Gamified activity for learning perspective taking

MIGUEL MONTEIRO, Instituto Superior Técnico, Portugal
ANA PAIVA, Instituto Superior Técnico, Portugal
ELMIRA YADOLLAHI, KTH Royal Institute of Technology, Sweden

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.

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.

Authors' addresses: Miguel Monteiro, miguel.alexandre.monteiro@tecnico.ulisboa.pt, Instituto Superior Técnico, Lisbon, Portugal; Ana Paiva, Instituto Superior Técnico, Lisbon, Portugal; Elmira Yadollahi, KTH Royal Institute of Technology, Stockholm, Sweden.

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 [29] and evolving it to this new context.

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.

3 THEORETICAL OVERVIEW

3.1 Perspective Taking

Perspective Taking (PT) 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. [26], 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 [15].

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

Visual Perspective Taking (VPT) 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. [8] 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.

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 Frame 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 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 [16], Surtees et al. [26] and in other literature are:

Lastly regarding FoRs, 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.

Similarly to Visual PT, one can also consider different levels of SPT, that develop at different times. Surtees et al. proposed a 2-level

model for SPT [26]. 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.

3.1.2 Cognitive PT. 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.

3.1.3 Affective PT. 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" [30]. Empathy and Affective PT are also related to the Cognitive dimension, as studies by Hinnant and O'Brien have found that Cognitive PT moderates the relation between both, and gender is impactful in this relation [10].

3.2 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.

4 RELATED WORK

4.1 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 [24]. 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 [21].

In line with this, children are less likely to use robots as tools, increasing the significance of the social part of the interaction [23]. 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' [21].

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 [7]. 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 [9]. A bigger perceived empathy can lead to a human more easily accepting and building with a relationship with the robot [19]. 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 [5]. 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].

4.1.1 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 [3] and to develop guidelines for it [11].

Teaching others has been shown to be a powerful learning method by research in education and cognitive sciences [1]. More recently, the same effect was suggested in computer-assisted learning environments [6]. 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 [18].

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 [14], 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 [13].

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 [4]. 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 [17].

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 [22].

4.1.2 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 [2]. The aforementioned study by Saerbeck et al. also mentions that the sociability of a learning agent can be enhanced by physical embodiment [22]. In the context of learning, a physical robot can also positively impact the learning gains, as suggested by Leyzberg et al. [17].

On the other hand, in a Towers of Hanoi related task, where the robot would tell the participant a stacking goal, Wainer and colleagues suggest physical embodiment can make a difference in perception of social capabilities and enjoyment [28].

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 [12].

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 [20].

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.

4.2 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 [29].

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 [29].

Finally, in the Virtual Maze game, a two-agent (one AI controlled) maze-guiding game, the goal was for players to navigate the maze to get to a safe zone. To do so, they had to apply perspective taking techniques to understand the agents' perspective. It was used in a couple of studies, with conditions regarding game mode: Cooperative, where one agent would help another get to the safe zone or Competitive, where the agents raced to the safe zone and SPT capabilities of the AI: no mistakes or some intentional mistakes. 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 version where the robot did not do PT mistakes [29]. 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.

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.

5 TECNICHAL DEVELOPMENT

5.1 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.

5.1.1 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.

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.

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.

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.
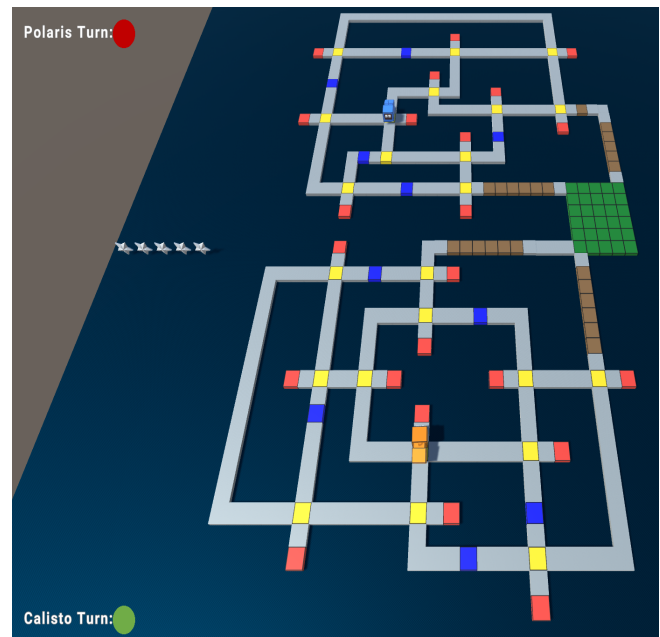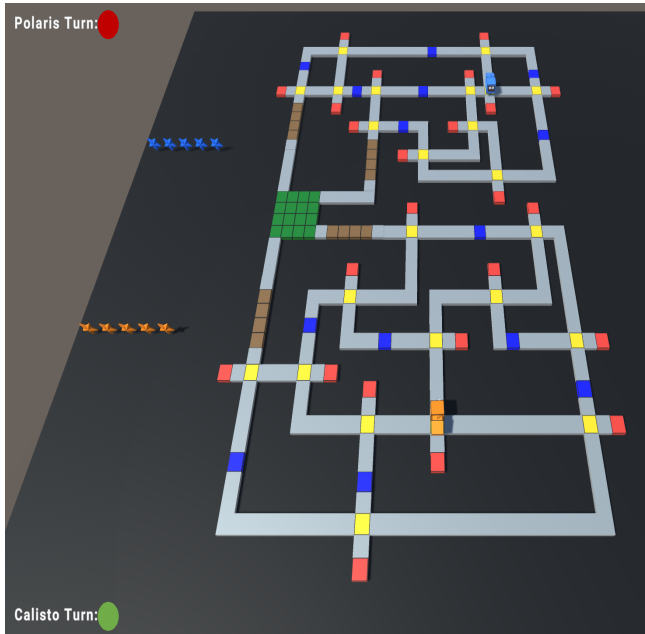

Fig. 1. New cooperative game mode.

Fig. 2. New competitive game mode.

5.1.2 Game implementation.
The different steps for the implementation of the game will now be described.

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, 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.

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 1.

Table 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 or 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%

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.

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 moves, and the experimenter would place them in their respective starting blocks and orientations, as seen in figure 4. 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.

5.1.3 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.

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.

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.

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.

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.

6 EXPERIMENTAL DESIGN

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 6.2.

6.1 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 [29]. 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 [20]. These are formulated, alongside with the hypotheses as follows:

6.1.1 R.Q. I - Overall SPT improvement. Do children improve their SPT abilities after the play session?

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

R.H. I.b Children in the Physical condition improve more in the post-tests than the children in the Virtual condition.

6.1.2 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?

R.H. II

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

6.1.3 R.Q. III - Impact of embodiment condition in game performance.
How does children's performance in the game change with the Physical/Virtual condition?

R.H. III

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

6.2 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 [27]. The full activity flow that will be described below can be seen in figure 3

Fig. 3. A flowchart of the activity.

An overview of the physical activity can be seen in figure 4. 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 4, 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.

6.2.1 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 ??.

6.2.2 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.

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.

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 2 and 1.
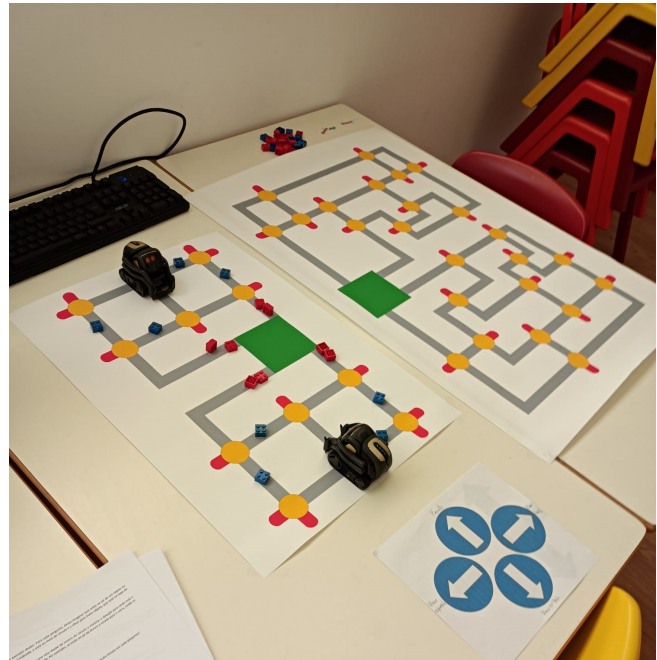

Fig. 4. Physical activity setup in one school. Tutorial (left) is ready to be played with Vectors in initial position, and the advanced level map is seen left.

6.2.3 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?

6.2.4 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.

6.3 Activity Results

6.3.1 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.

6.3.2 Research Hypotheses Results. The proposed hypotheses will now be discussed, based on the collected data.

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 2, 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 3.

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 2. 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 3. 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

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 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.
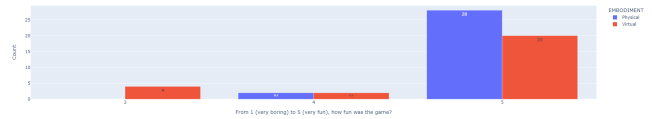

Fig. 5. Distribution of answers of how fun the game was, per embodiment condition (Blue - Physical, Red - Virtual).

R.Q. III: Performance between embodiment

The general metrics comparing the number of child moves needed to win is seen in table 4. 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 4. 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

6.3.3 Overall observations. 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 6, 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 7, 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.
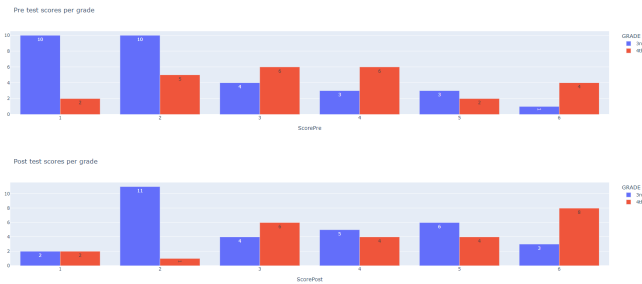
Fig. 6. Distribution of pre and post tests per grade (Blue - 3rd Grade, Red - 4th Grade).
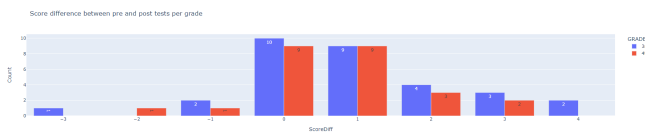

Fig. 7. Distribution of difference of scores in pre and post tests per grade (Blue - 3rd Grade, Red - 4th Grade).

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.

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.

7 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.

8 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.

REFERENCES

[1] John A. Bargh and Yaacov Schul. 1980. On the cognitive benefits of teaching. *Journal of Educational Psychology* 72, 5 (1980), 593–604. https://doi.org/10.1037/0022-0663.72.5.593

[2] Christoph Bartneck. 2003. Interacting with an embodied emotional character. *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces - DPPI '03* (2003), 55–60. https://doi.org/10.1145/782896.782911

[3] Tony Belpaeme, James Kennedy, Paul Baxter, Paul Vogt, Emiel E J Krahmer, Stefan Kopp, Kirsten Bergmann, Paul Leseman, Aylin C Küntay, Tilbe Göksun, Amit K Pandey, Rodolphe Gelin, Petra Koudelkova, and Tommy Deblieck. 2015. L2TOR - Second Language Tutoring using Social Robots. In *1st Int. Workshop on Educational Robotics at the Int. Conf. Social Robotics*. Paris, France. https://icsrwonder2015.wordpress.com/

[4] Maria Blancas, Vasiliki Vouloutsi, Klaudia Grechuta, and Paul F. M. J. Verschure. 2015. Effects of the Robot's Role on Human-Robot Interaction in an Educational Scenario. In *Biomimetic and Biohybrid Systems*, Stuart P. Wilson, Paul F.M.J. Verschure, Anna Mura, and Tony J. Prescott (Eds.). Springer International Publishing, Cham, 391–402.

[5] Laurianne Charrier, Alisa Rieger, Alexandre Galdeano, Amélie Cordier, Mathieu Lefort, and Salima Hassas. 2019. The RoPE Scale: a Measure of How Empathic a Robot is Perceived. https://doi.org/10.1109/HRI.2019.8673082

[6] Catherine C. Chase, Doris B. Chin, Marily A. Oppezzo, and Daniel L. Schwartz. 2009. Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology* 18, 4 (2009), 334–352. https://doi.org/10.1007/s10956-009-9180-4

[7] Levern Q. Currie and Eva Wiese. 2019. Mind Perception in a Competitive Human-Robot Interaction Game. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (2019), 1957–1961. https://doi.org/10.1177/1071181319631284 arXiv:https://doi.org/10.1177/1071181319631284

[8] John H. Flavell. 2000. Development of children's knowledge about the mental world. *International Journal of Behavioral Development* 24, 1 (2000), 15–23. https://doi.org/10.1080/016502500383421 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/016502500383421

[9] Heather M. Gray, Kurt Gray, and Daniel M. Wegner. 2007. Dimensions of Mind Perception. *Science* 315, 5812 (2007), 619–619. https://doi.org/10.1126/science.1134475 arXiv:https://www.science.org/doi/pdf/10.1126/science.1134475

[10] J. Benjamin Hinnant and Marion O'Brien. 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 (2007), 301–322. https://doi.org/10.3200/GNTP.168.3.301-322 arXiv:https://www.tandfonline.com/doi/pdf/10.3200/GNTP.168.3.301-322 PMID: 18200891.

[11] Eunja Hyun, Hyunmin Yoon, and Sooryun Son. 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* (Osaka, Japan) *(HRI '10)*. IEEE Press, 199–200.

[12] James Kennedy, Paul Baxter, and Tony Belpaeme. 2014. Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children. *International Journal of Social Robotics* 7 (12 2014). https://doi.org/10.1007/s12369-014-0277-4

[13] James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. 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)*. 451–452. https://doi.org/10.1109/HRI.2016.7451801

[14] Jacqueline Kory Westlund, Goren Gordon, Samuel Spaulding, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Lessons From Teachers on Performing HRI Studies with Young Children in Schools. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*

[15] Lawrence A. Kurdek and Maris M. Rodgon. 1975. Perceptual, cognitive, and affective perspective taking in kindergarten through sixth-grade children. *Developmental Psychology* 11, 5 (1975), 643–650. https://doi.org/10.1037/0012-1649.11.5.643

[16] Stephen C. Levinson. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511613609

[17] Dan Leyzberg, Brian Scassellati, Samuel Spaulding, and Mariya Toneva. 2012. The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains.

[18] Pengchao Li, Xinyi Jin, Yifan Liao, Yuanyuan Li, Mowei Shen, and Jie He. 2019. Cooperation turns preschoolers into flexible perspective takers. *Cognitive Development* 52 (10 2019), 100823. https://doi.org/10.1016/j.cogdev.2019.100823

[19] André Pereira, Iolanda Leite, Samuel Mascarenhas, Carlos Martinho, and Ana Paiva. 2011. Using Empathy to Improve Human-Robot Relationships. In *Human-Robot Personal Relationships*, Maarten H. Lamers and Fons J. Verbeek (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 130–138.

[20] André Pereira, Carlos Martinho, Iolanda Leite, and Ana Paiva. 2008. 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* (Estoril, Portugal) *(AAMAS '08)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1253–1256.

[21] Raquel Ros, Marco Nalin, Rachel Wood, Paul Baxter, Rosemarijn Looije, Yannis Demiris, Tony Belpaeme, Alessio Giusti, and Clara Pozzi. 2011. Child-Robot Interaction in the Wild: Advice to the Aspiring Experimenter. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante, Spain) *(ICMI '11)*. Association for Computing Machinery, New York, NY, USA, 335–342. https://doi.org/10.1145/2070481.2070545

[22] Martin Saerbeck, Tom Schut, Christoph Bartneck, and Maddy Janse. 2010. 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)*. ACM, Atlanta, 1613–1622. https://doi.org/10.1145/1753326.1753567

[23] Tamie Salter, Iain Werry, and François Michaud. 2008. Going into the Wild in Child-Robot Interaction Studies – Issues in Social Robotic Development. *Intelligent Service Robotics* 1 (01 2008), 93–108. https://doi.org/10.1007/s11370-007-0009-9

[24] Massimiliano Scopelliti, Maria Vittoria Giuliani, and Ferdinando Fornara. 2005. Robots in a domestic setting: A psychological approach. *Universal Access in the Information Society* 4 (12 2005), 146–155. https://doi.org/10.1007/s10209-005-0118-1

[25] Stela Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James Young. 2015. 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* 2015 (03 2015), 125–132. https://doi.org/10.1145/2696454.2696471

[26] Andrew Surtees, Ian Apperly, and Dana Samson. 2013. Similarities and differences in visual and spatial perspective-taking processes. *Cognition* 129, 2 (2013), 426–438. https://doi.org/10.1016/j.cognition.2013.06.008

[27] Jaroslav Vacek, Hana Vonková, and Roman Gabrhelík. 2017. A Successful Strategy for Linking Anonymous Data from Students' and Parents' Questionnaires Using Self-Generated Identification Codes. *Prevention Science* 18 (2017), 450–458.

[28] Joshua Wainer, David Feil-Seifer, Dylan Shell, and Maja Mataric. 2006. The role of physical embodiment in human-robot interaction. *IEEE Proceedings of the International Workshop on Robot and Human Interactive Communication*, 117 – 122. https://doi.org/10.1109/ROMAN.2006.314404

[29] Elmira Yadollahi. 2021. Exploring Spatial Perspective Taking in Human-Robot Interaction. (2021), 229. https://doi.org/10.5075/epfl-thesis-8352 Co-supervision with Instituto Superior Técnico (IST) da Universidade de Lisboa, Doutoramento em Engenharia Informática e de Computadores.

[30] Jamil Zaki. 2016. Moving beyond Stereotypes of Empathy. *Trends in Cognitive Sciences* 21 (12 2016). https://doi.org/10.1016/j.tics.2016.12.004