Gamified activity for learning perspective taking

Miguel Alexandre Barroso da Costa Cardoso Monteiro

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Supervisors: Prof. Ana Maria Severino de Almeida e Paiva
Prof. Elmira Yadollahi

Examination Committee

Chairperson: Prof. João António Madeiras Pereira
Supervisor: Prof. Ana Maria Severino de Almeida e Paiva
Member of the Committee: Prof. Hugo Miguel Aleixo Albuquerque Nicolau

November 2022

This work was created using LaTeX typesetting language
in the Overleaf environment (www.overleaf.com).

Acknowledgments

I would like to thank my parents, grandparents and brothers for all their support, love and care throughout all these years. Also thank my girlfriend for her love, support, encouragement and willingness to put up with me even through my toughest times. This would not have been possible without them, and I am forever grateful for every single one of them for always being there for me.

Thank you to my friends for the good times and talks and to my extended family for their help.

I would also like to extend my thanks to my supervisors Prof. Ana Paiva and Prof. Elmira Yadollahi, and to Marta Couto for their valuable insights, feedback and help on that made this work possible. I would also like to thank Colégio Bambino and Colégio S. Gonçalo and their staff for their availablity and kindful help to make the study possible.

Abstract

Perspective taking is a crucial ability for any human, using it in nearly every social interaction. The developments of the Human-Robot Interaction area and the increasing need for innovative learning ways have led to a increase in robot-assisted education. This thesis proposes an activity with a game that aims to explore and develop children's perspective taking abilities. This work continues and develops a new game from the work previously done in the development of the Virtual Maze two-player game, where the player must compete or cooperate with an AI agent while moving through a maze. While the original game was virtual, a physical version was created and the impact of embodiment and of the agent's perspective taking abilities in aspects such as the children's learning, enjoyment, perception of the robot or attention was tested. This was done by evaluating the children's pre and post gameplay perspective taking capabilities and collecting and opinions on the activity. This thesis explores the implications of developing perspective taking skills in children using virtual versus embodied interaction and by providing a fun, immersing way of practicing the skill. A study done with 56 children of the 3rd and 4th grade (7-10 years) showed a significant global improvement of spatial perspective taking skills after playing the game, alongside with the children that played the physical game indicating having significantly more fun than the virtual game children.

Keywords

Spatial Perspective Taking; Child-Robot Interaction; Gamified Interaction; Robot-Assisted Learning; Gamified Learning

Resumo

Tomada de perspetiva é uma competência crucial para qualquer humano, usada na maioria das interações sociais. Os desenvolvimentos na área da Interação Humano-Robô e a cada vez maior necessidade de maneiras de aprender inovativas levaram a um aumento na educação assistida por robôs. Esta tese propõe uma atividade com um jogo cujo objetivo é explorar e desenvolver das competências de tomada de perspetiva espacial nas crianças. Nesta tese foi extendido o trabalho previamente feito no desenvolvimento do jogo Virtual Maze, onde o jogador deve movimentar-se por um labirinto e competir ou cooperar com um agente. Enquanto que o jogo original era virtual, esta nova versão foi também adaptada para um ambiente físico, sendo um dos objetivos testar o impacto de ter o agente físico em aspetos como a aprendizagem e o divertimento da criança ou a sua perceção do robô. Isto foi feito avaliando as capacidades de tomada de perspetiva espacial da criança antes e depois da atividade e também recolhendo as suas opiniões da atividade. Esta tese explora o desenvolvimento da tomada de perspetiva espacial em crianças, conceptualizando e desenhando uma maneira divertida e imersiva de praticar esta competência. Um estudo realizado com 56 crianças dos 3º e 4º anos de escolaridade (7-10 anos de idade) mostrou um aumento global significativo das capacidades tomada de perspetiva espacial depois de jogar o jogo, assim como as crianças no jogo físico reportaram significativamente mais diversão que as crianças no jogo virtual.

Palavras Chave

Tomada de Perspetiva Espacial; Interação Criança-Robô; Interação Gamificada; Aprendizagem Robô-Assistida; Aprendizagem Gamificada

Contents

List of Figures

List of Tables

Acronyms

APT Affective Perspective Taking

CPT Cognitive Perspective Taking

CRI Child-Robot Interaction

FoR Frames of Reference

HRI Human-Robot Interaction

PT Perspective Taking

SPT Spatial Perspective Taking

ToM Theory of Mind

VPT Visual Perspective Taking

1

Introduction

Contents

1.1 Overview

Interacting and coexisting with robots is becoming ever more common. These can be facilitators of tasks, social companions, or learning assistants. Human-Robot Interaction (HRI) studies these interactions and one of its more prevalent fields is Child-Robot Interaction (CRI). Children are starting to learn with the help of robots, either by direct tutoring or by being part of activities to help them learn a specific subject or ability.

By creating gamified experiences with robot interaction, we can give children the opportunity to learn and develop skills while keeping them interested with the novelty of a robot. We aim to explore how we can help children develop their perspective taking skills with one of these activities. Perspective Taking (PT) is the ability of considering and understanding the world from other viewpoints, and it is fundamental for human development, as humans have a need to constantly use it in social interactions.

Despite both PT and embodiment being two hot topics in literature, with several studies being done in the last few years, studies overlapping both areas are fewer, especially in the context of Spatial Perspective Taking (SPT). As such, this study aims to explore this overlap in the context of CRI and children's development of SPT, taking as baseline the work done by Yadollahi in Virtual Maze [1] and evolving it to this new context.

1.2 Objectives

The first target is to explore the impact of different settings of the developed game in aspects like the learning, engagement and attention of the children. The game is two-player turn-based where children play either against or with an AI-controlled agent and must navigate a maze to win the game. If the study is successful, we can provide an entertaining activity where children will improve their SPT skills.

The study will be developed by focusing on assessing the impact of *the embodiment*, this is, playing the game with physical robots as agents or playing a computer version. We will evaluate the children's SPT capabilities with tests before and after playing the game and document children's perception about the game and the agent via a series of questions about their experience playing the game, namely regarding fun and difficulty.

The research questions are mainly motivated by how the activity can help the children and what is the impact of embodiment. Namely, regarding improvement of the SPT abilities after playing the game, it is hypothesized that there will be a significant improvement in post-test scores compared to the pre-test scores, as well as that this improvement will be significantly greater in the Physical condition. Then, regarding the impact of the embodiment condition in the children's perception of the game, it is hypothesized that children playing the physical version will find the game more fun than the ones playing the virtual version. Finally, regarding the impact of the embodiment in the game performance, it is hypothesized that the children in the Physical version will require less moves to complete the game, compared to the children in the Virtual version.

2

Theoretical Overview

Contents

Perspective Taking is the ability to consider and understand the world from other viewpoints. It can simply mean acknowledging others having a different perspective or computing and perceiving the perspective of others. There is several literature supporting that PT is crucial for human development. Being a complex topic, PT has several dimensions and levels and the development of each of these comes at different ages.

To better understand the task of PT, one could describe it with three basic components, as suggested by Surtees et al. [2], specifically:

- 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.

For instance, say person A (Self) sees a rabbit (Object) at their left. If A and B are standing in front of each other, person B sees the rabit at their right. Person A understanding Other's perspective, person B, is understanding from B's point of view the rabbit is at their right. These components give a good basis to better understand PT and its dimensions. The most commonly referred dimensions are Perceptual, Cognitive and Affective, as proposed by Kurdek and Rodgon in 1975 [3].

2.1 Perceptual Perspective Taking

The Perceptual dimension refers to Self's ability to perceive or imagine what the Other sees. It can be subdivided into Visual Perspective Taking (VPT) and SPT, though it should be noted that there is a part of the literature on visuospatial dimension that does not distinguish between these.

2.1.1 Visual Perspective Taking

Visual Perspective Taking refers to the Self's awareness of the Other's visual field of view. Literature on this dimension suggests that there are sub-levels inside it. Flavell et al. [4] define a 2-level division, where the major differences are with the what and the how relations.

In level-1, it is enough for Self to understand what Object is visually seen or accessible by Other in the world. In level-2, Self must understand how Other sees the object or the surrounding world, taking into consideration possible different viewpoints of representation. This level has been associated with the development of Theory of Mind (ToM) skills and it is developed at a later age than level-1. Level-2 requires more cognitive effort and it is linked to being embodied in the form of a deliberate movement simulation [5].

In addition to this, a later study by Moll and Meltzoff [6] suggests a level-0 that refers to the joint attentional abilities of infants at around one year of age. While no perspective knowledge is yet developed, infants have shared joint attention or engagement, as seen in activities such as pointing, alternating gaze or holding up and showing. The same authors define ”level-1 experimental PT” as the knowledge of what others are familiar with or not from past experiences. This particular level seems to be developed earlier than the level-1 PT. Furthermore, these also introduce a level-2A and a level-2B. Level-2A refers to when the child can recognize how Other sees the Object even if the way they see it differs from the child’s perspective. Level-2B is considered developed when the child is able to confront perspectives, this is, when the child can understand that an Object can be seen in different ways, given different viewpoints.

2.1.2 Spatial Perspective Taking

Spatial Perspective Taking refers to the Self’s ability to understand the Other’s spatial relation with the Object in the world. The concept of Frames of Reference is of upmost importance in this type of PT, as it is what one uses to encode the spatial information relative to the subject in question.

When producing descriptions of spatial relationships or when trying to understand them (this is, trying to perform a SPT task), a Frames of Reference (FoR) must be adopted. The modalities of these vary depending on the literature or discipline they are described in, such as philosophy, linguistics or brain sciences. The following are the most relevant for the SPT exercise frames of reference and their areas of use, adapted from Levinson [7] by Yadollahi [1]:

1. ’relative’ vs. ’absolute’
    - Philosophy, brain sciences, linguistics;
    - Relative is described by relations between objects, absolute is infinite void.
2. ’egocentric’ vs. ’allocentric’
    - Developmental and behavioural psychology, brain sciences;
    - Coordinate system originating in Self’s body (viewer-centered) versus anywhere else (environment-centered);
    - Also referred as subjective versus objective.
3. ’deictic’ vs. ‘intrinsic’
    - Linguistics;

- Also referred as speaker-centric versus non-speaker-centric;
- Centered on any of the speech participants versus not so centered;
- Sometimes extrinsic is used as a third term to include contribution of gravity (for instance, in words above or on).

4. ’viewer-centred’ vs. ‘object-centred’ vs. ‘environment-centred’

   - Psycholinguistics;
   - In viewer-centered, coordinate system is based on the perceiver’s perspective of the world;
   - In object-centered, objects are coded in respect to their intrinsic axes;
   - In environment-centered, objects are represented with respect to salient features of the environment.

Nevertheless, the most commonly used frames of reference for describing relationships between humans and/or objects in the world are: viewer-centered frames, object-centered frames and environment-centered frames.

Regarding Spacial Perspective, the frames adopted and used by Levinson [7], Surtees et al. [2] and in other literature are:

- relative FoR: relative to ourselves
- intrinsic FoR: relative to another person/thing
- absolute FoR: relative to some non-varying degree in the environment

Lastly regarding FoR, it is worth noting that there are factors that can induce differences in the use of frames of reference, namely if the Object involves another human or an inanimate object or if the one describing is an adult or a child. One example of this is SPT’s use in conversation: A study by Schober found, when performing a task describing the location of objects, adult participants that described these objects to another person in the room mostly used more egocentric perspectives than those describing them to an imaginary addressee [8]. The speakers would take turns describing the objects and speakers who described only one object at a time tended to be more explicit in their perspective than those who described a number of objects before switching roles.

Similarly to Visual PT, one can also consider different levels of SPT, that develop at different times and its usage depends highly on the task and on the development of the person [9]. Surtees et al. proposed a 2-level model for SPT [2]. The level-1 refers to the distinction between front and behind, while the level-2 refers to the left-right distinction.

As in VPT, the level-1 is considered to be developed earlier. It can be explained with the finding that level-2 judgements require egocentric mental rotation, in contrast to the level-1 judgements. In a study by Surtees et al., it was observed evidence of unintentional level-1 PT in adults, but not for level-2, suggesting the need for cognitive control [10].

2.2 Cognitive Perspective Taking

The Cognitive dimension refers to Self's ability to imagine what Other's experiencing. So, in this case, we can consider the Object as 'experience', for example, a stimulus, to which Self thinks about Other's knowledge of it and its perception, possibly aided by Self's own experiences to such stimulus.

This dimension has links to the ToM and the research on both subjects is related. The relation with pro-social behaviour has also been studied and the research on the topic contributes for the development of programs whose goal is to help populations such as those diagnosed with autism, whose PT abilities are often impaired, visible in their social behaviours.

2.3 Affective Perspective Taking

The Affective dimension refers to Self's ability to understand Other's feelings and emotional experiences. The term Empathy is significant and closely related in this context, meaning "the ability and tendency to share and understand other's internal states" [11]. Empathy and Affective Perspective Taking (APT) are also related to the Cognitive dimension, as studies by Hinnant and O'Brien have found that Cognitive Perspective Taking (CPT) moderates the relation between both, and gender is impactful in this relation [12]. Other studies by Eisenberg et al. have also shown a positive correlation between empathy and self-reported measure of CPT, in adults [13].

2.4 Theory of Mind

As mentioned, perspective taking is closely related to ToM, "which refers to the capacity to understand other people by ascribing mental states to them". According to research in this topic, there are five levels of understanding informational states that must be mastered for children to be able to take another perspective [14] and perspective taking plays a meaningful role in the development of the first two levels [15]. Specifically, level 1 involves simple VPT, relating to the fact that people can see different things, similar to the one of Flavell and collegues' model. Level 2 is referred as complex VPT and it is the "ability of knowing that people can see the same things differently". As for the remaining levels, level 3 is about understanding that "seeing leads to knowing". A common example scenario is questioning a child about an object that was put inside a box while the child had the eyes closed, and the child figuring out that it's not possible to know what is inside the box because he/she did not see it. Level 4 regards true beliefs and acting accordingly, this is, knowing only what has been seen and acting on it. Finally, level 5 addresses false beliefs, and predicting actions on its basis. A common scenario for testing this level is the candy box. A candy box is shown to a kid and he when asked, it is likely that he says candy is in it. When opened, the child finds pencils inside. Then, when asked what would someone else think the box had before opening, if the child answers candy it shows that this level is developed.

The false-belief level is expected to be developed by around 4 years of age [16], [17]. However, as pointed out, conditions such as those of the autistic spectrum can lead to deficits of the ToM abilities.

2.5 Perspective Taking in Robotics

Since scenarios can be designed where the interaction with agents or robots equipped with PT abilities brings benefits to individuals, it is important to investigate how one can equip agents with these abilities. Regarding this, it has been shown that humans make similar assumptions about robots as the ones they make for their human counterparts. This means that humans can take the robot's perspective almost as if it was another human.

Recently, several studies have developed agents with PT abilities and tried to understand how these can improve human and robot interaction. These were done in several different scenarios, including classic level-1 and level-2 perspective taking problems. Globally, PT is deemed to have an important role in collaborative and learning scenarios.

3

Related Work

Contents

3.1 Human-Robot and Child-Robot Interaction

Human-Robot Interaction (HRI) studies the interactions between humans and robots, and has seen a growth in relevance in the past years, as robotic agents are becoming more common in our lives, in a vast number of sectors. HRI is a interdisciplinary field, with contributions from fields like psychology, robotics, linguistics, computer science or design. Despite the advantages of this field and its increased use, there is some skepticism and lack of trust towards the everyday use of robots, such as in decision making and high stakes scenarios due to a lack of trust [18]. The number of studies in this area has also increased, and they cover a wide range of topics, such as how humans perceive the robot, how the human-robot relationship is developed, the impact of physical versus virtual or how HRI can be used in learning scenarios. The latter topic has had more research in the Child-Robot Interaction (CRI) field, a specific branch that focuses solely in the interactions with children.

Studies suggest that, when compared to adults, children react more strongly and more energetically when interacting with robots [19]. In line with this, children are less likely to use robots as tools, increasing the significance of the social part of the interaction [20]. As such, developers must take special care with this aspect when designing the agents. Factors like the facial expressions, movement, behaviour and embodiment of the robot can impact the social aspect, which can be assessed by evaluating metrics such as engagement, social interaction, empathy or 'believability'.

Ros and colleagues suggest that highly structured formats of activities may inhibit engagement, however, simple social behaviours like the robot asking for help proved at least as effective as methods whose objective was to directly manipulate engagement, concluding this metric rises 'as children are drawn into interaction' [19].

The human's perception of the robot may play a significant impact in the interaction. A study by Currie and Wiese put a human against a Cozmo robot in a competitive Go/No-Go game, where the participant was either told Cozmo was AI controlled or controlled by a human in a separate room [21]. The study suggests that this had a significant impact on the human's strategy and action execution.

One way of influencing perception is empathy. It is linked to being important in human communication and social behaviour. As such, interaction with robots may improve with empathy, both how much a human empathises with the robot and how the human perceives the robot's empathy, the latter being particularly important in cases such as companion robots, where there may be concerns of a lack of sense of agency and empathy [22]. A bigger perceived empathy can lead to a human more easily accepting and building with a relationship with the robot [23]. The RoPE Scale (Robot Perceived Empathy Scale) provides a way of reliably measuring perceived empathy with a set of questions which evaluate emphatic response and emphatic understanding that can be adapted to the particular problem and robot [24]. As for humans feeling empathy for a robot, a study by Seo et al. developed a scenario for inducing empathy for a robot and provides a generalizable instrument for measuring it [25].

3.1.1 HRI in Education

The impact of robots in education is increasing, and their role can range from being part of an activity that helps children learn a certain subject to being a tutor-like agent to assist children learning. Robot-assisted tutoring is a prevalent topic and there has been research to define its pedagogy, like the L2TOR project [26] and to develop guidelines for it [27].

Teaching others has been shown to be a powerful learning method by research in education and cognitive sciences [28]. More recently, the same effect was suggested in computer-assisted learning environments [29]. Similar to this, Li and colleagues suggest that a cooperation environment allows for children to develop their PT abilities more than in a competitive environment [30].

However, as in the whole HRI field, there are still some concerns about it's use, such as the teacher's pre-conceived notions that it will be disruptive, even if the experience afterwards does not match that expectation [31], or studies suggesting that children will still learn more with a human tutor, as opposed to a robot one, despite there being learning growth in both cases [32].

Blancas et al.,in a study where a robot gives a history lesson, point out knowledge improvement and capability of the robot to capture attention, but also raise awareness to gesture moderation since while attracting attention is important for the engagement of the student, he/she can be distracted by too many gestures by the robot [33]. This is inline with Leyzberg et al., that point out that the novelty of the robot can distract the students from knowledge acquisition, since they will focus on the robot itself instead of in the subject being taught [34].

As for the social and emotions impact, a study by Saerbeck et al. supports that the learning efficiency and motivation of students is bigger when the involved robots show more expressive and social behaviours [35].

3.1.2 Embodiment

There have been several studies regarding the impact of embodiment in aspects such as attention, enjoyment, performance in the activity or in learning, and results vary. The already mentioned study by Seo et al., where participants interacted with a robot that would exhibit a functional problem after showing its intelligent and autonomous capabilities to evaluate the participant's emphatic responses to this, found that people empathized more with the real embodied robot than with the simulated robot [25].

A study by Bartneck replicating a intelligent home context interaction found that there was no significant change in enjoyment between embodiment conditions, but that the presence of emotional expressions greatly increased the enjoyment [36]. The aforementioned study by Saerbeck et al. also mentions that the sociability of a learning agent can be enhanced by physical embodiment [35]. In the context of learning, a physical robot can also positively impact the learning gains, as suggested by Leyzberg et al. [34].

On the other hand, in a Towers of Hanoi related task, where the robot would tell the participant a stacking goal, Wainer and collegues suggest physical embodiment can make a difference in perception of social capabilities and enjoyment [37].

In a study by Kennedy and colleagues, where children performed a sorting task in a monitor with alien-like generated pictures with the help of a agent, with the condition being this agent being a robot or a virtual agent in a screen, no better performance in the task was found but authors suggest that physical robots carry an advantage in terms of social presence [38].

A study by Pereira and colleagues, where the participants played chess against a virtual or physical agent, reports that participants that played against the physically embodied robot had more fun [39].

While there may be different results that may be explained by the impact of the experiment's setting, the general direction seems to be that a physically embodied robot may have advantages, especially in the social aspect.

3.2 Spatial Perspective Taking

As for PT, several studies have also been carried out across all dimensions of PT as mentioned in the previous chapter, with the most relevant for this work being the ones in the context of learning and developing SPT, particularly in children.

The Objects Game study by Yadollahi et al., where each child had interact with a robot sitting in front of them by giving or receiving instructions and move around objects with different shapes (circles, rectangles) and colours shown in a tablet to reach a desired configuration aimed to study how the robot's perspective choices impact the children's own choices. It was observed that when the robot used a egocentric perspective, children mostly switched their perspective to implicit addressee-centric and that changing the robot's perspective tactic mid-interaction lead to confusion in the children, as this change was unjustified [1].

Also by Yadollahi et al., the Cozmo Maze study, where children played a maze guiding game such as the one developed in this study, but simpler and with only one agent on it and where children need to constantly adjust their perspective to the robot's, found some improvement in post-tests compared to pre-tests, although not statistically significant. More importantly, this study gave insights and set up the design of the Virtual Maze study, also a maze guiding game, but more complex with two agents on the board and with two conditions: game mode (Cooperative vs. Competitive) and agent's PT capabilities (Full-PT vs. Partial-PT). Due to the pandemic, the two studies were carried out on adults it was observed a significant improvement in post-tests compared to the pre-tests and a bigger improvement in the cooperative game mode [1]. Due to the importance of the Virtual Maze game for this work, it will be described in more detail in an upcoming section.

In conclusion, while there has been many studies of SPT and of the effects of embodiment, the area that combines the two has not yet been fully explored and contributing - the effects of embodiment in learning and developing SPT has not yet been fully explored and contributing to it is one of the main motivations of this work.

4

Technical Development

Contents

This chapter will describe the technical development of the game used in this study, starting by describing Virtual Maze, the baseline game and then the new game and the whole development cycle associated with it.

4.1 Virtual Maze game

The Virtual Maze game, developed by Yadollahi, will be evolved and adapted in this thesis. As such, it is relevant to understand the basic concept of it, as well as the conclusions taken from the study done with it [1].

4.1.1 General flow

The game, developed in Unity Game Engine, is a two-agent turn-based game, where each agent does one move at a time. Agent A, or Callisto, is controlled by the participant and Agent B, or Polaris is AI controlled using symbolic programming.

The general goal is navigating through a maze to get to an end zone while avoiding specific constraints and red zones (in which case a life will be lost and the agent will return to the starting point of that specific move).

Both agents can move forwards, backwards, left and right. Spatial perspective taking comes into play since the movement is not absolute but relative to the orientation of the agent. For instance, if the agent is facing north, the right command moves the agent east (no angular disparity). However, the same command with the agent facing east, does not make the agent move east but south (-90º angular disparity).

As such, in order to win, the participant needs to take into consideration both Callisto and Polaris's orientation, as it influences the movement of Polaris depending on the game mode, which will be described here.

4.1.2 Competitive

In this game mode, as illustrated in figure 4.1, the goal of the participant is to get Callisto to the green safe zone before Polaris does. To do so, it must avoid going into the red zones or replicating the last move of the opponent (the first move has free choice). For instance, if Polaris goes left, then Callisto can not perform the action Left. An error (red zone or same move) will result in losing one life. Each player has 3 lives and if one runs out of those, then the other agent will automatically win.

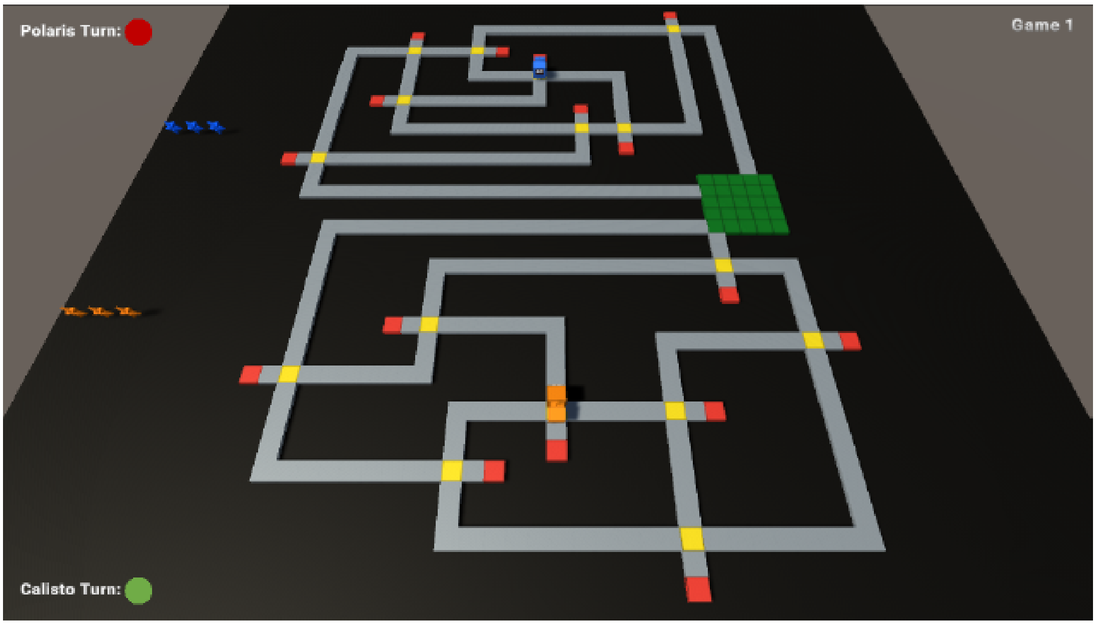

Figure 4.1: Screenshot of the game's competitive mode, with both agents (Polaris blue, Callisto orange).

4.1.3 Cooperative

In this game mode, seen in figure 4.2, the green safe zone is only accessible by one of the agents - the follower, and the goal is for the other agent - the leader - to guide the follower to the safe zone. This is similar to the Can You Guide Me? activity also developed by Yadollahi and colleagues [40], but slightly more complex since there may be angular disparity between the follower and the leader. The agents have a combined five lives, and similarly to the Competitive game, they lose one by either of them stepping on a red zone. However, this time the follower must replicate the leader's last move, and failing to do so will result in also losing one life.

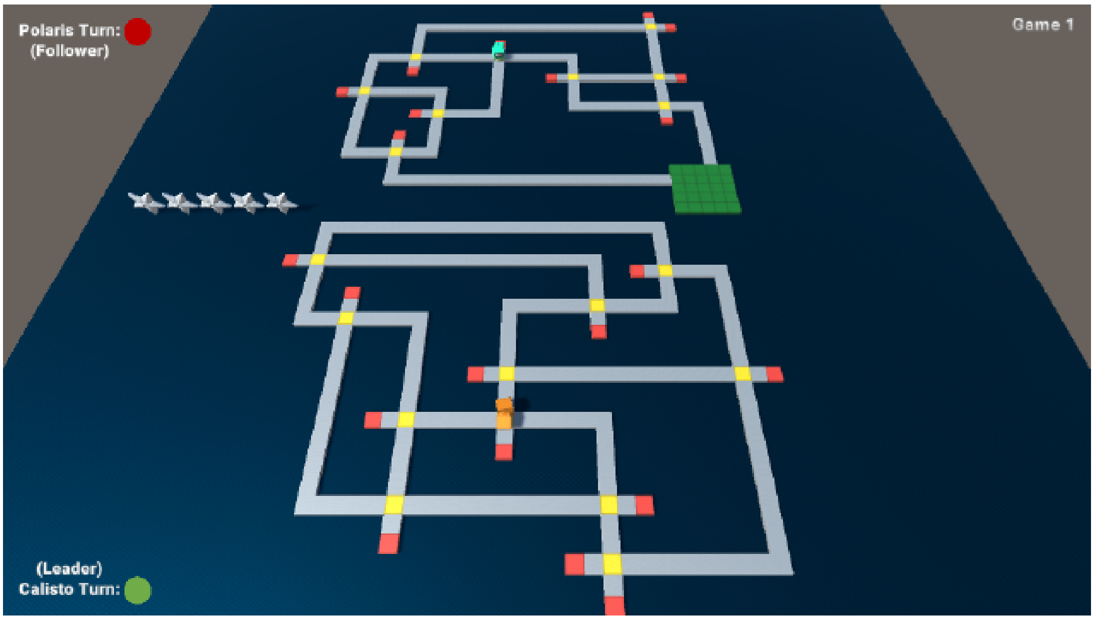

Figure 4.2: Screenshot of the game's cooperative mode, with both agents (Polaris blue, Callisto orange).

4.1.4 Agent Decision Making

There are two versions of Polaris' AI used in the game. In the Full-PT version the robot has perfect PT abilities while in the Limited-PT version the agent makes mistakes when there is a 180º angular discrepancy between the two agents. This replicates a common type of error children make, as found in the Cozmo Maze study [40].

4.1.5 Baseline Studies

There were made two studies with similar goals. The participants were in both cases, adults recruited in the Prolific platform with most in the 18-24 age group (66.2% and 63.15% respectively) followed by the 25-29 age group (23.9% and 16.84%, respectively). In both, the flow was as follows:

Pre-game tests to evaluate the participant's PT abilities prior to playing the game. **Tutorial** to ensure that the participant understood the rules of the game (for both competitive and cooperative). **Game play session** where the participant plays 4 games. In one of the studies, the Full-PT/Limited-PT condition of the robot was a between-subject variable (so each participant would only play once of these conditions) and the game type order was competitive-cooperative and cooperative-competitive, alternating the starting agent. On the following study, both the Full-PT/Limited-PT and the game type condition were treated as between-subject, so each subject would play 4 games of one out of four possible combinations (Full-PT/Coop, Full-PT/Comp, Limited-PT/Coop, Limited-PT/Comp). **Post-game tests** after the gameplay, tests were made to evaluate the participant's PT abilities and also the general perception of the game of factors such as fun, intelligence of the robot and difficulty.

These studies provided some interesting conclusions about the activity and perception of the robots abilities, the most relevant of which is that the participants significantly improved in the post-tests compared to the pre-tests, and this improvement was higher in the cooperative condition. However, the study also shows that there are more mistakes made in the cooperative mode. Adding to this, the participants deemed the game more fun and Polaris more intelligent in the Full-PT version.

The main limitations of these studies are the game being virtual and the participants being adults, as the desire was to run these studies in children with in person interaction, which was not possible due to the pandemic context at the time. With that in mind, this is one of the directions we want to expand this work on in this thesis, and its details will be explained in the following section.

4.2 Game Design

In this section the game will be explained in detail. Firstly, the overall rules of the game will be explained, followed by a description of the agent's decision making and how it was technically implemented and finally the design iterations will be discussed, providing some light on the process that took us from the Virtual Maze game to the current version.

4.2.1 Game rules

The game is a turn-based game with two agents in it, one being controlled by the child.

Since the moves in the game are regarded relative to each agent's orientation and not to the absolute frame of reference, this is where SPT comes into play. Namely, the player must use it twice when deciding the next move: Firstly the player must understand the other agent's last move and then the player must translate his selected move to the player's agent orientation.

4.2.1.A The goal

The goal is to navigate the maze and reach the end zone. However, at first the path to the goal is not completed. There are activation blocks between certain nodes (where the agent stops after a move). When an agent travels between two nodes with an activation block, this block will be deactivated and a part of the path to the end zone will be constructed. Thus, both agents must first travel the maze and activate a certain number of blocks to build the path.

4.2.1.B Losing lives

Each agent has a certain amount of lives per level. These can be lost two ways: By choosing a path to a red zone or by replicating the other agent's last move. For example, if the AI agent moved Right, then the player can not choose to move Right. As mentioned, this is where SPT is taken, as the moves are relative to each agent's orientation and not to the player's perspective. Four moves are possible: Right, Left, Front and Back.

Making a bad move will force the current agent to re select a move until a good move happens or all lives are lost.

Losing all lives will result in losing the game.

4.2.1.C Gamemodes

As mentioned, there are two gamemodes: **Cooperative** and **Competitive**. The rules and map designs are exactly the same for both modes, with one big exception: In the cooperative mode, if an agent activates a block, the path is built on both sides, whereas in the competitive version it only builds the path on the agent's own side. As a consequence, there are less activation blocks and more missing path blocks in the cooperative mode compared to the competitive mode. This was balanced in a way that it is necessary for both agents to activate blocks, as a way of inducing cooperation.

4.2.2 Game implementation

4.2.2.A Code refactoring

This project kicked off form the Virtual Maze source code, with the first part of the development being going through it and refactoring some classes. This aimed to make the game easier to implement the new version which, unlike the original game, had a lot of similarities and it would make sense for the classes with similar behaviours to be centralized, as opposed to having different classes for each game mode with very similar code except in a few parts. As such, classes that controlled the AI agent, the child agent and the state of the map and game were all joined irrespective of game mode (for example,

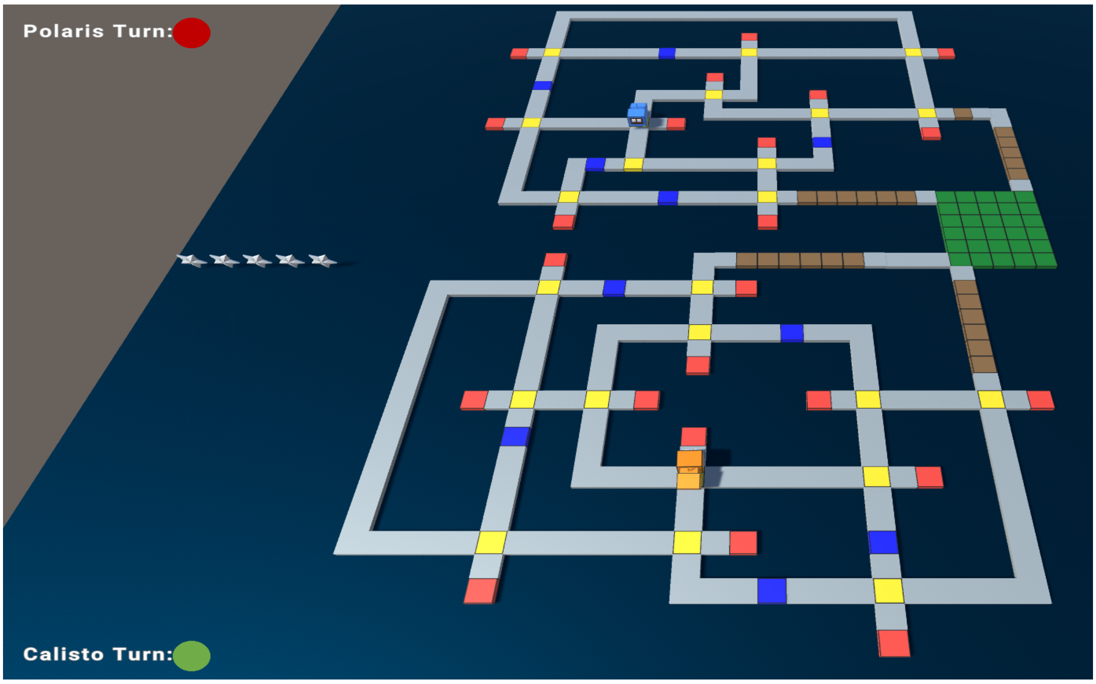

Figure 4.3: New cooperative game mode.

only a single script controlling the AI agent exists, instead of one for each game mode). Furthermore, the map was now dynamically loaded from a .csv file instead of being hard coded in each script per map. Again, this allowed for a greater flexibility in game and map design. To create a map, a new scene would be created where the different needed settings would be set directly in the map controller object in Unity Editor, avoiding having to create and tune different scripts in the code editor.

4.2.2.B Agent's Decision Making

The agent's decision making is based on a rule-based priority system. For each move, go through all the rules and add or subtract a certain amount of priority if a rule is true. Finally, choose the highest priority move and apply it. In case of a tie for highest priority, select at random.

Most rules depend on two variables: What the AI agent is getting by selecting the move and what the player's agent is losing by blocking that move. Some rules are game mode dependent, for example, blocking a path to the end zone or to an activation block loses priority in the cooperative version but gains priority in the competitive one.

Some rules in the competitive mode have a chance-based component. For example, a move that would block a activation zone path has a 25% chance of not applying its priority. This is to avoid the game of becoming too frustrating by making the AI agent not block the player's good moves every time. The full list of rules is seen in table 4.1.

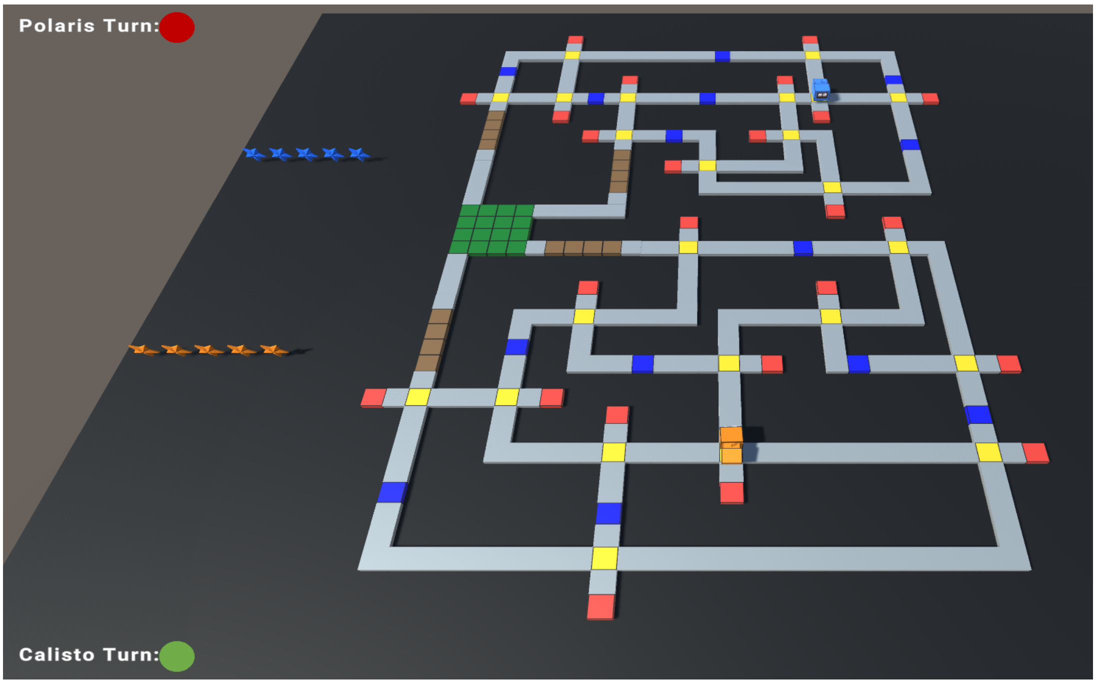

Figure 4.4: New competitive game mode.

4.2.2.C Virtual Version

The virtual version was made using Unity Game Engine, with one build per level being compiled. At the time of the activity, the selected build would be ran in the experimenter's private computer. Every gameplay session saved the selected moves data into a shared by all builds CSV file, for later analysis. While the initial project was an extension of Yadollahi's, a significant portion of code was changed or refactored to reflect the changes between the two versions, such as the path building mechanic, the changes to the AI or to the collection of data, and to allow for more flexibility in the development of different levels, such as the map being generated from reading directly a CSV file.

4.2.2.D Physical Version

The physical version was made using the Python programming language, using the anki_vector package to control the movement of the robots, among other auxiliary ones. To play the game, a game controller script would be ran, with two arguments that would describe game mode and level. The script would read a CSV file that encoded the map that was printed in a cardboard such as stopping points, paths between them (distance and turns), which routes had path building blocks, and so on, with this information then being stored in objects created for that effect, alongside other objects that would keep track of the state of the game (position and orientation of the robots, last moves, picked up activation blocks, number of lives). Then, the script would take control of the Vector robots, so they would not move around between

Table 4.1: The rule-based system for the decisions. For each move, it takes into account the result of that move for the robot or for the child (move being blocked) and game mode.

AI Destination	Child Destination	Game mode	Priority	Apply Chance
Red/Unbuilt Path	-	-	-100	100%
Building Block	-	Competitive	+3	75%
Building Block	-	Cooperative	+3	100%
Green Zone	-	-	+10	100%
Yellow with green access	-	-	+3	100%
-	Green Zone	Cooperative	-11	100%
-	Green Zone	Competitive	+5	65%
-	Red	Cooperative	+2	100%
-	Building Block	Cooperative	-2	100%
-	Building Block	Competitive	+2	65%

moves, and the experimenter would place them in their respective starting blocks and orientations, as seen in figure 5.2. Then the game would start, with the script replicating the behaviour of the virtual game (control turns, decide AI's move, validate moves). Each time a move was validated, the controller would request the respective robot to move to the next block by performing a sequence of movements as encoded in the CSV file (e.g. go forward x distance, turn y angles, go forward z distance). The data was collected and stored in a similar way as the virtual game.

Once the game ended, each robot would perform a set of animations depending on the game result, before releasing control of the robot.

4.2.3 Design Iterations

This section will describe and justify some decisions about this project and how the game evolved from the Virtual Maze to the current version. Keeping the goal of evaluating embodiment and gamemode in mind (Physical vs. Virtual and Cooperative vs Competitive) it was important to keep all versions as similar as possible.

4.2.3.A Cooperative vs. Competitive

Evaluating cooperation vs. competition in the Virtual Maze task would not be a trivial task. While the underlying SPT mechanic would be the same between the two versions, the flow and nuances of the game were different, especially in terms of map design and of the possible moves. Alongside this, some participants gave feedback that the game did not feel exactly cooperative as it was one agent guiding the other but the actions of the agent being guided didn't really impact the guiding agent other than in the overall result of the game.

The solution to this was the design of the activation blocks. By adding this, we were able to standardize the flow of the game as now the impact of actions and the calculation of wrong moves would be the same for both game modes, with the only slightly different thing being the number of activation blocks and missing path blocks, necessary to keep the balance. This difference in numbers is mostly to ensure that there is cooperation between the agents. While one could argue that one of the agents could cooperate simply by not standing in the other's path, we believe that the cooperation aspect increases by keeping the number of activation blocks on one side lower than the number of missing path blocks, making each agent to at least build a part of the other's path.

4.2.3.B Physical vs. Virtual

For this point, simplicity is key. Developing the virtual game first allowed for a bigger flexibility on testing ideas and on what was possible to do. However, every time an idea came around it was necessary to think if it was possible to do in the physical setting without over complicating things. One good example of this is the tutorial level. While the Virtual Maze tutorial was very well constructed and meant to be independent (no experimenter needed for clarification), designing it in the physical version would not yield the same results. Activation blocks also have proved a good idea here, as it is easily implemented in a simple Wizard-of-Oz style.

4.2.3.C Stalemates & Frustration

One early version of the current game had the number of activation blocks equal to the number of number of missing path blocks (for example, in competitive there would be 4 of each on each side and in cooperative 6 missing path blocks and 3 activation blocks on each side). Alongside this, the AI-controlled agent was also making all the correct choices. These facts resulted in longer game times and led to some frustration, especially noticed in the competitive version, where sometimes the game would even reach a stalemate where the players could never either complete the path or win the game as their moves were always blocked.

Since this is not meant to be a competition but a fun and engaging activity, it was imperative that this was fixed. Introducing more activation blocks than missing paths reduced game and stalemates. Additionally, the agent's decision making was also tweaked to allow for chance of some blocking moves not happening, allowing the player to do those moves. The idea was simple: at the end of the day, winning is more fun and constantly being blocked the winning move or going around in circles for some minutes is not.

4.2.4 Number of Lives

The Virtual Maze study, done in adults, had 3 lives per level. For the current study, since we'll be dealing with children, which are expected to make more errors, we settled on 5 lives per map. The first level is an exception, since the goal is learning the mechanics of the game. In this case, there will be unlimited lives to make sure the participant gets the necessary time to understand the game.

4.2.5 Pilot & activity changes

In the next section, the activity will be explained in detail. Its design had some changes due to early feedback and availability constraints, which will now be described. Early feedback was gotten from a session with 3 children in the 4th grade at the time (9 years).

4.2.5.A Tutorial

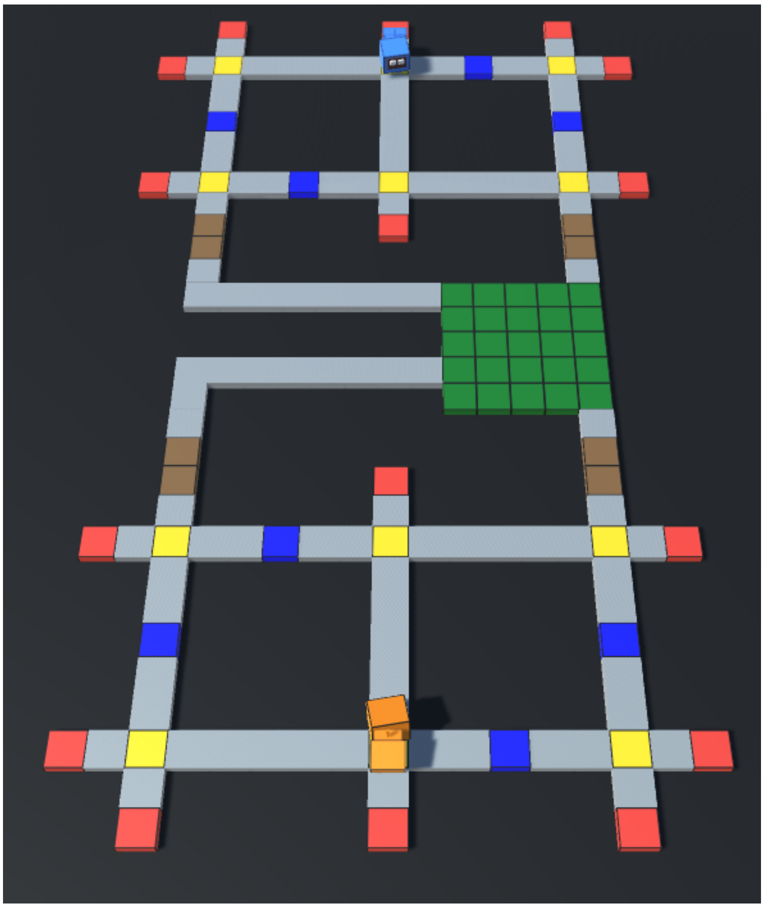

Figure 4.5: Tutorial level

The tutorial was changed to accommodate the changes in the target group (from adult to children) and the addition of the physical version. In this particular level, there will be no lives lost and the experimenter, after initially explaining the rules, will be actively engaging with the participant to ensure the rules and controls are understood. The design principle of this level was keeping it as symmetrical and as few nodes as possible, with no 90º degree turns between nodes, except on the path to victory, as seen in figure 4.5.

There are a few reasons why this route was chosen instead of a step-by-step tutorial level as in Yadollahi's work:

Consistency. While it is easy to create this step-by-step level in the virtual game, doing so with the physical version without the intervention of the experimenter is much more difficult. By doing it this way, consistency is kept between the two game modes. **Early Feedback.** Early feedback suggested that the step-by-step tutorial was not enough for participants to fully understand the rules without the experimenter's help. **Less post-tutorial games.** Since, due to time constraints, it was only possible for each participant to play one game after the tutorial (less than the previous study), a slightly longer, more complex tutorial was desirable as it would induce less errors by not knowing the rules/mechanics.

As for the activity it self, a the following adjustments were made: 1. Due to limited number of participants, it was decided to drop the conditions regarding gamemode, keeping the game Cooperative only. 2. Due to the activity being too long and the limited time where it was possible to do the activity, it was decided to drop the Path Test and to reduce the number of games played post tutorial from 2 to 1. 3. Early on, it was noticed that the children were struggling to even understand their own robot perspective, let alone the other robot. The game felt too difficult and as such, it was decided to explicitly block the other agents last move by not allowing the participant to choose it, but still clearly stating the reason we were doing so (you can't imitate the other robot) and that the same was true for the other robot.

5

Experimental Design

Contents

In this section the activity will be described. There are two conditions possible: Virtual and Physical. Conditions were randomly assigned to each participant but ensuring as even of a distribution as possible. The activity was carried out with children in the age group of 7 to 10 years and each participant went through each of the steps described in 5.2.

5.1 Research Questions and Hypotheses

The research questions are centered on the impact of the game and of the conditions (Physical vs. Virtual) on the children's perspective taking skills, performance and perception of the activity. The formulation of these questions draws motivation from Yadollahi's work, where it was found that the Virtual Maze game did help adults advance their SPT skills [1]. In this work, we are interested to study whether this is holds on a slightly different context and, especially, with a different target group: children. Furthermore, one of the main interests of this thesis is testing the effects of embodiment on both the activity as a whole and the development of SPT, and RQs regarding this aspect are in line with the aforementioned studies, such as the one by Pereira where it was deemed that the physical embodiment made the activity more fun [39]. These are formulated, alongside with the hypotheses as follows:

5.1.1 R.Q. I - Overall SPT improvement

Do children improve their SPT abilities after the play session?

5.1.1.A R.H. I.a

There will be a significant improvement in the post-test scores compared to the pre-test scores.

5.1.1.B R.H. I.b

Children in the Physical condition improve more in the post-tests than the children in the Virtual condition.

5.1.2 R.Q. II - Impact of embodiment condition in perception of the game

How does children's perception of the game change with the Physical/Virtual condition?

5.1.2.A R.H. II

Children that experienced the Physical condition find the game more fun than the ones on the Virtual condition.

5.1.3 R.Q. III - Impact of embodiment condition in game performance

How does children's performance in the game change with the Physical/Virtual condition?

5.1.3.A R.H. III

Children in the Physical condition conclude the game in less moves than the children in the Virtual condition.

5.2 Activity Flow

At the beginning of the activity, each participant would be told that they would be playing a game and answering a few questions. They would also be told that they should be completely at ease, ask any questions they had and that they could interrupt and leave the activity at any time if they chose to do so and that at no point would any decision they made impact them in terms of evaluation. Prior to starting, each participant created their own unique identifier code by answering a small set of questions, to comply with data regulations while preserving anonymity [41]. The full activity flow that will be described below can be seen in figure 5.1

Figure 5.1: A flowchart of the activity.

An overview of the physical activity can be seen in figure 5.2. The blue legos represented the path building blocks and the red legos represented the to be built blocks. To keep the activities similar, in both versions the participants were asked to control the buttons by pointing to the move they wanted to make in the sheet with the blue buttons, also seen in figure 5.2, and the experimenter would input their choice into the computer. To block a move, some red legos would be placed on top of one of the buttons, indicating the blocked move.

5.2.1 Pre-tests

This is to evaluate the SPT abilities of the child prior to playing the game. Specifically, these will be evaluated with the PT and Spatial Orientation Test, described in detail in appendix A.

5.2.2 Gameplay session

This is the core of the activity, where each participant played 2 games: one very basic level to get acquainted with the game and its mechanics (serving as a tutorial level) and one more advanced level once the participant understood the rules.

5.2.2.A Tutorial

As mentioned, the participant will play the tutorial level while being guided by the experimenter. The participant will be given the option to replay the level whatever many times he/she wants, to ensure rules are understood, only advancing to the next level once the participant is comfortable.

5.2.2.B Challenging level

After the tutorial, each participant plays one game in one of two slightly different layouts but with the same design principles and thus same level of difficulty. Each level contains on each side 12 nodes, 8 or 5 activation blocks, 4 or 6 path gaps depending on game mode (competitive and cooperative, respectively). The two different layouts can be seen in figures 4.4 and 4.3.

5.2.3 Post-tests

The post tests serve to evaluate the children's immediate SPT skills after the game. Similarly to the pre-tests, PT and Spatial Orientation Test was applied but with different contexts (new figures and order of the questions), as to not repeat the first set, but with the same level of difficulty. Additionally, participants were asked to answer three questions, quantifying their answer in a scale from 1 to 5: How difficult was it to understand the game? How difficult was it to play the game? How fun was the game?

5.2.4 Data Collection

Alongside the test results, data was be automatically collected by the game controllers. For each level, data such as the selected moves and the state of the game in those moves was collected. All data collection was approved by IST's Ethics Comitee and consented by each participant's sponsor of education. All data was kept anonymous and was only handled by the person responsible for the study.

5.3 Activity Results

5.3.1 Participants

A total of 59 children participated in the activity. The activity was only carried out after approval from the university's ethics committee and every participant had been given consent by the person responsible for them. At the beginning of every activity, it was clearly explained to the participant that he/she was free to ask any questions or quit at any time. 3 participants' results had to be invalidated due to technical problems or the children not wanting to answer the tests.

Out of the validated participants, 30 were assigned to the Physical condition and 26 to the Virtual condition. 25 were 4th graders and 31 were 3rd graders. As for age distribution, 6 (10.71%) children were 7 years old, 31 (55.36%) were 8 years old, 18 (32.14%) were 9 years old and a single child was 10 years old.

For all analysed data (pre-test scores, post-test scores, difference between pre-test and post-test scores, reported fun and number of moves), Shapiro-Wilk tests were ran to assess the normality of the data. In all instances, p-value ¡ 0.05 was observed, indicating the data does not follow a normal distribution and thus the statistical tests to apply were adjusted accordingly.

5.3.2 Research Hypotheses Results

5.3.2.A R.Q. I: Overall performance in pre and post tests

The general metrics for the results of the pre and post tests are available in table 5.1, where it can be seen an increase in the mean and median from the pre to the post test. The same applies when conditions are isolated, as it can be seen in table 5.2.

Applying a Wilcoxon signed-rank test to pre and post test scores, we get $W(55) = 91.0$, $p<0.001$, this together with the value increases on post scores shows a significant improvement from pre to post test globally. Therefore, R.H. I.a is accepted.

As for difference in improvement in performance on the post-tests, applying a Mann-Whitney U test to the post test scores separated by condition, we get $U(30, 26) = 306.0$, $p = 0.1630$. Furthermore, applying the same test to the difference in scores (post test score - pre test score) separated by conditions, it yields $U(30, 26) = 359.0$, $p = 0.6024$. In either case, while both conditions significantly increase their scores, there is no significant difference on the increase between conditions. Therefore, R.H. I.b is rejected.

Table 5.1: Metrics for the PT and SOT test

	PreTest	PostTests
mean	2.91	3.75
std	1.58	1.62
min	1.00	1.00
25%	2.00	2.00
50%	3.00	4.00
75%	4.00	5.00
max	6.00	6.00

Table 5.2: Metrics for the PT and SOT test, separated by embodiment condition

	PhysicalPre	PhysicalPost	VirtualPre	VirtualPost
mean	3.27	4.03	2.50	3.42
std	1.74	1.52	1.27	1.70
min	1.00	1.00	1.00	1.00
25%	2.00	3.00	1.25	2.00
50%	3.00	4.00	2.00	3.00
75%	4.75	5.00	3.00	5.00
max	6.00	6.00	5.00	6.00

5.3.2.B R.Q. II: Fun of the game

When asked this question, most children immediately answered very fun or 5 (or even numbers above that). While not measured directly, we observed that the children were much more excited and engaged with the novelty of the physical robots than those looking at the monitor of the virtual game. The distribution of answers, seen in 5.3 is as follows: in the physical condition, 28 participants (93.33%) answered 5 and 2 (6.67%) answered 4, whereas in the virtual condition, 20 participants (76.92%) answered 5, 2 (7.69%) answered 4 and 4 (15.38%) answered 3. Running a one-way ANOVA to this data, we get a $F(1, 54) = 4.7461$, $p = 0.0337$, showing a significant difference on the perception of fun between the two conditions. Therefore, R.H. II is accepted.

5.3.2.C R.Q. III: Performance between embodiment

The general metrics comparing the number of child moves needed to win is seen in table 5.3. Interestingly, the physical condition has a slightly higher average. Additionally, there is no significant difference when applying a Mann-Whitney U to the data, showing $U(30, 26) = 465.5$, $p = 0.2261$. Therefore, R.H. III is rejected.

Table 5.3: Metrics for the number of moves, separated by condition

	Virtual	Physical
mean	15.30	16.21
std	5.71	4.79
min	9.00	9.00
25%	11.00	13.00
50%	13.00	15.00
75%	18.00	19.00
max	30.00	26.00

5.3.3 Overall observations

5.3.3.A PT differences between grades

While it didn't impact the results, as the distribution was mostly even, during the activity, an unexpected significant difference between 3rd and 4th graders was noted, regarding their initial perspective taking skills. As seen in figure 5.4, the 3rd grade results are more skewed left (lower scores) than those of the 4th grade. This difference is significant for both pre and post tests, as confirmed in Mann-Whitney U tests for both pre tests $U(30, 26) = 549.0$, $p = 0.0068$ and post tests $U(30, 26) = 513.5$, $p = 0.0354$.

However, looking at figure 5.5, this significant starting skill difference did not seem to impact the amount of improvement. In fact, the Mann-Whitney U test for the scores' differences per grade does not show a significant different in improvement $U(30, 26) = 415.5$, $p = 0.6375$.

5.3.3.B PT and SOT test

In this test, a significant part of participants, specially 3rd graders and in the pre tests would just point in the direction the object asked was in the picture, despite the very first example given would be to point to the object that is North (direction the participant couldn't point in the circle as that was were the object the participant should be facing was.)

The improvement from pre to post tests was very clear, but then it also became apparent that despite most participants starting making adjustments, left-right distinction was a problem, as most of the post-test errors would be mistaking left with right and vice versa.

Some participants started rotating their heads and bodies to try to figure out the perspectives, something only a few were already doing in the pre-tests.

5.3.3.C Game questionnaire and children behaviour

While the first setup of the questionnaire contained some extra questions regarding what the children used to understand perspective and decide their moves, it was clear that it was hard for the children to explain their process, even giving by steering them in a direction with some predefined answers such as

”I rotated my head”. That being said, we opted to do notekeeping by observing the child and keeping track of their tactics. Mainly, three major observations were written down: hand gestures, head/body rotation and buttons rotation (rotating the input sheet to match the robot’s perspective). In total, 17 children used hand gestures, 26 used head and/or body rotation and 11 children rotated the buttons. It should be noted that this data is incomplete because of the way it was collected as it was not possible to collect it for all participants, therefore no major conclusions will be taken from it. One interesting fact is that all 5 children that scored perfectly in the pre-test used head/body rotation while playing the game.

Regarding the engagement of the children, while the Virtual condition ones were still happy to play the game, they lack in comparison to the Physical game’s reactions. The novelty of the robots was very interesting to them and they were much more engaged with the Vector robots, being very intrigued once they were taken out of the charger and started doing small interactions, with a big portion commenting on how ”cute” the robots were, or asking what their name was and wanting to play more at the end of the activity (though some also asked to play more in the virtual condition). That shows in the answers to how fun the game was, with only 2 out of 28 participants not giving it the maximum grade. Furthermore, children would often almost immediately answer 5 or very fun without second thoughts, contrasting with slower or hesitant answers to the other questions of the questionnaire (related to the difficulty of the game) or to the same question but in the Virtual condition.

5.3.3.D Child moves as a performance measure

One very common scenario was that the participants would select the button in the absolute direction instead of regarding the robot’s perspective but they wouldn’t necessarily lose a life (and thus not count as a mistake), since they would just follow another path than the one they wish to go. However, this would increases (in most situations) the amount of total moves the participant will need to do to win the game. As such, that is why it was decided to measure performance in terms of number of moves to finish the game instead of number of errors.

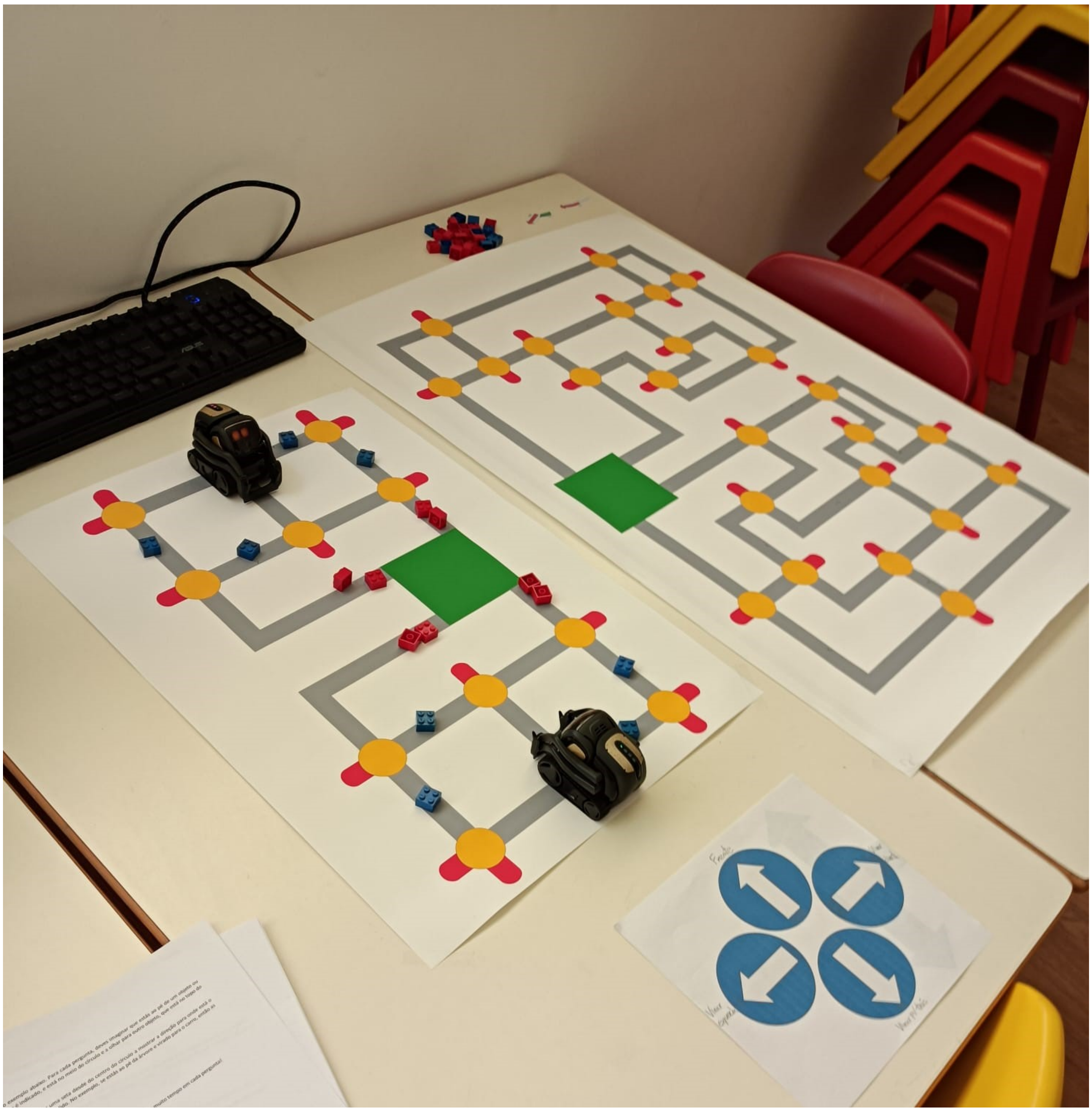

Figure 5.2: Physical activity setup in one school. Tutorial (left) is ready to be played with Vectors in initial position, and the advanced level map is seen left.

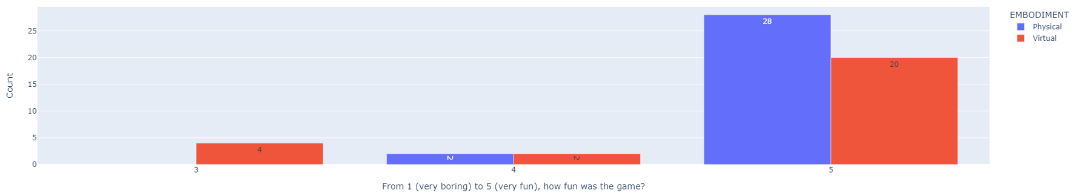

Figure 5.3: Distribution of answers of how fun the game was, per embodiment condition (Blue - Physical, Red - Virtual).

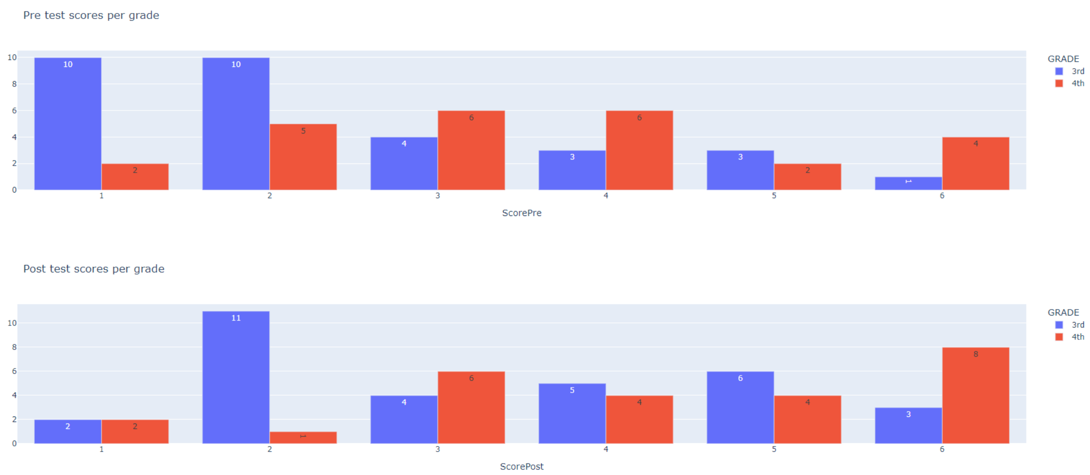

Figure 5.4: Distribution of pre and post tests per grade (Blue - 3rd Grade, Red - 4th Grade).

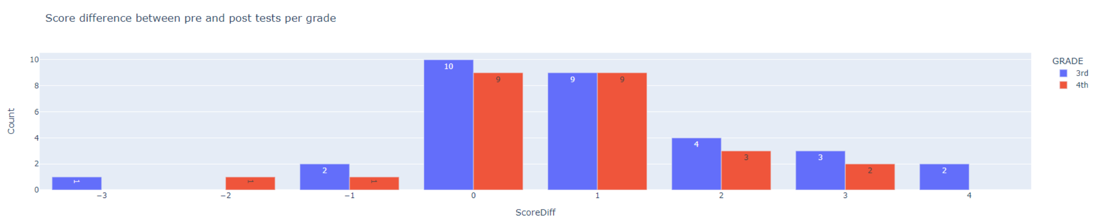

Figure 5.5: Distribution of difference of scores in pre and post tests per grade (Blue - 3rd Grade, Red - 4th Grade).

6

Conclusion

Contents

6.1 Limitations and Future Work

As mentioned, restrictions limited the scope of the study. While the initial activity was designed to also test the differences between Cooperative vs. Competitive, the latter had to be dropped. The other big limitation of the study was the unforeseen difference of skill between grades, gone undetected in the pilots since they were done with children at the end of the 4th grade, leading to the big adaptation of explicitly blocking the moves. Finally, to ensure statistical relevancy, the activity had to be shortened to get enough participants. Furthermore, the lack of a control group leads to some doubts about the actual impact of the game on the improvement of the post tests, as repetition could have an impact, despite the different context. It is likely that a study where these limitations are not present might help children even more and yield better, more ensuring results.

That being said, there are a few directions for future work, namely:

1. Designing ways of making the children have to take the other robot's perspective, without making the game too difficult, preferably piloting on 3rd graders. An example, as a untested suggestion, would be keeping the movement restriction, but not punishing the child by losing lives. 2. Carrying out the activity with a bigger number of participants, and possibly reintroducing the cooperative vs competitive condition. 3. Carrying out the activity with more time per participant (and thus more game play). 4. Introducing a control group, assuming more participants would be available. In this activity, children could for instance play a game that wouldn't require them to use SPT and as such not encouraging them to improve it. 5. Doing the pre-tests a few days in advance, as to reduce the impact of repetition, by spacing the tests apart in time. This would again assume availability to do so, which was not available in this activity.

It should be noted that, despite all the limitations and constraints, the children still improved their skills and any future work where these limitations would be reduced has a good possibility of helping the children develop their skills even more.

6.2 Conclusions

To summarize the activity, not only was the children's feedback overwhelmingly positive (both in the questionnaire and in the conversations with them), but also, as discussed in the previous chapter, children's SPT skills improved significantly (R.H.1a). As such, overall the activity can be deemed a success, with the main goal being achieved: developing a fun and interactive activity that allowed the children to develop their SPT skills.

As for the difference between conditions, there was no significant difference in the improvement of SPT skills (R.H.1b) or in difficulty (R.H.3). However, on top of the Physical condition being significantly more fun than the Virtual conditon (R.H.2), it was quite visible that children preferred the physical version, since they were much more engaged, intrigued and wanted to interact more with the Vector robots, as well as showing more feelings towards them.

As a final note, as an extension of Yadollahi's work, this study, despite its limitations, provided valuable insights in evaluating this type of activity not only on children but also in a physical setting, with good indicators of effectiveness and hinting that gamified activities like these can be valuable in children's education, contributing to its diversification through the use of robot-assisted learning.

Bibliography

[1] E. Yadollahi, "Exploring spatial perspective taking in human-robot interaction," p. 229, 2021, co-supervision with Instituto Superior Técnico (IST) da Universidade de Lisboa, Doutoramento em Engenharia Informática e de Computadores. [Online]. Available: http://infoscience.epfl.ch/record/289268

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

[3] L. A. Kurdek and M. M. Rodgon, "Perceptual, cognitive, and affective perspective taking in kindergarten through sixth-grade children." *Developmental Psychology*, vol. 11, no. 5, p. 643–650, 1975.

[4] J. H. Flavell, "Development of children's knowledge about the mental world," *International Journal of Behavioral Development*, vol. 24, no. 1, pp. 15–23, 2000. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/016502500383421

[5] K. Kessler and H. Rutherford, "The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions," *Frontiers in Psychology*, vol. 1, 2010. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00213

[6] H. Moll and A. Meltzoff, *Perspective-Taking and its Foundation in Joint Attention*, 08 2011, pp. 286–304.

[7] S. C. Levinson, *Space in Language and Cognition: Explorations in Cognitive Diversity*, ser. Language Culture and Cognition. Cambridge University Press, 2003.

[8] M. F. Schober, "Spatial perspective-taking in conversation," *Cognition*, vol. 47, no. 1, pp. 1–24, 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010027793900609

[9] D. Schachter and E. S. Gollin, "Spatial perspective taking in young children," *Journal of Experimental Child Psychology*, vol. 27, no. 3, pp. 467–478, 1979. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0022096579900365

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

[11] J. Zaki, "Moving beyond stereotypes of empathy," *Trends in Cognitive Sciences*, vol. 21, 12 2016.

[12] J. B. Hinnant and M. O'Brien, "Cognitive and emotional control and perspective taking and their relations to empathy in 5-year-old children," *The Journal of Genetic Psychology*, vol. 168, no. 3, pp. 301–322, 2007, pMID: 18200891. [Online]. Available: https://www.tandfonline.com/doi/abs/10.3200/GNTP.168.3.301-322

[13] N. Eisenberg, P. A. Miller, M. Schaller, R. A. Fabes, J. Fultz, R. Shell, and C. L. Shea, "The role of sympathy and altruistic personality traits in helping: A reexamination," *Journal of Personality*, vol. 57, no. 1, p. 41–67, 1989.

[14] F. Volkmar, "Teaching children with autism to mind read: A practical guide. by p. howlin, s. baron-cohen, amp; j. hadwin. wiley, chichester, 1999. pp. 290. £16.99 (pb)." *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 41, no. 8, p. 1077–1084, 2000.

[15] Y. Barnes-Holmes, L. McHugh, and D. Barnes-Holmes, "Perspective-taking and theory of mind: A relational frame account." *The Behavior Analyst Today*, vol. 5, no. 1, p. 15–25, 2004.

[16] J. H. Flavell, E. R. Flavell, and F. L. Green, "Development of the appearance-reality distinction," *Cognitive Psychology*, vol. 15, no. 1, pp. 95–120, 1983. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0010028583900051

[17] J. Perner, *Understanding the representational mind*. MIT Press, 1991.

[18] M. Scopelliti, M. V. Giuliani, and F. Fornara, "Robots in a domestic setting: A psychological approach," *Universal Access in the Information Society*, vol. 4, pp. 146–155, 12 2005.

[19] R. Ros, M. Nalin, R. Wood, P. Baxter, R. Looije, Y. Demiris, T. Belpaeme, A. Giusti, and C. Pozzi, "Child-robot interaction in the wild: Advice to the aspiring experimenter," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, ser. ICMI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 335–342. [Online]. Available: https://doi.org/10.1145/2070481.2070545

[20] T. Salter, I. Werry, and F. Michaud, "Going into the wild in child-robot interaction studies – issues in social robotic development," *Intelligent Service Robotics*, vol. 1, pp. 93–108, 01 2008.

[21] L. Q. Currie and E. Wiese, "Mind perception in a competitive human-robot interaction game," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 1957–1961, 2019. [Online]. Available: https://doi.org/10.1177/1071181319631284

[22] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," *Science*, vol. 315, no. 5812, pp. 619–619, 2007. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1134475

[23] A. Pereira, I. Leite, S. Mascarenhas, C. Martinho, and A. Paiva, "Using empathy to improve human-robot relationships," in *Human-Robot Personal Relationships*, M. H. Lamers and F. J. Verbeek, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 130–138.

[24] L. Charrier, A. Rieger, A. Galdeano, A. Cordier, M. Lefort, and S. Hassas, "The rope scale: a measure of how empathic a robot is perceived," 03 2019.

[25] S. Seo, D. Geiskkovitch, M. Nakane, C. King, and J. Young, "Poor thing! would you feel sorry for a simulated robot?: A comparison of empathy toward a physical and a simulated robot," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2015, pp. 125–132, 03 2015.

[26] T. Belpaeme, J. Kennedy, P. Baxter, P. Vogt, E. E. J. Krahmer, S. Kopp, K. Bergmann, P. Leseman, A. C. Küntay, T. Göksun, A. K. Pandey, R. Gelin, P. Koudelkova, and T. Deblieck, "L2TOR - Second Language Tutoring using Social Robots," in *1st Int. Workshop on Educational Robotics at the Int. Conf. Social Robotics*, no. January, Paris, France, 2015. [Online]. Available: https://icsrwonder2015.wordpress.com/

[27] E. Hyun, H. Yoon, and S. Son, "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*, ser. HRI '10. IEEE Press, 2010, p. 199–200.

[28] J. A. Bargh and Y. Schul, "On the cognitive benefits of teaching." *Journal of Educational Psychology*, vol. 72, no. 5, p. 593–604, 1980.

[29] C. C. Chase, D. B. Chin, M. A. Oppezzo, and D. L. Schwartz, "Teachable agents and the protégé effect: Increasing the effort towards learning," *Journal of Science Education and Technology*, vol. 18, no. 4, p. 334–352, 2009.

[30] P. Li, X. Jin, Y. Liao, Y. Li, M. Shen, and J. He, "Cooperation turns preschoolers into flexible perspective takers," *Cognitive Development*, vol. 52, p. 100823, 10 2019.

[31] J. Kory Westlund, G. Gordon, S. Spaulding, J. J. Lee, L. Plummer, M. Martinez, M. Das, and C. Breazeal, "Lessons from teachers on performing hri studies with young children in schools," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, ser. HRI '16. IEEE Press, 2016, p. 383–390.

[32] J. Kennedy, P. Baxter, E. Senft, and T. Belpaeme, "Heart vs hard drive: Children learn more from a human tutor than a social robot," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 451–452.

[33] M. Blancas, V. Vouloutsi, K. Grechuta, and P. F. M. J. Verschure, "Effects of the robot's role on human-robot interaction in an educational scenario," in *Biomimetic and Biohybrid Systems*, S. P. Wilson, P. F. Verschure, A. Mura, and T. J. Prescott, Eds. Cham: Springer International Publishing, 2015, pp. 391–402.

[34] D. Leyzberg, B. Scassellati, S. Spaulding, and M. Toneva, "The physical presence of a robot tutor increases cognitive learning gains," 01 2012.

[35] M. Saerbeck, T. Schut, C. Bartneck, and M. Janse, "Expressive robots in education - varying the degree of social supportive behavior of a robotic tutor," in *28th ACM Conference on Human Factors in Computing Systems (CHI2010)*. Atlanta: ACM, 2010, pp. 1613–1622.

[36] C. Bartneck, "Interacting with an embodied emotional character," *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces - DPPI '03*, p. 55–60, 2003.

[37] J. Wainer, D. Feil-Seifer, D. Shell, and M. Mataric, "The role of physical embodiment in human-robot interaction," 10 2006, pp. 117 – 122.

[38] J. Kennedy, P. Baxter, and T. Belpaeme, "Comparing robot embodiments in a guided discovery learning interaction with children," *International Journal of Social Robotics*, vol. 7, 12 2014.

[39] A. Pereira, C. Martinho, I. Leite, and A. Paiva, "Icat, the chess player: The influence of embodiment in the enjoyment of a game," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3*, ser. AAMAS '08. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2008, p. 1253–1256.

[40] E. Yadollahi, M. Couto, P. Dillenbourg, and A. Paiva, "Can you guide me? supporting children's spatial perspective taking through games with robots," in *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*, ser. IDC '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 217–222. [Online]. Available: https://doi.org/10.1145/3397617.3397831

[41] J. Vacek, H. Vonková, and R. Gabrhelík, "A successful strategy for linking anonymous data from students' and parents' questionnaires using self-generated identification codes," *Prevention Science*, vol. 18, pp. 450–458, 2017.

A

Activity tests

A.1 Path test

The figure A.1 shows examples of this test. The instructions for the 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).”

As described by Yadollahi, there are two approaches on how to describe the problem. One would be describing the path to the animal. For the provided dog example, this sequence would be ”Forward, Right, Forward, Left, Forward”. The second approach is as if one was walking the path with the dog. In this case the correct sequece would be ”Front, Right, Front, Left, Front, Left, Front, Right, Front” or simply ”Right, Left, Left, Right”, omitting the moving forward part.

Each participant will be given different pre and post tests, with different directions, for instance, the dog as pre test and the cat as post test.

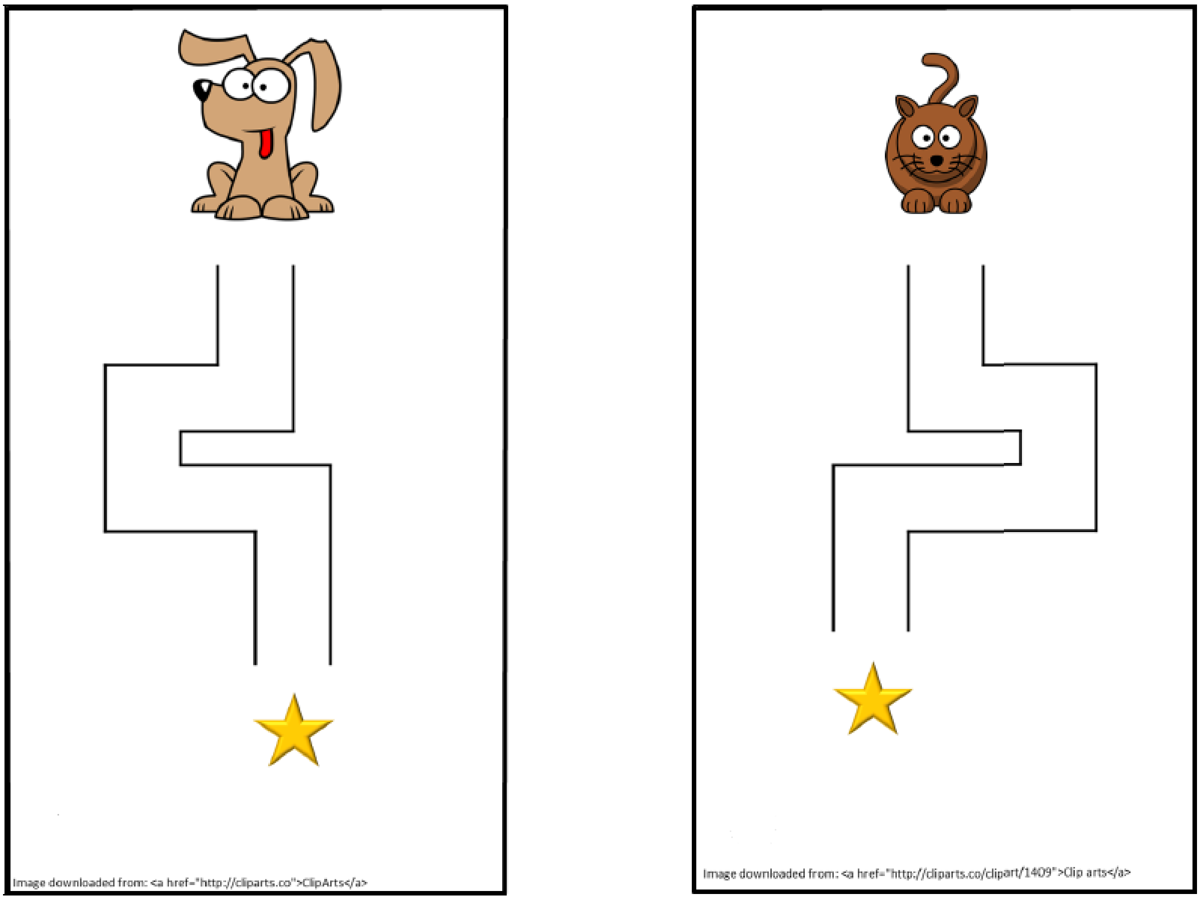

Figure A.1: Example of dog (left) and cat (right) path test.

A.2 PT and Spatial Orientation Test

Figure A.2 provides an example of the Spatial Orientation Test that will be applied. The children will be faced with an image containing objects. Then it's told to imagine being at one location looking at another object. Then it must indicate in the circle provided where it would point to another requested object, as described by the dotted line).

Compared to the adults' test, this contains less objects in a single test and less angular disparities.

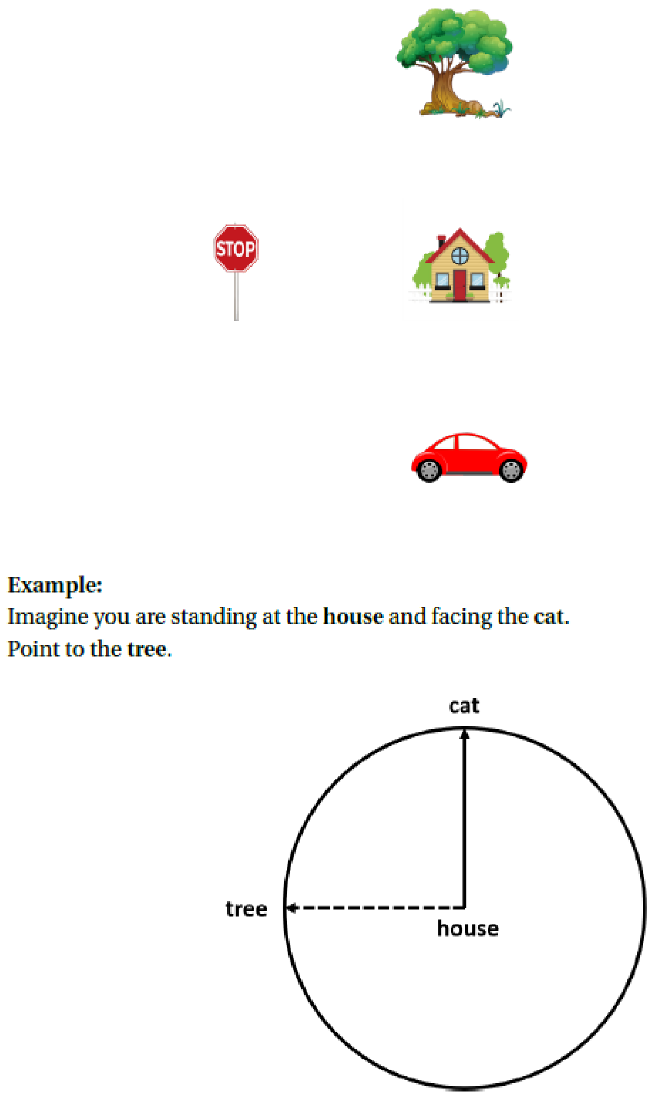

Figure A.2: Example of the PT and Spatial Orientation Test.