Moreover, we can also customize the mistakes to the types that the child does or does not recognize. Such adapting system can also help the robot to switch between different modes based on the child's strengths and weaknesses, to challenge them more. Second, to adjust the number of mistakes based on the child's performance in real-time. Especially to decrease the number of mistakes or certain types of them when the child has a good performance to give an illusion of improvement in the robot.

Integrating Eye-Trackers

Having adaptive pointing gestures call for providing more information to the robot regarding the child's real-time attentional state. Robot's knowledge of the child's attentional state can help it in achieving joint attention. There have been numerous studies on joint attention between human and a robot. A robot with such a knowledge is able to Initiate Joint Attention (IJA) with the child, Respond to Joint Attention (RJA) cues from the child, and ensure it (EJA) during the interaction. As a result, the interaction becomes more robust and autonomous.

For these reasons, we are interested in using eye-tracking glasses and integrating eye-tracking data into our robot. Such implementation consequently benefits our previous remarks on the future development of our platform.

4.6 Conclusions

This chapter presented a platform designed to support children's reading practices through a Reading While Listening (RWL) activity with a social robot. The presented experiment aimed at studying the effect of robot's pointing to the text on the child's performance as a measure of their joint attention. Furthermore, it tried to channel the learning by teaching paradigm as a way to keep children motivated to engage in the reading activity and feel more responsible for the robot. The experiment was designed to compare children's performance under two conditions for robot's pointing gestures (with and without pointing) in a within-subject study design.

We measured children's success in correcting the robot's mistakes, to evaluate their performance. The performance was analysed over the pointing conditions and different types of mistakes. No significant difference was observed in children's overall performance for both conditions. However, contrary to the expectation that pointing improves children's performance, some children were negatively affected by the robot's pointing in correcting the mistakes. The pointing gestures were deemed to be beneficial for children in directing their attention to illustrations and helping them correcting mistakes associated with that. On the other hand, children with lower reading proficiency had shown to be negatively affected by the pointing in finding the mistakes associated with the context. Overall, we could conclude that pointing gestures might have been distracting for children with lower reading proficiency by preventing them from comprehending the text and recognizing mistakes related to that.

In addition to the reported findings, other components of the interaction such as spatial arrangements and the methods used to study joint attention guided us in developing the next study in the thesis. While we were inclined to continue the developments with the platform used in this chapter, we found that the spatial arrangement of the child and the robot does not seem ideal for studying other components of spatial perspective taking. As a result, we decided to develop a new platform to study perspective taking. The new platform and research direction is discussed further in Chapter 5.

5 Objects Game: A Behavioural Study

In this chapter, we take the first step towards understanding children's perspective taking abilities and their tendency to adapt their perspective to a counterpart in the context of participating in a cooperative task with a robot. In the first section, we describe the motivation and expected contributions of the study. Then, we discuss the research questions that guided this study, followed by a description of the interaction, task, and platform designs. Then, we go through the details of the pilot and main studies and explain the study design, conditions, and result analyses. Finally, we wrap up this chapter by discussing the findings and limitations of the study, what we learned, and how it contributes to future chapters.

5.1 Scope and Research Goals

5.1.1 Motivation and Contribution

While the ability to recognize the correct perspective develops in humans from childhood and solidifies during school years, it needs to be developed in robotic and artificial agents as part of development of their cognitive framework. To develop such capabilities in robots, first, we decided to examine how children exhibit and adapt their perspective in an interaction with a robot. As a result, we designed a straightforward task composed of moving objects around using a touch screen to reach a goal presented to the players in the form of goal cards. The following study is designed with two goals in mind, understanding children's perspective choice when they initiate an interaction with a robot and their perspective adaption during the interaction. To evaluate children's perspective taking abilities and adaptive behaviour, the activity utilizes taking the spatial perspective of the other player as a key component for completing the task. This means each player needs to consider their counterpart's perspective to successfully complete the task. We analyze children's choices of *frame of reference* and *perspective marking* during the interaction and use it as an inspiration to design the perspective taking model.

5.1.2 Research Questions and Hypotheses

Considering the developmental timeline of perspective taking abilities and particularly spatial perspective, children tend to be still developing level-2 spatial perspective taking between the ages of 5 to 10 years old (Flavell et al., 1981; Moll & Meltzoff, 2011; Surtees et al., 2013). Before we start developing a perspective taking model for the robot in later chapters, we were interested to know more about children's perspective taking choices when they interact with a robot. First, we targeted frame of reference as a fundamental component of spatial perspective taking and observed how children choose and adapt their frame of reference when they interact with a robot. Second, as a consequence of using verbal instructions, we added perspective marking as the second component. To become familiar with these two components, *frame of reference* is a set of axes or origin points for addressing the position of the objects or their spatial relationships (Levinson, 1996; Mintz et al., 2004; Trafton et al., 2005). Various distinctions between frames across different disciplines are presented in Table 2.1. In this study, we are focusing on *egocentric* (from the Self point of view) versus *addressee-centric* frames of reference (from the Other point of view). On the other hand, *perspective marking* concerns how the speaker marks their perspective when they use natural language (Steels & Loetzsch, 2009). Here, we use the *explicit* category which corresponds to the use of possessive pronouns e.g., my, your and *implicit* category which corresponds to not using any of the possessive pronouns and rather the use of definite article e.g., the. Building upon these concepts, if a child seating in front of a robot tells the robot "give me a brick on *your* right", the child is addressee-centric and explicit. Whereas, if they tell the robot "give me the brick on the left", the child's utterance is marked as implicit, and it is not clear if the child is egocentric or addressee-centric meaning that the brick can be on the child's left or the robot's left. Consequently, we are interested to know if an activity focused on taking other's perspectives can help children to practice such skills with a robot. Our first research question looks at the impact of the activity on children's performance in a set of pretests and posttests that evaluate such skills.

- **Research Question 1:** Can interaction with a social robot with a focus on spatial perspective taking boost children's performance on tasks involving perspective taking?
  Hypothesis 1: children show better performance in the posttests compared to the pretests in taking the other character's perspective.

Our second research question is more focused on children's choice of perspective when they initialize the interaction with the robot. Understanding this helps to develop the robot's behaviour from two aspects, first is about composing the utterances to start the interaction that sounds natural and clear. The second is finding the best approach to comprehend children's initial utterances, for example by making initial assumptions or ask for clarification.

- **Research Question 2:** What is children's first perspective choice, when collaborating with a robot?

Hypothesis 2: Children use more "implicit egocentric" perspective (without using possessive pronouns) compared to "explicit and/or addressee-centric" perspectives When they instruct the robot for the first time.

The third research question, focuses on children's perspective adaptation, either when composing an instruction to the robot or comprehending the robot's instruction. Related to this research question, we have composed three different hypotheses that look at the perspective adaptation for the frame of reference and perspective marking separately. Furthermore, there is one hypothesis that corresponds to the experimental design of the study and children's perception of the robot change of perspective.

- **Research Question 3:** When working on a task that involves perspective taking, will children change their perspective and accommodate their counterparts to achieve the goal?
  Hypothesis 3 Children's overall choice of frame of reference shifts from egocentric in the first instruction to addressee-centric in the rest of the instructions.
  Hypothesis 4 Children's overall choice of perspective marking shifts from egocentric in the first instruction to addressee-centric in the rest of the instructions.
  Hypothesis 5 Children in the "Ego-Ego" condition (robot keeping the egocentric frame of reference) perform better in the last session compared to children in the "Ego-Add" condition (robot switching frame of reference).

It should be mentioned that before we carried out the experiment, in order to select the appropriate age group to participate in our study, we had run a small pilot study. In the pilot study we were interested in the following research questions:

- **Research Question A:** At what age group are children able to comprehend the task and carry it out without the help of the facilitator?
- **Research Question B:** At what age group are children able to correctly differentiate between their left/right and the robot's left/right?

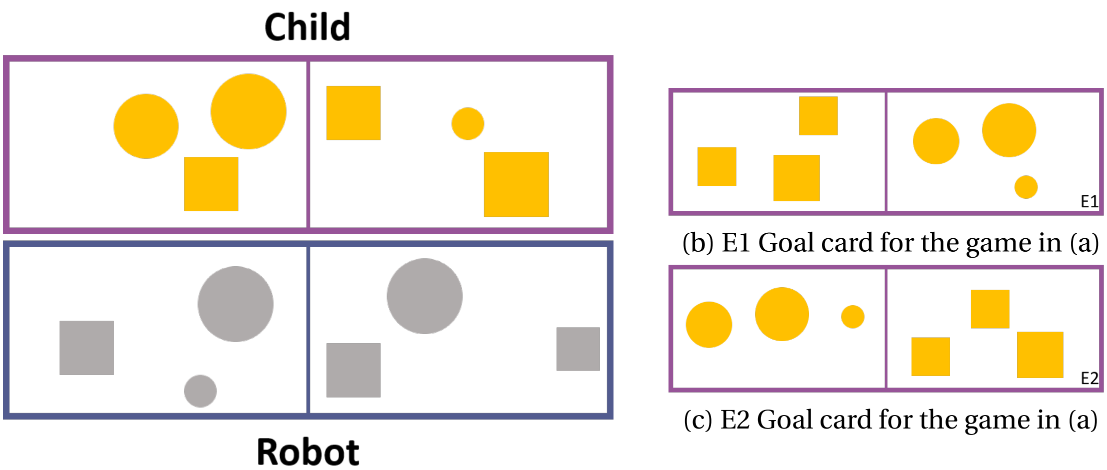
The response to research questions A and B were strictly with respect to the designed activity and for selecting the right participants, as a result, it cannot be generalized.

5.2 Development of the Object Game Platform

If we think about a child collaborating with a friend to assemble an object such as Lego bricks, we can consider a scenario where they are seated in front of each other, inherently, the child and their friend have mirrored egocentric frames of reference. When the child asks their friend to hand them an object or modify a part of the assembly, they might ask them using expressions that include left and right notions. Expressions such as *"Give me the red brick on the right"* or *Pass me the yellow brick on your right.* Now in such a scenario, the yellow brick is certainly on the child's left (as the speaker), however, the red brick might be on the child's left or their right, depending on which frame of reference they used.

5.2.1 Game Development

Our gamified task is called the *objects game*, and it includes simple geometrical objects such as circles and squares in various colours and children are supposed to move them from one side of the screen to the other. To get a glimpse of how the main screen looks, the easy version of the game is presented in Figure 5.1. The game includes a "main screen" which is what the child and the robot see on a touch screen during the interaction as shown in Figure 5.1a and "game cards" which are used to make instructions to play the game as shown in Figure 5.1b and 5.1c.



(b) E1 Goal card for the game in (a)

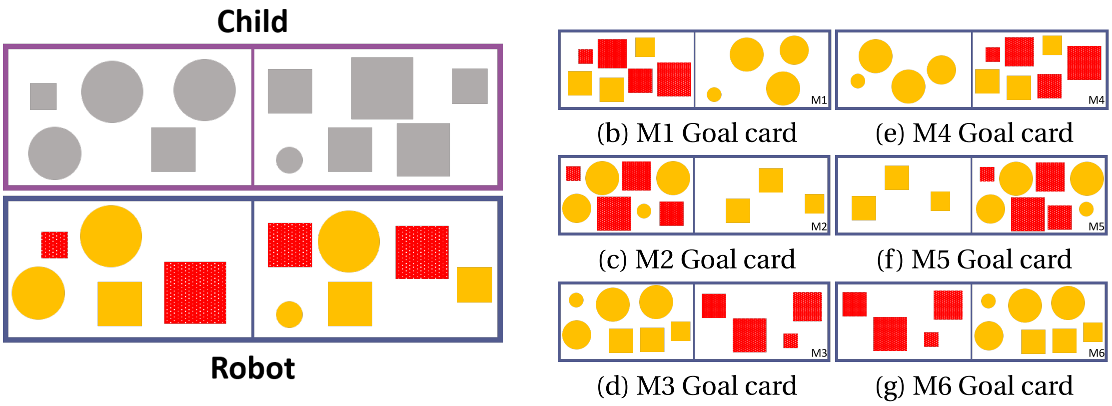
(c) E2 Goal card for the game in (a)

(a) main screen, child side activated

Figure 5.1 – Easy level of the game, including the main screen in (a) and goal cards in (b) & (c).

Main Screen

The main screen is divided into two main sections called the child side and the robot side. During the game, depending on the player's turn, one side is enabled, e.g. the objects are in colour and can be moved around while the other side is disabled, e.g. the objects are in grey and can't be moved. There is also a vertical division on each side, which is used for distinguishing left and right movements. The objects can be of any simple shape or colour, in the easy version of the game as shown in Figure 5.1a, we used two types of objects: *squares* and *circles* both only in *yellow* colour. The game is designed with two levels of difficulty, which is a function of the colour and the type of the objects presented in that level. The difficulty corresponds to the number of moves needed for a player to reach the state of the game presented in the goal card. As a result, the easy level which we considered as the practice level includes *yellow circles* and *yellow squares* and can be optimally solved in two moves. The medium level; which is used as the main game, has *red squares* in addition to the *yellow circles* and *yellow squares* and needs three optimal moves to be solved. Both levels are presented in Figures 5.1a and 5.2a.



(a) main screen, robot side activated

(b) M1 Goal card (e) M4 Goal card (c) M2 Goal card (f) M5 Goal card (d) M3 Goal card (g) M6 Goal card

Figure 5.2 – Medium level of the game, with main screen (a) and all the possible goal cards associated to it (b)-(g).

Goal Cards

The goal cards represent the desired final state of the game that players are supposed to recreate by moving the objects. I one round of the game, one player is given the goal card and its task is to guide the other player to reach a state similar to what is represented in the card. The number of goal cards available for each game depends on the combination of the objects and colours available in that game. For example, the easy version of the game has 2 goal cards and the medium level has 6 goal cards as shown in Figures 5.1b-5.1c and 5.2b-5.2g, respectively. In each round of the game, only one goal card is randomly selected and given to the player.

Player's Roles

The goal cards represent the desired final state of the game that players are supposed to recreate by moving the objects. At the beginning of each round, one player is given the goal card and it is supposed to guide the other player to reach a state similar to what is represented in the card. The number of goal cards available for each game depends on the combination of the objects and colours available in that game. For example, the easy version of the game has 2 goal cards and the medium level has 6 goal cards as shown in Figures 5.1b-5.1c and 5.2b-5.2g, respectively. In each round of the game, only one goal card is randomly selected and given to the player. When the game starts, one player has the task of guiding the other player to reach the state represented in the goal card without directly showing the card to them. The player with the goal card is called the *instructor* and the player moving the objects is called the *manipulator*. The instructions are composed of three components: the colour, the type of the object, and the moving direction. An example of a proper instruction is *"move the yellow* *circles to the right"*.

In the main screen shown in Figure 5.2a the robot side is activated and the child side of the game is disabled. This means that the robot can manipulate the objects in front of it and the child is supposed to instruct the robot. As mentioned earlier the instructor has one goal card and must guide the manipulator to reach the same state as the card. If the child holds the M4 goal card shown in Figure 5.2e, optimally they can guide the robot using three instructions. If the first instruction is *"Instruction 1: move the yellow circles to the right"*, it can be considered to be implicit, which means depending on the manipulator's choice of frame of reference, it can be interpreted differently. If the robot is egocentric, it would move the yellow circles to its own right. However, if the robot suspects that the child was egocentric when making the instruction, it can have an addressee-centric approach and move the yellow circles to the child's right which is the robot's left. During the interactions, we record these instructions and analyze the children's choice of frame of reference and utterances based on their goal cards. Figure 5.3 shows the setup of the experiment as part of study 2, the setup presents the robot in the instructor and the child in the manipulator roles.

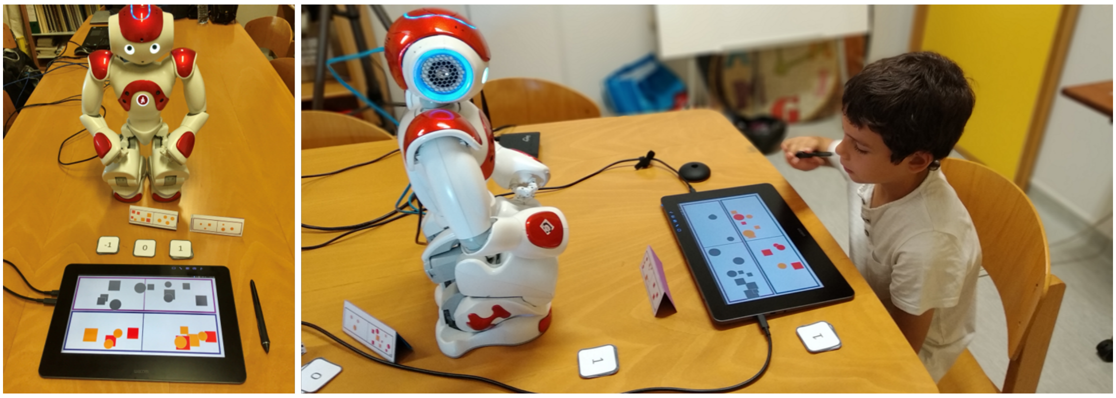


Figure 5.3 – The experimental setup when robot is the instructor and child is the manipulator.

5.2.2 Interaction Design

There are three elements involved in designing the activity and experiment. For instance, when someone is in the manipulator role, they can select an egocentric or addressee-centric frame of reference. However, when they are in the listener role, their choice of perspective partially becomes a function of the speaker's perspective. As previously mentioned, we are looking at children's choice of frame of reference and perspective marking in the context of the activity. To design the activity that involves spatial perspective taking, we recognized three elements that were observed in the utterances including spatial perspective taking. It is important to mention that in the current study, we do not implement complicated behaviour in the robot. The experiment is designed with a focus on observing and documenting children's perspective choices that can respond to our research questions. Implementing more complicated robot behaviour is part of future directions of this thesis. To clarify the study design, first we specify the following elements:

- **Perspective taker role:** e.g., Instructor (speaker) vs. manipulator (listener)
- **Frame of reference:** e.g., egocentric vs. addressee-centric
- **Perspective marking:** e.g., implicit vs. explicit

For the first element of the interaction, instructor vs. manipulator we decided to choreograph the interaction with children making the first instruction. The data from children's first instruction is used to answer our second research question. To give the child and the robot a chance to play both as an instructor and the manipulator, they play multiple sets of the game and switch roles. This means in one game the child has a goal card and instructs the robot to move the objects and in the next game, the robot has the card and instructs the child.

The second element of the interaction is the frame of reference that each player uses during the interaction. As part of the study design, we can control the robot's choice of frame of reference and observe the child's choice of frame of reference. When the robot is the instructor, it has the creative control over which frame of reference to use, in this case, we have decided to have the robot show the behaviour that we had presumed children show in their first utterance; *implicit egocentric perspective*. Since the robot is going to be implicit, if the child asks for clarification, the robot would update its instruction to an explicit egocentric utterance. On the other hand, in the manipulator role, where the robot perceives the child's instruction, the 'perception perspective' is a function of the child's perspective marking. If the child gives an implicit egocentric instruction, the robot is designed to perceive it egocentrically, which means the robot would make an incorrect move. However, if the child's instruction is explicit, either egocentric or addressee-centric the robot follows the instruction as is it. We expect that such behavioural design from the robot would create **confronting** which might lead to children's failure in reaching the goal, a possible effort to accommodate the robot, and hopefully adaptation behaviour.

As for the third element, again we can control the robot's perspective marking when the robot is instructing. We decided to keep the robot's instruction to be implicit egocentric in the first session as the instructor. However, to understand children's perception of the explicit behaviour of the robot, we designed the robot to be explicit in the second session as the instructor. The summary of the assumptions we made in designing the experiment is as follows:

- To evaluate children in both perspective roles, child and robot alternate between the instructor and manipulator roles,
- To document children's uninfluenced choice of perspective, children always start as the instructor in practice and session 1,

- To understand children's decision based on the robot's implicit instruction, the robot always gives *implicit egocentric* instructions in session 2 e.g. *"Move the squares to the right"*,
- To observe if and how children adjust their perspective to the robot after knowing the robot is egocentric, children instruct again in session 3,
- To understand how children react to the robot's change of frame of reference in session 4, robot's instruction perspective is divided into two conditions: *explicitly egocentric* (using *my*) and *explicitly addressee-centric* (using *your*).

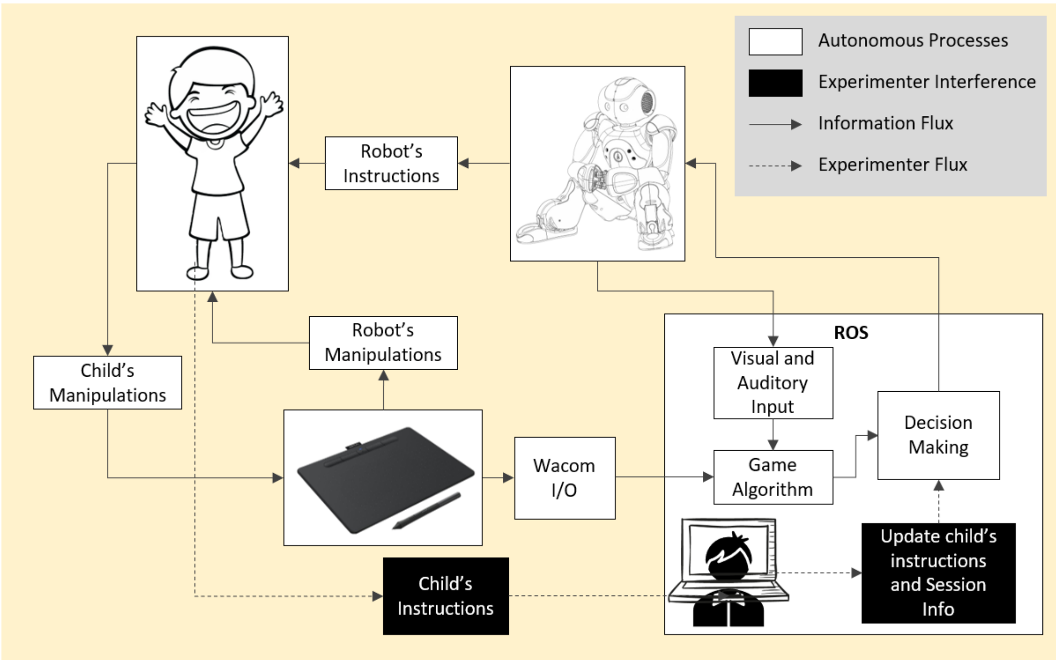


Figure 5.4 – Schematic of the study 2.

5.2.3 Experimental Design

The study uses a mixed-method design with two independent variables: *player's role* (instructor vs. manipulator) manipulated within-subjects and *robot's instructor perspective* (egocentric vs. addressee-centric and implicit or explicit) manipulated between-subjects. Table 5.1 provides the experimental design based on the assumptions and independent variables. It shows that children instruct in practice, session 1, and session 3 and the robot instructs in sessions 2 and 4. In the first condition called **Ego-Ego**, the robot is egocentric in both sessions 2 and 4, however, it is implicit in session 2 and explicit in session 4. In the second condition called **Ego-Add**, the robot is egocentric in session 2 and addressee-centric in session 4. As for perspective marking the robot is implicit in session 2 and explicit in session 4. As mentioned earlier, we decided to have the robot be implicit egocentric in session 2; the first time it instructs the child, so the robot has a similar behaviour as what we hypothesize children have. On the other hand, the reasoning behind changing the robot's perspective marking in session 4; the second time the robot instructs the child, is to have the robot show more adaptive behaviour in case children asked for clarification in session 2. Furthermore, we wanted to see how children perceive the change of frame of reference in **Ego-Add** condition.

As for dependent variables, we looked at children's performance in the tests, moves within the interaction, and their choice of perspective as an instructor and manipulator. In the instructor role, if children used possessive pronouns they were marked as *explicit* and if they did not use the pronouns, they were marked as *implicit*. Furthermore, their instruction was analyzed based on the goal card in their hand and marked as egocentric or addressee-centric accordingly. On the other hand, in the manipulator roles, we looked at the way they moved the object with respect to the robot's instruction and marked the move as correct or incorrect if it corresponds to the robot instruction.

5.2.4 Evaluation Methods

The activity was accompanied by two sets of pretests and post-tests. While we were interested in observing children's choices during the task, we also wanted to know if the task itself and interaction with the robot had any positive impact on children's learning. The design of the pretests and post-tests was inspired by the experiments that commonly investigate level 2 spatial perspective taking (Kessler & Rutherford, 2010; Kessler & Wang, 2012; Zacks et al., 2000).

Table 5.1 – Mixed experimental design with two conditions, the child is instructor in practice, session 1, and session 3, the robot is instructor in session 2 and session 4. "Ego" is short for egocentric, "Add" is for addressee-centric.

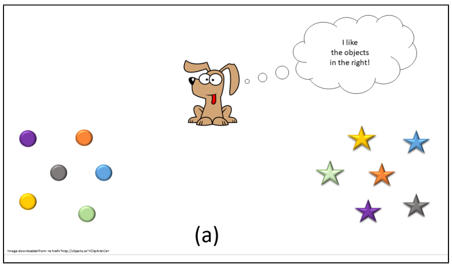
Conditions	**Pretests**	**Practice**	**Session 1**	**Session 2**	**Session 3**	**Session 4**	**Posttests**
Instructor		child	child	robot	child	robot	
Manipulator		robot	robot	child	robot	child	
Ego-Ego			Implicit Egocentric		Explicit Egocentric		
Ego-Add			Implicit Egocentric		Explicit Addressee-centric		

Left/Right Test or Toys Test

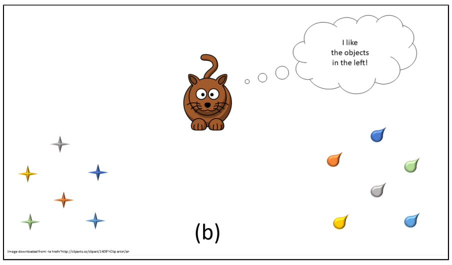
The Toys test was designed to evaluate children's recognition of the perspective difference between themselves and another agent (in this case an animal) that is facing them. Our evaluation is based on judging children's selection of the animal's favourite toy based on what the animal expresses in the test. If the child takes the animal's perspective, we consider it a correct answer, otherwise, it is incorrect. Two similar versions of this test have been designed with two different animals and different correct responses as shown in Figures 5.5a-5.5b. We alternated the tests as pretest and post-test between children. The instructions are as follows:

"The dog/cat thinks they like the right/left toy. Can you tell me which toy does the dog/cat likes? (wait for the response.)"

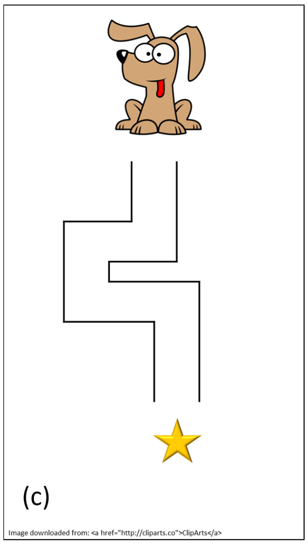
In the dog version of the test, the dog says *"I like the right side toys"*, and in the cat version, it says *"I like the left side toys"*. Hence, the correct answer for the dog version is *the balls* as they are on the right side of the dog and the correct answer for the cat is *"the drops"*. For example, if the child's answer is *the stars*, then the child is egocentric, and in this case, the answer is incorrect.



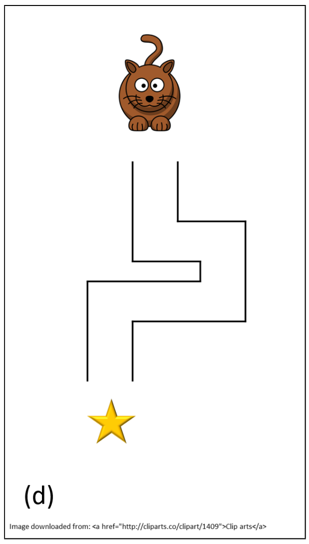
(a) toys test dog version



(b) toys test cate version



(c) path test dog version



(d) path test cat version

Figure 5.5 – Examples of pretests and posttests: V1 toys test (a) dog version and (b) cat version; V2 path test (c) dog version and (d) cat version.

Test of Direction Sense or Path Test

The path test shown in Figures 5.5c-5.5d is a simplified version of "Money Standardized Test of Direction Test" developed by Money et al. and modified by Zacks et al. which has been adapted for children. Again two versions of the test have been designed with different animals and directions and were alternated between the participants as pretests and post-tests. Similar to the previous test, this test is supposed to evaluate if children take the animal's perspective to guide them or not, and furthermore, if they do it correctly or not. Compared to the previous test, this test is more cognitively demanding. Looking at the test shown in Figure 5.5c, the child is supposed to guide the dog to reach the star. The instructions for this test are as follows:

"The dog/cat wants to reach the star at the end of the road, can you describe the path that the dog/cat needs to take to reach the star? (help the child by saying 'move forward' and let them complete the instructions)"

We do not provide children with the template to describe the paths which give them the freedom to describe it as they want. Any description within the following two approaches is considered correct. In one approach, the response is better described as defining the path. For example to guide the dog, a correct sequence for this approach in Figure 5.5c is *"Forward-Right-Forward-Left-Forward"*. In the second approach, the response is better described as walking with the animal in the path. For example the correct sequence for this approach is *"Front-Right-Front-Left-Front-Left-Front-Right-Front"*.

5.3 Pilot Study - Testing the Object Game and Platform

5.3.1 Research Questions

Before starting the main study, we wanted to make sure we were selecting the appropriate age group for our study. As a result, we formulated the following research questions and conducted a pilot study with 8 participants from different classes and age groups. As mentioned earlier, this pilot and its result cannot be generalized and its only purpose was to find the right age group for our study in the school we had access to.

The activity used is the same as what was described in the previous section. The only differences between the pilot and the main study are on some technical aspects and development side. Furthermore, in the pilot, the robot never asked the child if its move was correct or not, while this confirmation was added to the robot's interaction modality in the main study.

5.3.2 Participants

A total of 8 children were scheduled to participate in the study, however, one student was absent on the day of the study. As a result, 7 participants (4 female, 3 male) between the ages of 6 and 9 years old took part in this study. They were selected from four different age groups that were going to start 1st, 2nd, 3rd, and 4th grades in a Portuguese elementary school. The study had received ethical approval from the university's ethics committee and parental consent forms were collected from the parents of the participants before the main experiment.

5.3.3 Results

To determine the appropriate age group for participating in the main study, we looked at two criteria: children's ability to understand the task and to differentiate between their left/right and the robot's left/right. We wanted children to be able to understand the central concept, be challenged by the difference in perspectives, and make a decision to deal with the differences, either successfully or not. During the interaction, we noticed two of the participants of the ages of 6 and 7 years old (1st and 2nd grades) had fundamental problems distinguishing between their left/right. Furthermore, the 6 years old child had occasional problems identifying the shapes to produce the instructions. We had a plan to accommodate children with left/right issues by putting stickers on their hands and the robot's hand. Several psychology studies have used this technique in their perspective taking studies (Newcombe & Huttenlocher, 1992). However, it did not solve those children's issues and they were still confused about the robot's difference in perspective. We discussed this issue with the teachers, who advised us that the task was too difficult for children starting 1st and 2nd grades. On the other hand, we observed acceptable performances from children in the 3rd and 4th grade. The children in the 3rd grade were able to comprehend the task, they were egocentric at first, but one of them managed to recognize the discrepancy between theirs and the robot's perspective and update their instructions. With the 4th grade children, we observed that they effortlessly recognized the robot's different perspective and updated theirs. Based on our observation of children's performance and further discussions with the teachers, we decided to select children starting at 3rd and 4th grade to participate in the main study. We excluded younger children due to their issues with left/right and understanding of the task.

5.3.4 Findings from the Pilot

We were able to recognize a shortcoming in our interaction that was affecting children's perception of the robot. During the interaction, when the child instructed the robot in implicit egocentric instructions, considering the robot's egocentric perspective, the outcome of the move was opposite of the child's expectation. In such cases, some children were expecting the experimenter to explain why, and most just assumed the robot was faulty. To prevent this, we decided to add some level of transparency to the interaction for the future experiment by making the robot ask for feedback after every move. In case the child's feedback was negative the robot explains its decision by saying "but I moved them to my left/right".

5.4 Study 2 - Children's Perspective Choice and Adaptation

5.4.1 Hypotheses

Here, we briefly present the hypotheses for this study as described in section 5.1.2.

H1: children show better performance in the posttests compared to the pretests in taking the other character's perspective.

H2: Children use more "implicit egocentric" perspective (without using possessive pronouns) compared to "explicit and/or addressee-centric" perspectives When they instruct the robot for the first time.

H3: Children's overall choice of frame of reference shifts from egocentric in the first instruction to addressee-centric in the rest of the instructions.

H4: Children's overall choice of perspective marking shifts from egocentric in the first instruction to addressee-centric in the rest of the instructions.

H5: Children in the "Ego-Ego" condition (robot keeping the egocentric frame of reference) perform better in the last session compared to children in the "Ego-Add" condition (robot switching frame of reference).

5.4.2 Participants

A total of 35 participants (13 female, 22 male) between the ages of 7 to 9 took part in this study. They were selected from the 3rd and 4th grades of an elementary school to participate in the experiment. The study has received ethical approval from the university's ethics committee and a parental consent form from the parents of the participants before the main experiment. Moreover, the study was carried out after running a pilot with 7 children from 4 different age groups to test the system's functionality and to select the appropriate target age group.

5.4.3 Results

The following analyses were carried out after excluding the data from 2 children, with 33 participants (11 female and 22 male) between the ages of 7 to 9 years old ($M = 8.22$, $SE = 0.12$). Among the 33 participants, 18 of them were starting their 3rd grade and 15 were starting their 4th grade education.

H1: Children's Performance in Pretest and Posttest

To check if children improved in responding to the posttest in comparison to the pretest, first, we looked at the type of data collected from each test. To organize the responses when children did not specify the moving forward part, as long as they still used the correct sequence of turns, we considered their answers as correct. Furthermore, children were given no feedback on their responses in the pretest and posttest. Considering that for each test children either got 0 when they failed and 1 when they succeeded and the data for each test is dichotomous. We made a new variable by combining the result of both tests and looked at their overall performance in the tests. First, we analyzed the overall performance for normality using the Shapiro-Wilk test. A Shapiro-Wilk test for pretest and posttest of showed a significant departure from normality,

$W(32) = 0.78627, p = 1.786e-05$ and $W(32) = 0.70321, p = 7.306e-07$, respectively. Considering that we expect children's performance to improve from pretest to posttest and the data is skewed, we perform a one-tailed Wilcoxon test.

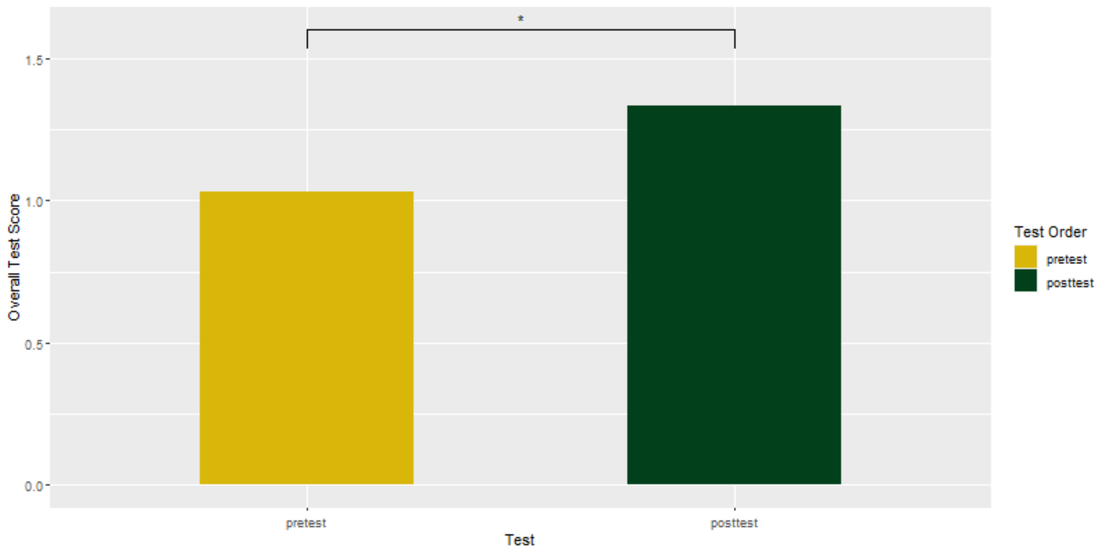


Figure 5.6 – H1. Children's overall performance in pretest and posttest with combined score from toys and path tests.

On average children performed better on the posttest ($Mdn = 2$) compared to the pretest ($Mdn = 1$). A One-tailed Wilcoxon signed rank test with continuity correction indicated that this difference was statistically significant, $V = 29.5, p-value = 0.03826$. The Overall test score is visualized in Figure 5.6.

H2: Children's First perspective Choice

To test this hypothesis we looked at children's first instruction in their instructor role, during the practice session. This instruction is before children receive any feedback or become aware of the robot's perspective taking abilities. As mentioned before, children start as the instructor for both conditions in practice and session 1, hence we can combine the participant data from both conditions to analyze this hypothesis. We have annotated children's instruction based on the frame of reference they used into *egocentric*, *egocentric-explicit*, *addressee*, and *addressee-explicit*. For simplicity, whenever we refer to egocentric it means implicit unless it stated otherwise. We noticed that none of the children used explicit utterances in their first instruction as presented in the first bar of the left in Figure 5.7. We used Chi-square goodness of fit to see if they significantly used more egocentric utterances compared to addressee-centric ones. On average children used more egocentric instructions compared to addressee-centric instructions, however, the test showed there is no significant difference between them ($\chi^2 = 3.6667, df = 1, p-value = 0.05551$). While the result rejects our hypothesis about significantly using egocentric utterances, it can be accepted when only considering the implicit utterances.

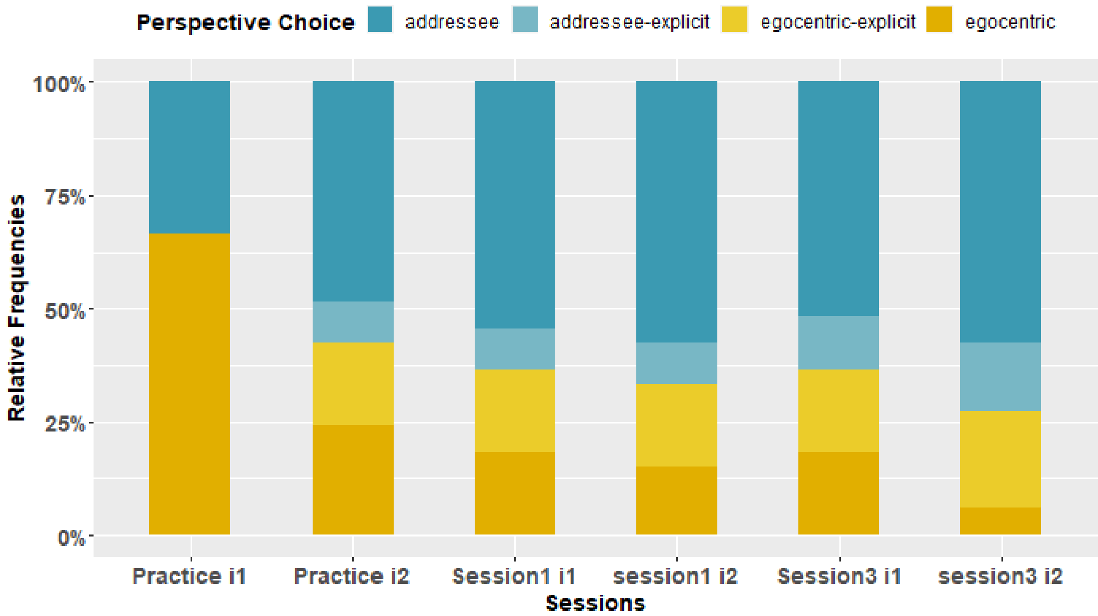


Figure 5.7 – H2. Children's first two instructions in practice, session 1, session 3 categorized based on frame of reference and perspective marking. ("i1", "i2" refer to instruction 1, instruction 2).

Furthermore, it shows how children were more prone to start with implicit egocentric utterances before their information about their counterpart was updated. The rest of the bars in Figure 5.7 shows children's first two instructions in practice, session 1, and session 3, where they were the instructor. The figure shows how children adapted their instructions to accommodate the robot along the whole experience.

H3: Children's Frame of reference Adaptation

In this part, we are interested to observe children's tendency to adapt their perspective to accommodate the robot. To have a better understanding of how children adapt, we have separated our hypothesis for the frame of reference into H3 and perspective marking into H4. To analyze H3, we have combined the implicit and explicit instructions for egocentric and addressee-centric perspectives into one variable. Figure 5.8 shows the percentage of using egocentric versus addressee-centric for the first instruction in practice, session 1, and session 3. As it can be seen in the figure, children made more egocentric instruction in the practice session (66.66%) compared to session 1 and session 3 (36.36%). Considering that the robot had an egocentric perception of the implicit instructions, this shows children had adapted their instructions to the robot's frame of reference. We used Cochran's Q test based on the variable being dichotomous with two levels and mutually exclusive and our need to compare between

Table 5.2 – H3. Pairwise McNemar's Chi-squared test with Bonferroni adjustment.

Group 1	Group 2	p	Adjusted p	p.significant
Practice i1	Session1 i1	0.0213	0.032	*
Practice i1	Session3 i1	0.0129	0.032	*
Session1 i1	Session3 i1	1	1.000	ns

three groups. Cochran's Q test determined that there was a statistically significant difference in the proportion of addressee-centric utterances over time $Q = 10.5263, df = 2, p - value = 0.005179$.

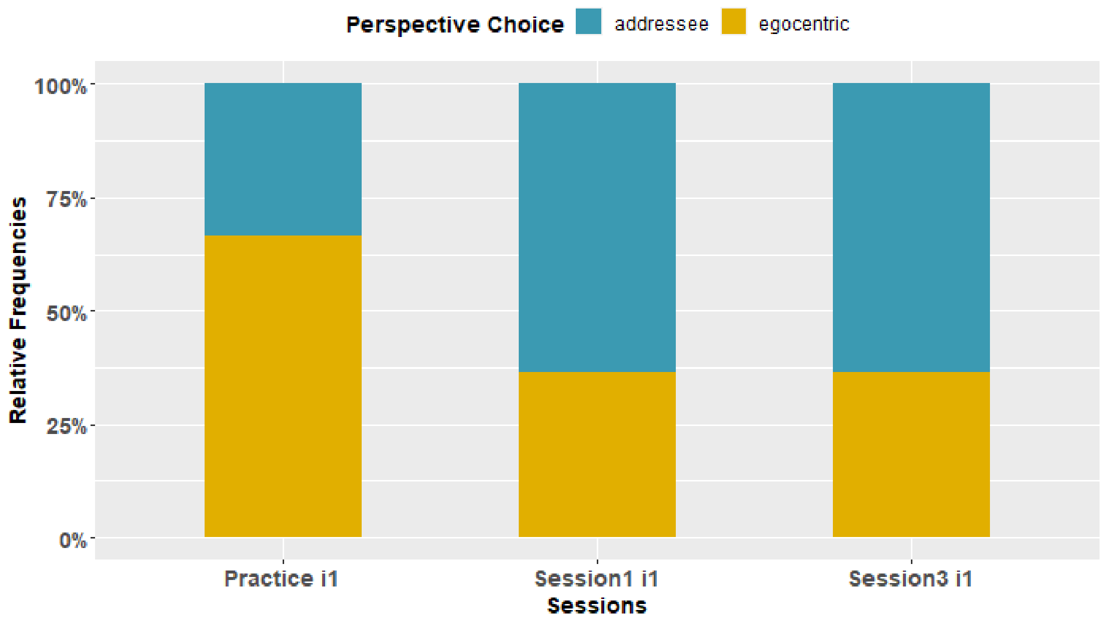


Figure 5.8 – H3. percentage of using egocentric vs. addressee-centric utterances in the first instruction of the practice, sessions 1 and 3.

We ran pairwise McNemar's Chi-squared test with Bonferroni adjustment which showed sessions 1 and 3 have significantly more addressee-centric utterances compared to practice. There is no significant difference between session 1 and session 3 as shown in Table 5.2. This means children made a significant change in their utterances after updating their mental model of the robot and then keep on instructing with that model.

H4: Children's Perspective marking Adaptation

To analyze H4, we have combined the "implicit egocentric" and "implicit addressee-centric" instructions into a variable called *implicit* and "explicit egocentric" and "explicit addressee-centric" instructions into *explicit* variable. Figure 5.9 shows the percentage of using implicit versus explicit utterances for the first instruction in practice, session 1, and session 3. The figure shows that children only used implicit utterances in the first instruction of the practice session, however, they started using explicit utterances in session 1 and session 3.

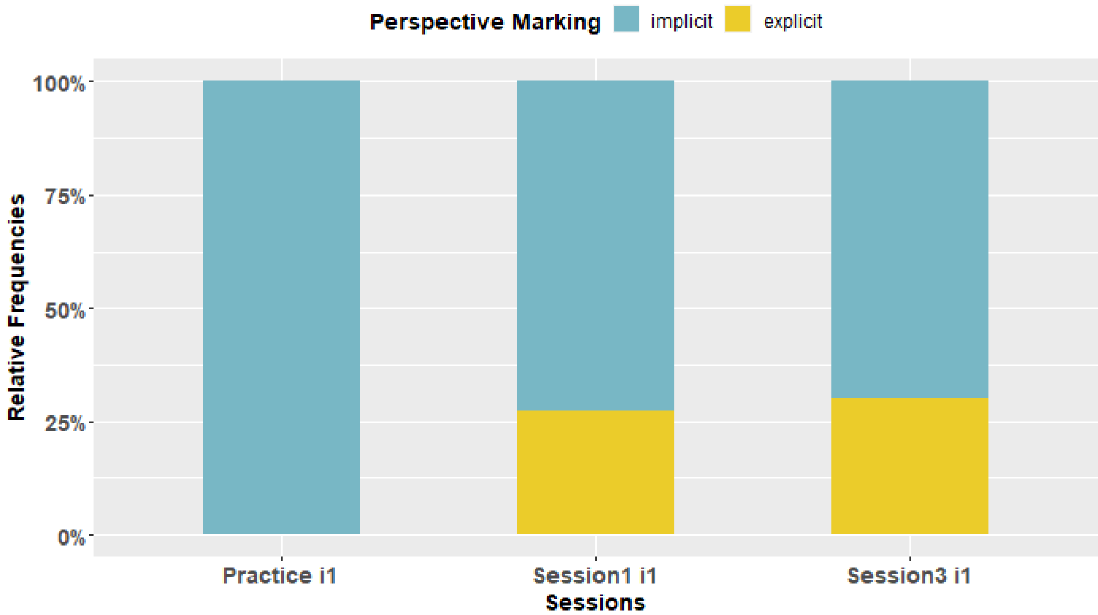


Figure 5.9 – H4. percentage of using implicit vs. explicit utterances in the first instruction of the practice, sessions 1 and 3.

A brief look at the Figure 5.9 shows, despite some children switching to explicit instructions in sessions 1 and 3 (27.28% and 30.31%), implicit instructions are still the dominant utterance. This shows a thought-provoking behaviour on children's side, that they'd rather just switch left and right in their instruction to accommodate the robot rather than explicitly addressing the robot. This result is perhaps due to the addition of extra cognitive processes when the child had to simultaneously switch the frame of reference and mark the perspective. Similar to H3, we had a dichotomous variable with two mutually exclusive levels and we decided to use Cochran's Q test. Cochran's Q test determined that there was a statistically significant difference in the proportion of addressee-centric utterances over time $Q = 16.5455, df = 2, p-value = 0.0002554$. We ran pairwise McNemar's Chi-squared test with Bonferroni adjustment which showed sessions 1 and 3 have significantly more addressee-centric utterances compared to practice. There is no significant difference between session 1 and session 3.

H5: Children's Performance vs. Conditions

In the last hypothesis, we compared children's performance in sessions 2 and 4. As mentioned in Table 5.1, in session 2 the robot instructs the child using an implicit egocentric perspective. Either, the child asks for clarification, in which the robot updates its instruction to an explicit one, or the child makes the move with their perception of the robot perspective. On the other hand, for session 4 the robot's instructions are explicit. In the Ego-Ego condition, the instructions are explicit egocentric and in the Ego-Add condition, they are explicit addressee-centric conditions. On one hand, we expect children to perform better in session 4 considering that the robot's frame of reference is explicitly expressed. On the other hand, we want to know if the switch of the robot's perspective from egocentric to addressee-centric will help children's performance, as they don't need to take the robot's perspective or confuse them as they have to update their mental model. Figure 5.10 shows children's performance in sessions 2 and 4 for Ego-Ego and Ego-Add conditions. The figure shows on average children performed better in session 4 compared to session 2. Considering the dichotomous and skewed data, we ran a McNemar Chi-squared test between sessions 4 and 2 for both conditions. For Ego-Ego condition McNemar's Chi-squared test with continuity correction showed ($\chi^2 = 4.9231, df = 1, p-value = 0.0265$).

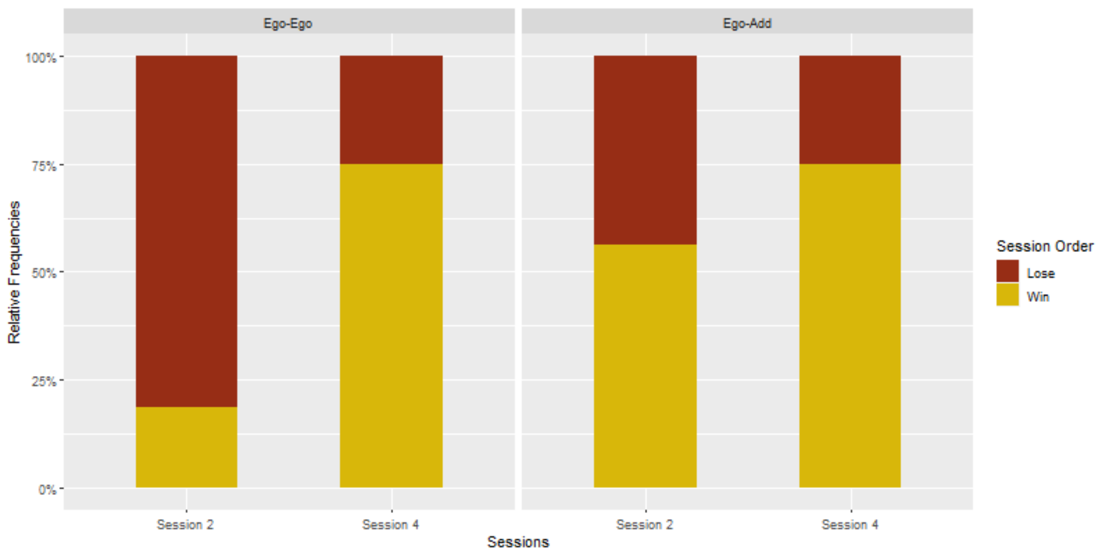


Figure 5.10 – H5. children's performance in session 2 and 4 for Ego-Ego and Ego-Add conditions.

5.5 Discussions

The analyses showed that children's first choice of perspective was *implicit egocentric* and they were able to correct their egocentric perspective to accommodate the robot after realizing the robot was egocentric. Studies support that not only children but also adults tend to have automatic moments of egocentric perspective, however, adults tend to correct immediately (Epley et al., 2004). The implications of children changing their perspective are highly valuable to us, particularly their choice of perspective when adapting to the robot's perspective. The robot's switching between frames of reference showed us that when children create a mental model of the robot they tend to stick to that model. As a result, changing the frame of reference in the middle of the interaction for no particular reason or necessity only has negative effects on children and can hinder the flow of the activity. This can be observed from the way children in the Ego-Add condition did not significantly improve in session 4 compared to session 2. One of the behaviours observed from children in the Ego-Add condition was that some children who successfully completed session 2 failed in session 4, as they were rather confused by the robot's change of perspective.

5.5.1 Limitations of the Study

Before we started this experiment, we had different experimental designs in mind. Considering the exploratory nature of this study, testing various designs and conditions in order would result in a better understanding of children's behaviour and a more diverse database for developing a comprehensive model. However, due to limitations on running experiments in schools, for our first experiment, we designed a more generally informative design. We plan to explore different designs in our future work that can inform and improve our cognitive model.

Robot not Providing Feedback in the Instructor Role
In this Experiment, robot asks for the child's feedback after every move it makes. In the practice or the first session of the game, when both the child and the robot maintain an egocentric perception, this feedback gives the child a chance to inform the robot of the wrong move. However, the robot's movement is correct but its perspective is different. Thus, the robot gives the child a hint of having an egocentric perspective by saying *"But I moved the yellow circles to my right"* in response to the child's instruction *"Move the yellow circles to the right"*. Unfortunately, during the design of our platform, we didn't consider a feedback mechanism for the robot, in case the child asks for the robot's feedback after their move. Due to the autonomous nature of our interaction design, we could not prompt feedback when the experiment was running. Hence, we decided not to add the robot feedback to the interaction after encountering the first child asking for feedback, to keep the experimental condition uniform for all participants. Only 6 children out of 33 participants (18%) asked for the robot feedback after their first move in the second session. Fortunately, this lack of feedback did not affect the experiment result, as most children who encountered a mirrored goal after finishing their first robot-instruct session, realized the robot was egocentric in giving the instructions.

5.5.2 Future Developments

We expect to use the result from this study to equip the robot with a model that can accommodate children's abilities or challenge them depending on the goal of the interaction. The next step for future development of the robot is to incorporate the model in the robot and evaluate how it performs in accommodating children's perspectives or perceiving them accurately. The platform can be used to challenge children to take different perspectives using more complicated and practical activities, such as the child and the robot collaborating to build something. Another future development for this study is to integrate the robot's behavioural model with affective computing models. Then investigate how the robot's cognitive-affective states, such as the robot showing frustration, affects children's perspective taking adaptation and perception of the robot (Yadollahi et al., 2019).

5.6 Conclusions

This chapter introduced a platform that studies children's perspective choice and perspective adaptation to the robot's throughout the interaction. The components used for studying children's behaviour were the frame of reference and perspective marking. The interaction consisted of several sessions, with the child and robot taking turns to instruct each other and move objects around. In one condition, the robot was egocentric in all the sessions, and in the other condition, the robot switched to the child's perspective in the last session.

It was observed that a considerable number of children, after realizing the robot's egocentric perspective did not switch to explicit expressions (with possessive adjectives), instead, they switched to an implicit addressee-centric perspective. This was particularly surprising as children were expected to switch their perspective to explicit instructions after the robot clarified to them that it was egocentric. The robot only clarified that it moved the object to its left and right if the child gave it negative feedback. This behaviour hinted at the fact that when children assigned a perspective model to the robot, they tend to keep making decisions using that model despite slight changes in the interaction. As a result, the robot's abrupt switch in perspective within one interaction and with no specific reason was not beneficial in keeping the interaction transparent and it confused children. The study provided a set of measures for keeping the interaction with children natural, such as switch between frames of reference only when it's necessary and being more explicit does not automatically make the interaction more understandable. The outcome of this chapter will be used in developing the cognitive model of perspective taking presented later in chapter 7.

6 Cozmo Maze: An Exploratory Study

In this chapter, we approach spatial perspective taking from a different angle. We study and analyze children's performance while guiding a robot through a maze, where they constantly need to take the robot's perspective in different angular disparities. First, we describe the goals of the study and the expected contribution. Then, we discuss the research questions that inspired the design of the experiment. This will be followed by a description of the activity, interaction, and platform. Later in the chapter, the details of the user study including the study design, conditions, and analyses of the results are presented. At the end of the chapter, we will discuss the general finding of the study and how it contributes to the rest of the thesis.

6.1 Scope and Research Goals

6.1.1 Motivation and Contribution

The focus of this chapter is twofold, evaluating children's spatial perspective taking abilities and assessing the potential of the designed interaction to practice perspective taking. The main difference between this study and the one in chapter 5 is the spatial positioning between the child and the robot. In chapter 5, the robot was seated in front of the child which required a 180° degree mental rotation or switching left and right to take the robot's perspective. In this chapter, the child and the robot's spatial positioning and consequently angular disparity are dynamic and change when the robot is moving in the maze. To reduce the cognitive load of the task, we have restricted the spatial disparity between the child and the robot to only 0°, 90°, 180°, and −90° degrees. To evaluate children's initial perspective taking abilities, we have used three different pretests and posttests. The tests are designed to evaluate their understanding and abilities concerning mental rotation and level-2 spatial perspective taking. Moreover, we allow children to play two games, then we measure their performance and evaluate their learning during the game. In the end, we test children with posttests to see if the games make a difference in their performance of taking other's perspectives.

6.1.2 Research Questions and Hypotheses

In this study, we have designed a robotic approach to improve the acquisition of spatial perspective taking skills in elementary school children using a game played with the Cozmo robot. To maximize our potential contribution to this learning process, we have developed a game inspired by racing games and remote-controlled cars with the difference that instead of controlling a car, children guide the Cozmo robot through a maze. In our first research question, we were curious to see if playing such an activity would improve children's performance on related tasks. We have used the same two tests that were introduced in Chapter 5 and further added a new one on mental rotation tasks.

- **Research Question 1:** Does taking the robot's perspective while guiding it boost children's performance on tasks involving perspective taking?
  Hypothesis 1: children show better performance in the posttests compared to the pretests in taking the other character's perspective.

While developing the activity, we were also thinking about the mental and cognitive processes involved in taking the robot's perspective. We wanted to know if showing children a video stream that showed them the robot's point of view, improved their performance in guiding the robot or not. The second hypothesis evaluates the video condition. It tries to see how the existence of the video stream helps children to understand the robot's perspective better. Therefore, the map is updated with some landmarks to increase the information flow from the robot's perception of left and right. The video partially recreates what people see when they play racing games from inside a car.

- **Research Question 2:** Does having a video stream of a robot's point of view can help children to guide the robot better?
  Hypothesis 2: Children make fewer mistakes in the "with video" condition compared to the "without video" condition.

Considering the different directions of the robot during the game, we were also interested to see if the angular disparity between the child and robot made a difference in children's performance. Research over the decade has shown that as the angular disparity between the perspective taker and the target's viewpoint increases, the speed and the accuracy in a spatial perspective task decreases (Kozhevnikov & Hegarty, 2001; Zacks & Michelon, 2005). In the initial portion of the game, we expect children to make extra effort to take the robot's perspective as the angular disparity between them and the robot increases. However, as they play more, we expect them to become more comfortable with taking the robot's perspective and make fewer mistakes.

- **Research Question 3:** How much the discrepancy between the robot's and the child's direction affect children's performance?

Hypothesis 3: Children make more mistakes as the angular disparity between theirs and the robot's perspective increases.
Hypothesis 4: Children make fewer mistakes in the second game compared to the first game, despite starting the second game in 180° degree condition.

6.2 Development of the Cozmo Maze Platform

If we consider a child playing with a remote-controlled car, when the child and the car are in the same direction; meaning that the car moves away from then (when going forward), controlling seems effortless. However, when the child and the car are in opposite directions; meaning that the car is moving towards them (when going forward), controlling suddenly become more complicated and they are more prone to have an accident. Inspired by remote-controlled cars, we have designed a game with the Cozmo robot, which requires the use of spatial perspective taking and similar scenarios for completing the game. In this section, we give a detailed description of the elements of the game and the perspective taking aspect of the design. Recent studies have looked at the design of robots behaviour such as Cozmo as a sidekick character (Luria, 2018) where it also inspired its use in this experiment.

6.2.1 Game Development

We designed a map based on three fundaments, the inclusion of the robot as a game character, a perspective-taking application, and an educational goal. Comprehensive School Mathematics Program (CSMP) [I] and other school practices inspired the initial idea of the map, where the child was supposed to move from a starting point to a designated goal by moving on directional paths from one node to another. Our study focuses on a scenario with the core idea of practising perspective taking and simultaneously understanding how the robot's intelligence can be developed for games in educational contexts.

The Cozmo robot[II] (Appendix A) is the main character of the game. We created a simple backstory that we presented to children: "Cozmo needs to collect stars to survive in a field. To collect the stars, the robot needs to take risks, get out of the safe zone (represented by green background), and move along the grey roads. If the robot goes to the red nodes within the danger zone (represented by black background), it loses the game." The nodes and the paths are designed to restrict the angular disparities between the robot and child to only a few angles. In the game, the robot only moves after receiving the child's instruction through the controller. It moves from one node to another and awaits the next instruction. The robot is also equipped with simple affective behaviour from Cozmo's emotion library to express its emotion upon winning or losing the game.

The controller consists of three buttons for moving to the *front*, *left*, and *right* directions, as

[I] http://stern.buffalostate.edu/CSMPProgram/Primary%20Disk/Start.html
[II] https://anki.com/en-us/cozmo.html

Figure 6.1 – Control Buttons and the screen to present the video feed.

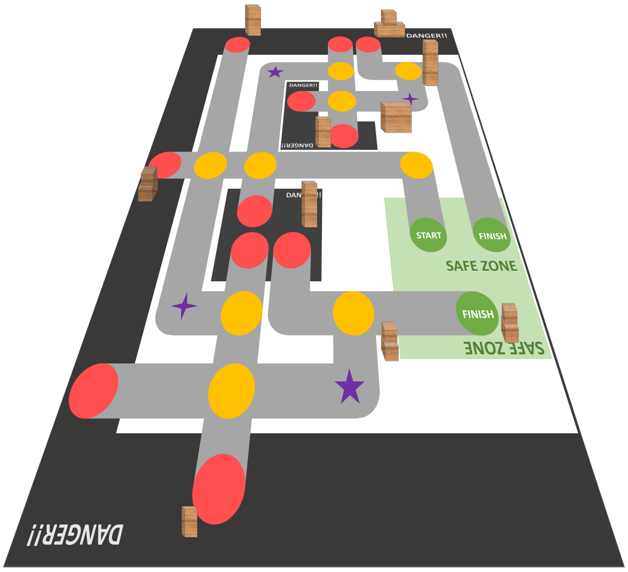


Figure 6.2 – Map of the game including the wooden blocks.

shown in Figure 6.1. It is always positioned in front of the child, regardless of which side of the map they are seated on. Children are informed that by pressing the button they can guide the robot and move it around the maze. Probably the most important aspect of the controller is that its directions correspond to the robot's perspective rather than the child's. This piece of information is not presented to the child upon starting the game. It's up to the child to discover this during the game, change their perspective to match the robot's, and choose the correct moving sequence. We deliberately design the controller to function with the robot's perspective. This way, children experience the possibility of seeing the maze from the robot's perspective. If the child instructs the robot to move in a direction without any available path, the robot acts confused and irritated for hitting a blocked road and waits for new instruction.

The video feed was added to the game as an experimental condition, to recreate the experience of having the first-person view in the game. We wanted to know if adding a first-person point of view could help children in guiding the robot or not. Children received the video feed using a tablet in front of them that showed the robot's view in real-time. After discovering the limitations of the robot's camera, we decided to add some landmarks to the game to help children have a better grasp of the robot's perspective during the game. The wooden landmarks positioned in the map are presented in Figure 6.2. Each landmark is carefully positioned in the robot's line of sight, either inclined to the left or right. It aims to assist children to have a better perception of the robot's perspective of the left and right.

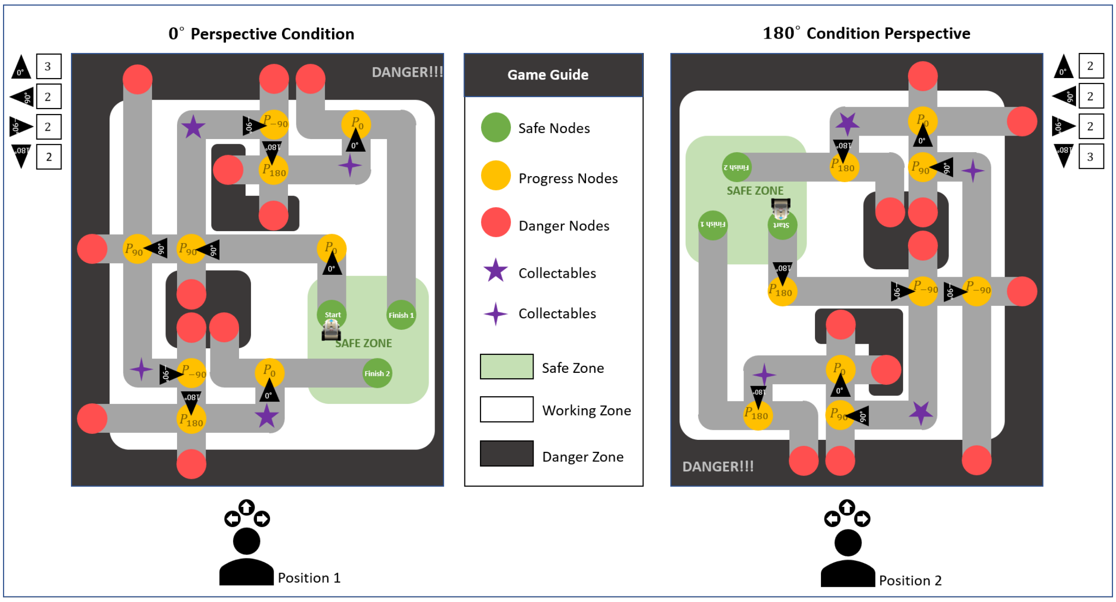


Figure 6.3 – The map with the game guide, plus outlining children's spatial position for the two games they played. The figure on the left shows 0° perspective condition (at the start point, robot and the child's perspective are aligned), and the figure on the right shows 180° perspective condition (at the start point robot and the child's perspective are 180° misaligned. The black triangles represent the robot and its rotation angle at each node.

6.2.2 Perspective Taking Task

The first principle in designing the map was creating a "perspective mismatch" between the child and the robot. The maze represented in Figure 6.3 has two finish nodes that are represented with green. To win the game with no mistakes, Finish1 requires a minimum of 5 moves and Finish2 requires a minimum of 6 moves (not counting the move where the robot departs from the start node). The black triangles represent the robot at different rotation angles with respect to the player's position. The sharp corner of the triangle represents the front of the robot. For this experiment, we decided to test four rotation angles of 0°, 90°, −90°, 180°. Furthermore, each progress node (yellow) is denoted with a rotation angle as a reference to the angle of mental rotation needed to take the robot's perspective at that node. For example, consider the right map in Figure 6.3, the first yellow node after the start node is denoted with $P_{180°}$. This means that if the child is seated in position 2, the robot always reaches that node with this rotation angle, which in turn results in 180° angular disparity between the child and the robot. We have deliberately assigned predefined rotation angles to each node as a means to have a better comparison ground for children's performance. Simultaneously, we have tried to give children enough freedom of choice to play while being evaluated, hence they can decide to use Finish1 or Finish2. The win sequence considering no mistake for position 1 and finish 1 is $P_{0°}$-$P_{90°}$-$P_{-90°}$-$P_{180°}$-$P_{0°}$, and for position 1 and finish 2 is $P_{0°}$-$P_{90°}$-$P_{90°}$-$P_{-90°}$-$P_{180°}$-$P_{0°}$. The win sequence considering no mistake for position 2 and finish 1 is $P_{180°}$-$P_{-90°}$-$P_{90°}$-$P_{0°}$-$P_{180°}$, and for position 2 and finish 2 is $P_{180°}$-$P_{-90°}$-$P_{-90°}$-$P_{90°}$-$P_{0°}$-$P_{180°}$.

6.2.3 Technical Development

The Cozmo robot was programmed using Cozmo SDK [III] and Python programming language with a ROS wrapper. The controller buttons are connected to Arduino boards which publish messages using ROS topics every time children press the buttons. The map shown in Figure 6.3 was simplified into nodes and branches and coded into the program. When the robot starts the game after receiving controller input, the program retrieves and updates the robot's current direction, node, and the distance to travel to reach the next node. The robot plays the game autonomously and the whole interaction is fully autonomous. When the child loses a game by reaching any of the red nodes, the game restarts from the last yellow node. Due to inaccuracies in the robot's positioning system, we place the robot in the last node, restart the game, and use a visual controller to inform the robot about the update in its current position. The detailed schematic of the experiment is presented in Figure 6.4.

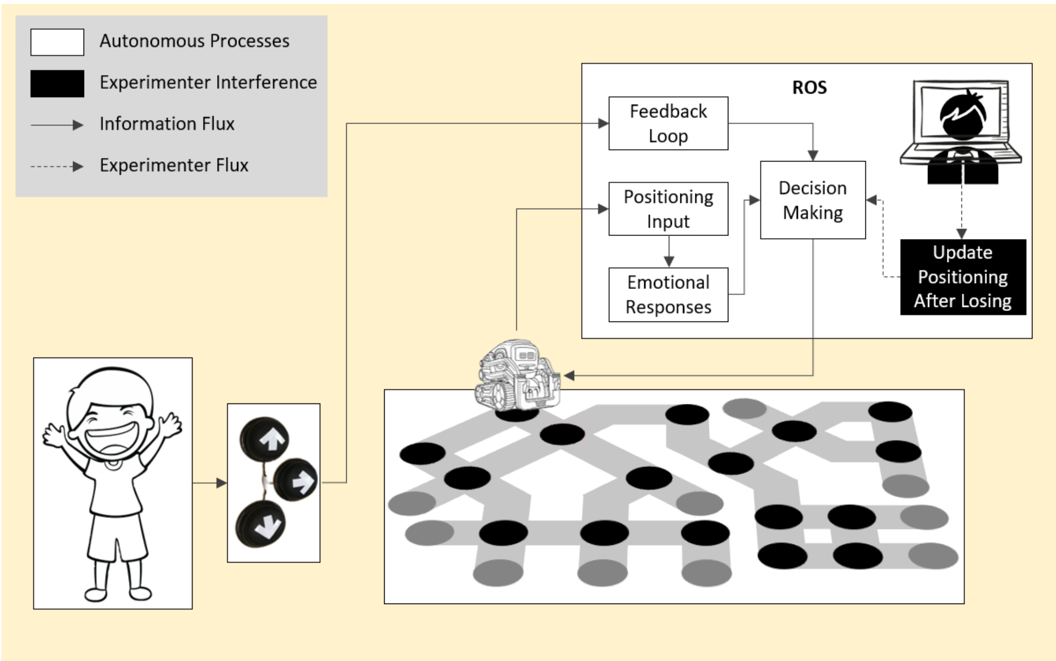


Figure 6.4 – Schematic of the study 3.

6.2.4 Evaluations Methods

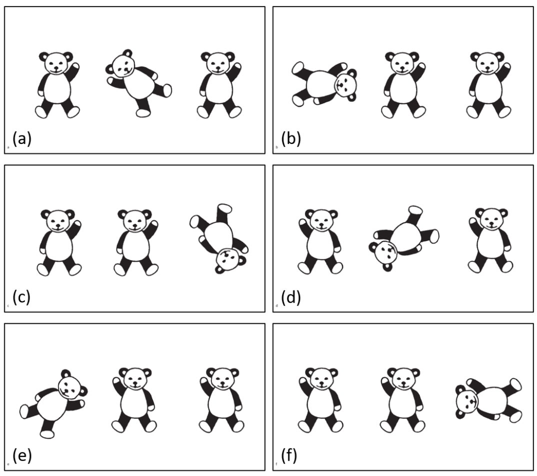
For evaluation, we have used three tests with each test trying to evaluate different skills associated with children's perspective taking abilities. The first two tests are "left/right test or toys test" and "test of direction sense or path test". Both of these tests were previously described in chapter 5 and are available in Appendix C.1 and C.2, respectively. The third test is called the "mental rotation test or panda test" and it is described below.

[III] https://developer.anki.com/

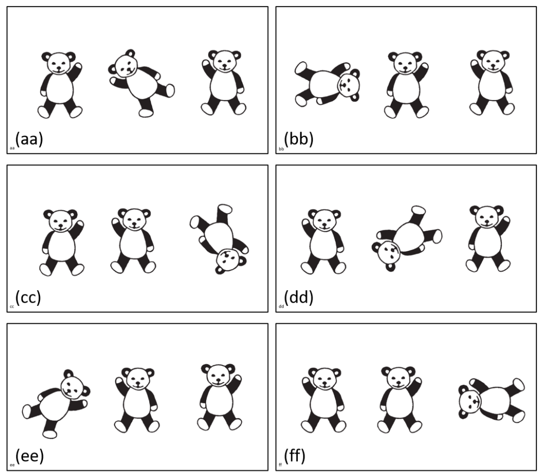
Mental Rotation Test or Panda Test

Several psychology studies specify the role of mental rotation in level 2 spatial perspective taking (Janczyk, 2013; Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Kessler & Wang, 2012). To see if playing this game has a positive impact on children's mental rotation abilities, we have used the test shown in Figures 6.5a and 6.5b inspired from a study by Perrucci et al. (Perrucci et al., 2008). This test focuses on children's object rotation skills which, based on some studies, has different cognitive processes compared to mental self-rotation used in perspective taking (Kessler & Rutherford, 2010; Zacks & Michelon, 2005). However, it also includes specifying the panda's tribe which requires perspective taking to find them from the images. The test starts with instructions (see Appendix C.3) that the experimenter explains to the child before taking the test.

The pretest and posttest sets are different in one aspect. In the pretests, the panda that is from a different tribe is always the rotated one, which means just by looking at two straight pandas the child can conclude that the rotated panda is the different one. However, they also need to specify which tribe the panda belongs to, which means they still need to recognize the left and right or use mental rotation. On the other hand, in the posttest, the one straight panda and one rotated panda are from the same tribe which means children need to use mental rotation to find the different panda, and then specify the panda's tribe.



(a) Mental rotation or panda pretest (one selected per participant)



(b) Mental rotation or panda posttest (one selected per participant)

Figure 6.5 – Examples of mental rotation pretests and posttests: 3 sequence of mental rotation tests were used for each child, the pretests were chosen from batch (a) in the left and the posttests were chosen from batch (b) on the right side of the figure.

6.3 Study 3 - Children's Performance in Cozmo Maze

6.3.1 Experimental Design

The game included two independent variables, video condition with two levels; "with video" and "without video" and robot's direction or the angular disparity with three levels; "0°", "90°", and "180°". The dependent variables include the children's responses to the test and performance in the game. The overall performance is evaluated based on the total number of wrong moves which are the moves resulting in falling off the maze. We also investigated the time children took to make a move and the techniques they used to move e.g. rotating their body. Those analyses are not presented here. The whole duration of the experiment (See Figure B.4), was around 20 to 30 minutes. The interaction started with the experimenters introducing themselves to the child and let the child sit in front of the buttons and the video display. Then children answered three sets of pretests, played two rounds of the game, answered the posttests and a godspeed questionnaire. There was no limit in how many times the game restarted after the child lost the game, which meant all children eventually finished both games. In the first set children played the game in *0° PT* condition (position1) and in the second set they played it in *180° PT* condition (position2). Both conditions are presented in Figure 6.3.

Table 6.1 – Between-subject experimental design with two conditions, "with video" and "without video".

Conditions	Pretests	Game 1	Game 2	Posttests
With Video		0° PT	180° PT	
Without Video		0° PT	180° PT	

6.3.2 Hypotheses

As mentioned in section 6.1.2, we have made four hypotheses to evaluate our game design, experimental condition, and children's learning gain. The hypotheses are:

H1: Children show better performance in the posttests compared to the pretests in taking the other character's perspective.

H2: Children make fewer mistakes in the "with video" condition compared to the "without video" condition.

H3: Children make more mistakes as the angular disparity between theirs and the robot's perspective increases.

H4: Children make fewer mistakes in the second game compared to the first game, despite starting the second game in 180° degree condition.

6.3.3 Participants

The study involved 22 typically developing children between the ages of 8 and 9 (15F, 7M) selected from third and fourth grade in elementary school. The experiment had a between-subject design consisting of two conditions of "with video" and "without video". Children were randomly assigned to each condition, 11 children (9F, 2M) played the game in "with video" condition and 11 (6F, 5M) in "without video" condition.

6.3.4 Results

H1: Children's Performance in Pretest and Posttest

The performance on the pretests and posttests were compared to see if children significantly improved after the games. We found no significant differences between the answers of the children to the panda test ($Z = 21; p = .857$); and no significant differences to the answers to the toys test (Mcnemar p=1). For analyzing the toys and path tests, similar to the previous study, we combined the result of both tests and looked at children's overall performance in the tests. First, we analyzed the overall performance for normality using the Shapiro-Wilk test. A Shapiro-Wilk test for pretest and posttest showed a significant departure from normality, $W(32) = 0.64022, p = 3.716e-06$ and $W(32) = 0.52227, p = 2.142e-07$, respectively. Considering that we expected children's performance to improve from pretest to posttest and the data is skewed, we perform a one-tailed Wilcoxon Test. On average children performed better on the posttest ($Mdn = 2$) compared to the pretest ($Mdn = 1$). A One-tailed Wilcoxon signed rank test with continuity correction indicated that this was not statistically significant, $V = 16.5, p - value = 0.05998$. The Overall test score is visualized in Figure 6.6.

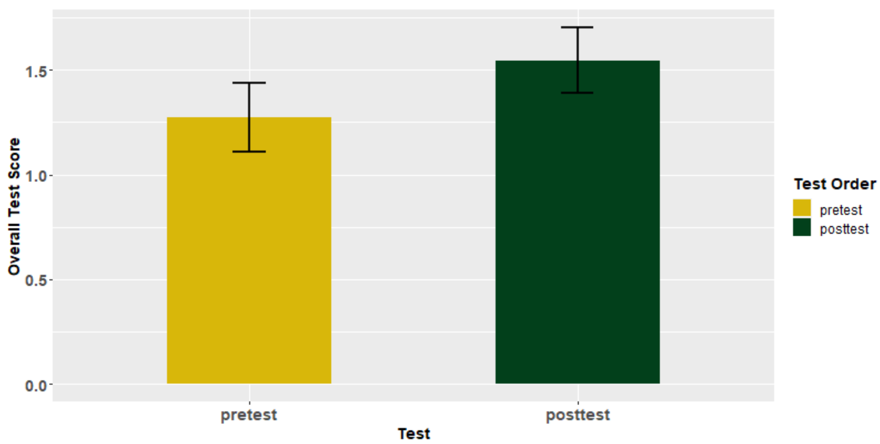


Figure 6.6 – H1. Children's overall performance in pretest and posttest with combined score from toys and path tests.

H2: Children's Performance vs. Video Conditions

Before starting the analyses, we had to process the data. Originally, we considered every move that didn't end up in the danger zone or a dead end as a correct move. However, when adjusted for children's learning, we realised the moves made immediately after an incorrect move should be considered as void. These moves were a response to an incorrect move and were not comparable to a move that was correct from the start. Consequently, we defined a new measure called adjusted correct move and used that in our analyses. Figure 6.7 shows that children in the without video condition make more mistakes than with video condition. We used the distribution of the overall number of wrong moves for both games between the video conditions and ran a Wilcoxon rank test with continuity correction $V = 70, p - value = 0.2658$. The test showed no significant difference between the two conditions, hence the hypothesis is rejected. As a result, for further analyses, we collapsed the data for both conditions.

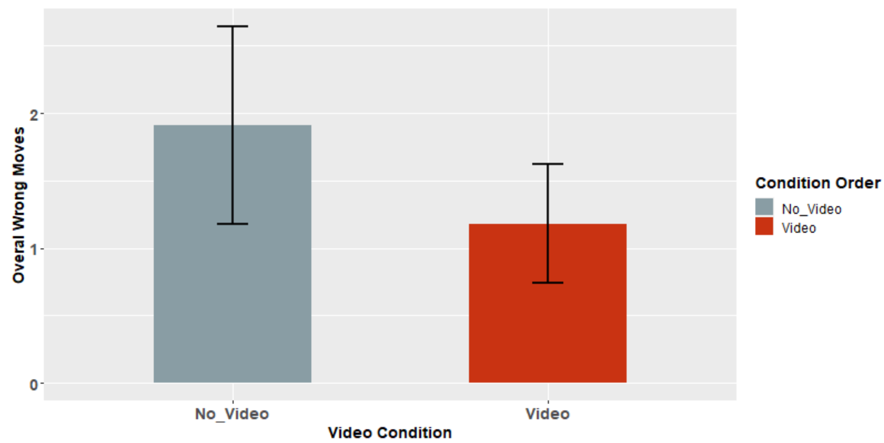


Figure 6.7 – H2. Percentage of the wrong moves based on angular disparity for 0°, 90°, and 180° rotation angles.

H3: Children's Performance vs. Rotation Angles

We tested the percentage of the wrong moves per angle (0°, 90°, or and 180°) with a Friedman's chi-square and found statistically significant differences ($\chi^2_f(2) = 10.186; p = 0.0061$). with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$ for comparisons between 0° and 180°. The differences between 0° and 90° reported at $p < 0.143$ and 90° and 180° reported at $p < 1$ were not significant. Figure 6.8 demonstrates how children performed when in different angular disparities. Children made almost zero mistakes in 0°. But the percentage of the mistakes they made had increased for 90° and 180° rotation angles. Quantitatively, children made more mistakes at 90° compared to 180°, however, they also encountered more 90° perspective conditions compared to 180°.

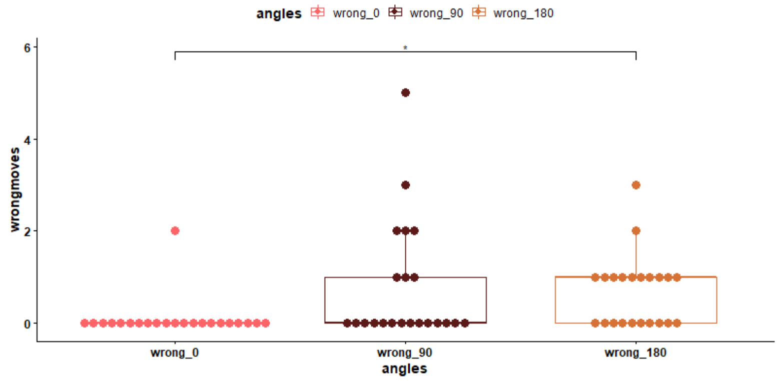


Figure 6.8 – H3. Percentage of the wrong moves based on angular disparity for 0°, 90°, and 180° rotation angles.

H4: Children's Performance vs. the Games

We analysed the distribution of the percentage of correct and incorrect moves for both games and found significant differences ($\chi^2_f(3) = 48.3; p \leq .001$). Pairwise comparisons with a Bonferroni adjustment show that children made significantly more correct moves than incorrect moves on both games (Game 1: $\chi^2_f(3) = 1.5; p \leq .001$; Game 2: $\chi^2_f(3) = 2.1; p \leq .001$). Children's performance between the two games is presented in Figure 6.9. Furthermore, children were overall faster on game 2 ($M = 162; SEM = 0.030$) than on game 1 ($M = 143; SEM = 0.034$) but this difference is not significant. Figure 6.9 shows that children performed better in the second game compared to the first game. This can be due to various reasons, such as children's initial profile, learning from the mistakes, learning from practice, or learning how to take the robot's spatial perspective. While we cannot decide which reason was the contributing factor to their improvement, we can still postulate that similar games with robots, maybe with more cognitively demanding elements and more challenging scenarios can be used in fostering skills such as perspective taking.

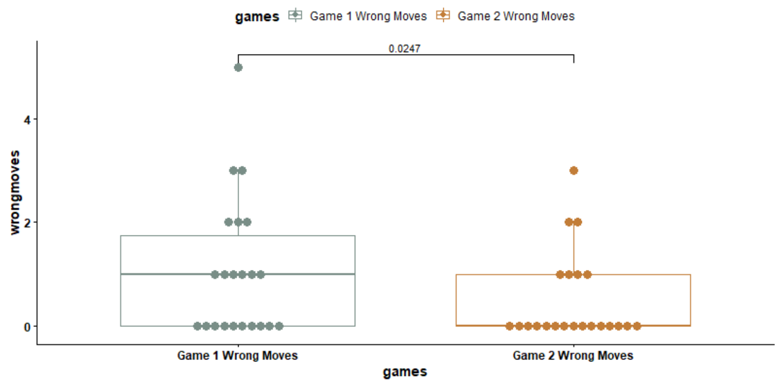


Figure 6.9 – H4. The comparison of percentages of correct and incorrect moves between game 1 and game 2.

6.4 Discussions

We did not observe any significant difference in children's performance between with and without video conditions. During the experiment, we observed that some children in the video condition paid almost no attention to the video feed. They were more interested in looking at the robot's movements than through the eyes of the robot. Except for the times that the robot was looking at them, in which case children could see themselves through the eyes of the robot and that was interesting to them. Nevertheless, this still did not guarantee that children used the video feed to make any perspective taking decision or adjustment. A similar phenomenon was previously studied by L. J. Wood et al. with ASD children. In their study, they used the Kaspar robot's camera and a screen to show children the robot's perspective e.g. what it sees through its eyes (L. J. Wood et al., 2019).

The result for the angular disparity is aligned with the previous research in perspective taking. As seen in Figure 6.8, children make more mistakes as the angular disparity between their perspective and the robot's perspective increases. If we compare their performance between the two games, we can see that children perform better in the second game compared to the first game, which is a sign of learning. During the experiment, we noticed that children had a different approach in dealing with angular disparity. The differences in the children's profiles along with the small sample size were the two contributing factors for not observing significant differences between the angles and other conditions.

Based on our observations during the experiment, we could divide children into three profiles, the ones who mentally rotated themselves or the robot, the ones who physically rotated themselves, and the ones who used trial and error to guide the robot. This last group of children did not show improvement in the posttest. Their performance was an indicator of their lack of grasping the main concept of the game and how to take the robot's perspective. Children who managed to make zero to no mistakes usually realized how the controller functioned from the robot's point of view in their second move in the first game. As shown in Figure 6.3(left), when the child reached the second yellow node, the robot had a 90° rotation. A group of children pressed the front button expecting the robot to move upward. However, when the robot moved to the next node on the left, some children immediately expressed that they have discovered the buttons correlate to the robot's perspective. This group effortlessly took the robot's perspective from the start and without needing any hint. Some children were still confused about the robot's weird move, but they slowly understood how to take the robot's perspective after making a few more mistakes. On the other hand, the last group of children were the ones who played based on trial and error.

6.4.1 Limitations

Regarding the children's lack of interest in the video feed, we thought about the limitations of this feature. First was Cozmo's restricted field of view in an almost monochromatic shade. As a result, even after adding the wooden blocks as landmarks, it was still difficult for children to understand what the robot is looking at in the video. We also noticed that except for children with an adequate command of perspective taking and mental rotation, most children were physically self-rotating to find the correct move sequence from the robot's perspective. We assume children probably had to use different cognitive processes to understand the robot's perspective through the video feed. Since using two different processes was more complicated and cognitively demanding, they opted for ignoring the video. Unfortunately, at the moment, we have no evidence to prove or disprove any of these assumptions.

On the other hand, we did not observe any performance improvement in the mental rotation test. We are aware of two limitations that might have caused this result. The first limitation was due to the difference between the first and the second test which made the second test a bit harder than the first one. The second limitation can be explained by the use of different processes between object rotation and spatial perspective taking. Research in psychology suggests that in spatial perspective taking tasks, people engage in mental rotation of the self rather than Object Rotation (OR) (Kessler & Rutherford, 2010; Zacks & Michelon, 2005). This means, to take other's perspectives, humans prefer to mentally rotate themselves rather than rotating the other person or object (Kozhevnikov & Hegarty, 2001; Zacks & Michelon, 2005). It has also been shown that self-rotation involves a different cognitive process compared to object rotation (Janczyk, 2013; Kessler & Wang, 2012).

As for the performance in the toys test, the majority of the children regardless of their performance in the game responded to this test correctly. We have used this test before in the object game presented in chapter 5 and with a similar target age group. Compared to this study, children in our previous study performed worse in the pretest and better in the posttest. We suspect that children's better performance in the current study was caused by taking the toys test right after the panda test. In the panda test, the panda tribes are introduced by the experimenter to the child. The tribes might have provided children with a hint about the switch of left and right for a character in the picture. Regarding the path test, while children's overall performance was not significantly different, most children who failed the pretest performed better in the posttest. We consider their improvement as a positive sign about them learning how to guide a character by taking their perspective. Overall the tests show that they have learned someone in front of them has a different perspective and spatial relation with objects compared to them. Similar to the main game, we have also observed some children were physically rotating their bodies to find the correct answer for guiding the animals.

Overall this study's biggest limitation was the small number of participants. Categorizing children based on their behavioural profiles and having a larger sample size could have contributed to a better evaluation of the system and children's performance.

6.4.2 Future Developments

Further developments include: improving the game design and adding a more dynamic interaction between the child and the robot. This task has the potential of being combined with other STEM subjects such as mathematics or used to improve affective computing models of behaviours for social robots. In the current design of the task, the robot was being guided by the child, meaning that it was not equipped with a complicated model, but rather a simple program to follow the child's instructions. In future, we would like to expand the interaction to include a robot equipped with an agent architecture that makes more complicated decisions and takes other's perspectives. To create such scenarios, we plan to develop a two-player version of the maze game. The new version includes a robot with agent architecture and a perspective taking model. The proposed development opens new opportunities to study the context of interaction, for example, competitive versus cooperative settings and will be discussed in more detail in the next two chapters.

6.5 Conclusions

This chapter introduced a gamified child-robot interaction scenario that studies children spatial perspective taking abilities through guiding the Cozmo robot in a maze. The interaction was designed to be game-based and through non-verbal channels e.g. pressing push buttons to guide the robot. Furthermore, the activity included a dynamic angular disparity between the child and robot, which required children to constantly switch their perspective to the robot's perspective.

To evaluate children's perspective taking abilities, they responded to three sets of pretests and posttests and their performance in guiding the robot in each angular disparity was evaluated. All children managed to successfully guide the robot to its destination, with some quickly adapting to switching to the robot's perspective, some taking time to switch their perspective, and few having problems with changing their perspective.

The experiment contributed to evaluating the platform's potential to practice perspective taking and children's perspective taking abilities in performing such tasks. Furthermore, the typical errors children made while taking the robot's perspective have inspired the development of a robot with limited perspective taking abilities for a future study that will be described in chapter 8. Overall, the findings of this study contribute to the design of a platform that will be used to evaluate the cognitive model of perspective taking.

Part IV

Modelling Perspective Taking and Evaluation

7 CogPeT: A Cognitive Model of Perspective Taking

In this chapter, we discuss the development of a model that can be incorporated into robots to improve interaction. Despite the importance of such topics, limited studies have considered and analyzed the importance of perspective taking in the interaction, specifically with children. Inspired by the previous studies presented in this thesis, we are proposing a model that generates behaviours and decisions for robots or virtual agents in interactions. The model particularly focuses on spatial perspective taking and its goal is to be more adaptive to the abilities of the human counterpart in child-robot interaction. The chapter provides a brief theoretical background ranging from developmental psychology to agent modelling. Then it discusses the cognitive model of perspective taking and provides details about the initial considerations, mechanisms, and integration with typical agent processes. In the last portion of the chapter, the adapted processes applied to the updated version of the Cozmo maze platform are discussed.

7.1 Theoretical Background

During the process of designing perspective taking mechanisms, it is fundamental to support modelling decisions based on the background provided by developmental psychology. There is an extensive body of literature and research on perspective taking that can inform our model and processes. To develop the conceptualization aspect of the model, we used insights from the theoretical background on perspective taking, recent findings in human-robot interaction, and our understanding of children's perspective taking abilities through the findings of our previous studies. We begin by presenting the elements that are deemed beneficial to perspective taking interaction through literature and our previous studies. This will be followed by a discussion exploring how the dynamics between a child and a robot have inspired the core of our perspective taking model for robots and agents.

7.1.1 Perspective Taking in Developmental Psychology

Over the past century, a great deal of research has been performed trying to categorize different types of perspective taking, the emergence of each level based on the age, and underlying mechanism for each perspective taking ability (Flavell et al., 1981; Moll & Tomasello, 2006; Piaget, 2013; Surtees et al., 2013). As such, perspective taking has been categorized differently in literature with some models already discussed and presented in Chapter 2. The child's social interaction both influences and is influenced by their ability to take another person's perspective. Kurdek and Rodgon have proposed three types or dimensions of perspective taking as in perceptual, cognitive, and affective (Kurdek & Rodgon, 1975). According to this categorization, perspective taking may be modelled at three distinct dimensions. Considering that each dimension focuses on different types of viewpoints, we expect to be able to develop the model for one dimension and generalize it to other dimensions by changing the input and output types albeit beyond the scope of this thesis. In this thesis, we are particularly interested in spatial perspective taking and its components. To conceptualize the model so that it works with our target interaction, we started with the three basic components involved in perspective taking tasks as proposed by Surtees et al. (Surtees et al., 2013). These components are:

- **a perspective-taker (Self)**, the person who judges, understands, or takes the other's perspective;
- **a target's (Other) perspective**, that is being judged, understood, or taken (it is commonly referred to another person, but it can also be a directional object, imagined self-perspective, or virtual or embodied entities such as agents or robots);
- **an object or circumstance (Object)**, upon which the perspective is taken.

These components can be used to explain the different perspective taking dimensions; perceptual, cognitive, and affective, and their sub-dimensions; spatial and visual for perceptual dimension; and levels; level-0, level-1, and level-2. By changing the object or circumstance we can change the dimensions and notions of perspective taking. For example, if the Object/Circumstance that Self tried to understand is the spatial relations or viewpoint of the Other with themselves or another object in the environment, they are exercising their spatial perspective taking abilities. On the other hand, if the Object is understanding the emotional experiences of the Other at a certain instance, then Self is exercising their affective perspective taking skills. As mentioned before, currently we are only focusing on a model for spatial sub-dimension, while we consider the possibilities and notions of expanding it to other dimensions and sub-dimensions as a proposal for future directions.

Looking at both sub-dimensions of perceptual perspective taking; spatial and visual, several studies and research have been conducted to understand the similarities and differences between their processes. Generally, the processes of perceptual perspective taking are differentiated on whether the Self needs to mentally rotate themselves into the position of Other or not (Michelon & Zacks, 2006). Within the field of social psychology, spatial perspective taking has received far less interest compared to the other dimensions (Surtees et al., 2013). This is due to the nature of this type of perspective taking which does not necessarily require understanding the other's mental content. Nevertheless, it is a skill well studied to understand its underlying mechanisms and developmental stages (Pickering et al., 2012; Surtees et al., 2012; Tversky & Hard, 2009). One of the main elements related to spatial perspective taking is frames of reference which "allows us to encode spatial information" (Levinson, 2003; Mintz et al., 2004). Different frames of reference in the literature concerning perspective taking are presented in Table 2.1. Besides activating frames of reference for processes that are operated mentally, there is also evidence of adults activating frame of references when making linguistic decisions (Carlson-Radvansky & Irwin, 1993; Carlson-Radvansky & Jiang, 1998). This can correspond to the use of spatial language which in Schober terms "involves spatial positions and movements of objects in the world".

7.1.2 From Developmental Psychology to Agent Modelling

Here, we reduce our assumptions of perspective taking cases to dyadic interactions. We also consider that depending on the task, one or both sides of the interaction get to exercise their perspective taking skill. In this sense, first, we assume that both sides of the interaction are called agents A and B. Now if we consider a task that can be completed when A and B are aware of each other's perspective, we can face the following four situations:

1. **Both A and B try to understand or take each other's perspective:**
   This implies that A needs to consider B's awareness of its perspective within its decision-making algorithm.
2. **Only A tries to understand or take Bs Perspective:**
   This implies that for the task to be completed A needs to constantly adjust its perspective to B's perspective since B is not aware of A's.
3. **Only B tries to understand or take As Perspective:**
   This implies that A makes a decision without considering B's perspective. It is B that is doing all the work.
4. **Neither A or B try to take the other's perspective:**
   This implies both A and B are making decisions without considering the perspective of the other, which usually results in failing the task.

Breaking down the situations that the agents might face when making a decision can help us to understand how the behaviours from developmental psychology can be translated into the agents. For example, we can map some of the scenarios from the object game in Chapter 5 to the situations mentioned above. In the objects game, the robot was designed to be egocentric so we can collect children's behavioural data. Hypothetically, if we consider the robot as agent A and the child as agent B, situations 3 and 4 apply to the interactions that happened in the game where the child was instructing. If the child successfully adapted their perspective to the robot's they were in situation 3 and if the child was implicitly egocentric they were in situation 4. On the other hand, in the sessions with the robot instructing, depending on the experimental condition, any of the four situations can apply. Our final aim with developing this model is to equip the robot in similar scenarios with capabilities to identify which situation they are in to find the best course of action. Furthermore, depending on the interaction goals, the robot can purposefully steer the interaction toward one of the situations mentioned above.

7.2 CogPeT: A Cognitive Model of Perspective Taking

7.2.1 Initial Considerations

To start conceptualizing, we consider the three components mentioned in the previous section that are fundamental to our model. As presented in Figure 7.1, Agent A (left) and Agent B (right) are interacting with each other. In each layer of the interaction, our first step is to identify the three components of Self, Other, and Circumstance. The second step critical to conceptualizing the model is to indicate the flow of the interaction. This is something that was not particularly covered in developmental psychology, but rather in the linguistics literature. When we consider one instance of the interaction, where the two agents interact with each other, the mental processes from both sides need to be addressed separately. In a given instance, when we are describing the mental processes of agent A, the perspective taking components differ from when we describe agent B. Imagine agent A, e.g. the acting agent, is performing an action such as speaking or instructing. When we consider agent A's processes, agent A is in the role of Self (S), agent B is Other (O), and understanding agent B's perspective is part of the Circumstance (C). While agent A is performing the action, agent B, e.g. the perceiving agent, starts to perceive the action in the form of listening or following the instruction. Looking at agent B's processes, agent B is in the role of Self (S), agent A is the Other (O), and understanding agent A's perspective is part of the Circumstance (C). To address all these roles, we need to model each agent's processes separately, first where Agent A "performs an action" and second where Agent B "perceives the action". Considering that in dynamic interactions, there is a back and forth between the two agents, each agent can become both an acting and perceiving agent in the course of interaction. The model represented in Figure 7.1, shows agents A and B and all the associated processes and roles. The simplified cognitive processes presented in the boxes on top of each agent are modelled based on the theoretical background and mechanisms observed in the object game and the Cozmo maze. In the next sections, we describe these mechanisms in more detail.

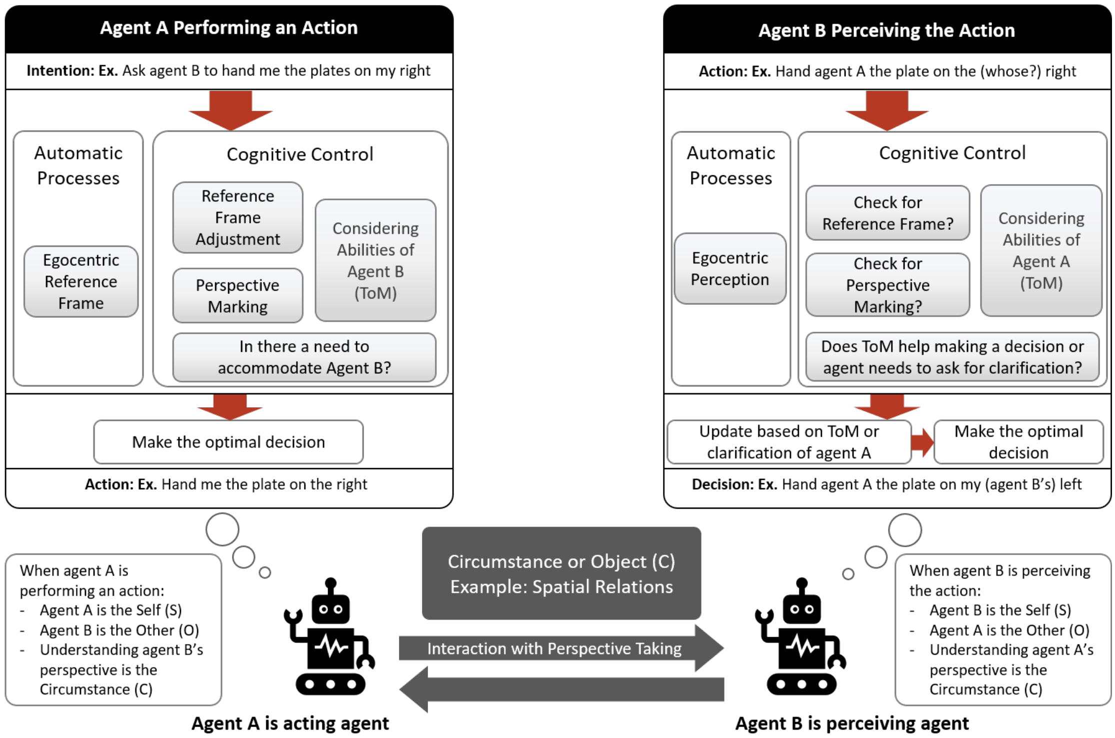


Figure 7.1 – Diagram representing the components of and context for an interaction that includes perspective taking mechanism. Note that the agent processes are for one instance of the interaction.

7.2.2 Perspective Decision Mechanisms

Studies show that egocentric biases are less common but not absent in adults compared to children (Epley et al., 2004). Epley et al. presents two explanations as to why adults have less egocentric biases. Either of these possibilities taps on the way the "automatic processes" and "cognitive control" work when humans (children or adults) encounter a perspective taking situation. This means, two processes are in play in perspective taking situations, with the "automatic process" to occur quickly and rapidly and the "corrective process" to activate through motivation and sustained attention (Epley et al., 2004). Various studies have tried to understand and decode the processes involved in different levels of perspective taking (Qureshi et al., 2010; Samson et al., 2010; Surtees et al., 2016). The finding on this topic is still limited, sometimes contradictory, and occasionally dependent on the task used. As a result, our model considers the existence of both "automatic processes" and "cognitive controlled processes". The model is designed to be implemented in virtual agents and robots, hence, we consider any process that helps the agent to adjust its perspective to the other agent as a "cognitive controlled" process.

Automatic Processes

There have been long arguments over finding the most appropriate definition of automaticity in cognitive science. Fiske and Macrae has proposed four features that can be used to evaluate the automaticity of a process (Fiske & Macrae, 2012). The features are; operation outside of cognitive control, efficiency, lack of awareness, and lack of intentionality. Understanding the degree to which operating outside cognitive control and efficiency are required in the automaticity of perspective taking is a potentially interesting line of research (Surtees et al., 2016). Both efficiency and operating outside cognitive control make their associated perspective taking abilities accessible to infants and non-human primates with limited cognitive resources. According to some studies, level-1 perspective taking might be processed both efficiently and outside of cognitive control (Samson et al., 2010; Santiesteban et al., 2014). Samson et al. has shown that their participants experienced involuntary intrusions of altercentric perspectives during the experiment where they were supposed to make perspective judgemtnes while being able to see an avatar's perspective at the same time. Another study by Qureshi et al. showed that while completing a secondary task that engaged the executive function, adults still computed perspectives that they didn't need. In our model, we attributed the notion of the automatic process to any egocentric perspective.

Cognitive Control

Cognitive control is defined as the ability to align thoughts and actions with one's intentions and goals (Miyake et al., 2000; Steinbeis & Crone, 2016). Within the cognitive domain, control is defined by the individuals' ability to inhibit their automatic response to a stimulus and replace it with a controlled response that can be more complex and cognitively demanding (Hinnant & O'Brien, 2007). Unlike Level-1 perspective, Level-2 perspective are not automatic and there is no evidence of it operating efficiently or outside of cognitive control. Previous studies in perspective taking have demonstrated that sometimes people spontaneously or automatically adopt other's perspectives regardless of if they were required to do that or not (Muto et al., 2019; Samson et al., 2010). We add any mechanism that helps the robot to decide and adjust its perspective to the other agent as a cognitive control process. As a result, we placed the following processes within the cognitive control framework; considering the abilities of the other agent, adjusting the frame of reference and adjusting perspective marking. Every time an agent needs to consider or adjust any of these mechanisms, it goes through cognitive control processes.

Frame of Reference is a set of axes or origin points for addressing position of the objects or their spatial relationships (Levinson, 1996; Mintz et al., 2004; Trafton et al., 2005). Different frames of reference in the literature concerning perspective taking are presented in Table 2.1. When we refer to a frame of reference in our model, fundamentally it can be any of the available frames in the works. However, since we are modelling for interactions with predefined dynamics, in our cases we are only considering *egocentric* (from the self point of view) versus *addressee-centric* frames of reference (from the other point of view).

Perspective marking relates to the speakers explicitly marking their perspective in natural language or not (Steels & Loetzsch, 2009). It consists of explicit or implicit inferences to the person the perspective is assigned. Perspective alignment takes less cognitive effort if the perspective is marked (Steels & Loetzsch, 2009). Perspective marking can be only considered where the interaction includes natural language, in interactions that do not include verbal communication, we can discard this process.

Considering the abilities of the other agent is a skill that can be incorporated in the agent as a way to adapt the interaction and the decision making especially to children's abilities. As a starting point, we are going to keep that as an ON/OFF feature, that can be activated or deactivated depending on the purpose of the interaction. In more complicated scenarios, the agent can decide when to turn the feature on or off at each instance of interaction. To use this feature, the agent particularly needs some prior interaction or few baseline interactions that help them to update their Theory of Mind (ToM) about their counterpart (Barnes-Holmes et al., 2004; Premack & Woodruff, 1978; Wellman, 1992).

7.3 Integrating CogPeT with Typical Agent Processes

To develop an agent that implements the model, it is important to understand the links between typical agent processes and the introduced model. From a computational perspective, our main inspiration for implementing the model is to develop an agent that can support perspective taking related decisions. The model can be integrated into any agent architecture that deals with decision making and reasoning, however, in this chapter we provide the example using the BDI architecture (Bratman et al., 1988; Rao, Georgeff, et al., 1995). Based on this architecture, the starting point of a typical agent corresponds with a set of Beliefs (knowledge about itself, others, and the environment), Desires (goals to pursue) and Intentions (actions or plans that the agent aims to achieve). The model can be integrated into the deliberate processes of the agent, uses and updates its beliefs, get direct influence by the desires, and modify the intentions. In this section, we present the cognitive agent architecture, focusing on the links between a typical BDI agent structure and perspective taking related concepts.

To create an agent architecture that can endow agents with perspective taking abilities and awareness of decision making plans that consider the abilities of other agents, our main inspiration is Surtees et al. perspective taking model. The advantage of using this model is that by changing the Circumstance or object (C), we can extend it to cognitive and affective dimensions of perspective taking, which is a suggestion beyond the scope of this thesis. In case of changing the circumstance to cognitive and affective situations, the relationships between the Self (S) and Other (O) need to be updated. Furthermore, the processes included in automatic processes and cognitive control need to be updated accordingly.

7.3.1 Perspective of the Acting Agent

The cognitive process of the agent when it makes a decision (Figure 7.2) starts from its goals, which in our case is a function of task design. For example, when the task requires taking the perspective of the other player e.g. the task cannot be completed without perspective taking, before perfomring any action the agent needs to assess the possible perspective choices. This assessment includes analyzing the possible ways to perform the action (e.g. instructing, moving, speaking) and evaluating if it requires taking perspective or not. Based on these initial analyses, the agent decides whether there is a perspective misalignment between themselves and the other agent. If the answer is YES, then the agent can start analyzing the other agent's perspective taking abilities. If the answer is No, the agent can proceed with their initial assessment of the best course of action and make the most common or convenient perspective choice. After the agent decides to assess the other agent's abilities, it can also check if there is a need to accommodate the other agent or not. This decision corresponds to the feature of the model called "Considering the Abilities of the Other Agent". If the feature is ON and there is a need to accommodate agent B, then agent A makes a decision accordingly. However, if there is no need and the feature is off, agent A makes a decision that is least ambiguous and more convenient to them. All these processes are presented in detail in Figure 7.2.

7.3.2 Perspective of the Perceiving Agent

The cognitive process of the perceiving agent (Figure 7.3) begins with a brief assessment of the perceived action. If the action is ambiguous especially from the perspective taking aspect then agent B can decide to ask for clarification or not. If it decides not to ask for clarification (e.g. it is not natural to the interaction to do so) it tries to search for clues to find the intended perspective. This search for clues can include using the knowledge base of agent A's abilities or agent A's behaviour in previous interactions. Then, agent B can proceed with perceiving the action as it aligns with the knowledge base. On the other hand, if the agent decides to ask for clarification, agent A can resolve the ambiguity and Agent B can perceive the action as it was intended. After the perception, depending on the course of interaction, agent B might receive an evaluation about its perception, e.g. wrong perception resulting in making a wrong move. If the perception was incorrect then Agent B can update its perception and respectively the knowledge base. On the other hand, if the perception was correct, Agent B can just adjust the knowledge base with the fact that the route resulted in a successful decision. The described processes are presented in detail in Figure 7.3.

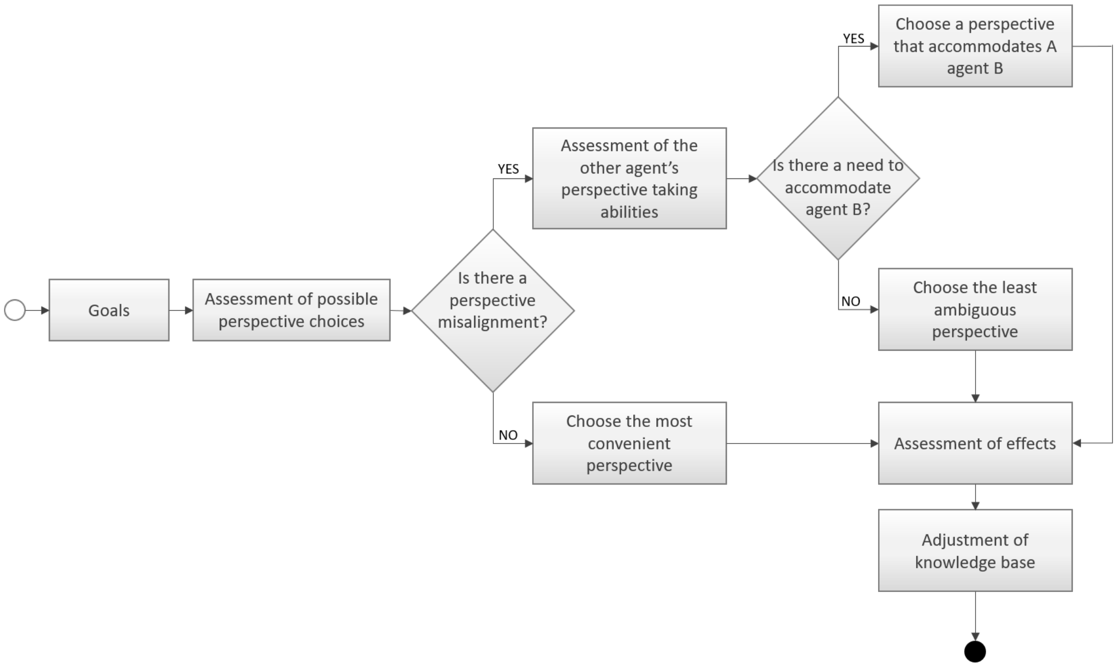


Figure 7.2 – The agent's cognitive process when performing an action.

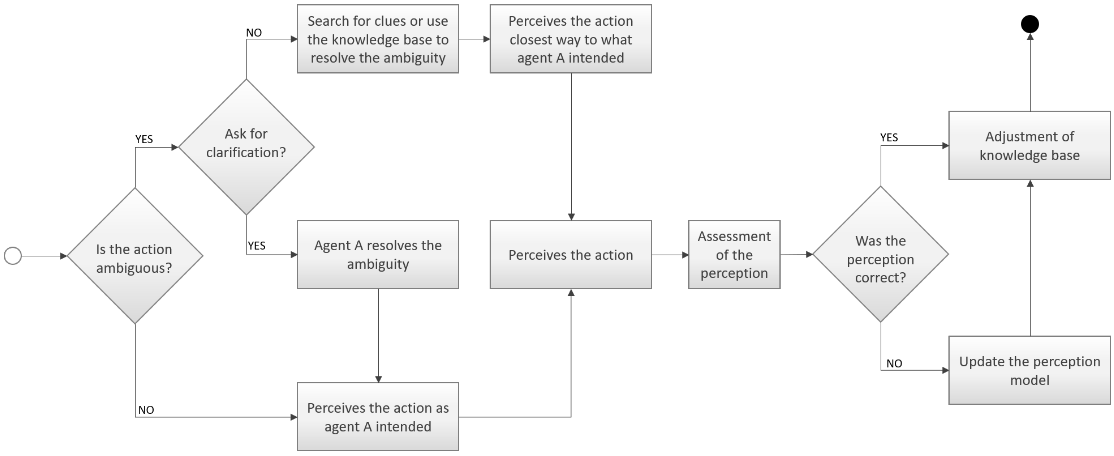


Figure 7.3 – The agent's cognitive process when perceiving the action.

7.4 Model Applied to the Two-Player Maze Game

We adapted the model to a two-player version of Cozmo originally described in Chapter 6. It include two virtual robots (agents) that play with each other in the same environment. The presented interaction in the maze game eliminates verbal communication and only focuses on decisions that are based on understanding the other agent's perspective and following the game rules. While a more comprehensive description of the game is presented in Chapter 8m in this section, we describe the updated processes in the model.

7.4.1 Initial Considerations

We called the agent controlled by the human as agent A and the one controlled by symbolic programming and equipped with CogPeT as agent B. Here, we present a brief and general overview of how the interaction works so we can update the agent processes accordingly. In the two-player version of the maze game, three steps that need to be considered are:

- Participant need to guide agent A that as a result of moving around in the maze has varying angular disparity compared to the participant, e.g. they constantly need to update their understanding of the agent A's frame of reference;
- Participant needs to understand agent B's moves which sometimes requires perceiving agent B's frame of reference (in case of mismatch with agent A). Agent B also moves around in the maze and has varying angular disparity compared to the participant;
- Participant needs to constantly consider the rules of the game (differs depending on the context of the game) when making decisions e.g. agent A loses a life if it repeats agent B's last move.

7.4.2 Modelling the Robots' Decisions

In the two-player maze, the interaction happens only through movements that eliminate the verbal communication and only focuses on decisions that are based on understanding the other agent's perspective. Considering the limited interaction modalities in the maze game, we only use parts of the model that relates to understanding the perspective of the other player and remove the parts that require clarification or ambiguity check. The model represented in Figure 7.4 shows agents A and B's cognitive model when engaging in an activity without verbal communication. The main difference between the model presented in Figure 7.1 and the one presented in Figure 7.4 is removing natural language-related processes which results in eliminating perspective marking. However, the model might still require accommodating other agent's options depending on the nature of the interaction. In the scenario we are using, the players should constantly evaluate the spatial relationship between themselves and the other player to make their move. They also need to evaluate the moves against the rules of

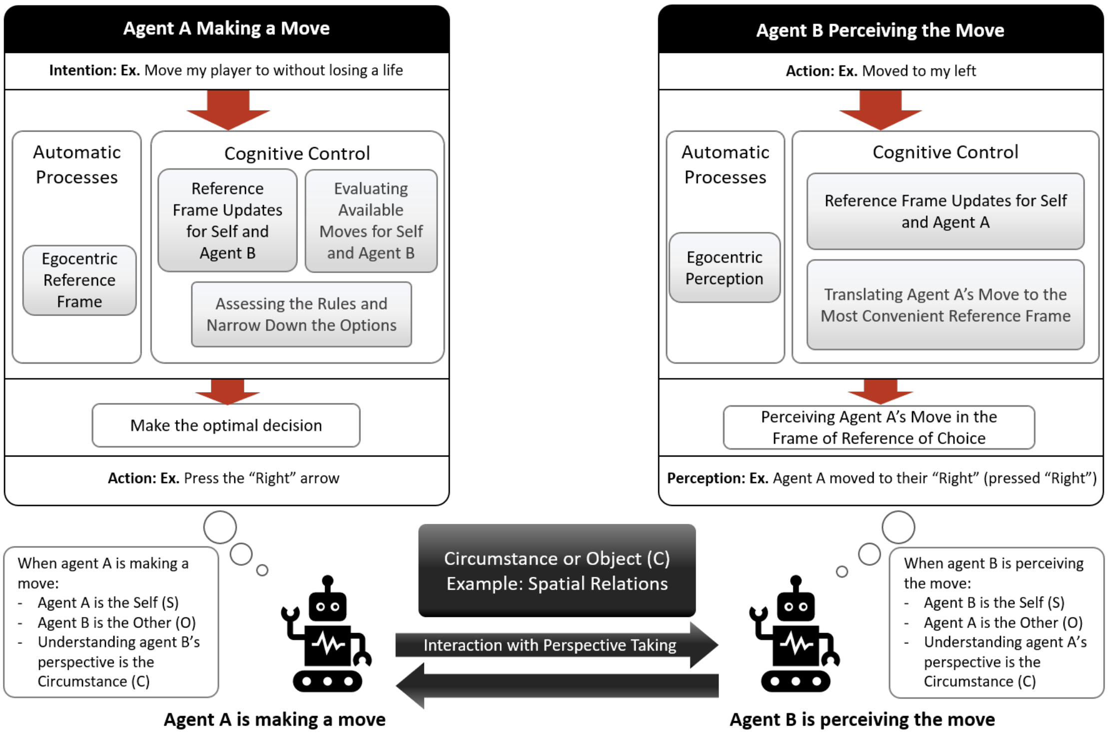


Figure 7.4 – Diagram representing the components and context for two agents playing in a maze game that includes their perspective taking mechanisms. Note that the agent processes are for one instance of the game.

the game. Furthermore, they need to map their and the other player's moves into a frame of reference that is most convenient for them. And finally, make a decision that is closer to their general goal by following the rules of the game.

7.5 Conclusions

This chapter presented the development of a cognitive model for perspective taking that contains a set of mechanisms placed as automatic and cognitive controlled processes. The model and mechanisms were inspired by the visuospatial models of perspective taking in developmental psychology and correspond to exercising spatial perspective taking abilities. The model is adaptable to the interaction context, e.g. verbal vs. non-verbal and expectations from the agent, e.g. challenge or accommodate the human. Its adaptability can be achieved by adding and removing processes to the model.

Furthermore, the chapter has investigated a range of processes that can be included in the robots, which were not replicated in this thesis. This includes the processes used in natural languages, such as perspective marking or adaptive behaviours for considering the abilities of the other agent. However, future works on both physical and virtual robots could shed some light on how the model performs in such scenarios.

The model has been applied to the two-player version of the maze game and the updated diagram and agent's cognitive processes have been presented in the chapter. While more details about the interaction and its adaptations to cooperative and competitive interaction will be presented in the next chapter, the discussed modifications represent how the model can be adapted depending on the task and context. Rather than evaluating the model for its performance, the study designed in Chapter 8 evaluates participants perception of a virtual robot with the model. To have a baseline, the interaction with the robot with full perspective taking abilities, e.g. having the model, is compared to a robot with limited perspective taking abilities, e.g. occasionally making child-level mistakes.

8 Virtual Maze: An Evaluation Study

This chapter is dedicated to evaluating the model developed in chapter 7. The evaluation is focused on the participants' perception of the robot with full perspective taking abilities e.g. with CogPeT in comparison to the one with limited abilities, e.g, making mistakes. In the first section, we describe the motivation for developing the game and the interaction that includes the research questions and the specific hypothesis we made for each study. Then, we describe the development of the game, virtual robots, and interaction modalities. This will be accompanied by details of the two studies carried out with the developed platforms including their study designs, conditions, and qualitative and quantitative analyses. Finally, we wrap up the chapter by discussing what we learned from the studies and how this work can be extended beyond the scope of this thesis.

8.1 Scope and Research Goals

8.1.1 Motivation and Contribution

Throughout this thesis, we looked at spatial perspective taking from different angles in Chapters 4, 5, and 6. Then we used some analogies from the research background, combined it with what we learned from our exploratory studies and developed a model of perspective taking that can be implemented in agents and robots. The model includes various elements that cover different interaction modalities. For our final study, we decided to implement some segments of the model in a robot and evaluate if the existence of perspective taking abilities in the robot affects the human-level perception of it. The initial plan for the final study was to run two different evaluations with children based on the platforms developed in Chapter 5 and 6. However, due to restrictions imposed by the pandemic, we had to redesign the activities to be suited for online testing. We decided to pilot the new virtual platform by recruiting adults through online channels. We were also planning to run an online experiment with children after recruiting them through schools. However, because of time constraints and difficulties in recruiting children, especially with schools closed and online classes, we decided to change the pilot study with adults into two full-scale studies.

After the adjustments, this chapter describes a new platform developed for online testing that is adaptable to experiments with robots in physical settings. The new platform is an extension of the Cozmo maze game previously described in Chapter 6. The main difference between the new version compared to the previous one is the number of players which are now two, and the addition of new contexts. In the new setting, one player is controlled by the human and the other player is controlled with an AI developed using symbolic programming. On the other hand, we were inspired to develop the interaction in two contexts of competition and cooperation based on a study by Li et al. In their study, they showed that preschoolers performed better in taking other's perspectives after priming with a cooperative task compared to a competitive one (Li et al., 2019). Their results motivated us to assess how the context of the game impacts participants' perception of the robots. In this chapter, we are primarily interested to observe how participants perceive the robot's spatial perspective taking abilities and its ability to predict their moves based on the experimental conditions. To create a baseline for comparison, we developed another agent with limited perspective taking abilities that makes occasional mistakes. The model evaluation is narrowed down to analyzing how participants' perception of the two robots differs.

8.1.2 Research Questions and Hypotheses

In this chapter, we are dealing with three research questions. Our main independent variable is the robot's perspective taking ability; PT Abilities, with two levels of "Full-PT" and "Limited-PT". The second independent variable was Game Type with two levels of "competitive" and "cooperative". We formulate our first research question based on our expectation that playing the game has a positive impact on the participants' spatial perspective taking abilities. To investigate this question, we evaluate participants' learning gain using a test before and after playing the games. We also made a hypothesis that investigates if the game type; cooperative vs. competitive, has any effect on their performance in the test.

- **Research Question 1:** Does playing games that require spatial perspective taking improve the participants' performance in a similar test?
  Hypothesis 1: Participants perform better in the posttest compared to the pretest.
  Hypothesis 2: Participants in the cooperative condition show more improvement in the posttest compared to the competitive condition.

The second research question is focused on how the robot's perspective taking abilities can affect the participant's performance. This question can be investigated by comparing the participant's performance based on the robot's perspective taking abilities and the context of interaction.

- **Research Question 2:** How much of the participants' performance, e.g. the number of mistakes, is a function of their teammate or opponent's perspective taking abilities?
  Hypothesis 3: Participants in the "Limited-PT" condition make more mistakes compared to participants in the "Full-PT" condition.
  Hypothesis 4: Participants in the cooperative condition make more mistakes compared to participants in the competitive condition.
  Hypothesis 5: Participants in the cooperative condition make more mistakes in the "Limited-PT" condition compared to the "Full-PT" condition; however, the number of mistakes in the competitive game is not affected by the PT condition.

In the final research question, we focused on the participant's perception of the robot's abilities. We were interested to know how they rated the robot's intelligence when it had the full perspective taking abilities compared to when it had limited abilities. Furthermore, we wanted to see if the robot's abilities affected the participant's perception of the difficulty of the game or the fun they had while playing it. For evaluating the participant's perception, we used self-assessment questionnaires in the format of rating the robot's intelligence, game difficulty, and game fun.

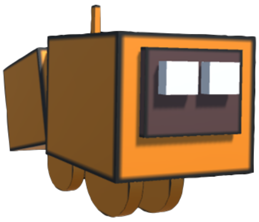
- **Research Question 3:** How participants perception differs when playing with/against an agent with/without perspective taking?
  Hypothesis 6: Participants rate the robot in the "Full-PT" condition higher in intelligence compared to the "Limited-PT" condition for both cooperative and competitive conditions.
  Hypothesis 7: Participants rate both games less difficult in the "Full-PT" condition compared to the "Limited-PT" condition.
  Hypothesis 8: Participants rate both games more fun in the "Full-PT" condition compared to the "Limited-PT" condition.

8.2 Development of the virtual Maze Platforms

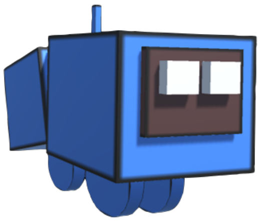
To develop the two games that were used in the following two studies, we have extended the maze game into a two-player game. The basics of moving around in the maze are the same as the game described in Chapter 6 with the robot moving around the maze using control buttons and with the addition of possibility to move backwards. However, the game-play is now more complicated and incorporates taking the other player's perspective to ensure making the right move.

Virtual robots

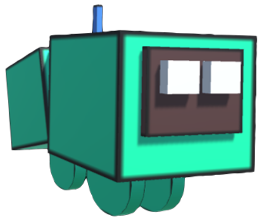
We have designed two virtual robots with similar features as the Cozmo robot and for simplicity, we are just calling them robots. The robot controlled by a human is named *Callisto* and the robot that is programmed to play autonomously is called *Polaris*. Participants can control Callisto using their keyboard's arrow key. The robots look exactly similar except for their colours and they are featured in Figure 8.1.



(a) Callisto



(b) Polaris (Competitive)



(c) Polaris (Cooperative)

Figure 8.1 – Virtual robots inspired by Cozmo robot with different colors according to their roles.

Competitive Game

The design of the competitive game is based on Callisto and Polaris playing against each other. One of the mazes used in study 5 is presented in Figure 8.2 with Callisto in orange (in the bottom half of the screen) and Polaris in blue (in the top half of the screen). The figure shows two independent mazes that are connected through a set of green blocks. The green blocks are called the safe zone and the robots' goal is to reach them. The rule is that whoever reaches the safe zone first wins the game. Each player has three lives that are represented in the form of stars on the left side of each maze. Every time a player loses a life, one of the stars disappears and every time they go on a red block, they fall off the maze and lose one life. The rules are very similar to the game in Chapter 6.

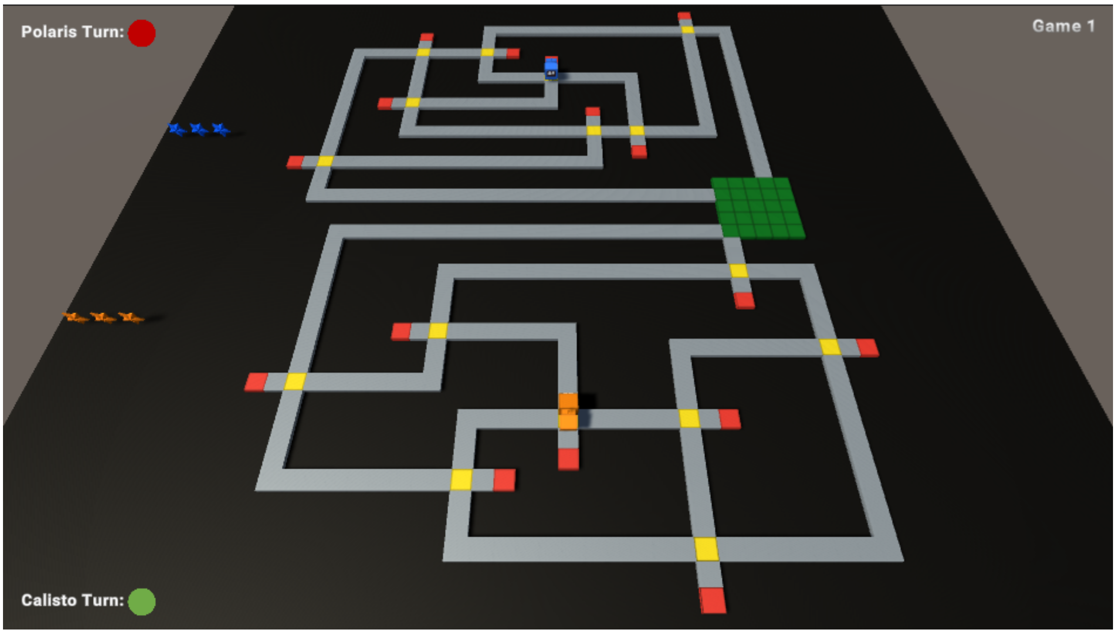


Figure 8.2 – Competitive Maze including Callisto and Polaris, (Screenshot of the game).

However, the two-player version includes an additional rule which is "if a player makes the same move as their opponent's last move, they lose a life". For example, if in Figure 8.2, Polaris starts the game by turning left, this means Callisto loses a life it it also turns left. Naturally, Callisto is free to move forward or turn right without losing a life and turn backwards and loses a life by falling off the maze. Before the game starts, the players are presented with a notice that explains how the arrow keys correspond to Callisto's perspective and not their worldview. However, if they fail to notice it during the instructions, they still have another chance to figure it out during the second tutorial and before starting the main game. The tutorials are used to ensure that they have understood all the rules before the main game starts. Furthermore, the players take turns to make their moves. This means when one player is moving the other one cannot move e.g. for Callisto the keys are disabled. We also have added some signals positioned beside the robot's names on the left side of the screen. The "turn signal" changes to green when it is that robot's turn to move and changes to red when it is not. If a player loses all of their lives before the other player reaches the safe zone, the opponent wins that game.

Cooperative Game

The design of the cooperative game is based on the players helping each other to win the game. In the cooperative version of the game, Polaris and Callisto are in the same team. Figure 8.3 shows one of the cooperative mazes with Callisto presented in Orange (bottom half of the screen) and Polaris in teal (top half of the screen). In the cooperative game, there is no connection between Callisto's maze and Polaris's maze. Considering that the robots are in the same team, together they have 5 lives which are visualized on the left side of the mazes. Furthermore, only one of the robots has access to the safe zone while the other robot does not, e.g. the maze shown in Figure 8.3.

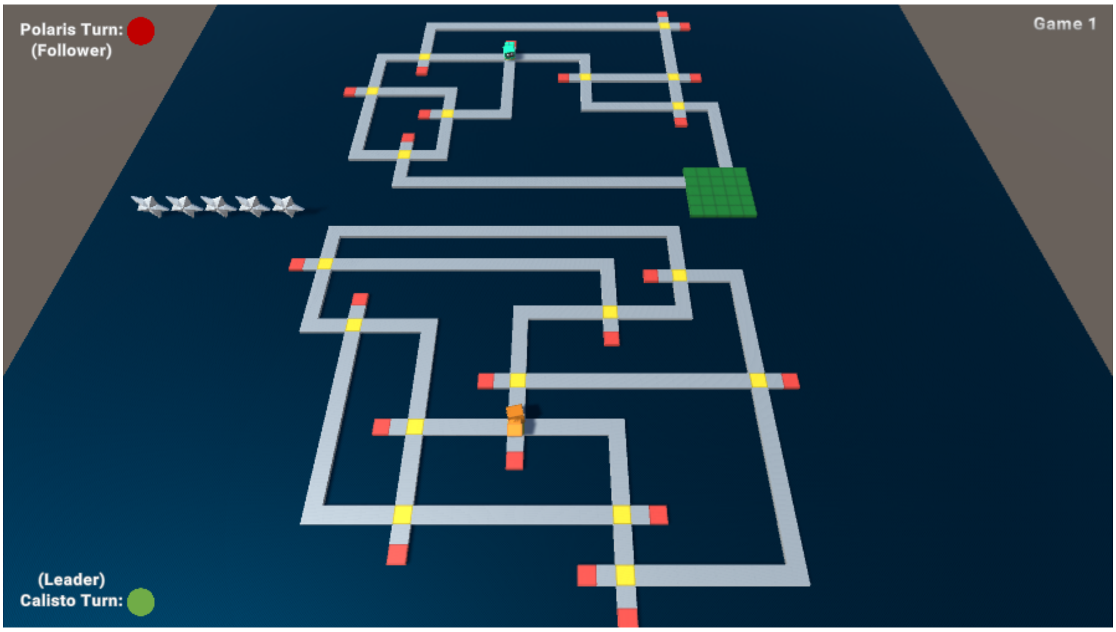
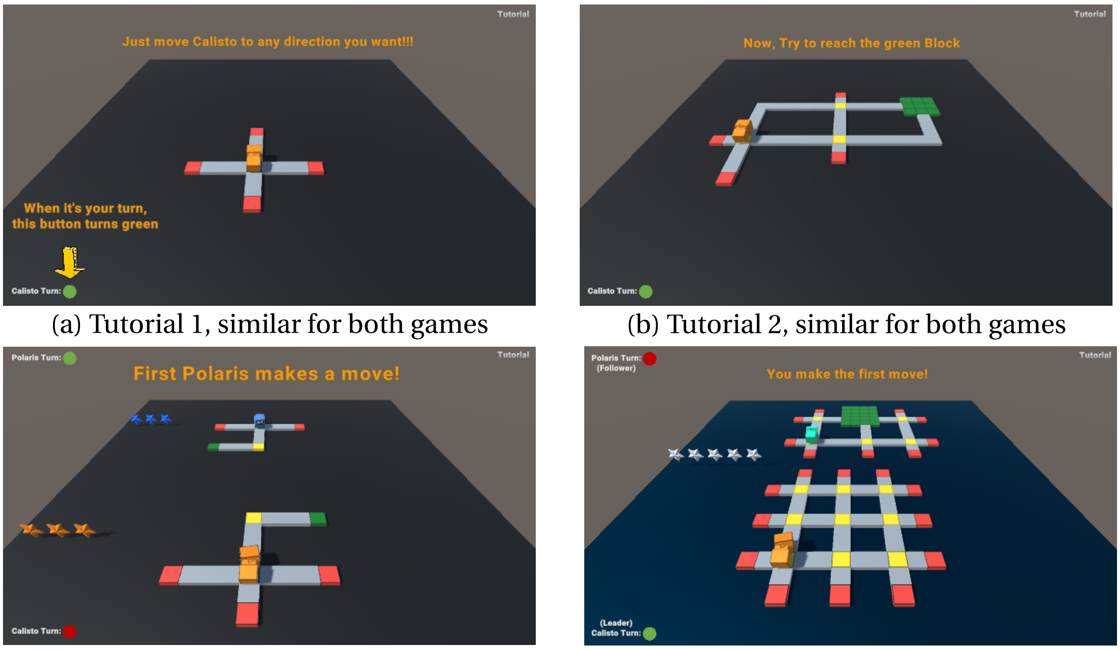


Figure 8.3 – Cooperative Maze including Callisto and Polaris, (Screenshot of the game).

Here, the basic of moving around and losing a life by going on a red block is the same as before. However, the cooperative aspect of the game comes from the rules the players should follow when moving around. The player without access to the safe zone has a role of a *leader* and it is supposed to guide their teammate to the safe zone. The player following the leader is called the *follower*. This means that the follower robot should make the exact same move as the leader and if it fails to do that the team loses a life. When the player with access to the safe zone reaches the safe zone the leader flies to the safe zone and the team wins.

Game Descriptions and Tutorials

Both games start with a description of the games and their rules, then the participants are directed to practice the rules in the tutorials. Each game has 3 tutorials, the first tutorial is designed to practice the basics of moving around using the arrow keys on the computer and it is the same for both games (Figure 8.4a). The second tutorial is used for practising and understanding that the arrow keys correspond to Callisto's perspective. As mentioned before this rule is communicated to the players earlier, but this was to ensure their understanding of the rule before the main game. As shown in Figure 8.4b, when Callisto starts its perspective is aligned with the participant, and the only successful moves are to go *forward* or *right*. After either move, Callisto is going to be positioned at a 90-degree angular disparity compared to the participant's perspective. As a result, at this point, if the participant makes a move from their own perspective and press *right*, the robot does not behave as they expected and usually after a few trials they understand this rule.



(a) Tutorial 1, similar for both games

(b) Tutorial 2, similar for both games

(c) Tutorial 3, only for Competitive game

(d) Tutorial 3, only for Cooperative game

Figure 8.4 – Snapshot of the tutorials before starting the main game: tutorial 1 (a) and tutorial 2 (b) are similar for both games, tutorial 3 is adapted to the game type, (c) shows the one used for competitive game and (d) shows the one used for cooperative games.

As for the third tutorial, they are dedicated to practising the specific rule of each game, as a result, we have different tutorials for the competitive and cooperative games. In tutorial 3 of the competitive game (Figure 8.4c), first Polaris makes a move and then the participant makes their move. The maze is simple and gives only the option of moving in the same direction as Polaris which results in losing the tutorial and then the player is prompted with a text explaining the rules again. On the other hand, tutorial 3 of the cooperative game (Figure 8.4d) can be played in two moves and simply is designed to show the participant how they can win the game in the cooperative game.

8.2.1 Technical Developments

The games have been developed using Unity Game Engine [I] and scripted in C#. The game sessions including the game description, tutorials, and main games were all developed in unity 3D. Besides the game, all the other parts of the experiments such as the consent form, questionnaires and the perspective taking tests were also developed in Unity 3D, however, they were represented in 2D format. The whole study was rendered and built using WebGL[II] (short for Web Graphics Library) which is a JavaScript API for rendering interactive 2D and 3D graphics within any compatible web browser without the use of plug-ins. The WebGL version of the whole study sequence was then uploaded to a secure and private server provided by VMCloud hosted at Instituto Superior Técnico, University of Lisbon. To provide the link of the game to online users, a domain name http://canyouguide.me/ was purchased and associated with the server. All the responses from the users to the consent form, questionnaires, tests and keystrokes from the game, in addition to the internal state of the game were collected through the participant's browser using PHP [III] scripts and saved into a CSV file located in the same server.

8.2.2 Agent Modelling

The agents were modelled with symbolic programming in C# and within the Unity system. Callisto, the human-controlled robot, just makes its move based on the keyboard input. As for Polaris, the developments of the robot slightly differs based on its perspective taking abilities condition. In the Full-PT condition, the robot demonstrates perfect perspective taking abilities, and in Limited-PT, the robot makes occasional mistakes inspired by the mistakes children make. Since each player has a limited number of lives, we have limited the number of Polaris's mistakes to only two mistakes in both Competitive and cooperative versions of the game. The reason for limiting the mistakes is to ensure that the Limited-PT robot still challenges Callisto in the competitive game and gives the team a chance to win in the cooperative game. The robot only makes mistakes when there is a 180-degree angular discrepancy between Polaris and Callisto, which is an error inspired by children's mistake from Chapter 6. We implemented

[I] https://unity.com/
[II] https://www.khronos.org/webgl/
[III] https://www.php.net/

the model presented in section 7.4 into Polaris's code. During the game, Polaris's code registers Callisto's position, direction and last move, then it uses an algorithm to translates that to its own perspective. It also ranks the possible moves for Polaris based on the probability that it blocks Callisto and/or takes Polaris closer to the safe zone. In the Limited-PT condition, the code includes an exception that only takes effect if Polaris and Callisto have $P_{180°}$ angular disparity.

8.3 Evaluation Methods

8.3.1 Interpersonal Reactivity Index Questionnaire

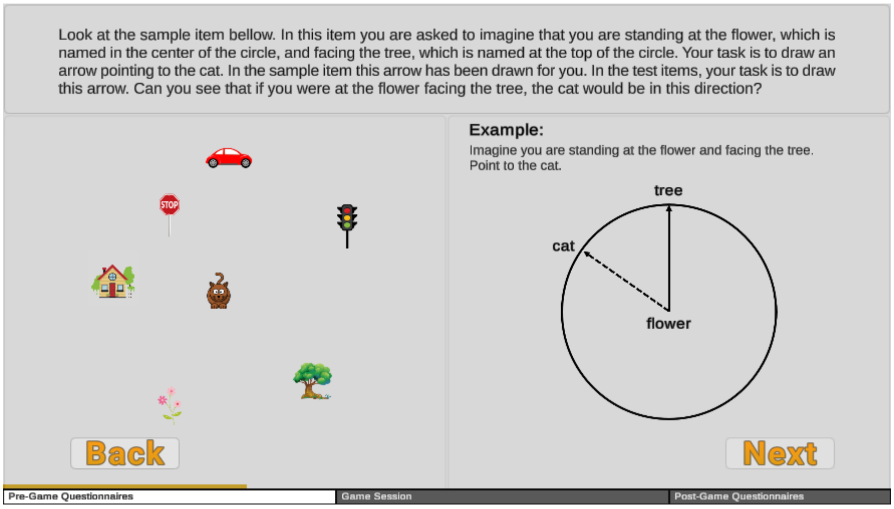
This questionnaire was designed by Davis et al. to test 4 dimensions organised in 4 sub-scales Perspective Taking (PT), Fantasy (FS), Empathic Concern (EC), Personal Distress (PD) with 7 questions each (Davis, 1983; Davis et al., 1980). This test has been widely used in perspective taking related research in recent years (Israelashvili et al., 2019; Mouw et al., 2020; Wolgast et al., 2020) For the studies described in this chapter, we have only used the 7 questions from the perspective taking subscale. More details about the tests and the questions used in our study are presented in Appendix D.1. The test provides us with participant self-evaluation of their cognitive perspective taking abilities. The result will be used in the qualitative analyses of the studies.

8.3.2 Spatial Representation Questionnaire

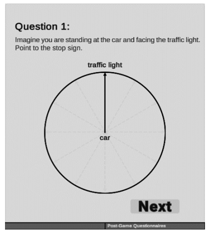
The Spatial Representation Questionnaire developed by Pazzaglia and De Beni (Pazzaglia & De Beni, 2001), provides a self measure of spatial orientation abilities. We used the results from this questionnaire in our qualitative analyses. Further details on this test and its questions are available in Appendix C.4.

8.3.3 Perspective Taking and Spatial Orientation Test

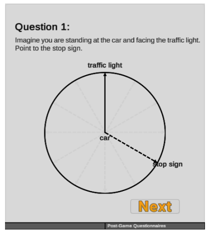
This test is designed to evaluate the participants' ability to imagine different perspectives and orientations and it is developed by Kozhevnikov and Hegarty (Kozhevnikov & Hegarty, 2001). The original test includes a description of the test, one solved example, and 12 questions which all are available in Appendix C.4. For study with adults, we have used the original test and created a modified version for children presented in Appendix C.4. The original test is designed to be answered with pen and paper and the participants have 5 minutes to respond to 12 questions. In our studies, we have divided the questions into two groups of 6, the first half is used as a pretest and the second half as a posttest. We have also provided the participants with 2 minutes and 30 seconds to respond to 6 questions. If a participant takes longer than the allocated time a screen will be prompted and guide them to the next section and they cannot respond to the rest of the questions. To adapt the test to our online platform, we have provided the participants with predefined arrows every 30° degree and they can select the closest angle as their response. Figures 8.5b and 8.5c show the question box before and after responding to the question. as it can be seen from the figure, the participant can only move to the next section after they select an answer when the “next” button activates. To evaluate participant’s responses, we used a range to account for their possible lack of accuracy to estimate the exact angle and limited angles available to choose from.



(a) Example as presented to the participants



(b) Before



(c) After

Figure 8.5 – (a) The example provided to participants before responding to perspective taking and spatial orientation test, (b) The question section of the test before selecting the answer and (c) after selecting the answer (All figures are screenshot from the study).

8.3.4 Game Evaluation Questionnaire

At the end of the study, after completing all the pretest, games, and posttest, participants were presented with a set of questions developed to understand their perception of Polaris and their self-evaluation of their performance. Furthermore, some of the questions asked participants about the techniques they used to play the game and take Callisto and Polaris’s perspective. Finally, we had an open-ended question asking about their feedback to improve the game for children. Full detail of the questions used for this questionnaire is provided in Appendix D.3.

8.4 Study 4 - Perception of the Full-PT vs. limited-PT Robot in Both Game Types

8.4.1 Experimental Design

For this study, we were interested in understanding the participants' perception and feedback about both cooperative and competitive games. We predominantly focused on receiving the participants' feedback after playing both games based on the robot perspective taking conditions; Full-PT and limited-PT. The robot perspective taking ability was considered to be a between-subject variable, so the robot either showed Full-PT or limited-PT as shown in Table 8.1. The game type was our within-subject variable which was counterbalanced for each condition. The game orders were competitive-cooperative and cooperative-competitive. The study was programmed to randomly assign the participants to a study condition and game order every time the WebGL was loaded. The participants in each condition went through all the pre-game tests, game set 1, game set 2, and responded to post-game tests. Each game set includes the game description and rules, 3 tutorials, and 2 games with the same maze. In the first game, Callisto starts the game and in the second game, Polaris starts the game.

The game and the platform used for this study are as described in the previous sections. There are only a few adjustments to the order and number of the games each participant plays to account for the experimental design. The game type is the within-subject variable and each participant is supposed to play two competitive and two cooperative games. We have only used one version of the maze for cooperative and one version for the competitive game. In the first game, Callisto starts the game and in the second game, Polaris starts it in the same maze. Furthermore, in the post-game questionnaire, participants respond to questions comparing the competitive versus cooperative games and also evaluate each game separately.

Table 8.1 – Mixed experimental design with two conditions of *Full-PT* and *Limited-PT*; between-subject variable is robot's perspective taking abilities and within-subject variable is game type, which is counterbalanced for both conditions.

Conditions	**Pre-game Tests**	**Game Set 1**	**Game Set 2**	**Post-Game Tests**
Full-PT		2 Games	2 Games	
Limited-PT		2 Games	2 Games	

8.4.2 Hypotheses

Here, we briefly present the hypotheses for this study based on the experimental design and our general research questions.

H1: Participants perform better in the posttest compared to the pretest.

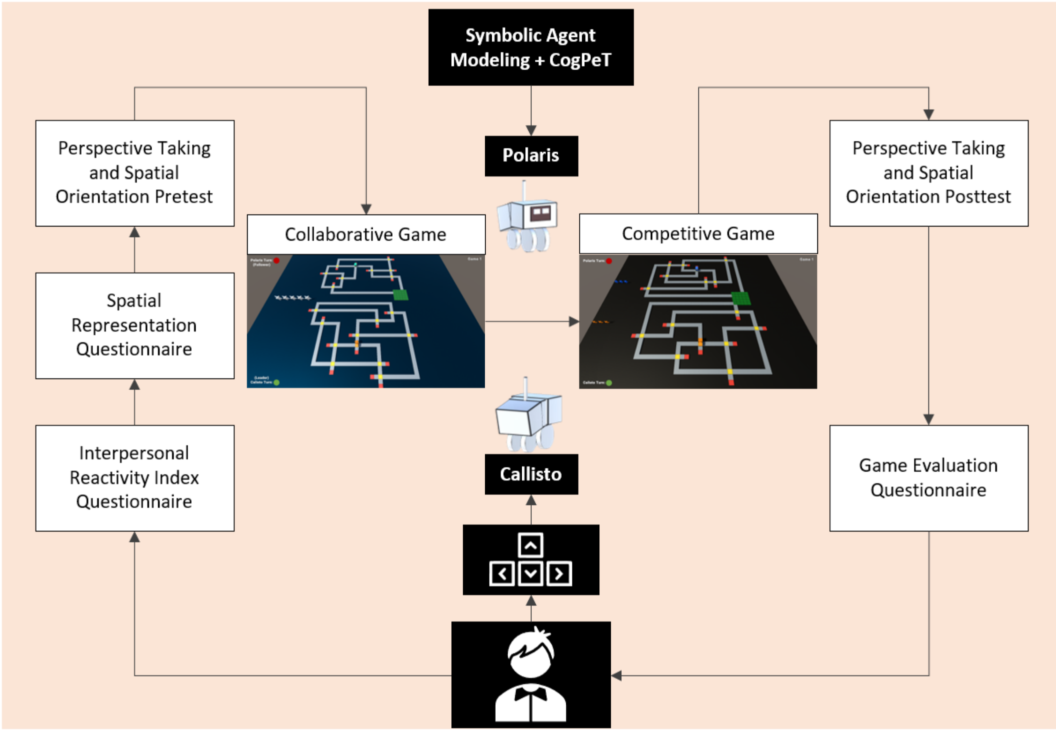


Figure 8.6 – Schematic of the study 4.

H2: Participants in the "Limited-PT" condition make more mistakes compared to participants in the "Full-PT" condition.

H3: Participants in the cooperative condition make more mistakes in the "Limited-PT" condition compared to the "Full-PT" condition; however, the number of mistakes in the competitive game is not affected by the PT condition.

H4: Participants rate the robot in the "Full-PT" condition higher in intelligence compared to the "Limited-PT" condition.

H5: Participants rate both games less difficult in the "Full-PT" condition compared to the "Limited-PT" condition.

8.4.3 Participants

A total of 80 Participants were recruited from the Prolific platform [IV]. Prolific is an alternative to other commercial crowdworking platforms such as Amazon's Mechanical Turk that is targeted at researchers (Palan & Schitter, 2018). All the participants who finished the study and a few who encountered technical problems while running the study were compensated for taking part in the study for an average of £7.5 per hour.

[IV] https://www.prolific.co/

8.4.4 Results

From the total of 80 participants, 71 were approved and analyzed. The 71 approved participants (49 male, 21 female, 1 other) were assigned to the two conditions randomly (35 Limited-PT, 36 Full-PT). Furthermore, the order of the games was counterbalanced with 35 participants in cooperative-competitive order and 36 participants in competitive-cooperative order. Participants' age ranged from 18 to 60 years old, with 66.2% of the participants being between 18-24 years old, 23.9% between 25-29 years old and the rest above 30. The age distribution of the participants is presented in Figure 8.7. The study had received ethical approval from the university's ethics committee and all participants were presented with the consent form before starting the experiment and could only start the game after accepting the terms. The following analyses were carried out for the 71 participants and reported for each hypothesis separately.

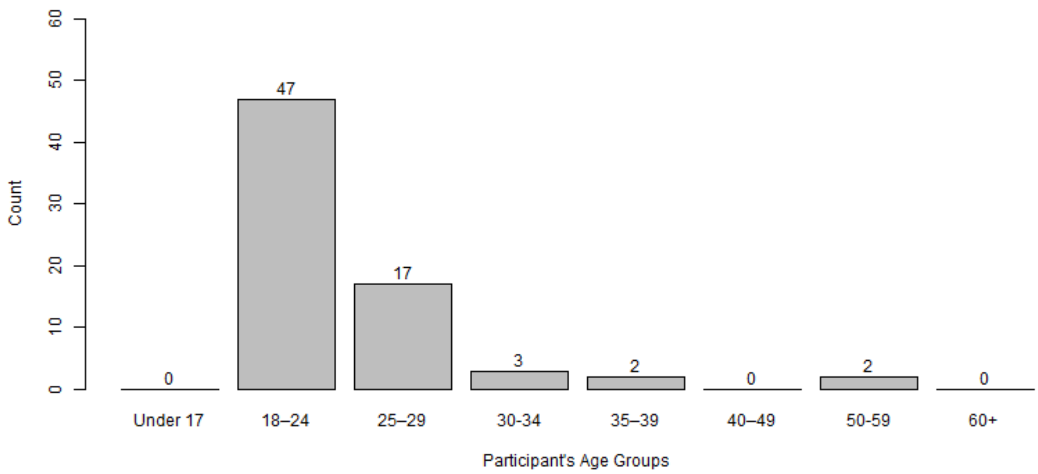


Figure 8.7 – Participants age distribution for study 4.

H1: Performance in Pretest vs. Posttest

As mentioned before, participants responded to 6 spatial orientation questions in the pretest and 6 in the posttest, and their overall score was computed by summing up their correct responses. The mean and standard deviation for the pretest and posttests are $M = 3.73; SEM = 0.22$ and $M = 4.24; SEM = 0.2$, respectively. Although it is not expected to think that playing 4 games would largely improve adults' spatial abilities, it is important to see if there was an effect. Considering the age of our participants, their perspective taking abilities should have been fully developed. However, we were interested to see if training perspective taking could have an immediate positive effect. As a result, we conducted a paired samples t-test to compare the pretest and posttest scores, and verified that participants performed significantly better on the posttest $t(70) = 2.595; p - value = 0.005758$, hence H1 is accepted. The participant's test score distribution for pretest and posttest is presented in Figure 8.8.

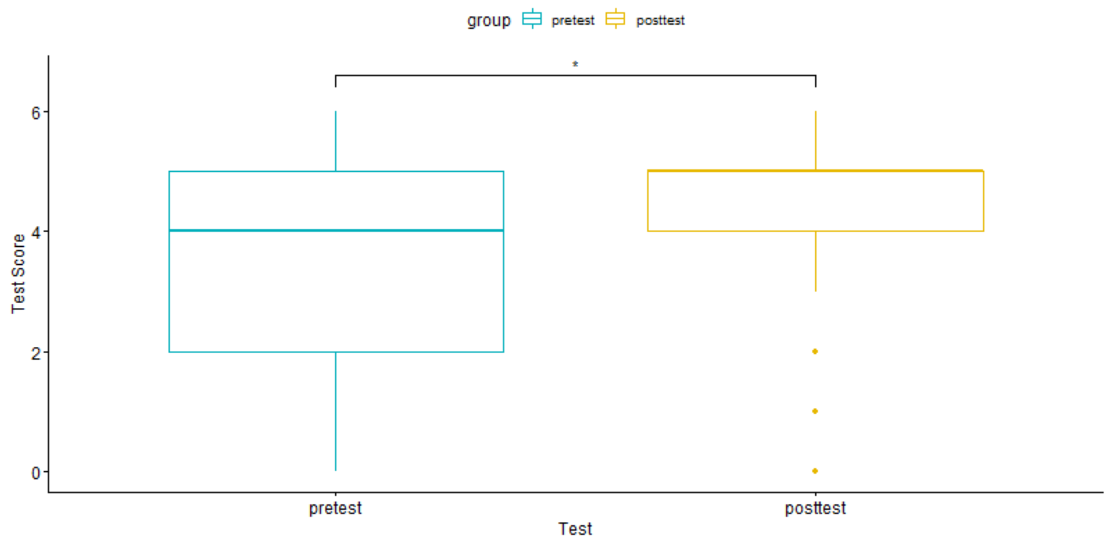


Figure 8.8 – H1. Participant's performance in pretest and posttest.

H2: Total Wrong Moves per PT Conditions

To evaluate participants performance in the games, we have noted if the participant won the game or not. The participants had 3 lives in the competitive game and shared 5 lives with the other robot in the cooperative games, so they could still win even if they lost a life or two. Whether they end up winning or not does not mean that they did not make any wrong move. As a result, we computed the number of wrong moves made throughout the games as a measure of overall performance. The distribution of the total number of mistakes that accounts for participants performance in both competitive and cooperative games shows they made more mistakes in the Limited-PT condition compared to the Full-PT condition. However, the ANOVA on the number of wrong moves for both games shows no significant difference between the Full-PT, Limited-PT conditions $F(1,69) = 14.49$; $p = 0.105$, hence H2 is rejected. This means that the robot's perspective taking abilities did not have a significant impact on the overall performance of the participants in terms of making mistakes.

H3: Total Wrong Moves per PT Conditions per Game Type

Considering the reasoning behind designing the competitive and cooperative games, our third hypothesis is about the within-subject variable Game Type. We wanted to know if there was any difference in participants performance concerning game type. ANOVA on the number of wrong moves shows the main effect of game type, as participants made more wrong moves for the cooperative games $F(1,69) = 29.14$; $p < .001$. This shows that the two variables interact $F(1,69) = 5.64$; $p = .02$ as participants made significantly more wrong moves on the

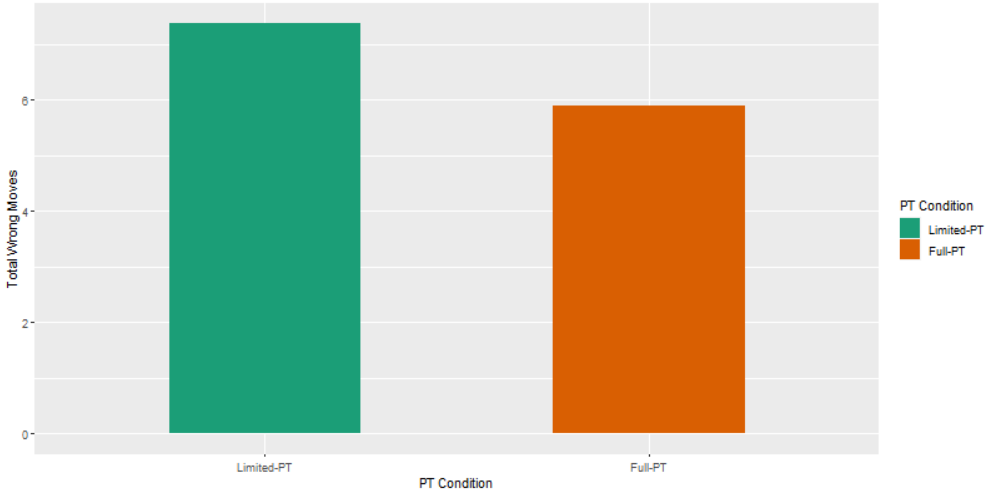


Figure 8.9 – H2. Total number of wrong moves versus robot's perspective taking abilities.

cooperative game when the robot had limited-PT as shown in Figure 8.10. With this result, we can accept H3, showing that showing limited perspective taking when the two players are in the cooperative setting can have more effect on the overall performance compared to when they compete against each other.

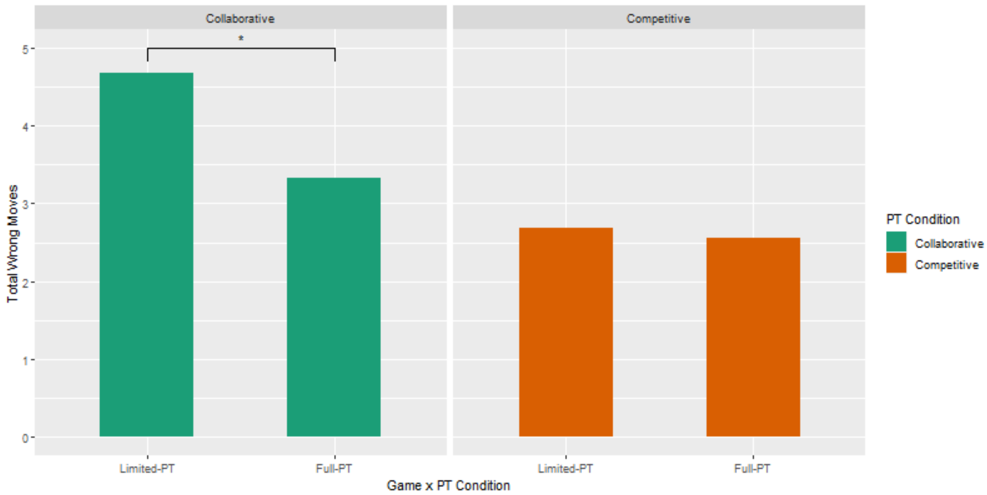


Figure 8.10 – H3. Total number of wrong moves versus robot's perspective taking abilities for each game type.

H4 and H5: Perception of Robot Intelligence and Game Difficulty per PT Conditions

In the first set of questions, we asked for the participant's perception of the robot's intelligence, the game's difficulty, and the game's fun. H4 focuses on their perception of the robots' intelligence and H5 deals with their perception of the difficulty of the game. For this study, we did not make any hypothesis regarding the participants rating of the fun they experienced while playing the games. However, we were interested to see if they rate the game as more fun in Full-PT condition compared to Limited-PT condition and which game they had more fun playing: competitive or cooperative. We ran a repeated measures ANOVA with the PT conditions (Full-PT vs. Limited-PT) as the within factor on the Perception (Intelligence; Difficulty; Fun), the results showed a main effect of perception $F(2,138) = 4.49; p - value = .013$. Although the 2 variables do not interact, we decomposed the interaction using simple effects contrast to see how this main effect was reflected across the groups and considering that the main effect by itself is not that informative. There is a significant difference between difficulty and fun for the cooperative in the PT condition ($F(2,68) = 4.92; p - value = 0.10$), however all in all we observed no significant difference between the participants' perception (Intelligence; Difficulty; Fun) based on their PT condition, hence both H4 and H5 are rejected as shown in Figure 8.11.

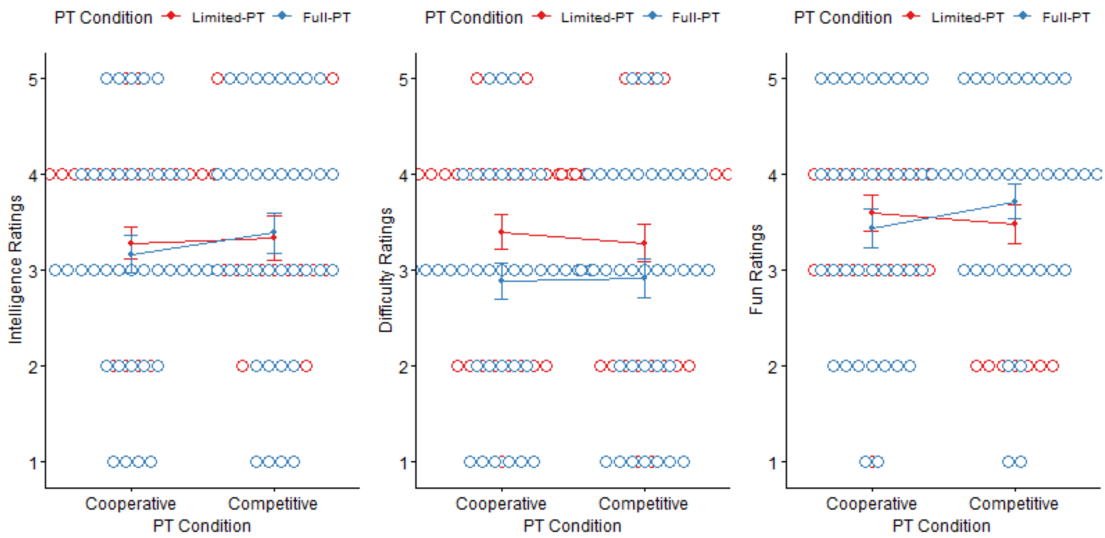


Figure 8.11 – H4 and H5. Participants perception (Intelligence, Difficulty, Fun) of the robot and the game in study 4.

8.4.5 Findings of the Study

As seen from the first hypothesis, participants significantly improved in the posttest compared to the pretest. This showed that the activity by itself helped them to practice their spatial perspective taking abilities independent of the robot's perspective taking abilities. Since the participants played both games in this study, we cannot evaluate the effect of game type on their improvement in the test. Another reason to have a mixed-design experiment was to see how participants compare their experiences in the two games. However, having the participants to go through learning to play two games with different contexts made it harder for us to evaluate their rating of the robot's intelligence and their experience in the games. As a result, we decided to run another study with the same platform but different experimental design that give the participant to only play one type of the games.

8.5 Study 5 - Perception of the Full-PT vs. limited-PT in Each Game Type

8.5.1 Experimental Design

After what we learned from study 4, we considered recruiting more participants and run a study with a different experimental design. To better evaluate the effect of context of interaction on the participants' perception of the robot, we treated the game type as a between-subject variable. As a result, the robot's perspective taking ability and game type were both considered to be between-subject variables, so the robot either had Full-PT or limited-PT and the participants played either cooperative or competitive games. This design resulted in four conditions as presented in Table 8.2. The participants were assigned randomly to each condition. This design gave the participant a better chance of learning and practising the game and get acquainted with the robot they compete against or cooperate with. To play four games, we used the same maze as the one used in the previous experiment for games 3 and 4 and added a new maze for games 1 and 2.

Table 8.2 – Study 5. Mixed experimental design with four conditions of *Full-PT* and *Limited-PT*; between-subject variable is robot's perspective taking abilities and within-subject variable is game type.

Condition 1	Condition 2	Pre-game Tests	Game Set	Post-Game Tests
Full PT	**Competitive**		4 Games	
Full PT	**Cooperative**		4 Games	
Limited PT	**Competitive**		4 Games	
Limited PT	**Cooperative**		4 Games	

8.5.2 Hypotheses

Here, we briefly present the hypotheses of this study based on the experimental design and our general research questions. By relying on the findings from study 4, we were able to update some of our hypotheses regarding the participant's perception of the game and the robot's perspective taking abilities.

H1: Participants perform better in the posttest compared to the pretest.

H2: Participants in the cooperative condition show more improvement in the posttest compared to the competitive condition.

H3: Participants in the cooperative condition make more mistakes compared to participants in the competitive condition.

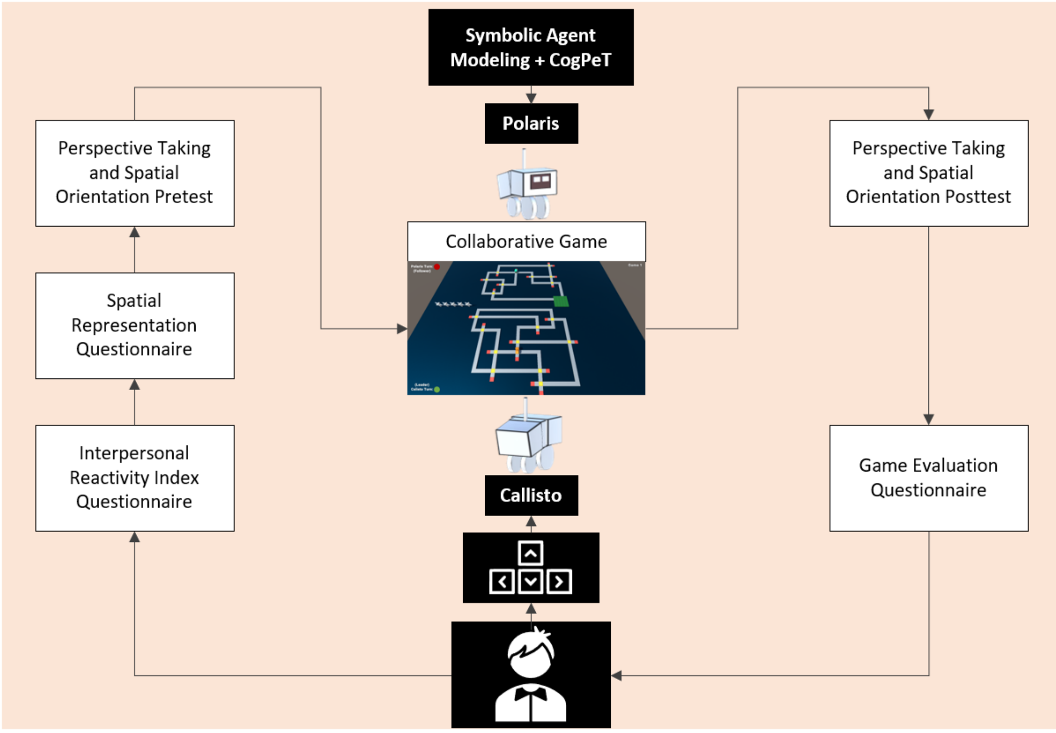


Figure 8.12 – Schematic of the study 5.

H4: Participants in the cooperative condition make more mistakes in the "Limited-PT" condition compared to the "Full-PT" condition; however, the number of mistakes in the competitive game is not affected by the PT condition.

H5: Participants rate the robot in the "Full-PT" condition higher in intelligence compared to the "Limited-PT" condition for both cooperative and competitive conditions.

H6: Participants rate the robot in the "Full-PT" condition higher in intelligence compared to the "Limited-PT" condition for both cooperative and competitive conditions.

H7: Participants rate both games more fun in the "Full-PT" condition compared to the "Limited-PT" condition.

8.5.3 Participants

A total of 100 participants have been recruited on the same platform as study 4; Prolific[V]. All the participants who finished the study and a few who encountered technical issues were all compensated for taking part in the study an average of £7.69 per hour.

[V] https://www.prolific.co/

8.5.4 Results

From 100 participants, a total of 95 participants were approved and their responses were used in the analyses. The 95 approved participants (60 male, 34 female, 1 other) were assigned randomly to the PT conditions (44 Limited-PT, 51 Full-PT) and game type condition (53 Cooperative, 42 Competitive). Participants' age ranged from 18 to 60 years old, with 63.15% of the participants being between 18-24 years old, 16.84% between 25-29 years old and the rest above 30. The age distribution of the participants is presented in Figure 8.13. The study has received ethical approval from the university's ethics committee and all the approved participants were presented with the consent form before starting the experiment and have signed it.

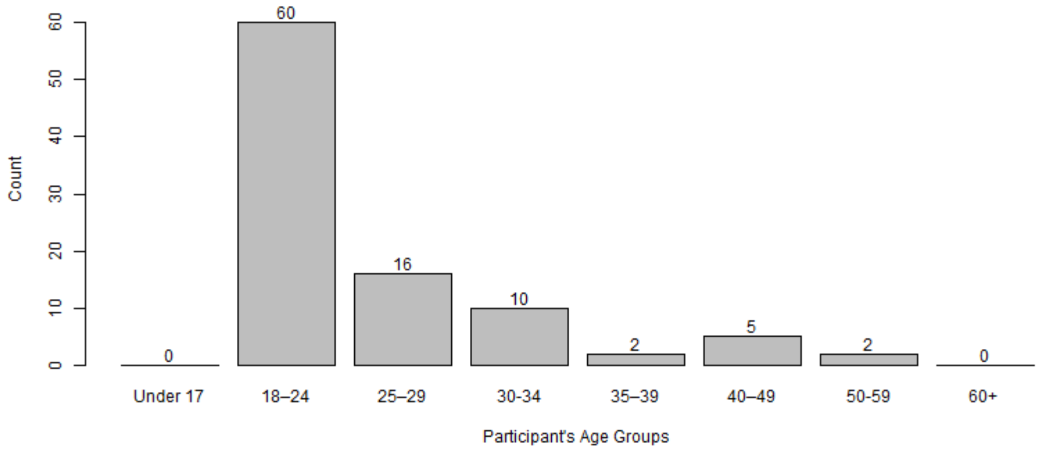


Figure 8.13 – Participants age distribution for study 5.

H1: Overall Performance in Pretest and Posttest

We looked at the participant's overall performance in the pretest and posttest to see if they showed any immediate improvement after playing the game. The mean and standard deviation of the pretest and posttest are $M = 3.326; SEM = 0.19$ and $M = 4.24; SEM = 0.18$, respectively. We conducted a one-tailed paired sample t-test and compared the performances in the pretest and posttest $t(94) = -4.8442, p-value = 0.0000025$ which shows that the participants significantly improved in the pretest compared to the posttest, and H1 is accepted. Figure 8.14 shows the distribution of overall performances in the pretest and posttest.

H2: Performance in Pretest and Posttest per Game Type

In the current study design, participants only played one game type for a longer duration. This gave us the possibility to isolate the effect of the game type. The mean and standard

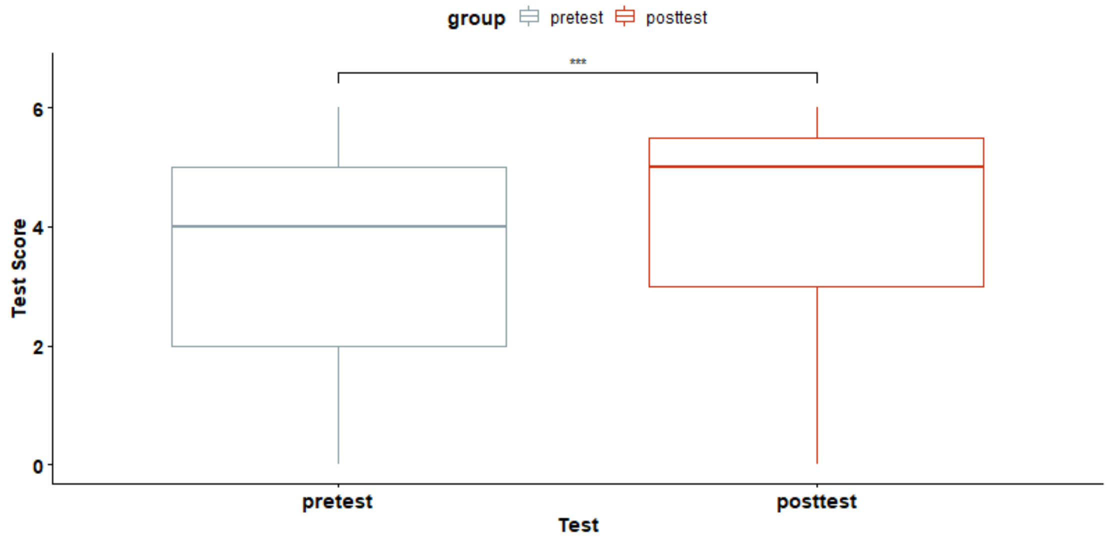


Figure 8.14 – H1. Participant's performance in pretest and posttest.

deviation for competitive condition's pretest and posttest are $M = 3.167$; $SEM = 0.3027$ and $M = 4.024$; $SEM = 0.2901$ and for cooperative game $M = 3.453$; $SEM = 0.2447$ and $M = 4.34$; $SEM = 0.2362$. We looked at the performance in the tests based on the game type condition. We ran a one-tailed paired sample t-test for competitive and cooperative conditions which showed $t(41) = -2.9939, p - value = 0.002326$ and $t(52) = -2.9939, p - value = 0.0001847$, respectively. Based on the analyses and as shown in Figure 8.15, participants significantly improved in both conditions. We also compared the posttest scores between the competitive and cooperative games and the sampled t-test showed no significant difference between posttest scores between the game type conditions $t(84.21) = -2.9939, p - value = 0.20$.

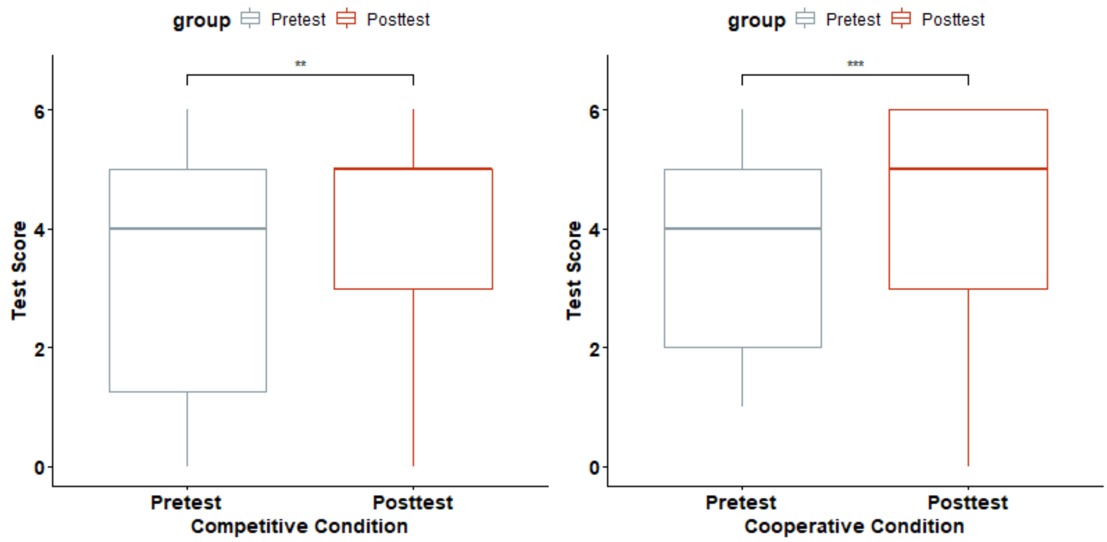


Figure 8.15 – H2. Participant's performance in pretest and posttest based on the game types.

It can be observed that playing either of the competitive and cooperative games helped the participant to improve their immediate spatial perspective taking skills. Furthermore, the improvement in the cooperative condition was more significant than the competitive condition.

H3: Total Wrong Moves per Each Condition

Here, we looked at the total number of wrong moves per game type and perspective taking conditions. A one-way ANOVA was performed to compare the effect of game type on the total number of wrong moves. The analyses revealed that there was a statistically significant difference in the total number of wrong moves between the cooperative and competitive conditions $F(1,93) = [5.268], p - value = 0.024$). As shown in Figure 8.16 (left), participants made significantly more wrong moves in the cooperative games compared to the competitive game. The effect of PT condition on the total wrong moves was evaluated using a one-way ANOVA which shows a statistically significant effect. The result shows that $F(1,93) = [8.806], p - value = 0.004$) the participants made significantly more wrong moves in the Limited-PT condition compared to the Full-PT condition. This result is visualized in Figure 8.16 (right). Based on the analyzes we can accept H3 and then have a closer look at the interaction between the two variables in the next hypothesis.

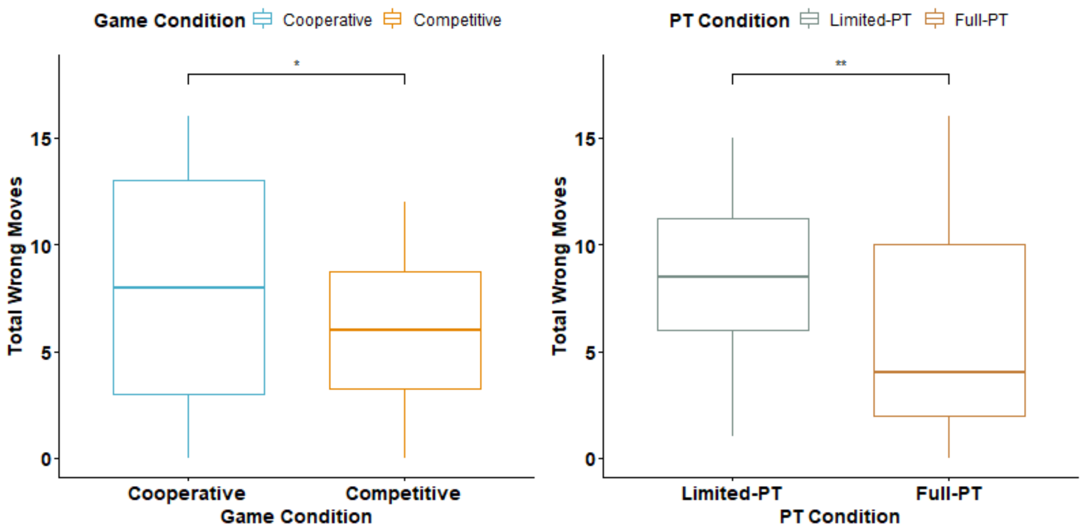


Figure 8.16 – H3. Number of wrong moves based on the game types (left) and perspective taking abilities (right).

H4: Total Wrong Moves per Game Types Condition and PT Condition

A two-way ANOVA was conducted to examine the effect of game type and perspective taking on the total wrong moves. There was a statistically significant interaction between the effects of game type and perspective taking for total wrong moves, $F(1,91) = 4.09, p = 0.046$ as shown in Figure 8.17. Data supports our fourth hypothesis and shows that a player with full perspective taking abilities has a higher impact when two players are cooperating than when the players are competing against each other. Pairwise comparisons with a Bonferroni adjustment shows $F(1,91) = 3.70$, $p = 0.00036$ that in the cooperative condition, participants made significantly more wrong moves in Limited-PT condition compared to Full-PT. The analyses for the competitive condition showed no significant difference $F(1,91) = 0.59$, $p = 0.594$ in the wrong moves made in the Limited-PT versus Full-PT.

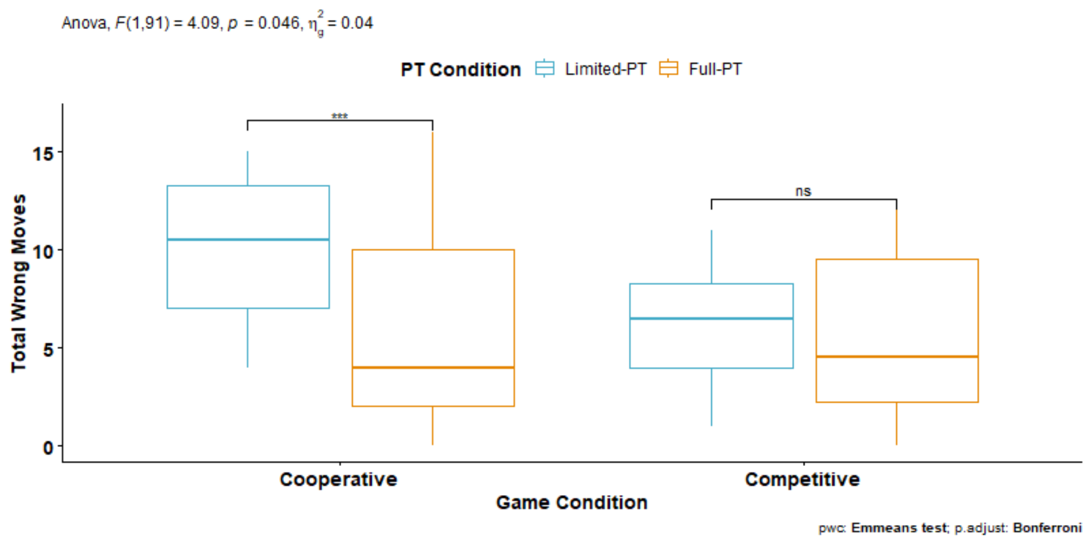


Figure 8.17 – H4. Interaction between the game types and perspective taking conditions for total number of wrong moves.

H5, H6, H7: Perception of Intelligence, Difficulty and Fun per PT Conditions

In this section, we analyzed the participants perception of the robot and the game based on the robot's perspective taking abilities and game type. H5 focuses on participants' perception of the robots intelligence and H6 and H7 deal with their perception of difficulty of the game and fun. We ran a two-way ANOVA that examined the effect of game type and perspective taking on the rating of intelligence. The analyses shows no interaction between the game and PT condition $F(1,91) = 1.602$; $p - value = 0.209$. Although the two variables did not interact, we decomposed the interaction using simple effects contrast to see how this main effect was reflected across the groups. the ANOVA shows $F(1,91) = 6.824$; $p - value = 0.0107$ significant difference of means in the rating of intelligence for the PT condition. Participants have rated the robot with Full-PT higher in intelligence compared to the one with Limited-PT. Pairwise comparisons with a Bonferroni adjustment shows that the rating was only significantly higher in the cooperative condition $F(1,91) = -2.79$; $p - value = 0.0063$ and not in the competitive condition $F(1,91) = -0.79$; $p - value = 0.42$. Figure 8.18 (left) shows the rating for intelligence per conditions and the result shows that we can accept H5.

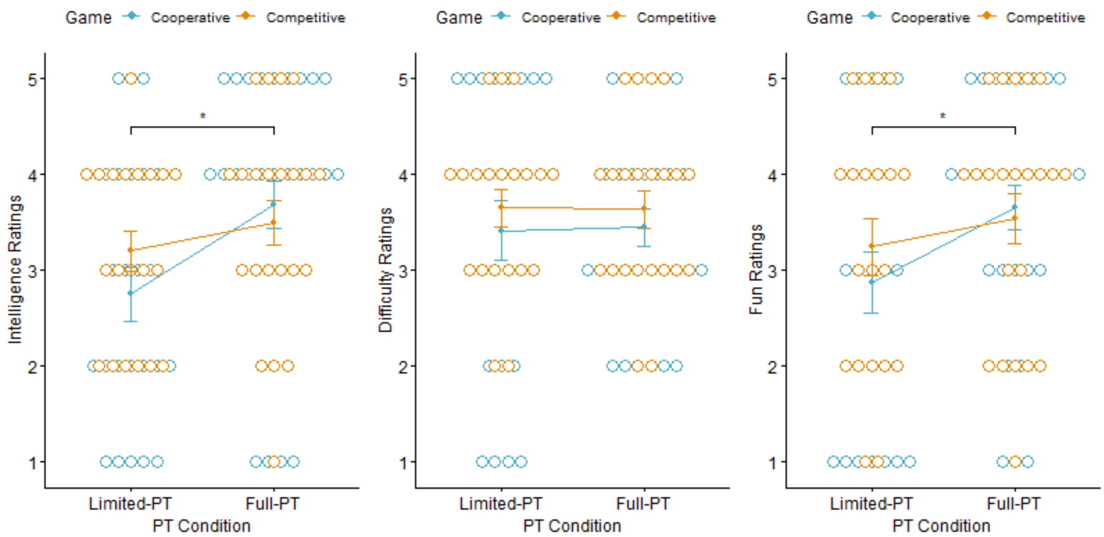


Figure 8.18 – H5, H6, H7. Participants perception (Intelligence, Difficulty, Fun) of the robot and the game in study 5.

Similarly, we performed an ANOVA for the rating of fun and difficulty. The detail of all the analyses are presented in Table 8.3. As shown form the table there is no difference in the rating of the game difficulty per game type and perspective taking condition and H6 is rejected. However, participants in the cooperative condition have rated the game significantly more fun when the robot had Full-PT compared to when it had Limited-PT. Using pairwise comparison with Bonferroni adjustments $F(1,91) = -2.10; p - value = 00.038$ proves that H7 can be accepted. Generally, participants playing the cooperative game were more prone to rate the robot as higher in intelligence or the game as more fun when the robot showed better perspective taking abilities compared to the ones playing the competitive game.

Table 8.3 – ANOVA Table (type II tests)

Perception	**Effect**	**DFn**	**DFd**	**F**	**p**	**p<0.05**	**ges**
	Game Condition	1	91	0.185	0.668		0.002
Intelligence	PT condition	1	91	6.824	0.011	*	0.070
	Game Condition:PT condition	1	91	1.602	0.209		0.017
	Game Condition	1	91	0.813	0.370		$9.00e-03$
Difficulty	PT condition	1	91	0.002	0.960		$2.74e-05$
	Game Condition:PT condition	1	91	0.009	0.923		$1.04e-04$
	Game Condition	1	91	0.185	0.175		0.002
Fun	PT condition	1	91	4.173	0.044	*	0.044
	Game Condition:PT condition	1	91	0.758	0.386		0.008

8.5.5 Qualitative Analyses

Analyses of Interpersonal Reactivity Index Questionnaire

As mentioned before, we only used the perspective taking subscale of this test. The descriptive statistics from the participants responses shows $Min = 1.43, M = 2.86; Max = 4.000$. The scores split based on the gender of participants shows a similar distribution for both genders. Test scores for male participant are $Min = 1.43, M = 2.78; Max = 4.00$ and for female participants are $Min = 1.71, M = 2.86; Max = 3.71$.

Analyses of Spatial Representation Questionnaire

To evaluate the participant's responses to this question, we used the evaluation methods provided by Pazzaglia and De Beni (Pazzaglia & De Beni, 2001). They explained a method to divide the participants into *high-survey* and *low-survey* groups based on the sum of their responses to some of the questions in the questionnaire. We followed their evaluation method and split the participant into two groups, and analyzed their performance based on these two groups. In their study, the high-survey participants have shown better performance in adopting spatial-holistic strategies to perform spatial tasks. The scores for all participant were $Min = -1.00, Median = 8.00; Max = 11.00$. After the median split, 67.36% of participants scored below the median and were placed in the low-survey group (44 Males, 16 Females) and 32.63% of participants were placed in a high-survey group (19 Male, 5 Female).

Analyses of the Methods used by Participants to Take the robot's Perspective

As part of the post-game questionnaires, participants were asked about the techniques they used to make their move and to perceive Polaris's move. The questions marked as questions 10 and 11 in the Appendix D.3 are:

- Q10. When playing the games which technique did you use more frequently to make your move?
- Q11. Which technique did you use more frequently to understand your teammate or opponent's move?

Figure 8.19 shows the percentage that each option was selected for taking Calisto and Polaris's perspective. The participants are categorized based on their spatial representation scores into low-survey and high-survey. The majority of the participants selected the same technique for taking both robots' perspectives. But as shown in the figure, some techniques were more popular in one group than the other. For example to take Callisto's perspective, 42% of participants in the high-survey group selected "rotating my head" compared to 30% in the low-survey population. The popular methods among the low-survey group were "rotating my head" and "closing my eyes and rotate in my head". Less than 10Furthermore, around 25% of participants in each group selected "other, please specify" and provided their own method. The responses to this option are presented in Table 8.4.

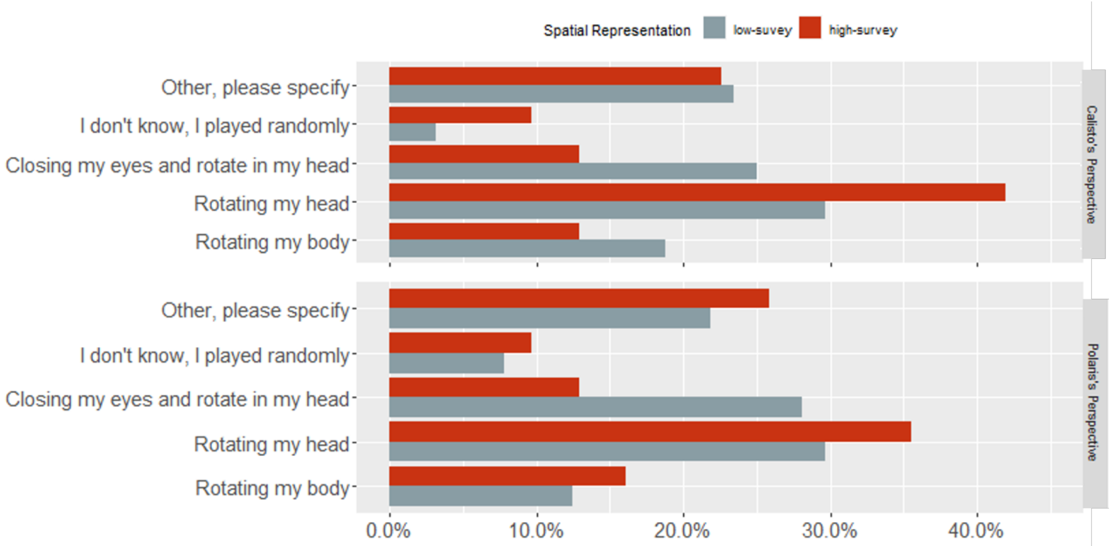


Figure 8.19 – Participants responses regarding the techniques they used to make their move and take Polaris's perspective. The data is organized based on their spatial representation categories.

Analyses of Multiple Choice Questions

As part of the post-game questionnaires, participants were asked about the methods they used to take Callisto's and Polaris's perspectives. The questions correspond to number 12 and 13 in Appendix D.3 and it is shown below.

- Q12. In my opinion, Polaris was... [multiple selection open]
- Q13. In my opinion, I was... [multiple selection open]

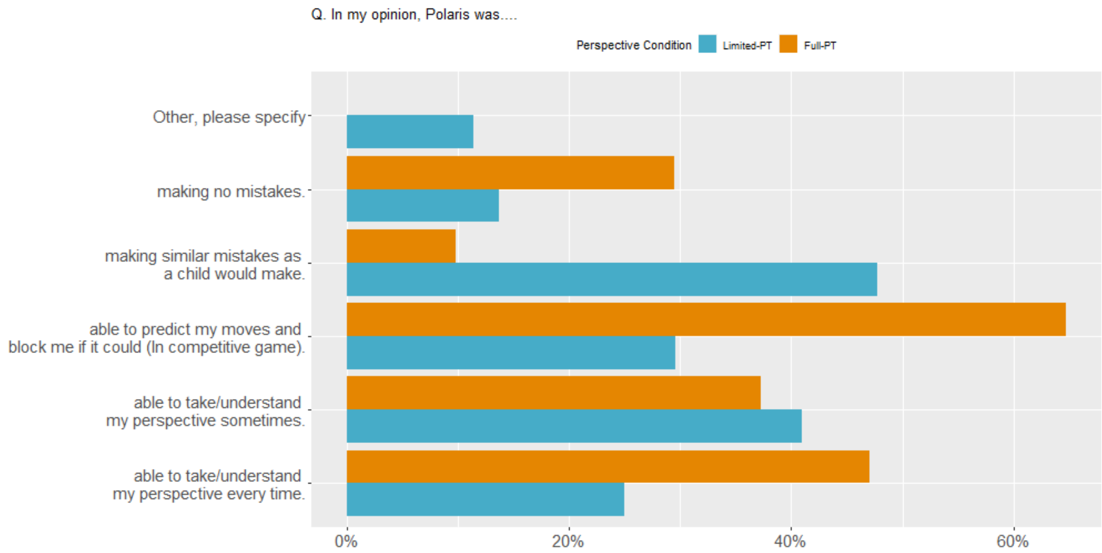
Figure 8.20a shows the participants evaluation of Polaris's performance presented based on Polaris's perspective taking abilities. The responses show that the participants evaluated the Full-PT robot higher in the following categories "making no mistakes", "able to predict my moves", and "able to understand my move every time". This evaluation aligns with the robot's abilities and proves the model works. Participants also rated the Limited-PT Polaris higher in "making similar mistakes as a child would make" and "able to take my perspective sometimes". Generally, the participants rating of the Polaris's abilities in both PT aligns with our initial assumptions when designing the robots.

Figure 8.20b shows the participants' evaluation of their performance in taking the other robot's perspective. The most interesting trend observed from this figure is the responses to the option

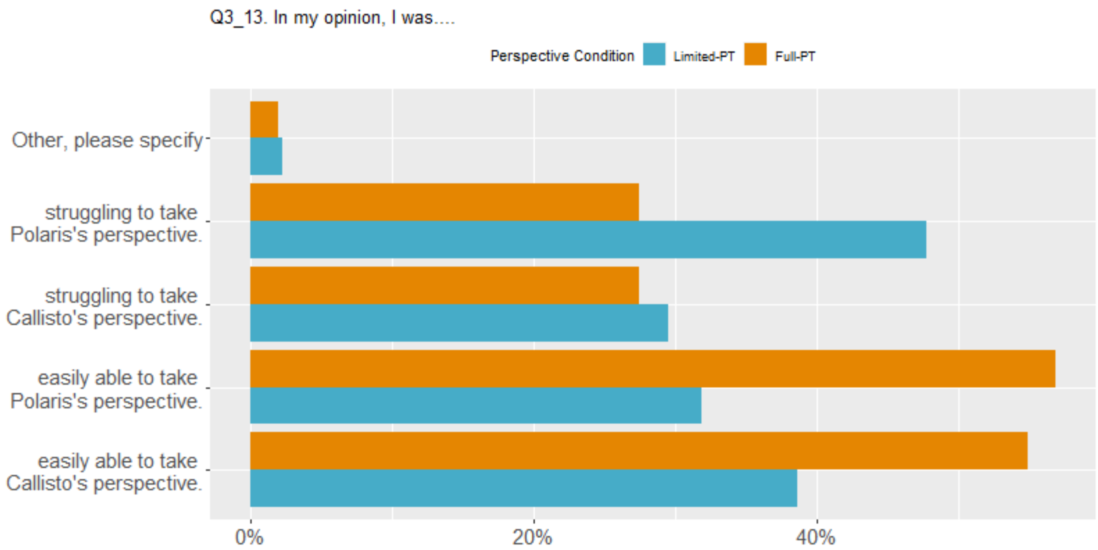
Table 8.4 – Responses to “other, please specify” option in Questions 10 and 11 where they are asked about the methods they used to take the robot’s perspective.

ID	Q10: When playing the games which technique did you use more frequently to make your move?	Q11: Which technique did you use more frequently to understand your teammate or opponent’s move?
1	visualization	visualization
2	Imagination	Carefully observing and imagining it from their point of view
3	using my hand	using my hand
4	Rotate in my head without closing my eyes	
5	I can just see the directions	Just watched the bots movement
6	just imaging in my head	imagining in my head
7	I imagined the the point of vierw of the robot	i tried predicting the opponents easiest path to the goal and tried to make it difficult for it to win easily or for me to win
8	observe the map of my teammate choosing the best road	just see where he go
9	i don’t used any	none
10	I did not rotate.I just followed the left/right directions	I did not rotate.I just followed the left/right directions
11	I rotated in my head but did not close my eyes	I rotated in my head but did not close my eyes
12	i imagined my “up arrow” was the robots face and i clicked the arrow on wich direction the robot turned to	same as before
13	imagining the route that leads to the green field and following every one step.	remembering the way my teammate turned (or not)
14	rotating in my head with open eyes plus trying to foresee the opponent move	open eyes rotating + study of the map
15		keeping in mind left and right from the robot’s perspective
16	none of the above	same as in question 4
17	I placed myself in the position of the robot	I placed myself in the position of the robot
18	I just try to look from the character eyes	I saw where he was looking at
19	rotate in my head without closing my eyes	rotate in my head without closing my eyes
20	Visualize with open eyes	Visualize with open eyes
21	imagining in my head without closing eyes	i didnt understand this part that much
22	i looked at the other robot point of view to choose directions	
24	vectors	putting myself in his shoes

"struggling to take Polaris's perspective". It seems that more participant in the Limited-PT condition struggled to take Polaris's perspective. As a comparison, almost equal number of participants selected "struggling to take Callisto's perspective" option. This was not expected at all, however, one explanation that comes to mind is the robot that makes mistake is perceived as more unpredictable.



(a) Participants evaluating Polaris's performance



(b) Participants evaluating their performance in the game

Figure 8.20 – Participants evaluation of the Polaris's performance in the game (a) and evaluation of their performance in taking Callisto's perspective.

8.6 Discussions

Figure 8.17 shows that participants in the cooperative and Limited-PT condition made the highest number of wrong moves compared to the other groups. Another interesting finding illustrated in the Figure is that the participants in Full-PT condition had similar distributions of total wrong moves in either the cooperative or competitive game. This shows that while the games had different goals and concepts, the general performance was still a function of players understanding of the game and perspective taking abilities as long as the other player showed a near perfect perspective taking behaviour. On the other hand, as soon as the robot cooperating with the human showed a limited and faulty performance, it affected the overall performance of the participant. It can be intuitive that when we cooperate with others, our abilities and performance can affect other members of the group and overall performance. In the case of our cooperative game, the cooperation came in the format of guiding and following the other teammate. This means, the participant's performance was not just a function of their perspective taking abilities but also the robot's abilities. This can explain why the total number of wrong moves were significantly higher than in Limited-PT condition compared to the Full-PT condition. If your teammate has limited perspective taking abilities and guides you in the wrong direction, you are more likely to make a wrong move compared to the one that has perfect abilities. On the other hand, in the competitive game there is no significant difference in the total wrong moves made between Limited-PT and Full-PT conditions. Looking at Figure 8.17, we can see participants in the competitive condition have not been affected by the robot's perspective taking abilities, while, the design of the task and the competition can have a large impact in how the participants.

8.6.1 Limitations of the Study

Perhaps the biggest two limitation of the studies presented in this chapter are being virtual and with adult participants. As we mentioned, the course of this thesis and particularly Chapters 7 and 8 was modified due to the restrictions the prevented us to run in person studies with children. That was also the biggest contributor to the way we designed the virtual platform and evaluated the model through participant's perception. An overall overview of the limitations:

Ideally, we wanted the experiment to happen using embodied robots and in person interaction. Using embodied robot would have let us to keep the continuity with the previous studies. Furthermore, it is considered more intuitive to study spatial perspective taking using embodiment rather than virtual interaction.

The initial goal of this thesis was to develop a perspective taking model for robots in child-robot interaction. All of our exploratory studies were carried out with children and in principle, we would have preferred to test the final platform and evaluate the model using children as well. However, as mentioned before, we were unable to perform our final experiments with children and we consider this as one of the biggest limitation of this chapter and perhaps thesis. Despite that, we still received valuable insights from adults and it still helped us to make some conclusions with respect to the model and the platform.

To recruit participants, we used the crowdworking platform called Prolific. The participants were screened for having high approval rates and only the data from the ones who completed the whole interaction and did not report any bug were approved for analyses. We did not put any limit on the age and gender of the participants, which resulted in having more young adults age groups and more male than females. This did not affect the results as the statistics of the self reported measures from Interpersonal Reactivity Index (IRI) and Spatial Representation Questionnaires showed no significant difference between the scores. However, since the study were carried out online using a browser, we could not screen the approaches the participants used to take perspectives or respond to the questions.

8.6.2 Future Developments

In both study 4 and 5, the participants we asked about their suggestion for adapting the games to children. Their responses and comments ranged from reporting bugs in the game and how to make the game more child friendly to how a simple interaction like this has helped them to practice spatial perspective taking. We are planning to incorporate these comments and what we learned from the analyses to improve the game for future studies in this topic.

Running the experiment with children
We have already adjusted the pretest and posttests to children and the games are also designed in different difficulty formats. We are also planning to run an online experiment with children aligned with the research presented here, yet beyond the scope of this thesis.

Developing the same interaction with embodied robots
The game and platform developed in this chapter is inspired by the one presented in Chapter 6. In the Cozmo Maze platform, we used Cozmo robot and printed the maze on A4 paper. While designing the virtual maze, we kept the possibility of designing a version with Cozmo or Vector robot in mind. As a future research direction, the results from the embodied and virtual developments can be compared for evaluating spatial perspective taking behaviours in these two settings.

Improving the robot's perspective taking model
The model implemented in the robot can be improved to adapt the player's perspective abilities, particularly in the cooperative version of the game. As observed in the results, participants rated the robot with full-PT abilities as more intelligence, and the game as more fun. This showed, the participants were more perceptive of the the other robot's abilities in the cooperative game compared to the competitive one. This provides us with insights on studying and developing perspective taking abilities for robots in cooperative interaction.

Improving the game-play
During the study 4 and 5, we received some valuable feedback from the participants to help us improve the game. Furthermore, the game design gives us flexibility to experiment and develop

different interaction modalities. Considering the participants significant improvement in the posttest, the interaction can be used as a perspective taking intervention and for practices.

8.7 Conclusions

This chapter presented two user studies that aimed at evaluating user's perception of a virtual robot's perspective taking abilities. First, we equipped a virtual robot with CogPeT and called its perspective taking abilities, "Full-PT". To have a ground for comparison, we developed another virtual robot that makes occasional mistakes when there is a perspective mismatch and called it "Limited-PT". Then, we designed a between-subject experiment, where the participants either interact with the Full-PT or Limited-PT robot. To make sure the perception is not completely a function of the type of interaction, we also developed a platform that encompasses two modes of interaction in the context of competitive and cooperative games. In study 4, we let participants experience both competitive and cooperative interactions and we analyzed their performance in pretest and posttest and their performance in both interactions. Then we evaluated their perception of the robot and the games. We noted a difference in the participant's performance between the two types of games. However, their evaluation of the robot's abilities and the games were confounded due to playing both games in one set. As a result, we designed a second experiment, where the participants only experienced playing either the competitive or cooperative game. Same as previous study the robot's perspective taking abilities was considered to be a between-subject variable. In the second experiment, we were able to observe significant differences in the way the robot's PT abilities affected the participants' performance and perception depending on the context of the game. In the competitive condition, participants performance was not affected by the robot's PT abilities and that also reflect on their perception of the robot. participants in this condition perceived both robots as equally intelligent and perceived the games as equally fun and difficult. On the other hand, in the cooperative condition; where wining the game was a function of the participant and the robot cooperating with each other, the robot's PT abilities affected the participants' performance. This in turn affected their perception of the robot's intelligence as participants rated the robot with Full-PT significantly more intelligent than the one with Limited-PT. Similarly, this affected their perception of the game fun, where the game with Full-PT robot was rated significantly more fun than the one with limited-PT. Surprisingly, the participants perception of the game difficulty was not affected by the robot's PT abilities.

On the other hand, we observed that participants performance improved significantly in both studies. The improvement was significant for both PT conditions and game contexts. While the improvement after playing the cooperative game was statically more significant than playing the competitive game, we can confirm that both contexts can be used for training spatial perspective taking abilities in adults. Furthermore, the results from the participants perception of the robot's perspective taking abilities can contribute to the modelling and developments of robot's abilities in cooperative tasks.

Future work in this topic and using the developed model and platforms can test the system with children and observe how the competitive vs. cooperative priming affects their performance and perceptions. Moreover, the model and platform described in this study is compatible to be incorporated in embodied robots and physical maze. This can provides us with a ground for comparing the result form the virtual maze with a physical one.

Summary and Conclusions Part V

9 Conclusions and Outlook

The purpose of this chapter is threefold. First, it gives an overview of the findings and contributions of this thesis, followed up by a discussion that highlights the limitations of the work and potential solutions. Finally, it presents the research emerging from this thesis and its future directions.

9.1 Overview and Contributions of the Thesis

The main contribution of this thesis is the study of spatial perspective taking using the interaction between children and robots, which led to a cognitive model of perspective taking (CogPeT) that can be integrated with any symbolic agent architecture. To improve the agent's perspective taking abilities, the model can include more processes and functions and can be expanded to cognitive and affective perspective taking dimensions. This thesis has made an effort to address the topic of perspective taking in various domains for developing interactions with children and adults.

First, the cognitive abilities required to improve the interaction; particularly the spatial perspective taking, were detailed and examined. Before any modelling or implementation in the robot started, three exploratory studies were carried out to study the required components, document children's behaviour, and discover the needs for improvements. The first study's contribution to the research was more related to the interaction and robot behaviours rather than the perspective taking aspect. The second and third studies were heavily focused on activities with a spatial perspective taking core and studied the prospective components to be included in the model. The main limitation of the user studies was the limited number of participants which restricted the analyses and results. Nevertheless, the overall results showed the potential of developed activities for improving children spatial perspective taking abilities, given the robot is equipped with a perspective model that can foster or challenge the child's abilities.

The research in developmental psychology contributed a lot in formulating the basic processes of the model. The underlying mechanisms of perspective taking in adults and children have shown to be highly dependent on the level of perspective comprehension. Regardless of the degree of difficulty to take other's perspective, there is always some processes that emerge spontaneously and some after deliberation. Furthermore, there is no consensus on how exactly some perspective taking behaviours emerge in children and adults. To address this issue, we decided to develop a model with flexible format for the agent. The model includes the mechanisms for automatic and cognitive controlled processes. Any process that is used for deliberate adaptation of the agent's perspective is categorized as a cognitive control process. All the processes were inspired by the literature on the topic and the exploratory studies.

The exploratory approach to find and select components to be implemented in the model is just one way of developing a cognitive model of perspective taking and one of the contributions of this thesis. The other contribution is to investigate and develop interactions that can foster perspective taking abilities in children and adults. While the reported developments were focused on the perceptual domain, the main ambition of this thesis is to contribute to the research on fostering cognitive and affective perspective taking developments in children through practices with robots and agents.

The research presented in this thesis contributes to the theoretical and modelling of the robot's behaviour. Additionally, it provides a set of activities developed for children and robots with the potential to be used as standalone platforms in both educational and game-based learning scenarios. While the thesis presentation is focused on describing the user studies, model development, and evaluation, the studies are heavily embedded in considerable research to find and develop the activities that can provide the most appropriate answers to each research question. Respectively, We hope this research opens up future directions on studying perspective taking in various research areas, including child-robot interaction, cognitive science, prosociality in robotics, and transparency in robotics and AI with particular attention to utilizing or getting inspired from the platforms and activities developed throughout this thesis.

9.2 Limitations

Before presenting the limitations, it is critical to mention that the original plan for evaluating the model was to incorporate the model in the objects game and Cozmo maze platforms presented in Chapters 5 an 6, respectively. Essentially, the goal was to evaluate how the model behaves when there is a need to adapt frame of reference and perspective marking while interacting with children. However, we had to reconsider the modelling section and completely restructure the evaluation method due to the limitation imposed by the pandemic that prevented us from running studies in schools. The move from using embodiment which is a fundamental factor in utilizing spatial perspective taking to virtual made us redesign the evaluation study to what was presented in Chapter 8.

9.2.1 Designing Experiments with Children

The presented work aims at developing the robot's cognitive abilities for interaction with children. Our main focus was to develop the activities and the platforms with children in mind and to communicate with teachers and parents to adapt the designs to children's abilities and cognitive level. We also made sure to pilot the systems with children and to incorporate the comments received from children, parents and teachers. Nevertheless, developing technologies and activities for children is far more challenging than it is for adults. One of our main limitations was designing the interaction and technologies before discussion with teachers and children. Generally, we designed the activity first, then reiterated and modified it after running a pilot and discussing it with teachers and parents. To include children in the design phase required an infrastructure that facilitates collaboration with schools and children, an option that was not available to us. However, such considerations can be beneficial for future researchers in the field of child-robot interaction.

9.2.2 Experiments with Small Sample Sizes

Another limitation of running studies with children was having a small sample size in all the first three studies. We tried to recruit more children and contact more schools, but depending on the school curricula and their availability to run experiments, we were met with certain constraints regarding the time slots that we could go to school. For example, in the CoReader study (Chapter 4), children were recruited from an international school in Geneva. The experiment was carried out just before the start of the Christmas holidays per school availability. Furthermore, the study required the student to participate in two sessions on two different days. Being close to the holidays, some students missed the second session which resulted in exclusion from the analyses. In the second study (Object Game, Chapter 5), we had to recruit the participants from two grades to be able to have an acceptable number of participants. Furthermore, one of the common issues we faced was the students who could not participate in the experiment because their parents did not sign the consent form. Some of those students were repeatedly asking the teacher to let them participate in the study, while we were unable to accommodate their request due to the ethical protocols. In the Cozmo Maze study (Chapter 6), for the second time, we were given a time slot to run the study before the Christmas school holidays. The school preferred to only allocate that time slot to us since it was after the end of the coursework. In this case, while we had consent approval from more than 30 students, around a third of the participants could not attend the school on the last day of the experiment due to a heavy storm. Having a large sample size with children is even more important than studies with adults. The age group that we targeted includes children that are still in the developmental stages of their cognitive development. Even if we select the participants from a narrow age groups, they still show high variability in their behaviour. Not being able to group them based on these abilities before or after running the experiment, put a constraint on evaluating their performance and behaviour on the given task. Just as an example, the main findings of the CoReader study is observed after accounting for children's reading abilities which was a measure provided by the teacher before the start of the experiment.

9.2.3 Limitations of the Model

The model is not as comprehensive as we expected it to be. It can be considered as a preliminary on the topic with the potential to be expanded by adding more processes required for taking the spatial perspective of others. It also does not detail how the processes work which was a research direction we originally had in mind. However, after opting for designing the virtual maze, which required an extensive amount of work, we decided to simplify the model and only focus on what is needed in the final study. Furthermore, the model does not dissociate between the processes that emerge as a result of verbal instructions and natural language. This is something that can be added to the model in future developments.

9.2.4 Evaluating the Model only with the Virtual Setting

The processes used in taking the perspective of others differ in a physical setting compared to a virtual one. While we made optimum effort to map the features used in the physical world to the virtual game, we are still aware that it cannot be a one-to-one mapping. During the testing period of the virtual platform, we had mixed comments on changing the way the participants saw participants. We also experimented with adding a zooming in and out feature and including an extra screen with a top view. Ultimately, we selected the current design to limit the cognitive demand on the player and give them a chance to only focus on the game and the robot's perspective. Despite all, we are well aware that the current design is still not optimal and there is always room for improvement.

9.2.5 Evaluating the Model Only with Adults Participants

The behaviour that adults showed cannot be generalized to children. Both for the way they performed in the tests and the game and for their perception of the robots. Furthermore, evaluating the model with children would have provided us with more data points to compare with the exploratory studies.

9.3 Future Directions

All the chapters that included a user study detailed the possible future directions of the research presented in that chapter, this section focuses on the overall future research directions of the thesis.

9.3.1 Perspective Taking Interventions for Children

The activities designed for the exploratory studies all had the ultimate goal of being used as standalone platforms to assist children in developing their skills. In the first study, the context was reading with the ultimate goal of discovering features and robot behaviours that increase children's joint attention. Additionally, we were hoping to integrate the activity with interventions aimed at helping children with reading difficulties. The perspective taking studies can be developed into game-based interventions with the goal of practising skills associated with spatial perspective taking. The virtual maze has already shown a significant improvement in adult's performance in the test. If such results persist after running the experiment with children, we might be able to consider such platforms for children's practices and as an intervention method.

9.3.2 Expanding the Model to Cognitive and Affective Dimensions

One of the model limitations was not considering all the processes used in visual and spatial perspective taking. Consequently, adding such processes to the model can be considered as a future direction to improve the model. This can particularly target the different processes used in visual in comparison to spatial perspective taking. Furthermore, by including the processes needed in cognitive and affective perspective taking, the model can be tailored for socially demanding scenarios.

9.3.3 Relating Perspective Taking with Prosocial Behaviour in Robots

Recent advances in social cognitive theory and child-robot interaction have tried to understand the extent to which social robots can trigger prosocial behaviour in children (Peter et al., 2021). Paiva et al. proposes application cases where autonomous agents and robots can be used to foster prosocial behaviour in societies (Paiva et al., 2018). While more research is needed to connect perspective taking with prosocial behaviour in robotics, the early studies in psychology can be a good starting point (Bengtsson & Johnson, 1992; Zahn-Waxler et al., 1977). A study by Grant and Berry have researched the link between prosocial behaviour and perspective taking on creativity, They found that perspective taking mediates the moderating effects of prosocial behaviour (Grant & Berry, 2011). Children interacting with robots that can take their perspective and might encourage them to take the perspective of others can inspire studies that investigate the links with developing prosocial behaviours in children. Furthermore, perspective taking abilities can contribute to developing robots with more prosocial functionalities.

9.3.4 Relating Perspective Taking with Transparency in Robotics and AI

Transparency in robotics is defined as the robot's ability to explain its actions (Kim & Hinds, 2006). In recent years, the research has shifted toward creating more transparent robots and artificial intelligence (Felzmann et al., 2019). Research suggests that if a robot provides more explanations about its actions, it can result in humans gaining a better understanding of its actions (Kim & Hinds, 2006). Several recent studies have focused on developing technologies and cognitive models to equip robots with transparent behaviours, particularly in the interaction with children (Charisi et al., 2021; Johal, 2020; Westlund & Breazeal, 2016). Having a robot with practical cognitive perspective taking abilities can help the robot evaluate the child's understanding of the interaction; through verbal and non-verbal features, and decide where to be more transparent or reduce the ambiguity in the interaction. Transparency is one of the major applications that explains the need for developing better cognitive models in robots and artificial agents.

9.4 Epilogue

To make decisions, humans not only rely on the input from the environment but also on their perceptions of others within that environment. Much like humans, we need robots that can make decisions by understanding humans intentions, in addition to using their sensory information from the physical world. This thesis has made few preliminary steps toward actualizing this goal, by getting inspiration from psychology, human-robot interaction, and cognitive modelling. While the core of the work is aimed at spatial perspective taking, its exploration and development can be extended to other disciplines and opens up new opportunities overall.

Appendices Part

A Robots and Platforms

A.1 Robot Operating System (ROS)

Robot Operating System (ROS)[I] is an open-source middleware for developing robotic applications (Quigley et al., 2001). ROS is not an operating system but a collection of frameworks with libraries and tools used to design, implement, and execute robotic applications. The power of ROS lies in its ability to provide developers with hardware abstraction, device drivers, libraries, message-passing processes, ready-to-use implementation of commonly-used robot functionalities, and package management. ROS is well-integrated with the majority of robots and sensors and allows for integration of various hardware and processes (within ROS called nodes) in a standardized manner. The first three studies described in Chapters 4, 5, and 6 all used ROS to incorporate the robots and other hardware in their platforms.

A.2 NAO Robot

NAO[II] is a commercially available humanoid robot developed by the French-based Aldebaran Robotics in 2008 and acquired by a Japanese-based company and rebranded as SoftBank Robotics in 2015. The robot is programmed with a specialised Linux-based operating system called NAOqi and it is accompanied with a graphical programming tools called Choregraphe and Monitor. So far more than 13,000 versions of the robot has been used in more than 70 countries around the world. NAO has been used as a standardized platform in several domains such as education, healthcare, autism therapy, and so on. The ROS driver for NAO robot was originally developed by Freiburg's Humanoid Robots Lab and Armin Hornung[III] and it provides the essential wrappers for NAOqi API in ROS. In this thesis, NAO robot (Figure A.1) has been used in developing interaction with children in studies presented in Chapter 4; the *CoReader Platform*, and Chapter 5; the *Objects Game Platform*.

[I] http://wiki.ros.org

[II] https://www.softbankrobotics.com/emea/en/nao

[III] http://wiki.ros.org/nao

Figure A.1 – NAO robot by Softbank robotics.

A.3 Cozmo and Vector Robots

Cozmo[IV] is a robot developed by Anki robotics and launched in 2016. Anki robotics shut down production in 2018 due to the lack of funding and later was acquired by Digital Dream Labs. The robot's design was inspired by Wall-E and Eve characters from Pixar and it was marketed as a toy. Cozmo comes with a free SDK that offers access to different robot's library to program its emotions and behaviour. The robot's SDK can be programmed in Python. No official ROS wrapper for Cozmo SDK has been released, however, the SDK can be used with ROS nodes with some extra developments and unofficial wrappers. While Cozmo is not equipped with speech recognition abilities, another robot released by Anki Called Vector[V] include this feature besides more open-source development functionalities. Both robots are shown in Figure A.2.

Figure A.2 – Cozmo and Vector robots by Anki Robotics.

[IV]https://anki.com/en-us/cozmo.html
[V]https://www.digitaldreamlabs.com/collections/vector-products

A.4 Unity Game Engine

Unity[VI] is a cross-platform game engine and integrated development environment (IDE) developed by Unity Technologies. The first version of Unity was launched in 2005 with the goal was to provide professional game development tools to amateur game developers. Unity gives developers the possibility of creating games in both 2D and 3D environments, with primary scripting done in c#. With Unity being a cross-platform engine, it supports several platforms for mobile, desktops, web, console, and virtual/extended reality platforms. The game developed in Chapter 8 was developed in Unity 3D and rendered for web platfrom using Unity WebGL[VII]. Example of the game and the robot developed with Unity environment is presented in Figure A.3.

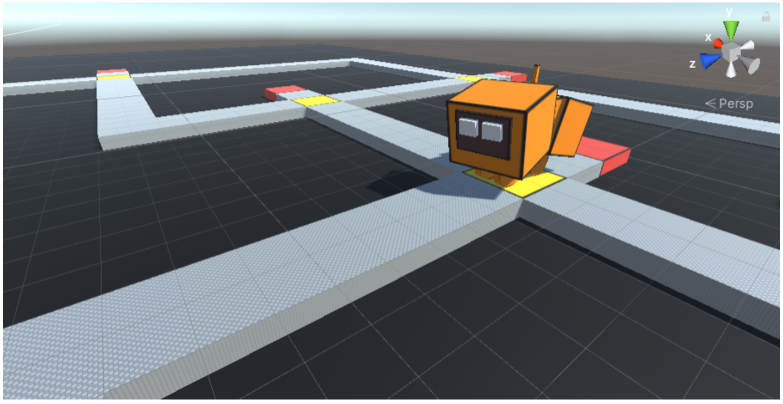


Figure A.3 – Game developed in Unity 3D environment.

A.5 Arduino

Arduino[VIII] is an open-source electronics platform used for building digital devices. The single-board microcontrollers and kits are great for developing tools and programmable hardware. They can be used to read inputs such as light on a sensor or finger on a button and turn that to outputs such as activating an actuator or turning of a LED. Arduino project provides an integrated development environment (IDE) that can be directly used with ROS. The push buttons used in Chapters 4 and 6 were developed using an Arduino board for receiving their input. The input from the Arduino was integrated with ROS to publish the press signal in real-time to ROS nodes, where it was delivered to the robot's program.

[VI]https://unity.com/
[VII]https://www.khronos.org/webgl/
[VIII]https://www.arduino.cc/

B Tasks and Games

B.1 CoReader

The schematic of the CoReader platform in addition to extra figures are provided here.

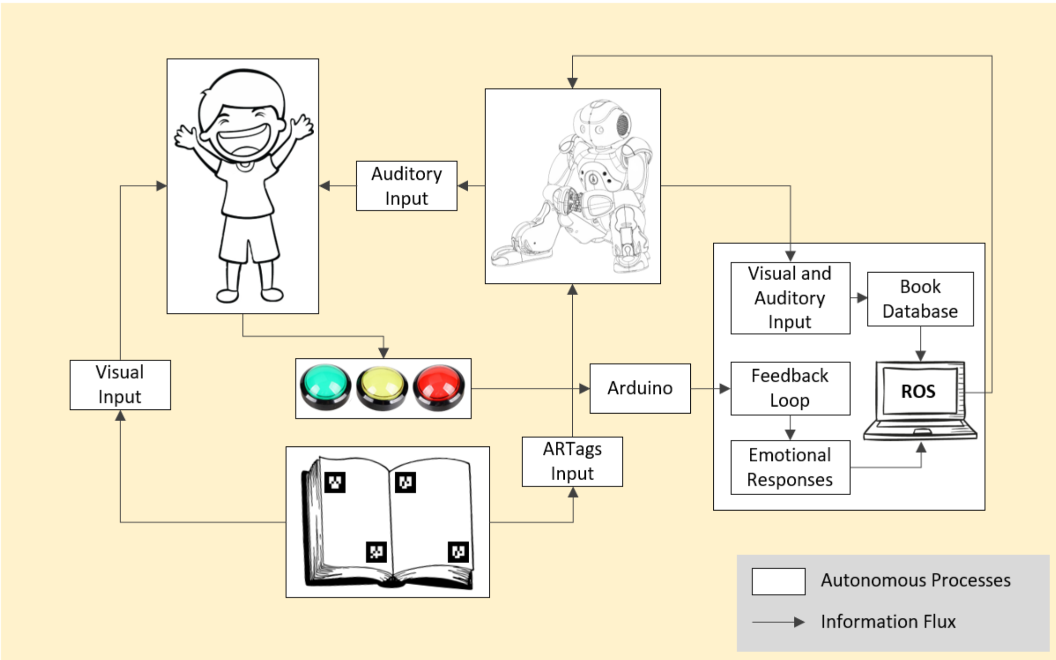


Figure B.1 – Schematic of the study 1.

B.2 Objects Game

The schematic of the Objects game platform in addition to extra figures are provided here.

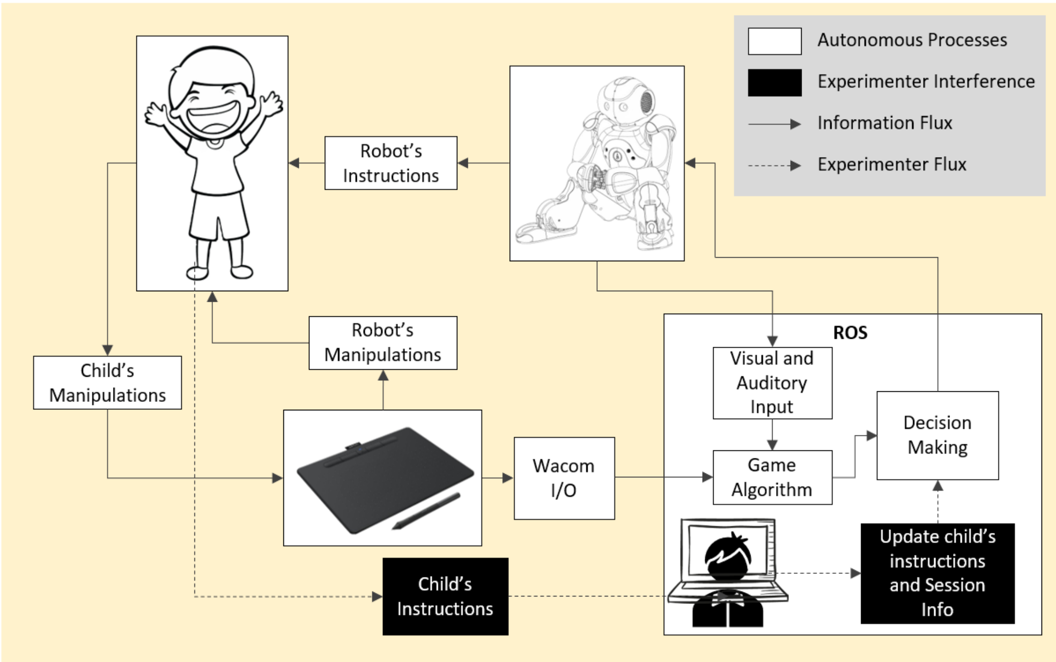


Figure B.2 – Schematic of the study 2.

B.3 Cozmo Maze

The schematic of the Objects game platform in addition to extra figures are provided here.

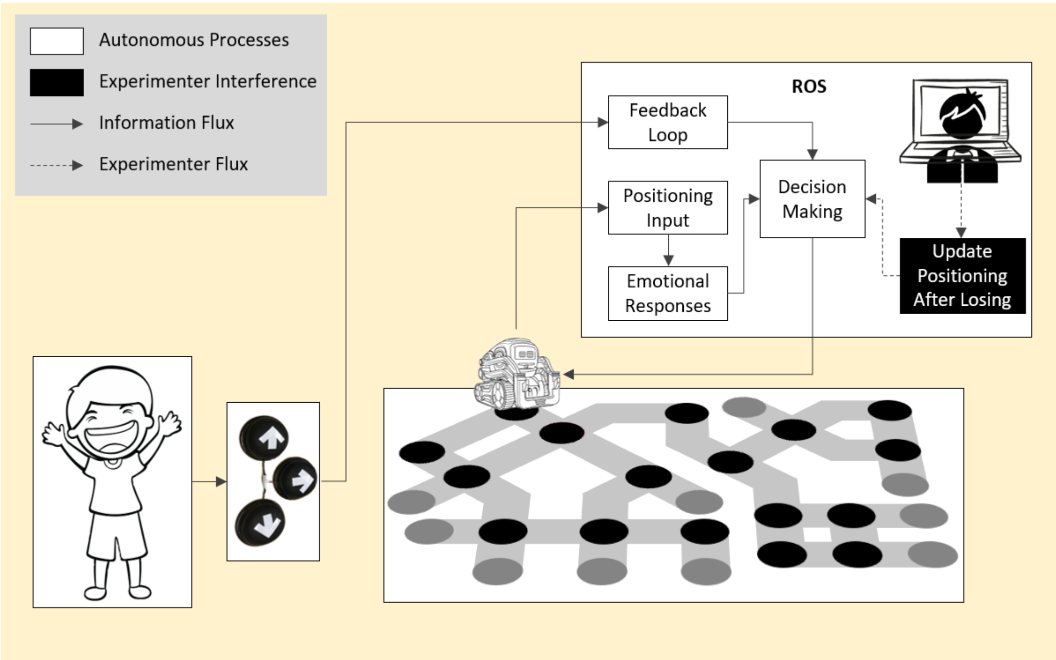


Figure B.3 – Schematic of the study 3.

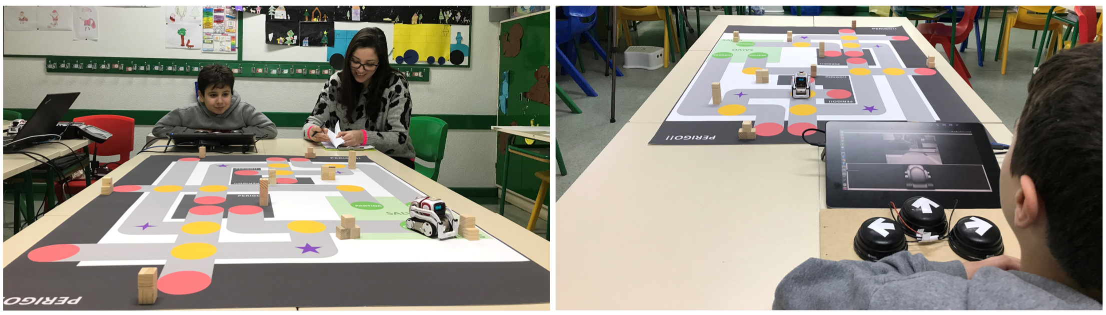


Figure B.4 – The experimental setup with the Cozmo maze.

B.4 Virtual Maze

Details given to Prolific participants to take part in the study are shown below.

In this study, we will ask you to play a game with a virtual robot, but before and after that we would like you to respond to some questionnaires related to the game.
Notes about the Game:

1. Before starting the game, please make sure your browser supports WebGL using the following link (you should see the cube spinning) https://get.webgl.org/
    - Please do not use Firefox as the game is more compatible with chromium-based browsers such as Chrome.
    - If you are using macOS Big Sur, please do not use chrome and opt for using Safari instead.
2. Please do not use Reload, Forward and Backward button on your browser.
3. To have a full-screen experience press the blue button on the right corner and press ESC to normal.
4. You will be asked about your Prolific ID, unfortunately, you can't paste it and need to type it down.

The interaction sequence provided to the participants for study 4 and study 5 are http://canyouguide.me/letsplaytogetherandcompete/, and http://canyouguide.me/, respectively. Both links are live and available for testing.

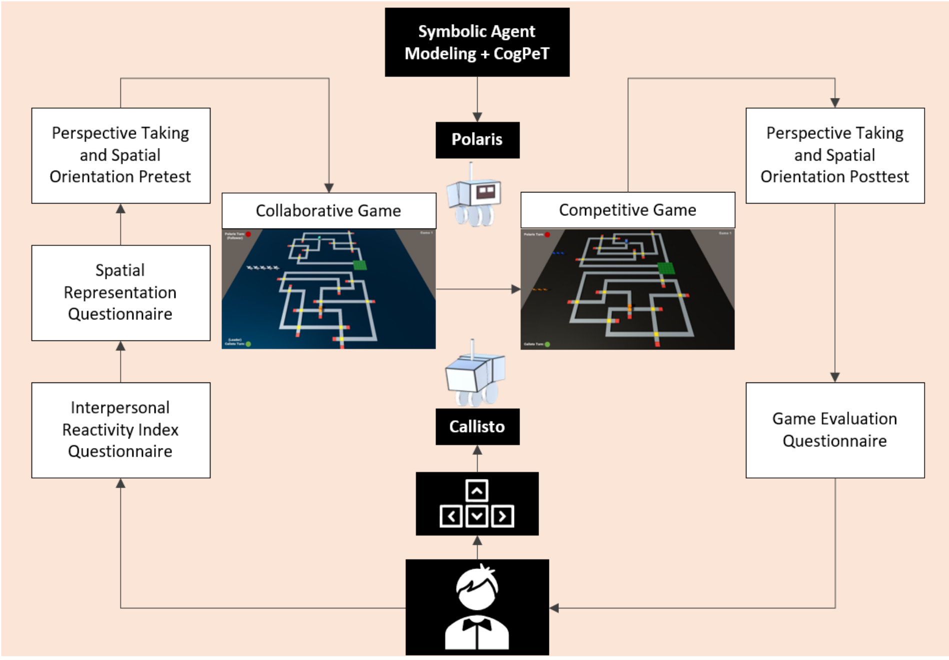


Figure B.5 – Schematic of the study 4.

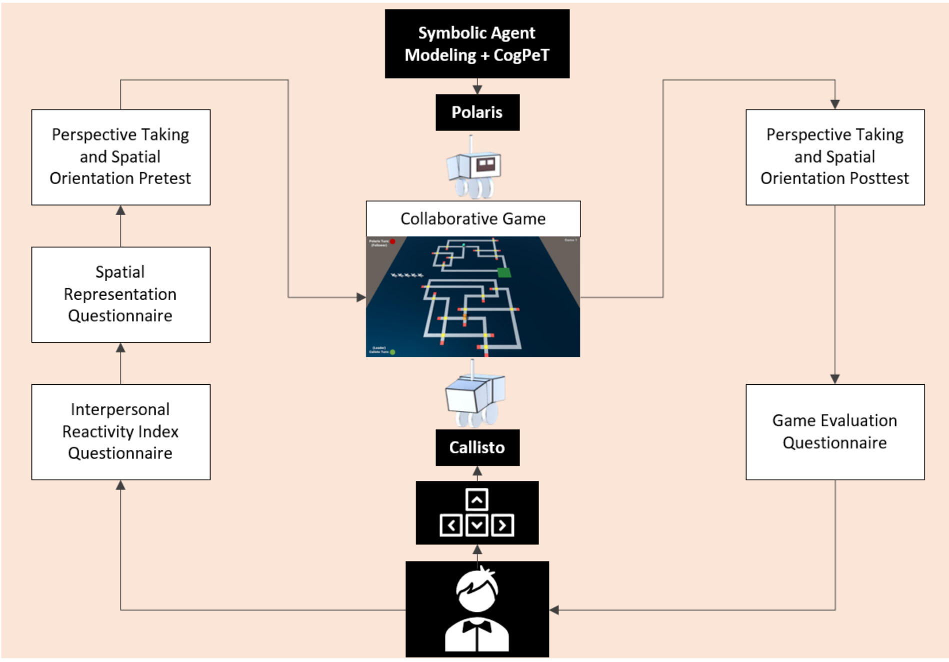


Figure B.6 – Schematic of the study 5.

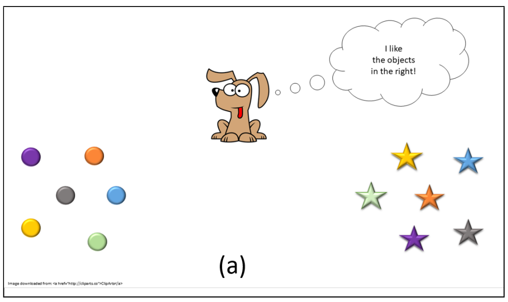
C Tests

C.1 Left/Right Test or Toys Test

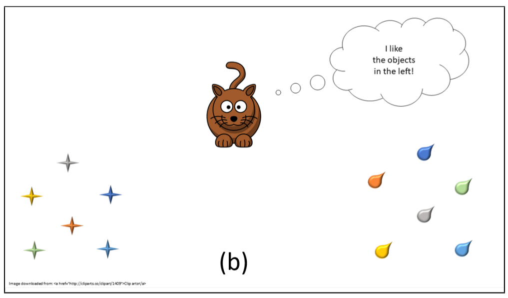
This test is designed to evaluate children's recognition of perspective difference between them and another agent (in this case an animal) that is facing them. Our evaluation is based on judging children's selection of the animal's favorite toy based on what animal expresses in the test. If the child takes the animal's perspective, we consider it a correct answer, otherwise, it is incorrect. Two similar versions of this test has been designed with two different animals and different correct responses. We alternated the tests as pretest and post-test between children. The instructions are as follows:

The dog/cat thinks they like the right/left toy . Can you tell me which toy does the dog/cat likes? (wait for the response).

In the dog version of the test (Figure C.1a), the dog says *"I like the right side toys"*, and in the cat version (Figure C.1b) it says *"I like the left side toys"*. Hence, the correct answer for the dog version is *the balls* as they are in the right side of the dog and the correct answer for the cat is *"the drops"*. For example, if the child's answer is *the stars*, then the child is egocentric, and in this case the answer is incorrect.



(a) Dog version



(b) Cat version

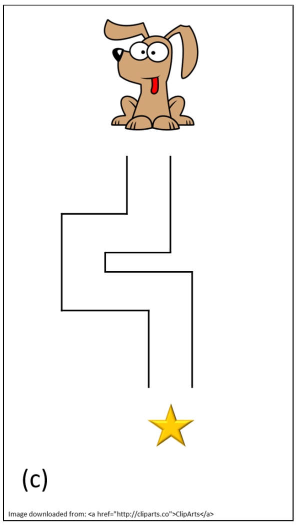
Figure C.1 – Examples of the pretests and posttests for left/right test or toys test.

C.2 Test of Direction Sense or Path Test

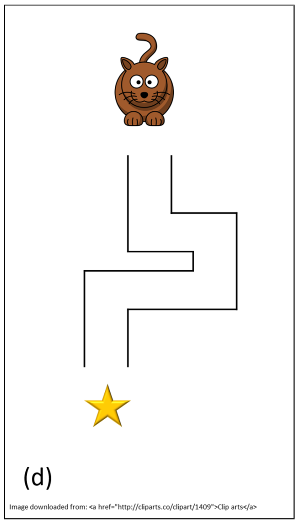
The path test shown in Figures 5.5c-5.5d is a simplified version of "Money Standardized Test of Direction Test" developed by Money et al. and modified by Zacks et al. which has been adapted for children. Again two versions of the test have been designed with different animals and directions and were alternated between the participants as pretests and post-tests. Similar to the toys test, this test is supposed to evaluate if children take the animal's perspective to guide them or not, and furthermore, if they do it correctly or not. Compared to the previous test, this test is more cognitively demanding. Looking at the test shown in Figure 5.5c, the child is supposed to guide the dog to reach the star. The instructions for this test are as follows:

> "The dog/cat wants to reach the star at the end of the road, can you describe the path that the dog/cat needs to take to reach the star? (help the child by saying 'move forward' and let them complete the instructions)."

Considering that we didn't give any specifics to children as to how to describe the path, we observed children use two approaches for guiding the animal. In one approach, their response is a combination that is better described as defining the path. For example to guide the dog, a correct sequence for this approach in Figure C.2a is *"Forward-Right-Forward-Left-Forward"*. In the second approach, they give the combination that is better described as walking with the animal in the path. For example the correct sequence for this approach is *"Front-Right-Front-Left-Front-Left-Front-Right-Front"*. Some children did not specify the moving forward part, nevertheless, as long as they still used the correct sequence of turns, we considered their answer as correct.



(a) Dog version



(b) Cat version

Figure C.2 – Examples of the pretests and posttests for test of direction sense or path test.

C.3 Mental Rotation Test

Several psychology studies specify the role of mental rotation in level 2 spatial perspective taking (Janczyk, 2013; Kessler & Rutherford, 2010; Kessler & Thomson, 2010; Kessler & Wang, 2012). To see if playing this game has a positive impact on children's mental rotation abilities, we have used the test shown in Figures 6.5a and 6.5b inspired from a study by Perrucci et al. (Perrucci et al., 2008). This test focuses on children's object rotation skills which, based on some studies, has different cognitive processes compared to mental self-rotation used in perspective taking (Kessler & Rutherford, 2010; Zacks & Michelon, 2005). However, it also includes specifying the panda's tribe which requires perspective taking to find them from the images. The test starts with instructions that the experimenter explains to the child before taking the test.

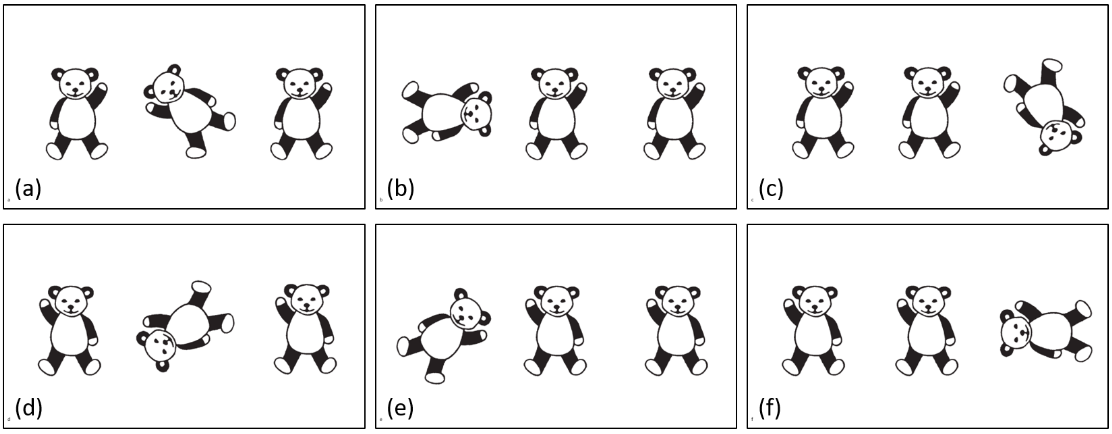


Figure C.3 – Pretest for mental rotation or panda test.

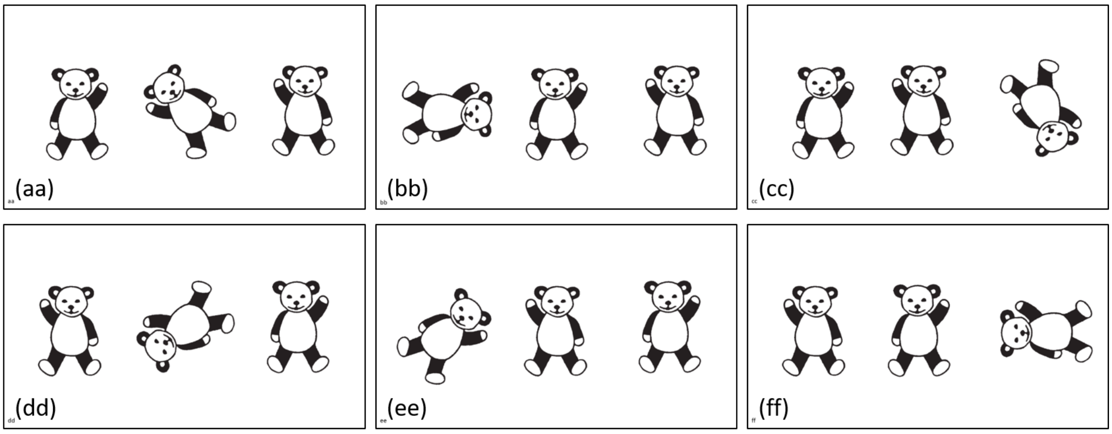


Figure C.4 – Posttest for mental rotation or panda test.

"There are two panda tribes. One panda tribe (show the children the panda R cards) always walks around with this arm raised (point to the three panda's right arm), and the pandas in the other tribe (leave the panda R cards in sight, and show the three panda L cards) always walk around with this arm raised (point to the panda's left arm). Can you see the difference? (Wait for the child to nod or say 'yes'.) Now, the pandas never hang out by themselves. They usually hang out with other pandas from their same tribe (show the two separate card groups), but not always. Sometimes, pandas from different tribes do hang out together. What you will do is tell me when pandas are from the same tribe or from different tribes. For example (show the three cards used in the first pretraining trial), two of these pandas are the same; they're from the same tribe, and one panda is from the other tribe. Which panda is different? (Wait for the child to respond. If the child responds correctly, ask the reason for his/her choice and continue with the second trial. If the child responds incorrectly, ask . . .) 'Why? Be careful! Which pandas are the same?' (Wait for the response.) 'And so, the different panda is. . . .' (Wait for the response and ask the reason for the child's choice.)"
Perrucci et al., 2008.

The pretest and post-test sets are different in one aspect. In the pretests, the panda that is from a different tribe is always the rotated one, which means just by looking at two straight pandas the child can conclude that the rotated panda is the different one. However, they also need to specify which tribe the panda belongs to, which means they still need to use mental rotation. On the other hand, in the post-test the one straight panda and one rotated panda are from the same tribe which means children definitely need to use mental rotation to find the different panda, and then specify the panda's tribe.

C.4 Perspective Taking/Spatial Orientation Test

Developed by Kozhevnikov and Hegarty, this test is designed to evaluate the participants' ability to imagine different perspective and orientations (Kozhevnikov & Hegarty, 2001). The instructions of the test as presented to the participants are presented below.

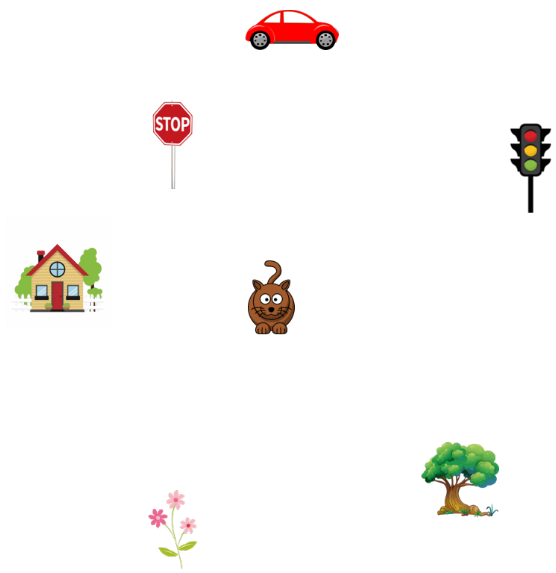
Adult Variation

This is a test of your ability to imagine different perspectives or orientations in space. On each of the following pages you will see a picture of an array of objects and an "arrow circle" with a question about the direction between some of the objects. For the question on each page, you should imagine that you are standing at one object in the array (which will be named in the center of the circle) and facing another object, named at the top of the circle. Your task is to draw an arrow from the center object showing the direction to a third object from this facing orientation.

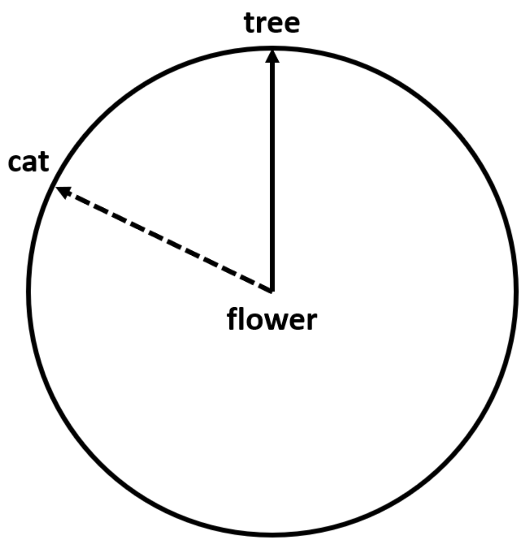
Look at the sample item on the next page. In this item you are asked to imagine that you are standing at the flower, which is named in the center of the circle, and facing the tree, which is named at the top of the circle. Your task is to draw an arrow pointing to the cat. In the sample item this arrow has been drawn for you. In the test items, your task is to draw this arrow. Can you see that if you were at the flower facing the tree, the cat would be in this direction? Please ask the experimenter now if you have any questions about what you are required to do.

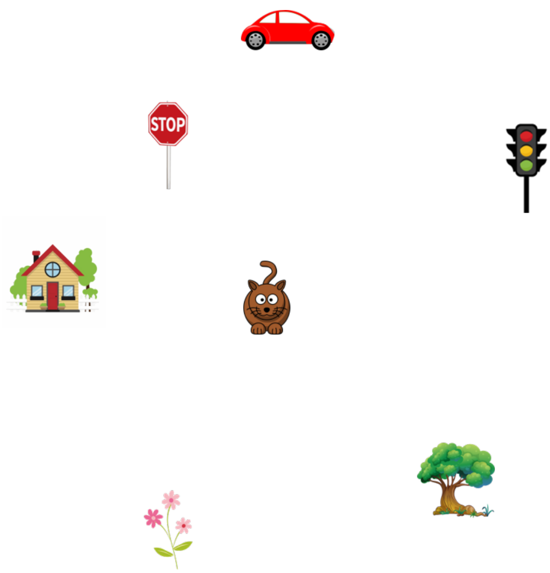
There are 12 items in this test, one on each page. For each item, the array of objects is shown at the top of the page and the arrow circle is shown at the bottom. Please do not pick up or turn the test book¬let, and do not make any marks on the maps. Try to mark the correct directions but do not spend too much time on any one question.

You will have 5 minutes for this test.

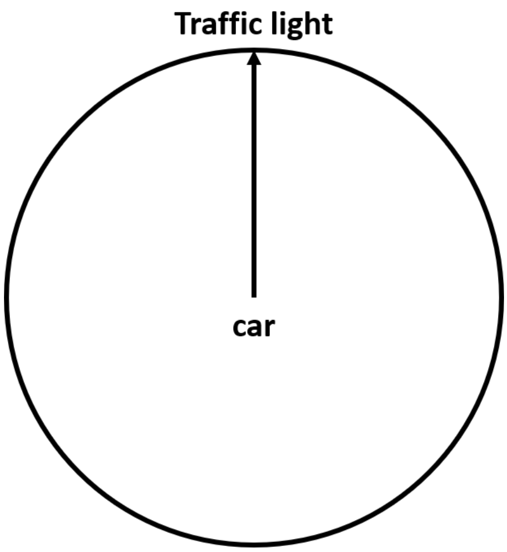


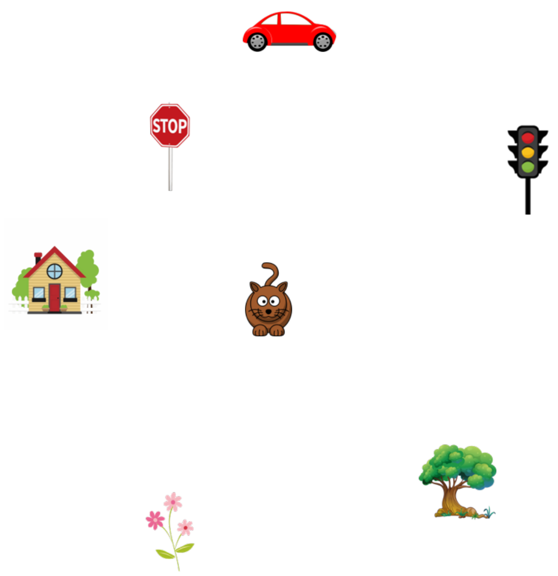
Example:
Imagine you are standing at the **flower** and facing the **tree**.
Point to the **cat**.



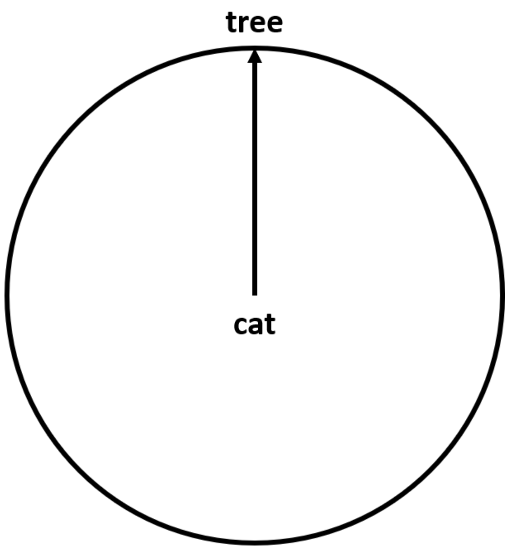


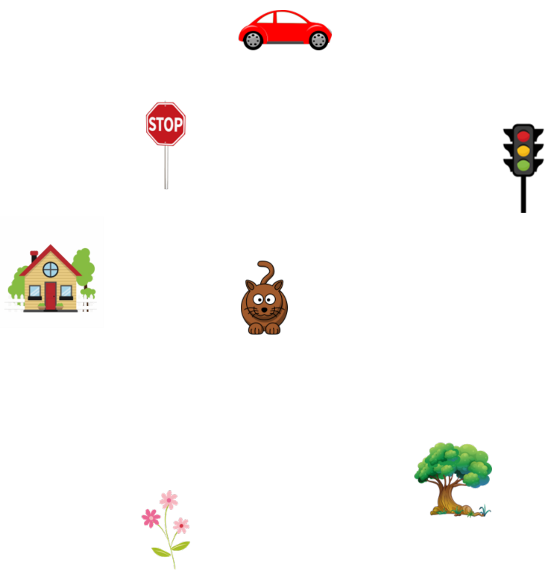
1. Imagine you are standing at the **car** and facing the **traffic light**.
Point to the **stop sign**.



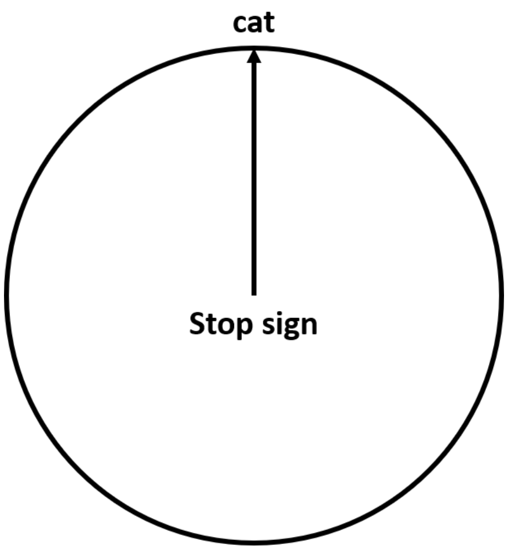


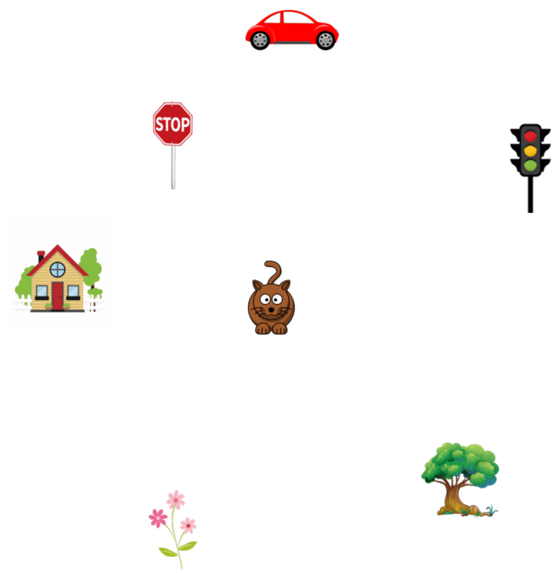
2. Imagine you are standing at the **cat** and facing the **tree**.
Point to the **car**.



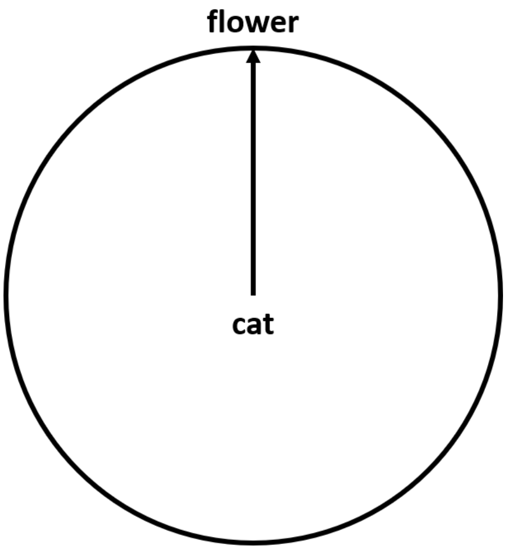


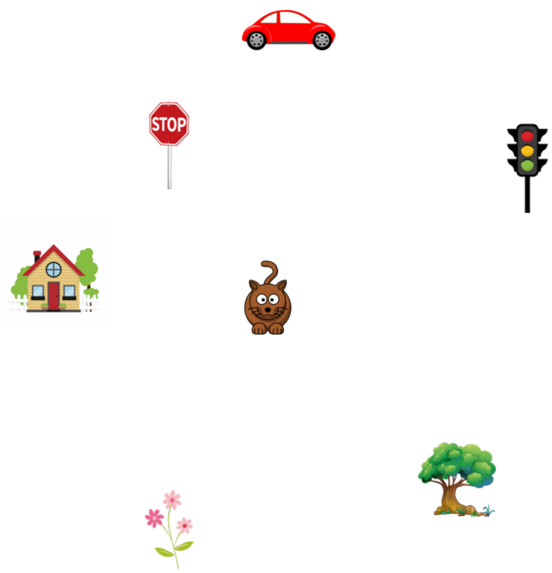
3. Imagine you are standing at the **stop sign** and facing the **cat**.
Point to the **house**.



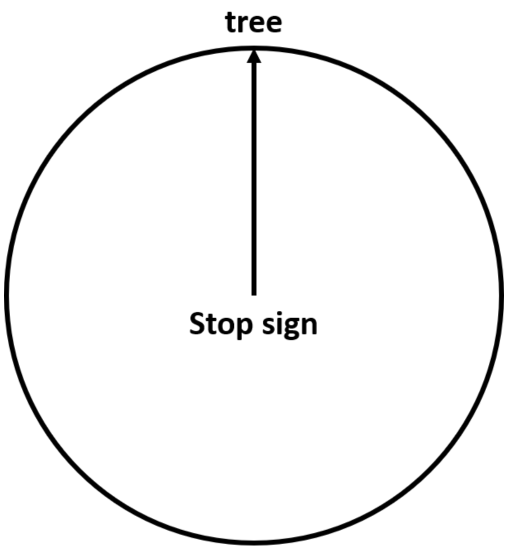


4. Imagine you are standing at the **cat** and facing the **flower**.
Point to the **car**.

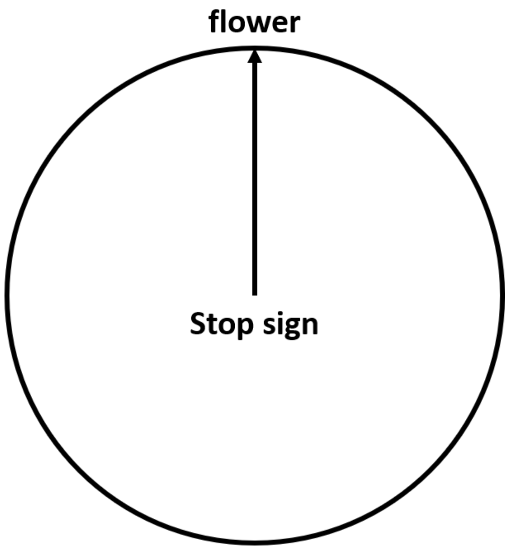


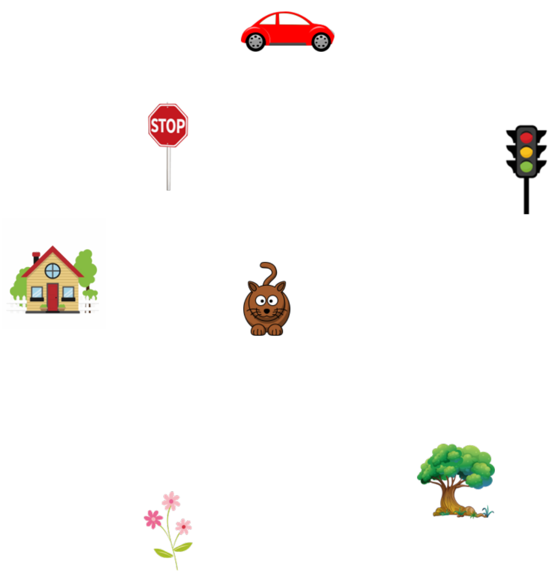


5. Imagine you are standing at the **stop sign** and facing the **tree**.
Point to the **traffic light**.

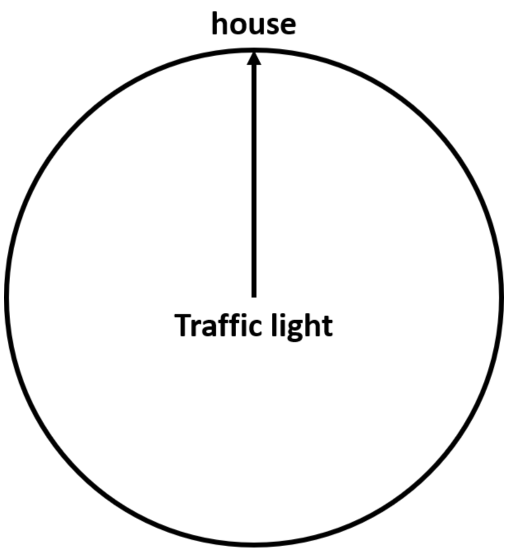


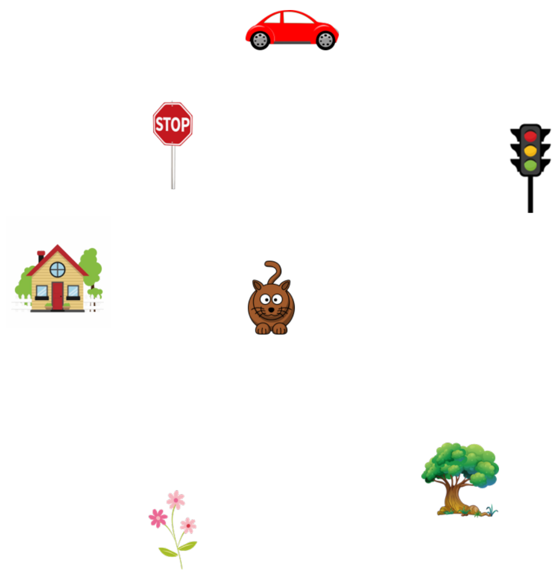
6. Imagine you are standing at the **stop sign** and facing the **flower**.
Point to the **car**.



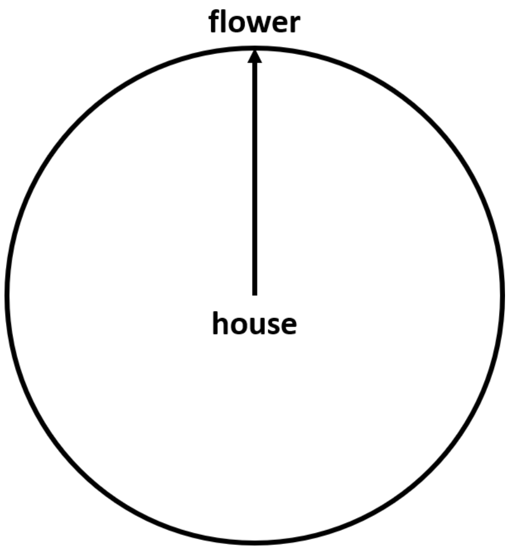


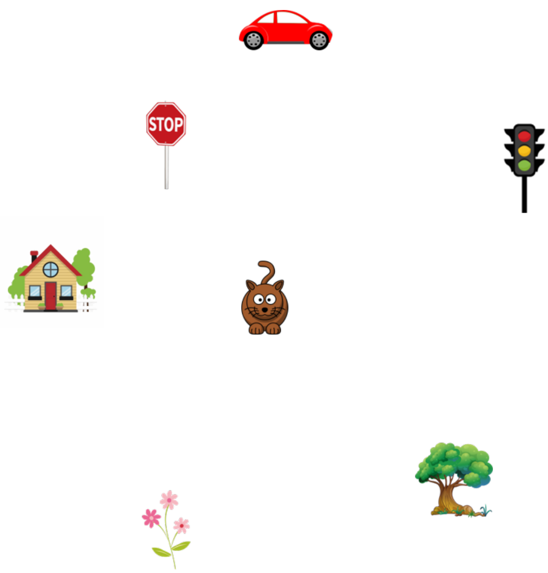
7. Imagine you are standing at the **traffic light** and facing the **house**.
Point to the **flower**.



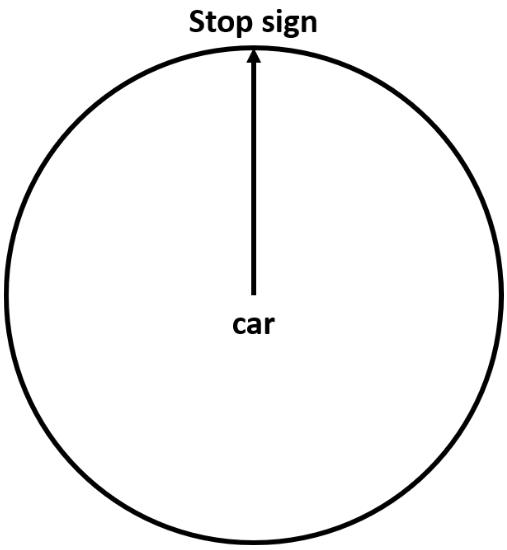


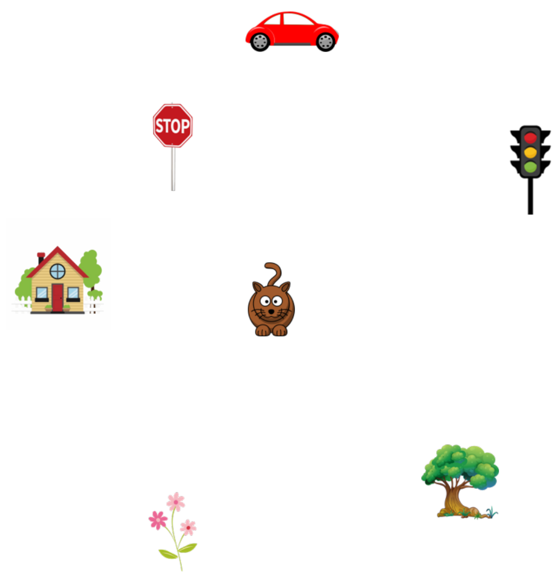
8. Imagine you are standing at the **house** and facing the **flower**.
Point to the **stop sign**.



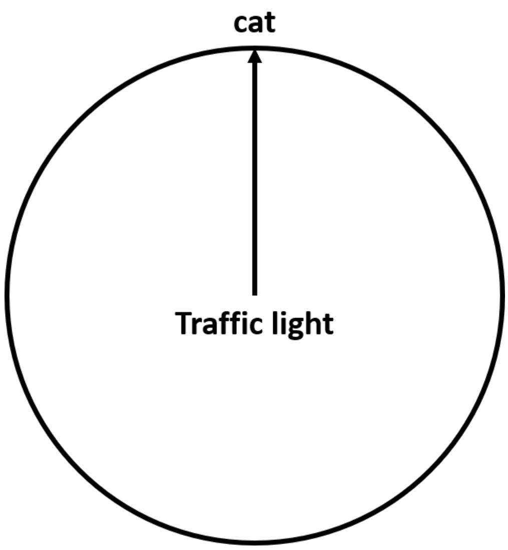


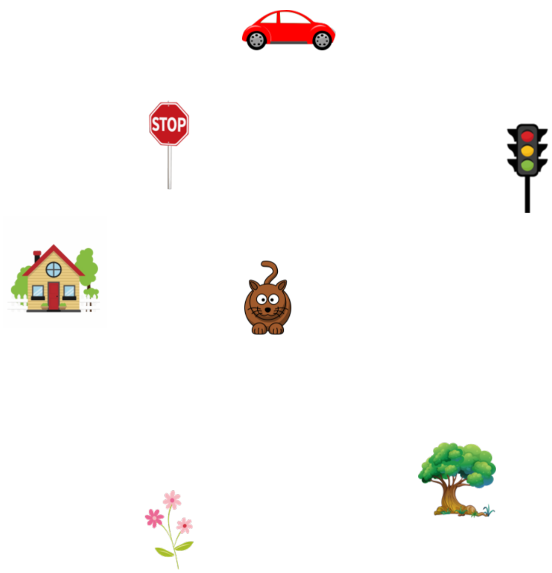
9. Imagine you are standing at the **car** and facing the **stop sign**.
Point to the **tree**.



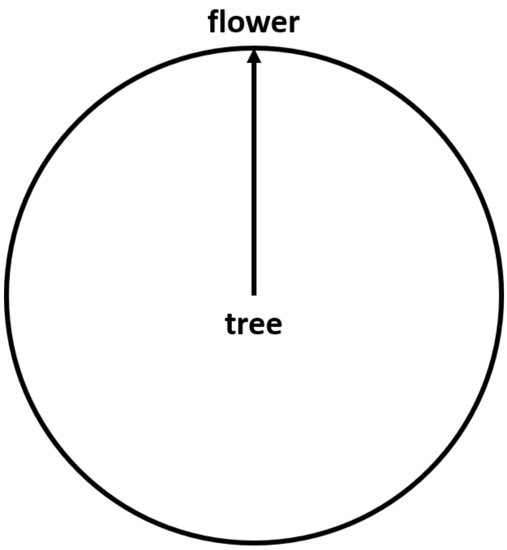


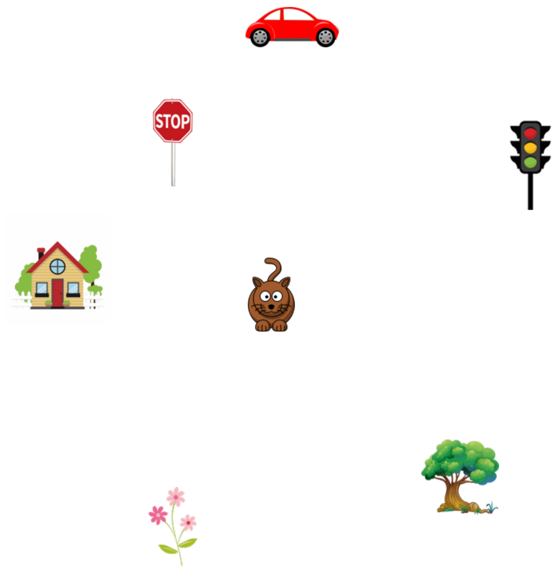
10. Imagine you are standing at the **traffic light** and facing the **cat**.
Point to the **car**.



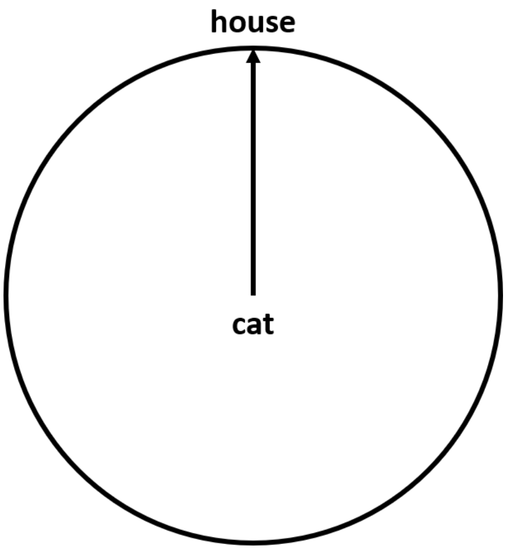


11. Imagine you are standing at the **tree** and facing the **flower**.
Point to the **house**.





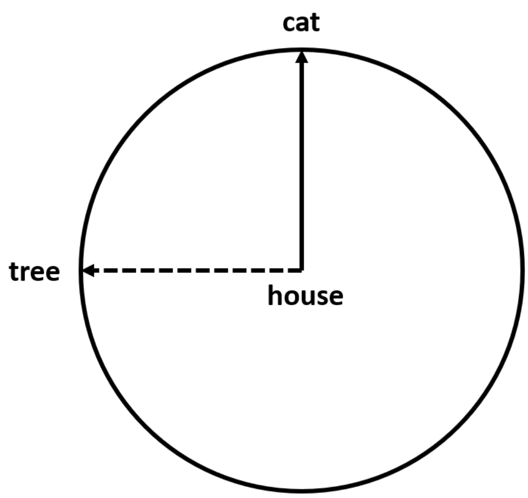
12. Imagine you are standing at the **cat** and facing the **house**.
Point to the **traffic light**.



Children Variation

Developed by Mary Hegarty, Maria Kozhevnikov, David Waller and modified for children by Elmira Yadollahi and Marta Couto Kozhevnikov and Hegarty. We present one sample of the test here.

Example:
Imagine you are standing at the **house** and facing the **cat**.
Point to the **tree**.



D Questionnaires

D.1 Interpersonal Reactivity Index Questionnaire

This test was developed by Davis et al. and it consists of 4 sub-scales of Perspective Taking (PT), Fantasy (FS), Empathic Concern (EC), Personal Distress (PD) with 7 questions each (Davis et al., 1980). The test comes with 28-items on a 5-point Likert scale that ranges from “Does not describe me well” to “Describes me well”. For our experiments, we have only used the PT scale. The following section shows the test as presented in the experiment.

NOTE: (-) denotes item to be scored in reverse fashion
A = 0
B = 1
C = 2
D = 3
E = 4
Except for reversed-scored items, which are scored:
A = 4
B = 3
C = 2
D = 1
E = 0

Perspective Taking Sub-Scale
The following statements inquire about your thoughts and feelings in a variety of situations. For each item, indicate how well it describes you by choosing the appropriate letter on the scale at the top of the page: A, B, C, D, or E. When you have decided on your answer, fill in the letter next to the item number. READ EACH ITEM CAREFULLY BEFORE RESPONDING. Answer as honestly as you can. Thank you.

1. I sometimes find it difficult to see things from the "other guy’s" point of view. (PT) (-)

☐ A ☐ B ☐ C ☐ D ☐ E

Does not describes me well — Describes me well

2. I try to look at everybody's side of a disagreement before I make a decision. (PT)

☐ A ☐ B ☐ C ☐ D ☐ E

Does not describes me well — Describes me well

3. I sometimes try to understand my friends better by imagining how things look from their perspective. (PT)

☐ A ☐ B ☐ C ☐ D ☐ E

Does not describes me well — Describes me well

4. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (PT) (-)

☐ A ☐ B ☐ C ☐ D ☐ E

Does not describes me well — Describes me well

5. I believe that there are two sides to every question and try to look at them both. (PT)

☐ A ☐ B ☐ C ☐ D ☐ E

Does not describes me well — Describes me well

6. When I'm upset at someone, I usually try to "put myself in his shoes" for a while. (PT)

☐ A ☐ B ☐ C ☐ D ☐ E

Does not describes me well — Describes me well

7. Before criticizing somebody, I try to imagine how I would feel if I were in their place. (PT)

☐ A	☐ B	☐ C	☐ D	☐ E
Does not describes me well				Describes me well

D.2 Spatial Representation Questionnaire

This test was developed by Pazzaglia and De Beni (Pazzaglia & De Beni, 2001). For our experiments, we have only used the PT scale. The following section shows the test as presented in the experiment.

1. Do you think you have a good sense of direction?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

2. Are you considered by your family or friends to have a good sense of direction?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

3. Think about the way you orient yourself in different environments around you. Would you describe yourself as a person:

a. who orients him/herself by remembering routes connecting one place to another?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

b. who orients him/herself by looking for well-known landmarks?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

c. who tries to create a mental map of the environment?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

4. Think of an unfamiliar city. Write the name . . . .
Now try to classify your representation of the city:

a. survey representation, that is a map-like representation

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

b. route representation, based on memorising routes

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

c. landmark-centred representation, based on memorising single salient landmarks (such as monuments, buildings, crossroads, etc.)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

5. When you are in a natural, open environment (mountains, seaside, country) do you naturally individuate cardinal points, that is where north, south, east, and west are?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

6. When you are in your city do you naturally individuate cardinal points, that is do you and easily where north, south, east, and west are?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

7. Someone is describing for you the route to reach an unfamiliar place. Do you prefer:

a. to make an image of the route?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

b. to remember the description verbally?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

8. In a complex building (store, museum) do you think spontaneously and easily about your direction in relation to the general structure of the building and the external environment?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

9. When you are inside a building can you easily visualise what there is outside the building in the direction you are looking?

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Not at all Very good

10. When you are in an open space and you are required to indicate a compass direction (north-south-east-west), do you:

☐ point immediately?

☐ need to think before pointing?

☐ have difficulty?

11. You are in a complex building (many floors, stairs, corridors) and you have to indicate where the entrance is, do you:

☐ point immediately?

☐ need to think before pointing?

☐ have difficulty?

D.3 Virtual Maze Post-Game Questionnaire

1. How do you rate the robot's Intelligence in the competitive game?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Not intelligent		Neutral		Highly intelligent

2. How do you rate the difficulty of the competitive game?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Easy		Neutral		Difficult

3. How do you rate the fun you had playing the competitive game?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Boring		Neutral		Very Fun

4. How do you rate the robot's Intelligence in the collaborative game?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Not intelligent		Neutral		Highly intelligent

5. How do you rate the difficulty of the collaborative game?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Easy		Neutral		Difficult

6. How do you rate the fun you had playing the collaborative game?

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Boring		Neutral		Very Fun

7. Which robot was more intelligent?

☐ 1	☐ 2	☐ 3	☐ 4
Competitive	Collaborative	Both	None

8. Which game was more difficult to play?

☐ 1 ☐ 2 ☐ 3 ☐ 4

Competitive Collaborative Both None

9. Which game was more fun to play?

☐ 1 ☐ 2 ☐ 3 ☐ 4

Competitive Collaborative Both None

10. When playing the games which technique did you use more frequently to make your move?

☐ Rotating my body

☐ Rotating my head

☐ Closing my eyes and rotate in my head

☐ I don't know, I played randomly

☐ Other, please specify

11. Which technique did you use more frequently to understand your teammate or opponent's move?

☐ Rotating my body

☐ Rotating my head

☐ Closing my eyes and rotate in my head

☐ I don't know, I played randomly

☐ Other, please specify

12. In my opinion, Polaris was... [multiple selection open]

☐ able to take/understand my perspective every time.

☐ able to take/understand my perspective sometimes.

☐ able to predict my moves and block me if it could (In competitive game).

☐ making similar mistakes as a child would make.

☐ making no mistakes.

☐ Other, please specify

13. In my opinion, I was... [multiple selection open]

☐ easily able to take Callisto's perspective.

☐ easily able to take Polaris's perspective.

☐ struggling to take Callisto's perspective.

☐ struggling to take Polaris's perspective.

☐ Other, please specify

14. How much do you think a child would enjoy the game? (with more levels)

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Not at all				Very much

15. Open-ended question: As you know this is a pretest for a game we will adapt for children. Any comments you may have, that can help us improve are greatly appreciated.

E Author’s Publications

This appendix presents the references and brief summaries of all peer-reviewed publications and contributions that were published during the course of this PhD. The first part is dedicated to the publications that directly contribute to this thesis with the link to their associated chapters. The second part presents all the co-authored publications that are not part of the main work of this thesis .

E. Yadollahi, W. Johal, A. Paiva, P. Dillenbourg, “When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration,” In Proceedings of the 17th ACM Conference on Interaction Design and Children, pages 195–206. ACM, 2018.

- This paper describes research aimed at supporting children’s reading practices using a robot designed to interact with children as their reading companion. The result of the user study shows that, deictic gestures such as pointing might be distracting for children with low reading proficiency, preventing them from comprehending the text and recognizing mistakes related to that.
- Chapter 4 is based on this article.

E. Yadollahi, W. Johal, J. Dias, P. Dillenbourg, A. Paiva “Studying the Effect of Robot Frustration on Children’s Change of Perspectiv” Workshop of Social emotions in the 8th ACM Conference on Affective Computing and Intelligent Interaction, 2019.

- Proposes the development of an interaction based on the objects game that evaluates how changes in the robot’s cognitive-affective state e.g. frustration affect children’s perspective taking adaptation and perception of the robot.
- The article is an extension of the platform presented in 5.

E. Yadollahi, M. Couto, W. Johal, P. Dillenbourg, A. Paiva "Exploring the role of perspective taking in educational child-robot interaction" International Conference on Artificial Intelligence in Education, 346-351.

- Presents the the design of the objects game and preliminary results from the pilot study to select the appropriate age group to participate in the main study.
- Parts of Chapter 5 is based on this article.

E. Yadollahi, M. Couto, P. Dillenbourg, A. Paiva "Can you guide me? supporting children's spatial perspective taking through games with robots" Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts.

- Describes the design and implementation of a gamified platform that evaluates children's perspective taking ability while interacting with a robot. The game is designed with different levels of difficulty with educational implications such as practicing mathematics.
- Parts of Chapter 6 is based on this article.

E. Yadollahi, P. Dillenbourg, A. Paiva "Changing Perspective as a Learning Mechanism" Companion of the 2020 ACM/IEEE Inter-national Conference on Human-Robot Interaction, 612-614

- Summarizes the perspective taking tasks, experimental studies, and future works for developing a comprehensive model of perspective taking for social robots.

This section presents all the co-authored publications through collaborations:

T. Asselborn, A. Guneysu, K. Mrini, **E. Yadollahi**, A. Ozgur, W. Johal, P. Dillenbourg, "Bringing Leters to Life: Handwriting with Haptic-Enabled Tangible Robots," In Proceedings of the 17th ACM Conference on Interaction Design and Children, pages 219–230. ACM, 2018.

- Presents a robotic approach to improve the teaching of handwriting using the tangible, haptic-enabled and classroom-friendly Cellulo robots.
- as part of Digital Learning and Analytic course project and in collaboration with researchers from CHILI Lab at EPFL.

A. Güneysu Özgür, A. Özgür, T. Asselborn, W. Johal, **E. Yadollahi**, B. Bruno, M. Skeweres, P. Dillenbourg "Iterative Design and Evaluation of a Tangible Robot-Assisted Handwriting Activity for Special Education" Frontiers in Robotics and AI7 (2020), 29.

- Investigates the role of interactive haptic-enabled tangible robots in supporting the learning of cursive letter writing for children with attention and visuomotor coordination issues.
- In collaboration with researchers from CHILI Lab at EPFL.

S. Tulli, M. Couto, M. Vasco, **E. Yadollahi**, F. Melo, A. Paiva "Explainable Agency by Revealing Suboptimality in Child-Robot Learning Scenarios" International Conference on Social Robotics, 23-35.

- Proposes a search-based approach to generate contrastive explanations using optimal and sub-optimal plans and implement it in a scenario for children.
- In collaboration with the researchers from Instituto Superior Técnico.

E. Yadollahi, S. Chandra, M. Couto, A. Lim, and A. Sandygulova. "Children, robots, and virtual agents: Present and future challenges" In Interaction Design and Children, pages 682–686, 2021.

- Presents a proposal for a full-day workshop that makes efforts to broaden our understanding and perspectives of how virtual agents, affect and potentially improve the well-being of children. l It also provides an opportunity for an interdisciplinary debate about the present and future of child-agent interactions.

Bibliography

Ackermann, E. (2001). Piaget's constructivism, papert's constructionism: what's the difference. *Future of learning group publication*, *5*(3), 438.

Ahmed, I., Lubold, N., & Walker, E. (2018). Robin: using a programmable robot to provide feedback and encouragement on programming tasks, In *International conference on artificial intelligence in education*. Springer.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). Act-r: a theory of higher level cognition and its relation to visual attention. *Human–Computer Interaction*, *12*(4), 439–462.

Asada, M. (2014). Towards artificial empathy: how can artifcial empathy follow the developmental pathway of natural empathy. *Artificial empathy–An interdisciplinary investigations. Special Issue Int J Soc Robot*.

Asimov, I. (2004). *I, robot* (Vol. 1). Spectra.

Baraka, K., Couto, M., Melo, F. S., & Veloso, M. (2019). An optimization approach for structured agent-based provider/receiver tasks, In *Proceedings of the 18th international conference on autonomous agents and multiagent systems*.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology*, *72*(5), 593.

Barnes-Holmes, Y., McHugh, L., & Barnes-Holmes, D. (2004). Perspective-taking and theory of mind: a relational frame account. *The Behavior Analyst Today*, *5*(1), 15.

Baron-Cohen, S. (1997). *Mindblindness: an essay on autism and theory of mind*. MIT press.

Baron-Cohen, S. E., Tager-Flusberg, H. E., & Cohen, D. J. (2000). *Understanding other minds: perspectives from developmental cognitive neuroscience*. Oxford University Press.

Baron-Cohen, S., & Hammer, J. (1997). Parents of children with asperger syndrome: what is the cognitive phenotype? *Journal of cognitive neuroscience*, *9*(4), 548–554.

Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., Neerincx, M., Baroni, I., & Zelati, M. C. (2013). Child-robot interaction: perspectives and challenges, In *International conference on social robotics*. Springer.

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., Bergmann, K., Leseman, P., Küntay, A. C., Göksun, T., Et al. (2015). L2tor-second language tutoring using social robots, In *Proceedings of the icsr 2015 wonder workshop*.

Bengtsson, H., & Johnson, L. (1992). Perspective taking, empathy, and prosocial behavior in late childhood. *Child Study Journal*.

Berlin, M., Gray, J., Thomaz, A. L., & Breazeal, C. (2006). Perspective taking: an organizing principle for learning in human-robot interaction, In *Aaai*.

Bialystok, E., & Codd, J. (1987). Children's interpretations of ambiguous spatial descriptions. *British Journal of Developmental Psychology*, *5*(3), 205–211.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & at Vanderbilt, T. T. A. G. (2005). Learning by teaching: a new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*(3-4), 363–392.

Bloom, P. (1999). *Language and space*. MIT press.

Boland Jr, R. J., & Tenkasi, R. V. (1995). Perspective making and perspective taking in communities of knowing. *Organization science*, *6*(4), 350–372.

Borenstein, J., & Arkin, R. C. (2017). Nudging for good: robots and the ethical appropriateness of nurturing empathy and charitable behavior. *Ai & Society*, *32*(4), 499–507.

Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational intelligence*, *4*(3), 349–355.

Breazeal, C., Berlin, M., Brooks, A., Gray, J., & Thomaz, A. L. (2006). Using perspective taking to learn from ambiguous demonstrations. *Robotics and autonomous systems*, *54*(5), 385–393.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork, In *Intelligent robots and systems, 2005.(iros 2005). 2005 ieee/rsj international conference on*. IEEE.

Brown, L., & Howard, A. M. (2013). Engaging children in math education using a socially interactive humanoid robot, In *Humanoid robots (humanoids), 2013 13th ieee-ras international conference on*. IEEE.

Bus, A. G., Van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: a meta-analysis on intergenerational transmission of literacy. *Review of educational research*, *65*(1), 1–21.

Call, J., & Tomasello, M. (2011). Does the chimpanzee have a theory of mind? 30 years later. *Human Nature and Self Design*, 83–96.

Carbo, M. (1978). Teaching reading with talking books. *The Reading Teacher*, *32*(3), 267–273.

Carlson-Radvansky, L. A., & Irwin, D. E. (1993). Frames of reference in vision and language: where is above? *Cognition*, *46*(3), 223–244.

Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference-frame selection. *Psychological Science*, *9*(5), 386–391.

Carlson-Radvansky, L. A., & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science*, *7*(1), 56–60.

Castellano, G., Paiva, A., Kappas, A., Aylett, R., Hastie, H., Barendregt, W., Nabais, F., & Bull, S. (2013). Towards empathic virtual and robotic tutors, In *International conference on artificial intelligence in education*. Springer.

Charisi, V., Davison, D., Wijnen, F., Van Der Meij, J., Reidsma, D., Prescott, T., Van Joolingen, W., & Evers, V. (2015). Towards a child-robot symbiotic co-development: a theoretical approach, In *Aisb convention 2015*. The Society for the Study of Artificial Intelligence and the Simulation of . . .

Charisi, V., Sabanović, S., Cangelosi, A., & Gomez, R. (2021). Designing and developing better robots for children: a fundamental human rights perspective, In *Companion of the 2021 acm/ieee international conference on human-robot interaction*.

Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable agents and the protégé effect: increasing the effort towards learning. *Journal of Science Education and Technology*, *18*(4), 334–352.

Chomsky, C. (1976). After decoding: what? *Language Arts*, *53*(3), 288–314.

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.

Clark, H. H., & Marshall, C. R. (2002). Definite reference and mutual knowledge. *Psycholinguistics: critical concepts in psychology*, *414*.

Cunningham, J. W. (1984). A simplified miscue analysis for classroom and clinic. *Reading Horizons*, *24*(2), 2.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.

Davis, M. H. (2018). *Empathy: a social psychological approach*. Routledge.

Davis, M. H. Et al. (1980). A multidimensional approach to individual differences in empathy.

de Ruirer, J. P. (2000). 14 the production of gesture and speech. *Language and gesture*, *2*, 284.

Devin, S., & Alami, R. (2016). An implemented theory of mind to improve human-robot shared plans execution, In *2016 11th acm/ieee international conference on human-robot interaction (hri)*. IEEE.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological bulletin*, *101*(1), 91.

Eisenberg, N., Miller, P. A., Schaller, M., Fabes, R. A., Fultz, J., Shell, R., & Shea, C. L. (1989). The role of sympathy and altruistic personality traits in helping: a reexamination. *Journal of personality*, *57*(1), 41–67.

El Hamamsy, L., Johal, W., Asselborn, T., Nasir, J., & Dillenbourg, P. (2019). Learning by collaborative teaching: an engaging multi-party cowriter activity, In *2019 28th ieee international conference on robot and human interactive communication (ro-man)*. IEEE.

Elekes, F., Varga, M., & Király, I. (2017). Level-2 perspectives computed quickly and spontaneously: evidence from eight-to 9.5-year-old children. *British Journal of Developmental Psychology*, *35*(4), 609–622.

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: equivalent egocentrism but differential correction. *Journal of experimental social psychology*, *40*(6), 760–768.

Fan, M., Antle, A. N., & Cramer, E. S. (2016). Design rationale: opportunities and recommendations for tangible reading systems for children, In *Proceedings of the the 15th international conference on interaction design and children*. ACM.

Farrant, B. M., Devine, T. A., Maybery, M. T., & Fletcher, J. (2012). Empathy, perspective taking and prosocial behaviour: the importance of parenting practices. *Infant and Child Development*, *21*(2), 175–188.

Feil-Seifer, D. J., & Matarić, M. J. (2009). Human-robot interaction. *Encyclopedia of Complexity and Systems Science*, 4643–4659.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Robots and transparency: the multiple dimensions of transparency in the context of robot technologies. *IEEE Robotics & Automation Magazine*, *26*(2), 71–78.

Fischer, T. (2018). Perspective taking in robots: a framework and computational model.

Fischer, T., & Demiris, Y. (2019). Computational modelling of embodied visual perspective-taking. *IEEE Transactions on Cognitive and Developmental Systems*.

Fiske, S. T., & Macrae, C. N. (2012). *The sage handbook of social cognition*. Sage.

Flavell, J. H. (1977). The development of knowledge about visual perception., In *Nebraska symposium on motivation*. University of Nebraska Press.

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International journal of behavioral development*, *24*(1), 15–23.

Flavell, J. H. (2004). Theory-of-mind development: retrospect and prospect. *Merrill-Palmer Quarterly (1982-)*, 274–290.

Flavell, J. H., Beilin, H., & Pufall, P. (1992). *Perspectives on perspective taking*. Erlbaum Hillsdale, NJ.

Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive psychology*, *15*(1), 95–120.

Flavell, J. H., Flavell, E. R., Green, F. L., & Wilcox, S. A. (1981). The development of three spatial perspective-taking rules. *Child Development*, 356–358.

Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance–reality distinction. *Monographs of the society for research in child development*.

Fong, T., Thorpe, C., & Baur, C. (2003). Collaboration, dialogue, human-robot interaction, In *Robotics research*. Springer.

Friederici, A. D., & Levelt, W. J. (1990). Spatial reference in weightlessness: perceptual factors and mental representations. *Perception & Psychophysics*, *47*(3), 253–266.

Frith, C., & Frith, U. (2005). Theory of mind. *Current biology*, *15*(17), R644–R645.

Galinsky, A. D., Maddux, W. W., Gilin, D., & White, J. B. (2008). Why it pays to get inside the head of your opponent: the differential effects of perspective taking and empathy in negotiations. *Psychological science*, *19*(4), 378–384.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of personality and social psychology*, *78*(4), 708.

Gentili, R. J., Oh, H., Huang, D.-W., Katz, G. E., Miller, R. H., & Reggia, J. A. (2015). A neural architecture for performing actual and mentally simulated movements during self-intended and observed bimanual arm reaching movements. *International Journal of Social Robotics*, *7*(3), 371–392.

Gerbier, E., Bailly, G., & Bosse, M. L. (2018). Audio–visual synchronization in reading while listening to texts: effects on visual behavior and verbal learning. *Computer Speech & Language*, *47*, 74–92.

Goodman, K. S. (1969). Analysis of oral reading miscues: applied psycholinguistics. *Reading Research Quarterly*, 9–30.

Goodman, K. S. (1973). Miscue analysis: applications to reading instruction.

Goodman, Y. M., & Burke, C. L. (1972). *Reading miscue inventory manual: procedure for diagnosis and evaluation: readings for taping*. Macmillan.

Goodrich, M. A., & Schultz, A. C. (2008). *Human-robot interaction: a survey*. Now Publishers Inc.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory.

Gordon, G., & Breazeal, C. (2015). Bayesian active learning-based robot tutor for children's word-reading skills, In *Twenty-ninth aaai conference on artificial intelligence*.

Grant, A. M., & Berry, J. W. (2011). The necessity of others is the mother of invention: intrinsic and prosocial motivations, perspective taking, and creativity. *Academy of management journal*, *54*(1), 73–96.

Hadwin, J. A., Howlin, P., & Baron-Cohen, S. (2015). *Teaching children with autism to mind-read: the workbook*. John Wiley & Sons.

Hamilton, A. F. d. C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition*, *113*(1), 37–44.

Harris, L. J. (1972). Discrimination of left and right, and development of the logic of relations. *Merrill-Palmer Quarterly of Behavior and Development*, *18*(4), 307–320.

Harris, L. J., & Strommen, E. A. (1972). The role of front-back features in children's "front," "back," and "beside" placements of objects. *Merrill-Palmer Quarterly of Behavior and Development*, *18*(3), 259–271.

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition & Emotion*, *3*(4), 379–400.

Harrison, A. M., Schunn, C. D. Et al. (2003). Act-r/s: look ma, no"cognitive-map, In *International conference on cognitive modeling*.

Hiatt, L. M., Trafton, J. G., Harrison, A. M., & Schultz, A. C. (2004). A cognitive model for spatial perspective taking., In *Iccm*. Citeseer.

Hinnant, J. B., & O'Brien, M. (2007). Cognitive and emotional control and perspective taking and their relations to empathy in 5-year-old children. *The Journal of Genetic Psychology*, *168*(3), 301–322.

Hood, D., Lemaignan, S., & Dillenbourg, P. (2015). When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting, In *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*. ACM.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind-read: a practical guide for teachers and parents*. J. Wiley & Sons Chichester.

Huang, C.-M., & Thomaz, A. L. (2010). Joint attention in human-robot interaction., In *Aaai fall symposium: dialog with robots*.

Hyun, E., Yoon, H., & Son, S. (2010). Relationships between user experiences and children's perceptions of the education robot, In *Proceedings of the 5th acm/ieee international conference on human-robot interaction*. IEEE Press.

Imai, M., Ono, T., & Ishiguro, H. (2003). Physical relation and expression: joint attention for human-robot interaction. *IEEE Transactions on Industrial Electronics*, *50*(4), 636–643.

Israelashvili, J., Sauter, D., & Fischer, A. (2019). How well can we assess our ability to understand others' feelings? beliefs about taking others' perspectives and actual understanding of others' emotions. *Frontiers in psychology*, *10*, 2475.

Janczyk, M. (2013). Level 2 perspective taking entails two processes: evidence from prp experiments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1878.

Johal, W. (2020). Research trends in social robots for learning. *Current Robotics Reports*, 1–9.

Johnson, D. O., Cuijpers, R. H., & van der Pol, D. (2013). Imitating human emotions with artificial facial expressions. *International Journal of Social Robotics*, *5*(4), 503–513.

Johnson, M., & Demiris, Y. (2005). Perceptual perspective taking and action recognition. *International Journal of Advanced Robotic Systems*, *2*(4), 32.

Johnson, M., & Demiris, Y. (2007). Visuo-cognitive perspective taking for action recognition. AISB.

Justice, L. M., Pullen, P. C., & Pence, K. (2008). Influence of verbal and nonverbal references to print on preschoolers' visual attention to print during storybook reading. *Developmental Psychology*, *44*(3), 855.

Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: a field trial. *Human-computer interaction*, *19*(1), 61–84.

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, *7*(2), 135–169.

Kendon, A. (1990). *Conducting interaction: patterns of behavior in focused encounters* (Vol. 7). CUP Archive.

Kennedy, W. G., Bugajska, M. D., Harrison, A. M., & Trafton, J. G. (2009). ''like-me" simulation as an effective and cognitively plausible basis for social robotics. *International Journal of Social Robotics*, *1*(2), 181–194.

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, 213.

Kessler, K., & Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition*, *114*(1), 72–88.

Kessler, K., & Wang, H. (2012). Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spatial Cognition & Computation*, *12*(2-3), 133–158.

Kim, T., & Hinds, P. (2006). Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction, In *Roman 2006-the 15th ieee international symposium on robot and human interactive communication*. IEEE.

Kory, J., & Breazeal, C. (2014). Storytelling with robots: learning companions for preschool children's language development, In *Robot and human interactive communication, 2014 ro-man: the 23rd ieee international symposium on*. IEEE.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition*, *29*(5), 745–756.

Kozima, H., & Yano, H. (2001). A robot that learns to communicate with human caregivers, In *Proceedings of the first international workshop on epigenetic robotics*.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? interaction and perspective taking with robots investigated via fmri. *PloS one*, *3*(7), e2597.

Kurdek, L. A., & Rodgon, M. M. (1975). Perceptual, cognitive, and affective perspective taking in kindergarten through sixth-grade children. *Developmental psychology*, *11*(5), 643.

Lee, S.-l., Lau, I. Y.-m., Kiesler, S., & Chiu, C.-Y. (2005). Human mental models of humanoid robots, In *Robotics and automation, 2005. icra 2005. proceedings of the 2005 ieee international conference on*. IEEE.

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, *5*(2), 291–308.

Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., & Dillenbourg, P. (2016). Learning by teaching a robot: the case of handwriting. *IEEE Robotics & Automation Magazine*, *23*(2), 56–66.

Lemaignan, S., Ros, R., Alami, R., & Beetz, M. (2011). What are you talking about? grounding dialogue in a perspective-aware robotic architecture, In *2011 ro-man*. IEEE.

Levinson, S. C. (1996). Language and space. *Annual review of Anthropology*, *25*(1), 353–382.

Levinson, S. C. (2003). *Space in language and cognition: explorations in cognitive diversity*. Cambridge University Press.

Leyzberg, D., Spaulding, S., & Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences, In *Proceedings of the 2014 acm/ieee international conference on human-robot interaction*. ACM.

Li, P., Jin, X., Liao, Y., Li, Y., Shen, M., & He, J. (2019). Cooperation turns preschoolers into flexible perspective takers. *Cognitive Development*, *52*, 100823.

Lim, A., & Okuno, H. G. (2015). A recipe for empathy. *International Journal of Social Robotics*, *7*(1), 35–49.

Lu, Y., Chen, C., Chen, P., Chen, X., & Zhuang, Z. (2018). Smart learning partner: an interactive robot for education, In *International conference on artificial intelligence in education*. Springer.

Luria, M. (2018). Designing robot personality based on fictional sidekick characters, In *Companion of the 2018 acm/ieee international conference on human-robot interaction*.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The early development of inferences about the visual percepts of others. *Child development*, 357–366.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. University of Chicago press.

Michaelis, J. E., & Mutlu, B. (2017). Someone to read with: design of and experiences with an in-home learning companion robot for reading, In *Proceedings of the 2017 chi conference on human factors in computing systems*. ACM.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & psychophysics*, *68*(2), 327–337.

Mintz, F. E., Trafton, J. G., Marsh, E., & Perzanowski, D. (2004). Choosing frames of referenece: perspective-taking in a 2d and 3d navigational task, In *Proceedings of the human factors and ergonomics society annual meeting*. SAGE Publications Sage CA: Los Angeles, CA.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive psychology*, *41*(1), 49–100.

Moll, H., & Meltzoff, A. N. (2011). Foundation in joint attention. *Perception, causation, and objectivity*, *286*, 286–304.

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*(3), 603–613.

Money, J., Alexander, D., & Walker, H. (1965). *A standardized road-map test of direction sense: manual*. Johns Hopkins Press.

Mouw, J., Saab, N., Gijlers, H., Hickendorff, M., Paridon, Y. v., & Broek, P. (2020). The differential effect of perspective-taking ability on profiles of cooperative behaviours and learning outcomes. *Frontline Learning Research*, *8*(6), 88–113.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, *1*, 209–0015.

Muldner, K., Lozano, C., Girotto, V., Burleson, W., & Walker, E. (2013). Designing a tangible learning environment with a teachable agent, In *International conference on artificial intelligence in education*. Springer.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child development*, *78*(3), 938–954.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current directions in psychological science*, *16*(5), 269–274.

Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: modeling and evaluation of human-like gaze behavior, In *Humanoid robots, 2006 6th ieee-ras international conference on*. IEEE.

Muto, H., Matsushita, S., & Morikawa, K. (2019). Object’s symmetry alters spatial perspective-taking processes. *Cognition*, *191*, 103987.

Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). A constructive model for the development of joint attention. *Connection Science*, *15*(4), 211–229.

Nakajo, R., Murata, S., Arie, H., & Ogata, T. (2015). Acquisition of viewpoint representation in imitative learning from own sensory-motor experiences, In *2015 joint ieee international conference on development and learning and epigenetic robotics (icdl-epirob)*. IEEE.

Newcombe, N. (1989). The development of spatial perspective taking, In *Advances in child development and behavior*. Elsevier.

Newcombe, N., & Huttenlocher, J. (1992). Children's early ability to solve perspective-taking problems. *Developmental psychology*, *28*(4), 635.

Ogata, T., Yokoya, R., Tani, J., Komatani, K., & Okuno, H. G. (2009). Prediction and imitation of other's motions by reusing own forward-inverse model in robots, In *2009 ieee international conference on robotics and automation*. IEEE.

Paiva, A., Santos, F., & Santos, F. (2018). Engineering pro-sociality with autonomous agents, In *Proceedings of the aaai conference on artificial intelligence*.

Palan, S., & Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27.

Pandey, A. K., Ali, M., & Alami, R. (2013). Towards a task-aware proactive sociable robot based on multi-state perspective-taking. *International Journal of Social Robotics*, *5*(2), 215–236.

Papadopoulos, F., Dautenhahn, K., & Ho, W. C. (2013). Aibostory–autonomous robots supporting interactive, collaborative story-telling. *Paladyn, Journal of Behavioral Robotics*, *4*(1), 10–22.

Papert, S. A. (2020). *Mindstorms: children, computers, and powerful ideas*. Basic books.

Parde, N., & Nielsen, R. D. (2019). Ai meets austen: towards human-robot discussions of literary metaphor, In *International conference on artificial intelligence in education*. Springer.

Pazzaglia, F., & De Beni, R. (2001). Strategies of processing spatial information in survey and landmark-centred individuals. *European journal of cognitive psychology*, *13*(4), 493–508.

Pellegrini, A. D., Brody, G. H., & Sigel, I. E. (1985). Parents' book-reading habits with their children. *Journal of Educational Psychology*, *77*(3), 332.

Perner, J. (1991). *Understanding the representational mind*. The MIT Press.

Perrucci, V., Agnoli, F., & Albiero, P. (2008). Children's performance in mental rotation tasks: orientation-free features flatten the slope. *Developmental Science*, *11*(5), 732–742.

Peter, J., Kühne, R., & Barco, A. (2021). Can social robots affect children's prosocial behavior? an experimental study on prosocial robot models. *Computers in Human Behavior*, *120*, 106712.

Piaget, J. (2013). *Child's conception of space: selected works vol 4* (Vol. 4). Routledge.

Piaget, J., & Inhelder, B. (1967). The child's conception of space.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). The child's conception of geometry.

Pickering, M. J., McLean, J. F., & Gambi, C. (2012). Do addressees adopt the perspective of the speaker? *Acta psychologica*, *141*(2), 261–269.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(4), 515–526.

Quigley, M., Faust, J., Foote, T., Leibs, J., Et al. (2001). Ros: an open-source robot operating system.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition*, *117*(2), 230–236.

Raffle, H., Ballagas, R., Revelle, G., Horii, H., Follmer, S., Go, J., Reardon, E., Mori, K., Kaye, J., & Spasojevic, M. (2010). Family story play: reading with young children (and elmo) over a distance, In *Proceedings of the sigchi conference on human factors in computing systems*. ACM.

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology*, *30*(1), 59–74.

Rao, A. S., Georgeff, M. P. Et al. (1995). Bdi agents: from theory to practice., In *Icmas*.

Rich, C., Ponsler, B., Holroyd, A., & Sidner, C. L. (2010). Recognizing engagement in human-robot interaction, In *Human-robot interaction (hri), 2010 5th acm/ieee international conference on*. IEEE.

Robins, B., Dautenhahn, K., Wood, L., & Zaraki, A. (2017). Developing interaction scenarios with a humanoid robot to encourage visual perspective taking skills in children with autism–preliminary proof of concept tests, In *International conference on social robotics*. Springer.

Ros, R., Lemaignan, S., Sisbot, E. A., Alami, R., Steinwender, J., Hamann, K., & Warneken, F. (2010). Which one? grounding the referent based on efficient human-robot interaction, In *19th international symposium in robot and human interactive communication*. IEEE.

Ros, R., Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., Belpaeme, T., Giusti, A., & Pozzi, C. (2011). Child-robot interaction in the wild: advice to the aspiring experimenter, In *Proceedings of the 13th international conference on multimodal interfaces*.

Roy, D., Hsiao, K.-Y., & Mavridis, N. (2004). Mental imagery for a conversational robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *34*(3), 1374–1383.

Rubegni, E., & Landoni, M. (2016). Evaluating engagement in reading: comparing children and adult assessors, In *Proceedings of the the 15th international conference on interaction design and children*. ACM.

Salter, T., Werry, I., & Michaud, F. (2008). Going into the wild in child–robot interaction studies: issues in social robotic development. *Intelligent Service Robotics*, *1*(2), 93–108.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929.

Scassellati, B. (1999). Imitation and mechanisms of joint attention: a developmental structure for building social skills on a humanoid robot, In *Computation for metaphors, analogy, and agents*. Springer.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition*, *47*(1), 1–24.

Schodde, T., Bergmann, K., & Kopp, S. (2017). Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making, In *Proceedings of the 2017 acm/ieee international conference on human-robot interaction*.

Schrodt, F., Layher, G., Neumann, H., & Butz, M. V. (2015). Embodied learning of a generative neural model for biological motion perception and inference. *Frontiers in computational neuroscience*, *9*, 79.

Shantz, C. U. (1975). The development of social cognition.

Sharma, K., D'Angelo, S., Gergle, D., & Dillenbourg, P. (2016). Visual augmentation of deictic gestures in mooc videos. Singapore: International Society of the Learning Sciences.

Sisbot, E. A., Marin-Urias, L. F., Alami, R., & Simeon, T. (2007). A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, *23*(5), 874–883.

Steels, L., & Loetzsch, M. (2009). Perspective alignment in spatial language.

Steinbeis, N., & Crone, E. A. (2016). The link between cognitive control and decision-making across child and adolescent development. *Current Opinion in Behavioral Sciences*, *10*, 28–32.

Surtees, A., Noordzij, M., & Apperly, I. (2012). How to lose yourself in space: children and adults' use of frames of reference and perspectives. *Dev. Psychol*, *48*, 185–191.

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426–438.

Surtees, A., Samson, D., & Apperly, I. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105.

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, *104*(46), 17954–17958.

Tapus, A., & Mataric, M. J. (2007). Emulating empathy in socially assistive robotics., In *Aaai spring symposium: multidisciplinary collaboration for socially assistive robotics*.

Taylor, M. (1988). Conceptual perspective taking: children's ability to distinguish what they know from what they see. *Child development*, 703–718.

Thomaz, A. L., Berlin, M., & Breazeal, C. (2005). An embodied computational model of social referencing, In *Robot and human interactive communication, 2005. roman 2005. ieee international workshop on*. IEEE.

Todd, A. R., Cameron, C. D., & Simpson, A. J. (2017). Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition*, *159*, 97–101.

Tomasello, M. (2010). *Origins of human communication*. MIT press.

Tomasello, M. Et al. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development*, *103130*, 103–130.

Torrey, C., Fussell, S. R., & Kiesler, S. (2009). What robots could teach us about perspective-taking. *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*, 93–106.

Trafton, J. G., Cassimatis, N. L., Bugajska, M. D., Brock, D. P., Mintz, F. E., & Schultz, A. C. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *35*(4), 460–470.

Tversky, B., & Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition*, *110*(1), 124–129.

Van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PloS one*, *13*(8), e0202442.

Vander Heyden, K. M., Huizinga, M., Raijmakers, M. E., & Jolles, J. (2017). Children's representations of another person's spatial perspective: different strategies for different viewpoints? *Journal of experimental child psychology*, *153*, 57–73.

Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Wallach, W., Allen, C., & Smit, I. (2008). Machine morality: bottom-up and top-down approaches for modelling human moral faculties. *Ai & Society*, *22*(4), 565–582.

Warnier, M., Guitton, J., Lemaignan, S., & Alami, R. (2012). When the robot puts itself in your shoes. managing and exploiting human and robot beliefs, In *2012 ieee ro-man: the 21st ieee international symposium on robot and human interactive communication*. IEEE.

Weisz, E., & Zaki, J. (2017). Empathy building interventions: a review of existing work and suggestions for future directions. *The Oxford handbook of compassion science*, 205–217.

Wellman, H. M. (1992). *The child's theory of mind*. The MIT Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child development*, *72*(3), 655–684.

Westlund, J. M. K., & Breazeal, C. (2016). Transparency, teleoperation, and children's understanding of social robots, In *2016 11th acm/ieee international conference on human-robot interaction (hri)*. IEEE.

Westlund, J. M. K., Park, H. W., Williams, R., & Breazeal, C. (2018). Measuring young children's long-term relationships with social robots, In *Proceedings of the 17th acm conference on interaction design and children*.

Williams, R., Park, H. W., & Breazeal, C. (2019). A is for artificial intelligence: the impact of artificial intelligence activities on young children's perceptions of robots, In *Proceedings of the 2019 chi conference on human factors in computing systems*.

Wolgast, A., Tandler, N., Harrison, L., & Umlauft, S. (2020). Adults' dispositional and situational perspective-taking: a systematic review. *Educational Psychology Review*, *32*(2), 353–389.

Wood, L. J., Robins, B., Lakatos, G., Syrdal, D. S., Zaraki, A., & Dautenhahn, K. (2018). Piloting scenarios for children with autism to learn about visual perspective taking, In *Annual conference towards autonomous robotic systems*. Springer.

Wood, L. J., Robins, B., Lakatos, G., Syrdal, D. S., Zaraki, A., & Dautenhahn, K. (2019). Developing a protocol and experimental setup for using a humanoid robot to assist children with autism to develop visual perspective taking skills. *Paladyn, Journal of Behavioral Robotics*, *10*(1), 167–179.

Wood, L., Dautenhahn, K., Robins, B., & Zaraki, A. (2017). Developing child-robot interaction scenarios with a humanoid robot to assist children with autism in developing visual perspective taking skills, In *2017 26th ieee international symposium on robot and human interactive communication (ro-man)*. IEEE.

Yadollahi, E., Johal, W., Dias, J., Dillenbourg, P., & Paiva, A. (2019). 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.

Yadollahi, E., Johal, W., Paiva, A., & Dillenbourg, P. (2018). When deictic gestures in a robot can harm child-robot collaboration, In *Proceedings of the 17th acm conference on interaction design and children*. ACM.

Yirmiya, N., Sigman, M., & Zacks, D. (1994). Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Development and psychopathology*, *6*(2), 263–272.

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and cognitive neuroscience reviews*, *4*(2), 96–118.

Zacks, J. M., Mires, J., Tversky, B., & Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation*, *2*(4), 315–332.

Zahn-Waxler, C., Radke-Yarrow, M., & Brady-Smith, J. (1977). Perspective-taking and prosocial behavior. *Developmental Psychology*, *13*(1), 87.

Zaki, J. (2017). Moving beyond stereotypes of empathy.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience*, *15*(5), 675–680.

Zhao, X., Cusimano, C., & Malle, B. F. (2016). Do people spontaneously take a robot's visual perspective?, In *2016 11th acm/ieee international conference on human-robot interaction (hri)*. IEEE.

Elmira **Yadollahi**

DOCTORAL RESEARCHER · ROBOTICS

Instituto Superior Técnico, Lisbon, Portugal

(+351) 910 538 321 | elmira.yadollahi@epfl.ch | www.elmirayadollahi.com | ElmiraYadollahi

elmirayadollahi | @ElmiraYad | ElmiraYadollahi

Keywords: *Human-Robot Interaction, Child-Robot Interaction, Theory of Mind, Perspective Taking, Robots in Education*

Education

Ph.D. in Robotics, Brain and Cognition — *Lisbon, Portugal & Lausanne, Switzerland*

INSTITUTO SUPERIOR TÉCNICO (IST), UNIVERSITY OF LISBON — *Feb. 2017 - Oct. 2021*

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL) — *Nov. 2016 - Oct. 2021*

- The doctoral program RBCog – Robotics, Brain and Cognition is funded by the Portuguese Foundation for Science and Technology (FCT) PD/BD/135150/2017. The IST – EPFL track grants a Joint degree from EPFL and IST.
- Under supervision of Prof. Ana Paiva and Prof. Pierre Dillenbourg with research carried out in GAIPS and CHILI Laboratories.
- In EPFL the degree is carried out in Robotics, Control and Intelligent Systems (EDRS) department. | GPA: 5.75 out of 6
- In IST the degree is carried out in Information Systems and Computer Engineering (DEIC) department. | GPA: 18 out of 20

M.Sc. in Mechanical Engineering — *Daejeon, South Korea*

KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (KAIST) — *Sep. 2012 - Feb. 2015*

- Degree Recieved in Mechanical Engineering with focus on Acoustics, Noise, and Vibration.
- Thesis: Acoustic Localization of Small Leak Holes in Long Pipeline
- Under Supervision of Porf. Ih Jeoung Gong | GPA: 3.75 out of 4.3

B.Sc. in Mechanical Engineering — *Tehran, Iran*

SHARIF UNIVERSITY OF TECHNOLOGY — *Sep. 2007 - Aug. 2012*

- Thesis: Design of Knee Exoskeleton Mechanism to Assist Walking
- Under supervision of Prof. Hassan Zohoor | GPA: 15.27 out of 20

Research Experience

Artificial Intelligence for People and Society (GAIPS), INESC-ID — *Lisbon, Portugal*

DOCTORAL ASSISTANT — *Jun. 2018 - Sep. 2021*

- My research includes generating a decision-making model of perspective-taking that helps robots to make informed decisions while collaborating with children. It is carried out under supervision of Prof. Ana Paiva and guidance of Dr. Marta Couto.

Computer-Human Interaction In Learning Laboratory (CHILI) — *Lausanne, Switzerland*

DOCTORAL ASSISTANT — *Nov. 2016 - May 2018*

- Worked on developing the CoReader platform for using robot in educational scenarios with children. Research was carried out with supervision of Prof. Pierre Dillenbourg and guidance of Dr. Wafa Johal.

TeleRobotics and Control Laboratory (TCL Lab) — *Daejeon, South Korea*

POST GRADUATE RESEARCHER — *Apr. 2015 - Feb. 2016*

- Started my research in human-robot interaction projects with supervision of Prof. Kwon Dong Soo by developing interaction modalities using sound source localization and visual inputs for robots while incorporating Kinect as an external input.

Acoustics Lab | Noise and Vibration Center (NoViC) — *Lisbon, Portugal*

GRADUATE RESEARCHER — *Sep. 2012 - Feb. 2015*

- Worked with prof. Jeong-Guon Ih, on projects related to acoustics and sound source localization.

Work Experience

Artificial Intelligence for People and Society (GAIPS) — *Lisbon, Portugal*

LAB MANAGER | PART TIME — *Feb. 2020 - Sep. 2021*

- Managing the servers, robots, equipment purchases, and general organizations.

SM Instruments Inc. — *Daejeon, South Korea*

INTERNATIONAL MARKETING ENGINEER | FULL TIME — *Feb. 2016 - Oct. 2016*

- Promoting Sound Camera with our US partner National Instruments Corporation.
- Example video tutorial created by me.

English and Second Language Tutor *Daejeon, South Korea*
TUTOR | FREELANCE *Nov. 2014 - Sep. 2016*
- English one to one tutoring to children between 6- 9 years old.

Scania Bus Assembly Plant, Iran *Semnan, Iran*
R & D ENGINEER | INTERN *Jun. 2010 - Sep. 2010*
- Research and development on the endurance of the bodies designed and produced for the Scania chassis.

Teaching Experience

Social Robotics and Human-Robot Interaction *Lisbon, Portugal*
INSTITUTO SUPERIOR TÉCNICO (IST), UNIVERSITY OF LISBON *Fall 2020*
- The theoretical classes were taught by Prof. Rui Prada with me teaching the practical classes.
- As the only TA, my responsibilities included creating material, teaching the practical classes in workshop format, guide the students in developing their projects, and grading the projects. The workshops were organized on the following topics: Workshops on *Design of Social Robots*, *Technical Development*, *Experimental Design*, and *Statistical Analyses.*

Introduction to Visual Informatics *Lausanne, Switzerland*
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL) *Spring 2017, Spring 2018*
- The course was taught by Prof. Pierre Dillenbourg.
- As part of the TAs team, my tasks included preparing the homework and moderating the practical classes using Blender, and grading.

Programming C *Lausanne, Switzerland*
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL), SWITZERLAND *Fall 2017*
- The course was taught by Prof. Ronan Boulic.
- As part of the TAs team, my tasks included preparing the homework, guiding the students in programming, designing the exams, and grading.

Projects

WP6- Social AI: Learning and Reasoning in Social Contexts *Europe*
EU HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME GA NO 952215 *Jan. 2021 - Present*
- TAILOR is an ICT-48 Network of AI Research Excellence Centers funded by EU Horizon 2020. This work package will establish a roadmap in this field drawing new directions for research exploring new avenues for social AI focusing on its foundations techniques and algorithms.

Ecosystem of Child-Robot Interaction *Portugal, Canada, United States*
INTERNATIONAL COLLABORATION *Feb. 2020 - Present*
- In collaboration with researchers from University of Washington in USA, University of Waterloo in Canada and INESC-ID in Portugal. This project aims at creating a clear picture of how ethics in Child-Robot Interaction by providing guidelines for researchers and other stakeholders .

An Extended Framework for Characterizing Robots in Movies *Portugal*
COLLABORATION WITHIN PORTUGAL *Nov. 2020 - Present*
- In collaboration with researchers from Instituto Universitário de Lisboa (ISCTE) in Portugal.

An Extended Survey of Technical and Experimental Developments using Nao Robot *Portugal, Australia, Kazakhstan*
INTERNATIONAL COLLABORATION *Jan. 2021 - Present*
- In collaboration with researchers from University of New South Wales in Australia and Nazarbayev University in Kazakhstan. The research provide insights into the deployment of Nao, the most sought-after social robot, and discuss the progress in its application across various domains.

Studying the Effect of Transparency in Interaction with Robots *Portugal*
COLLABORATION WITHIN IST *Aug. 2019 - Nov 2020*
- In collaboration with the researchers from Instituto Superior Técnico and research published at ICSR 2020 receiving *Best Student Paper Award*.

CogPeT: Cognitive Model of Perspective Taking *Portugal*
PHD RESEARCH PROJECT *Sep. 2018 - Sep. 2021*
- As part of my doctoral research, the research aims at developing a Cognitive Model of Perspective Taking , including the use of NAO (partly published in AIED 2020 with presentation) and Cozmo robots (partly published in IDC 2020 with presentation) and virtual robotic agents.

Bringing Letters to Life: Handwriting with Haptic-Enabled Tangible Robots *Switzerland*
COLLABORATION WITHIN EPFL *Sep 2017 - Jun. 2018*
- as part of *Digital Learning and Analytic course* project and in collaboration with researchers from CHILI Lab at EPFL published in IDC 2018.

CoReader: Studying Joint Attention in a Reading Activity with a Learner Robot 202 *Switzerland*
PHD RESEARCH PROJECT *Feb. 2017 - Jun. 2018*
- as part of my doctoral research, presented in my EPFL Candidacy Exam and published in IDC 2018 receiving *Best Student Paper Award* and *Honorable Mention in Best Paper Award*

Publications

2020 doi • S. Tulli, M. Couto, M. Vasco, **E. Yadollahi**, F. Melo, A. Paiva "Explainable Agency by Revealing Suboptimality in Child-Robot Learning Scenarios" International Conference on Social Robotics, 23-35.

2020 doi • **E. Yadollahi**, M. Couto, W. Johal, P. Dillenbourg, A. Paiva "Exploring the role of perspective taking in educational child-robot interaction" International Conference on Artificial Intelligence in Education, 346-351.

2020 doi • **E. Yadollahi**, M. Couto, P. Dillenbourg, A. Paiva "Can you guide me? supporting children's spatial perspective taking through games with robots" Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts.

2020 doi • **E. Yadollahi**, P. Dillenbourg, A. Paiva "Changing Perspective as a Learning Mechanism" Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 612-614.

2020 doi • A. Güneysu Özgür, A. Özgür, T. Asselborn, W. Johal, **E. Yadollahi**, B. Bruno, M. Skeweres, P. Dillenbourg "Iterative Design and Evaluation of a Tangible Robot-Assisted Handwriting Activity for Special Education" Frontiers in Robotics and AI7 (2020), 29.

2019 • **E. Yadollahi**, W. Johal, J. Dias, P. Dillenbourg, A. Paiva "Studying the Effect of Robot Frustration on Children's Change of Perspective" Workshop of Social emotions in the 8th ACM Conference on Affective Computing and Intelligent Interaction, 2019.

2018 doi • **E. Yadollahi**, W. Johal, A. Paiva, P. Dillenbourg, "When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration," In Proceedings of the 17th ACM Conference on Interaction Design and Children, pages 195–206. ACM, 2018.

2018 doi • T. Asselborn, A. Guneysu, K. Mrini, **E. Yadollahi**, A. Ozgur, W. Johal, P. Dillenbourg, "Bringing Leters to Life: Handwriting with Haptic-Enabled Tangible Robots," In Proceedings of the 17th ACM Conference on Interaction Design and Children, pages 219–230. ACM, 2018.

2016 • **E. Yadollahi**, J.G. Ih, "Acoustic Characterization of Leaks in a Pipeline," In proceedings of Inter-Noise 2016.

Fellowships

2017 **Joint PhD Scholarship.** RBCog-PhD – Robotics, Brain and Cognition PhD program between EPFL in Switzerlandand IST in Portugal supported by Foundation for Science and Technology of Portugal (FCT)PD/BD/135150/2017. *FCT, Portugal*

2012 **Korean Government Scholarship.** Master of Science program in Mechanical Engineering at KAIST including tuition and monthly stipend. *KAIST, South Korea*

Honors, Awards, & Grants

2020 **Best Student Paper Award.** Awarded to "Explainable Agency by Revealing Sub-optimality in Child-Robot Learning Scenarios" at the International Conference on Social Robotics, ICSR 2020. *Colorado, U.S.A.*

2020 **Future Digileaders 2020.** Grant to travel to Sweden after travel restrictions are lifted. *Stockholm, Sweden*

2020 **HRI Pioneers Workshop 2020 Travel Grant.** A premiere forum for graduate students in HRI *Cambridge, U.K.*

2018 **Norman Foster Foundation Robotic Atelier Travel Grant.** Only 10 scholars from around the world were selected and funded to travel to the one week event. *Madrid, Spain*

2018 **CCI Student Best Paper Award.** Awarded to "When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration" to at Interaction Design and Children Conference, IDC 2018. *Trondheim, Norway*

2018 **Honorable Mention in Best Paper Award.** Awarded to "When Deictic Gestures in a Robot Can Harm Child-Robot Collaboration" to at Interaction Design and Children Conferenc, IDC 2018. *Trondheim, Norway*

2007 **Ranked 202nd (top 0.1%).** National University Admission Examination in Physics and Mathematics Discipline *Semnan, Iran*

2007 **Ranked 245th (top 0.1%).** National University Admission Examination in Foreign Languages Discipline *Semnan, Iran*

Service and Leadership

Special Issue on Child-Robot Interaction: Design, Evaluation, and Novel Solutions

Guest Editor *Mar. 2021 - PRESENT*

- Accepted to Interaction Studies Journal (John Benjamins Publishing)

Children, Robot, and Virtual Agents: Present and Future Challenges *Athens, Greece*

Co-Organizer *Mar. 2021 - Jun. 2021*

- Accepted to Interaction Design with Children Conference, IDC 2021

203

HRI Pioneers Workshop 2021 — *Colorado, USA*
GENERAL CHAIR (NON-US) — *Sep. 2020 - Apr. 2021*
- in conjugation with International Conference on Human-Robot Interaction, HRI 2021

Child-Robot Interaction: Present and Future Relationships Workshop — *Colorado, USA*
CO-ORGANIZER — *Aug. 2020 - Nov. 2020*
- Accepted to International Conference on Social Robotics, ICSR 2020

International Conference on Human-Robot Interaction, HRI 2020 — *Cambridge, U.K.*
STUDENT VOLUNTEER — *Jan. 2020 - Mar. 2020*

EPFelles: EPFL Female Student Association — *EPFL, Switzerland*
MEMBER — *2020 - PRESENT*

The consulting Society EPFL — *EPFL, Switzerland*
MEMBER — *2019 - PRESENT*

NCCR Robotics — *Switzerland*
MEMBER — *2016 - PRESENT*

Additional training

Mar. 2021	**The Road to a successful HRI: AI, Trust and ethicS - TRAITS Workshop!**.	*Colorado, U.S.A.*
Nov. 2020	**Future Digileaders**.	*Stockholm, Sweden*
Jun. 2020	**Children's Critical Reflections on AI and Robotics Workshop**.	*London, U.K.*
Mar. 2020	**HRI Pioneers Workshop 2020**. Presentation Video available at	*Cambridge, U.K.*
Sep. 2019	**Social Emotions, Theories and Model Workshop**.	*Cambridge, U.K.*
Jun. 2019	**Serious Games and Casual Free-to-Play Games Workshop**.	*Lisbon, Portugal*
Nov. 2018	**Norman Foster Foundation Robotics Atelier**. A video of the one week event is available at	*Madrid, Spain*
Sep. 2018	**SMART School on Computational Social, and Behavioral Sciences**.	*Paris, France*
Jul. 2018	**Methods and Research on Gaze Tracking Workshop**.	*Lisbon, Portugal*
Jun. 2018	**Near Future of Child-Robot Interaction Workshop**.	*Trondheim, Norway*
Nov. 2017	**Swiss Robotics Industry Day 2018**. Poster presented at the event is available here	*Lausanne, Switzerland*
Oct. 2017	**Robots for Learning (R4L) Workshop**.	*Switzerland*
Sep. 2017	**Social Human-Robot Interaction Summer School**.	*Alentejo, Portugal*

Skills

Programming	**Advanced:**	• Python • C# • Java • QML • MATLAB • LaTeX
	Intermediate:	• QTQuick • Processing • C • C++ • R • PHP
	Beginner:	• QBasic • Fortan • Pascal
Operating Systems	**Advanced:**	• ROS • Linux • Windows
Software	**Advanced:**	• Unity Game Engine • IBM SPSS • SolidWorks • AutoCAD • CATIA
	Intermediate:	• Blender • InDesign • Illustrator • Photoshop • Unreal Engine

Reviewing

Journals:
- ACM Transactions on Human-Robot Interaction (THRI)
- Journal of Engineering and Technology Management
- International Journal of Child-Computer Interaction
- Frontiers in Psychology
- User Modeling and User-Adapted Interaction
- Paladyn, Journal of Behavioral Robotics

Conferences:
- ACM/IEEE **HRI**
- ACM **IDC**
- ACM **CHI**
- ACM and SIGCHI **HAI**
- SIGCHI **CHI PLAY**
- ASME Manufacturing Science and Engineering **MSEC**

Languages

English	Full professional proficiency	C2
	TOEFL iBT: 105 & **GRE:** 312	Dec. 2014
Persian	Native or bilingual proficiency	C2
Korean	Limited Working Proficiency	B2
French	Elementary proficiency	A2
Portuguese	Elementary proficiency	A1

Hobbies

- Photography
- Traveling & Backpacking
- Drawing & Painting
- Fitness & Strength Training
- Extreme Sports

Personal

Nationality: Iranian
Date of Birth: 12 May 1989
Places Lived: *Semnan*, *Tehran*, Iran
Daejeon, South Korea
Lausanne, Switzerland
Lisbon, Portugal