Do Children Adapt Their Perspective to a Robot When They Fail to Complete a Task?

Elmira Yadollahi
KTH Royal Institute of Technology
Stockholm, Sweden
elmiray@kth.se

Marta Couto
INESC-ID Lisboa
Lisboa, Portugal
marta.couto@gaips.inesc-id.pt

Pierre Dillenbourg
École polytechnique fédérale de Lausanne (EPFL)
Lausanne, Switzerland
pierre.dillenbourg@epfl.ch

Ana Paiva
Instituto Superior Técnico, University of Lisbon
Lisboa, Portugal
ana.paiva@inesc-id.pt

ABSTRACT

Spatial understanding and communication are essential skills in human interaction. An adequate understanding of others' spatial perspectives can increase the quality of the interaction, both perceptually and cognitively. In this paper, we take the first step towards understanding children's perspective-taking abilities and their tendency to adapt their perspective to a counterpart while completing a task with a robot. The elements used for studying children's behaviours are the frame of reference and perspective marking, which we evaluated through a task where players needed to compose instructions to guide each other to complete the task. We developed the interaction with an NAO robot and analyzed the children's instructions and their performance throughout the game. Our initial findings demonstrated that children tend to compose their first instruction by following the principle of least collaborative effort. Children significantly changed and adapted their perspective, i.e. frame of reference and perspective marking to the robot, mainly when the robot failed to follow their instructions correctly. Additionally, results show that children tend to create a mental model of their counterparts and the robot changing that frame of reference might affect their performance or the flow of the interaction.

CCS CONCEPTS

• **Social and professional topics** → *Children*; • **Computing methodologies** → **Cognitive science**; • **Human-centered computing** → **User studies**; • **Applied computing** → *Collaborative learning*.

KEYWORDS

perspective-taking, child-robot interaction, spatial cognition, spatial descriptions, gamification

ACM Reference Format:
Elmira Yadollahi, Marta Couto, Pierre Dillenbourg, and Ana Paiva. 2022. Do Children Adapt Their Perspective to a Robot When They Fail to Complete a Task?. In *Interaction Design and Children (IDC '22), June 27–30, 2022, Braga, Portugal.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3501712.3529719

IDC '22, June 27–30, 2022, Braga, Portugal

ACM ISBN 978-1-4503-9197-9/22/06.
https://doi.org/10.1145/3501712.3529719

1 INTRODUCTION

From the moment children learn that their parents' face does not disappear behind their hand while playing peak-a-boo to the day they perceive why their parents were so strict about their education, they experience different forms of perspective-taking. The first instance refers to learning object permanence, which children develop between the ages of 6 to 24 months. Understanding their parents decisions regarding their education and well-being happens years later when they have developed the ability to take others' perspectives cognitively and affectively, with true realization only occurring to them when they become parents themselves. The field of developmental psychology has a long history of investigating perspective-taking abilities and their underlying mechanisms in children and adults. In general terms, perspective-taking (PT) is the capacity to perceive and understand the world from other viewpoints, with the perception ranging from acknowledging that others see things differently to computing others' different perspectives. Humans use this ability daily and arguably without realizing that they repeatedly employ their perspective-taking skills. Nevertheless, the repeated usage is no indication that humans always correctly perceive or estimate other's perspectives.

The different dimensions and levels of perspective-taking abilities develop in humans and some mammals (monkeys) at different ages [40]. The three main dimensions of perspective-taking are perceptual, cognitive, and affective, each one corresponding to the type of perspective that people need to take. In recent years, the literature on the topic has covered research on all dimensions and sub-dimensions, from understanding the developmental stages from infancy to adulthood [12, 27, 29] to dissociating the underlying processes of different levels of perspective-taking [39, 42]. A brief overview of the components and processes involved in perspective-taking helps in understanding the importance of these skills in human communication and daily survival. This, in turn, can lead to a better understanding of why incorporating such skills in agents and robots should be the next step in developing them. To date, several research fields have made discussions on the importance of developing the theory of mind abilities in robots and AI [4]. Theory of mind (ToM) is the ability to attribute mental states, beliefs, and s to self and others [2, 14, 33, 43]. Having a theory of mind corresponds to acknowledging that others have mental states that may differ from the self. To master taking others' perspectives, children need to master five levels of understanding informational states [2, 16]. The first three levels correspond to different levels of perceptual and cognitive perspective-taking. To develop such capabilities in robotic and artificial agents for interaction with children, as a first step, we decided to examine how children exhibit and adapt their perspective during the interaction with a robot.

In this work, we explore children's spatial perspective-taking (SPT) abilities and their tendency to adapt their perspective to a counterpart in the context of interacting with a robot. To create a baseline for evaluating children's performance, we designed a straightforward task composed of moving 2D geometrical objects around a touch screen to replicate a state presented in a goal card. The experimental study has been designed with two objectives; (1) understanding children's perspective choice when they initiate an interaction with the robot and (2) their perspective adaptation during the interaction. To evaluate children's perspective-taking abilities and adaptive behaviour, the task is designed to be successfully completed only when each player understands and/or takes the perspective of the other player. Meanwhile, in the initial sessions, the robot acts more egocentric, meaning that it does not adapt to the child's perspective. This behaviour is meant to provide us with a testing ground to observe and evaluate children's adaptation. Our first research question addresses children's first perspective choice before observing the robot's behaviour.

RQ1: What is children's first perspective choice, when collaborating with a robot? The second research question aims at understanding how children perceive the robot and adapt their perspective for the benefit of the task and the robot.

RQ2: Do children adapt their perspective to the robot when they fail to complete the task?

The final research question looks at how such interactions and the activity itself can benefit children in developing their perspective-taking skills or in a broader sense their spatial thinking skills.

RQ3: Does interacting with a robot improve children's skills in adopting others perspectives?

The whole research aims to contribute to the field on two grounds, (1) providing an overview of how targeted tasks can help children develop cognitive abilities such as perspective-taking and (2) understanding children's perception of social robots through their tendency to adapt their perspective to robots. In the following section, we provide the relevant state of the art research and our motivation to pursue this research direction. We describe the development of the task and the interaction design in section 3. The details related to the user study such as experimental design and procedures are provided in section 4, followed by the analysis and results in section 5. The last two sections discuss the findings of the study and its implications in developing robots and activities for children that encompasses perspective-taking abilities in robots and fosters their development it in children.

2 RELATED WORK

The term perspective-taking appears in a wide variety of fields, from developmental and cognitive psychology to social sciences and linguistics, from simple daily human interactions to preventing a nuclear war between two countries in a harrowing conflict [15]. The word "perspective" can be extensive and cover various dimensions and domains. Kurdek and Rodgon proposed three dimensions to the type of perspectives that can be taken, perceptual, cognitive, and affective [24]. The perceptual domain covers what people can either understand or compute from other's visual or spatial perspectives. Visual perspective taking (VPT) consists of the self's awareness of the other's visual field of view and spatial perspective-taking (SPT) deals with the self's spatial understanding of the other's perspective or spatial relation with the objects in the environment. Looking at SPT from a developmental point of view, this ability has not shown a uniform developmental pattern. Some evidence indicates an earlier developmental timeline for notions of front and back [3, 18]. By 3-4 years of age children consistently use the words "in front" and "behind", however, it is much later when they exhibit the same consistent use for "to the right of" and "to the left of" [17, 18]. Looking at Flavell and colleagues and Moll and Meltzoff's developmental stages of Visual perspective-taking (VPT), Surtees et al. had proposed a 2 level developmental model for SPT [12, 28, 36]. Their model is based on the developmental delays of the abilities associated with each level and is meant to facilitate describing similarities and differences between visual and spatial perspective-taking. Surtees et al. consider in front and behind judgement as "level-1 type" and to the left and to the right judgement as "level-2 type". In both visual and spatial sub-dimensions, the level-1 perspective can be considered early-developing and level-2 as the later-developing skill. Surtees et al. have demonstrated that different processes are involved in the early-developing and later-developing perspective-taking independent of whether the judgments are visual or spatial [36]. One important component of spatial perspective-taking is "Frames of Reference" which allows us to encode spatial information relative to self/other/object [26]. When perceiving and understanding the spatial relations with another person or object, one needs to adopt a frame of reference. The same adoption is required when one is producing expressions that describe the spatial relationships. For communication to occur the perceptual cues and the verbal cues that describe the spatial relationships should be mapped into a mental representation [5, 13]. Retrospectively, the frame of reference with respect to spatial positions can build bridges between perception and language. This is important to consider, as the definitions and distinctions in the literature include the linguistics approach to the topic. Levinson emphasizes that essentially the distinctions between frames of reference correlate to the distinction between their underlying coordinate systems and not the objects that invoke them [26]. In this research, we have opted for using the 'egocentric' versus 'allocentric' classification which is commonly used in developmental and behavioural psychology and brain sciences.

The prospect of humans and robots interacting with each other on different topics calls for the need to study and investigate how to develop and incorporate this cognitive skill into robots and agents. In recent years, an emerging body of research has been dedicated to developing robots and agents with perspective-taking abilities and understanding how these abilities help improve human and robot interaction. One of the pioneering works on perspective-taking by Trafton et al. shows how equipping the robot with visual perspective-taking abilities can help to resolve ambiguous situations [41]. The ambiguous situation involved the robot having visual access to two objects where one was occluded from the human counterpart, a classic level-1 visual perspective-taking situation. The study initially analyzed human-human interaction scenarios involving perceptual perspective-taking. Then it proceeded to provide three important conceptual guidelines in building robotic systems in human-robot interaction. The authors evaluated their system in a collaborative interaction and with various frames of reference [41]. In another related work, Kennedy et al. studied level-2 perspective-taking abilities in robot using "like-me" simulation which included the robot applying its reasoning capabilities to the imagined situations [20]. Another study by Ros et al. incorporated the object ontology into the resolution of ambiguity [34].

The emerging body of research in human-robot interaction had discussed perspective-taking in robotics, demonstrating that perspective-taking played an important role in collaborative and learning scenarios with robots [20, 34, 41]. One of the factors to consider in studying perspective-taking development in robots is understanding how humans perceive robots and their agency. It has been shown that the assumptions humans make about robots are similar to the assumptions they make about their human counterparts [25]. For example, only showing certain nonverbal behaviours from the robot was enough for humans to attribute mental models to robots [48]. As a result, people tend to take the robot's perspective almost as much as they take other people's perspectives. In an effort to answer Alan Turing's pivotal question "Can machines think?", Krach et al. investigated perspective-taking with robots using fMRI, demonstrating that "the tendency to build a model of other's mind linearly increases with its perceived human-likeness" [23]. All these studies show that perspective-taking plays an important role in collaborative and learning scenarios. Another factor that contributes to the robot's ability to collaborate and cooperate with humans is the robot's understanding of spatial language and human's spatial language toward them. A recent study by Xiao et al. looked at the human speaker's perspective choices toward other humans versus robots addressee [44]. Their results indicated that humans assumptions about the addressee's capabilities to understand spatial descriptions did not differ significantly between humans and robots, however, they did not regard the robot as a human-like social partner. This shows that, even if a certain robot is still far from being a social partner in the human's mind, people still have certain expectations of the robot's understanding of the spatial world. Several other studies have revealed inconsistent differences in the way human speakers treat other humans and robots when addressing them [10, 31, 38]. The inconsistent findings can be attributed to the differences in the differences in the tasks, the speakers, and the addresses in particular robot types. Despite the differences in the setup and results, all the aforementioned studies agreed that human speakers followed the principle of least collaborative effort [6].

3 DESIGN

In this section, we present our approach for designing the task, interaction, and system architecture. The robot used in this study is the NAO robot, an affordable social robot, widely used in various fields of human-robot interaction [1]. The robot was originally developed by the French company, Aldebaran Robotics, later acquired by a Japanese company called Softbank robotics. It includes two cameras and four microphones with LED lights around the eyes and has 25 degrees of freedom. To program the robot a specialized NAOqi framework is used, which can be accessed either through the graphical interface Choregraphe or Robot Operating System in ROS. The technical development of the system and architecture is presented later in the section.

3.1 Object Game Task Design

Our gamified task is called the *objects game*, and it includes simple geometrical objects such as circles and squares in different colours and children are supposed to move them from one side to the other side of the screen. To get a glimpse of how the game looks, the easy version of the game is presented in Figure 1. The game includes a "main screen" (Figure 1a) and "goal cards" (Figures 1b and 1c). The main game is visualized on a touch screen and can be manipulated by the players, whereas a goal card is a mall physical card provided to one player and used to produce instructions to play the game.

3.1.1 Main Screen. The main screen is divided into two halves called the child side and the robot side. During the game, depending on the player's turn, one side is enabled, i.e. the objects are represented in colour and can be moved around, while the other

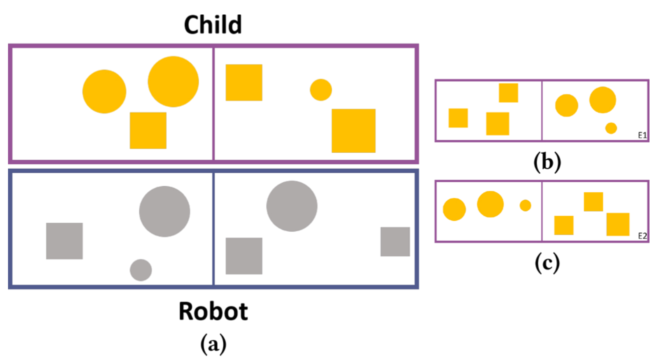

Figure 1: Easy level of the game, (a) main screen, child side activated, (b) E1 goal card, and (c) E2 goad card.

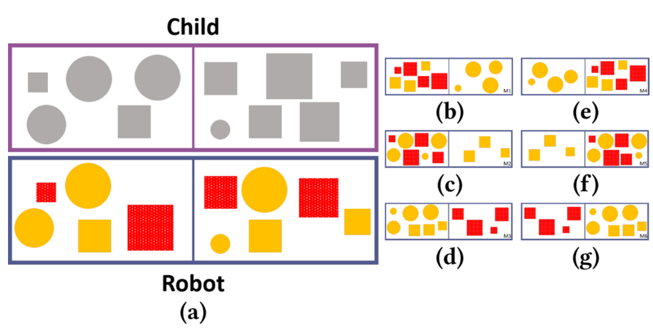

Figure 2: Medium level of the game, (a) main screen, robot side activated, (b) M1 goal card, (c) M2 goad card, (d) M3 goal card, (e) M4 goal card, (f) M5 goal card, (g) M6 goal card.

side is disabled, i.e. the objects are in grey and can't be moved. There is also a vertical division on each half used for distinguishing the left and right sides for each player. The game is designed with two levels of difficulty, which is a function of the colour and the type of the objects presented in that level. The difficulty corresponds to the number of moves needed for a player to reach the state of the game presented on the goal card. The objects can be of any shape or colour, in the easy version of the game as shown in Figure 1a, we used two types of objects: *squares* and *circles* both only in *yellow* colour. This level includes *yellow circles* and *yellow squares* and can be optimally solved in two moves. The medium level, i.e. the main game, has *red squares* in addition to the *yellow circles* and *yellow squares* and needs three optimal moves to be solved. Both levels are presented in Figures 1a and 2a.

3.1.2 Goal Cards. The goal cards represent the desired final state of the game that players are supposed to recreate by moving the objects. In one round of the game, one player is given the goal card and their task is to guide the other player to reach a state similar to what is represented in the card. In each round, as long as the correct number of objects of the same shape and colour are placed on the side represented in the card, the goal is achieved. The number of goal cards available for each game depends on the combination of the objects and colours available in that game. For example, the easy version of the game has 2 goal cards and the medium level has 6 goal cards as shown in Figures 1b-1c and 2b-2g, respectively. In each round of the game, one goal card is randomly selected and given to the player.

3.1.3 Player's Roles. When the game starts, one player has the task of guiding the other player to reach the state represented in the goal card without directly showing the card to them. The player with the goal card is called the *instructor* and the player moving the objects is called the *manipulator*. The instructions are composed of three components: colour, type of the object, and moving direction. An example of proper instruction is *"move the yellow circles to the right"*. For example, in the main screen shown in Figure 2a, the robot side is activated and the child side of the game is disabled. This means that the robot can manipulate the objects in front of it and the child is supposed to instruct the robot.

3.2 Interaction Design

There are three elements involved in designing the activity and experiment:

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

The perspective taker role refers to the players role while playing the game, which was described in the previous section as instructor and manipulator. To be comparable with the terms used in the related work, the instructor role refers to the speaker and the manipulator role refers to the listener. The importance of this element corresponds to the player's level of autonomy in selecting their frame of reference. A player in the instructor role takes the lead on which frame of reference to use for their instruction. While to perceive the instruction, the manipulator's choice of perspective

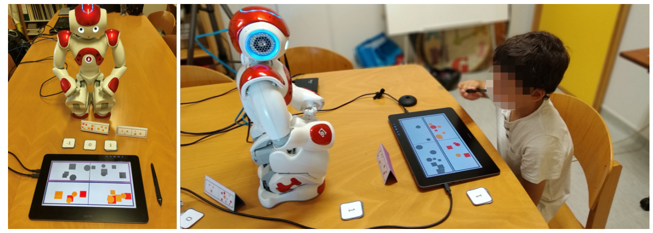

Figure 3: The experimental setup when robot is the instructor and child is the manipulator.

might be more limited and a function of the instructor's instructions. For the second element, frame of reference, we have restricted the choice to two, *egocentric* (from the self point of view) versus *addressee-centric* frames of reference (from the other point of view). The spatial positioning between the child and robot has been selected accordingly to accommodate this choice, which is Kendon's vis-à-vis arrangement, where the two players face each other directly [19]. In this setting, each player can either compose their instruction from the self's perspective or the other player's perspective. The last element factored in the interaction design is called *perspective marking* and it corresponds to how the speaker marks their perspective when they use natural language [35]. This element contributes to how the frame of reference is conveyed. Inspired by the Steels and Loetzsch's work, we used *explicit* marking which corresponds to the use of possessive pronouns e.g. my, your, and their, and *implicit* marking which corresponds to not using any of the possessive pronouns and rather the use of definite article e.g. the. When using implicit marking, it is the manipulator's responsibility to comprehend who's frame of reference is being used or ask for clarification. Whereas, using the explicit marking ease the comprehension of the instruction. Aligned with the principle of least collaborative effort, humans tend to use more implicit marking until there is a need for an explicit one [6]. On the other hand, according to Steels and Loetzsch, a marked perspective i.e. explicit requires less cognitive effort for aligning perspectives. Building upon these concepts, if a child is seated in a vis-à-vis with the robot and asks "give me a brick on *your* right", the child's utterance is addressee-centric and explicit. Whereas, if they tell the robot "give me the brick on the left", the child's utterance is marked as implicit, and it is not clear if the child is egocentric or addressee-centric meaning that the brick can be on the child's left or the robot's left.

For the first element of the interaction, instructor vs. manipulator, we decided to choreograph the interaction with children always making the first instruction. The data from children's first instruction is used to answer our first research question. Then, to give the child and the robot a chance to play as both instructor and manipulator, they play multiple sets of the game and switch roles. This means in one game the child has a goal card and instructs the robot and in the next game, the robot has the card and instructs the child. During the interaction, we control the robot's choice of frame of reference and observe the child's choice of frame of reference. Since the robot as the instructor has the creative control over which frame of reference to use, we decided the robot to show the behaviour that we had presumed children show in their first utterance, i.e. 'implicit egocentric perspective'. However, if the child asks for clarification, the robot would update its instruction to an explicit egocentric utterance. On the other hand, in the manipulator role, where the robot perceives the child's instruction, the perception perspective is a function of the child's perspective marking. If the child gives an implicit instruction, the robot is designed to perceive it egocentrically, which means if the child was egocentric, the robot makes an incorrect move. However, if the child's instruction is explicit, either egocentric or addressee-centric, the robot follows the instruction as it is. We expect that this behavioural design creates a certain level of perspective confronting which leads to children's failure in reaching the goal, a possible effort to accommodate the robot, and hopefully adaptation behaviour. As for the robot's perspective marking, we decided to keep the robot's instruction to be implicit egocentric in the first session as the instructor. However, to understand children's perception of the explicit behaviour of the robot, we designed the robot to be explicit in the second session as the instructor. The summary of the assumptions we made in designing the experiment is as follows:

- To evaluate children in both perspective roles, child and robot alternate between the instructor and manipulator roles,
- To document children's uninfluenced choice of perspective, children always start as the instructor in practice and session 1,
- To understand children's decision based on the robot's implicit instruction, the robot always gives *implicit egocentric* instructions in session 2 e.g. *"Move the squares to the right"*, and make it explicit if the child asks for clarification,
- To observe if and how children adjust their perspective to the robot after knowing the robot is egocentric, children instruct again in session 3,
- To understand how children react to the robot's change of frame of reference and perspective marking in session 4, robot's instruction perspective is divided into two conditions: *explicitly egocentric* (using *my*) and *explicitly addressee-centric* (using *your*).

To go through a round of interaction, for example, if the robot holds the M4 goal card shown in Figure 2e, optimally it can guide the child using three instructions. If the first instruction is *"move the yellow circles to the right"*, it is an implicit instruction, which means depending on the manipulator's choice of frame of reference, it can be interpreted differently. If the child is egocentric, it would move the yellow circles to its own right. However, if the child suspects that the robot was egocentric when making the instruction, it can have an addressee-centric approach and move the yellow circles to the robot's right which is the child's left. If the child has a doubt, they can ask the robot for clarification. During the interactions, we record these instructions and analyze the children's choice of frame of reference and utterances based on their goal cards. Figure 3 shows the setup of the experiment, the setup presents the robot as the instructor and the child in the manipulator role.

3.3 Technical Development

In this study, the NAO robot was programmed using Python API and ROS. The object game was designed using QML (Qt Modeling Language) [1] and was visualized and played through a Wacom

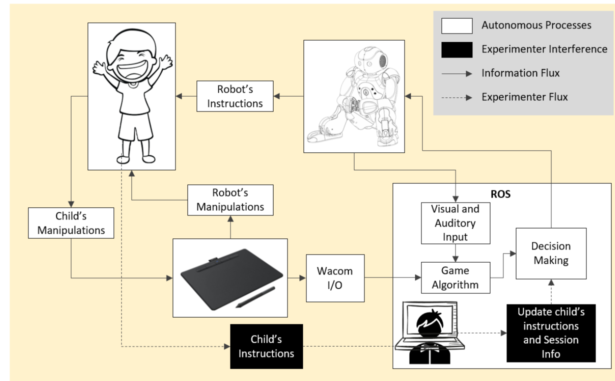

Figure 4: System Architecture

Cintiq Pro [2], where children could interact with it using its stylus. The robot interacted with the game using ROS nodes, where it received the current state of the game and sent the movement orders through ROS messages and simultaneously pointing at the objects on the Wacom to give the illusion of moving them with its hand. Figure 4 illustrates our system's architecture which represents the flow of the interaction in a more visual format. The system was semi-autonomous, with the experimenter only interfering to enter the child's instructions through a visual interface instead of using speech recognition. The reasoning behind avoiding speech recognition was the difficulties with recognizing speech for children in a language other than English and in this case Portuguese.

4 METHOD

4.1 Experimental Design

The study used a mixed design with two independent variables: *player's role* (instructor vs. manipulator) manipulated within-subjects and *robot's instructor perspective* (egocentric vs. addressee-centric and implicit vs. explicit) manipulated between-subjects. Table 1 provides the experimental design based on the assumptions and independent variables. It shows that children instruct in practice, session 1, and session 3 and the robot instructs in sessions 2 and 4. In the first condition called **Ego-Ego**, the robot is egocentric both times it instructs, i.e. sessions 2 and 4, and it is implicit in session 2 and explicit in session 4. In the second condition called **Ego-Add**, the robot is egocentric in session 2 and addressee-centric in session 4, where it is implicit in session 2 and explicit in session 4. The robot's implicit egocentric behaviour in session 2 was inspired by our first hypothesis about the behaviour children would show in their first instruction. The change of robot's perspective marking to explicit in session 4 was designed to evaluate how children perceive the robot's change of frame of reference and perspective marking. In cases where children asked for clarification in session 2, this change could give them the impression that the robot is accommodating them. It is worth mentioning that our initial experimental design involved more conditions, however, due to the limited number of

[1] https://doc.qt.io/qt-5/qtqml-index.html

[2] https://www.wacom.com/en-us/products/pen-displays/wacom-cintiq-pro-overview

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

Condition	Practice	Session 1	Session 2	Session 3	Session 4
Ego-Ego			Implicit Egocentric		Explicit Egocentric
Ego-Add			Implicit Egocentric		Explicit Addressee-centric

participants we could recruit, we opted for these two conditions that could cover our main research questions.

Regarding the dependent variables, we collected children's performance in the tests, moves within the interaction, and their choice of perspective as the instructor and manipulator. In the instructor role, if children used possessive pronouns they were marked as *explicit* and if they did not use any pronouns, they were marked as *implicit*. Furthermore, their instructions were analyzed based on the goal card in their hand and marked as egocentric or addressee-centric accordingly. When children were manipulating the objects, we evaluated their moves with respect to the robot's instruction and marked them as correct or incorrect if they corresponded to the robot's instruction.

4.2 Hypotheses

To find answers to our research questions and informed by the design of the task and interaction, we have developed the following hypotheses. Our first hypothesis is inspired by two research domains, theories of children's egocentric perspective in psychology and theories of least effort in linguistic [6, 11, 32]:

H1: Children use more "implicit egocentric" perspectives (without using possessive pronouns) compared to "explicit and/or addressee-centric" perspectives when they instruct the robot for the first time.

The next three hypotheses address our second research question which deals with children's perspective adaptation. The hypotheses look at the changes in the children's choice of frame of reference (H2a) and perspective marking (H2b) when they instruct the robot and the children's performance when the robot changes its frame of reference and perspective marking (H2c).

H2a: Children's overall choice of frame of reference shifts from egocentric in the first instruction to addressee-centric in the rest of the instructions.

H2b: Children's overall choice of perspective marking shifts from egocentric in the first instruction to addressee-centric in the rest of the instructions.

H2c: Children in the "Ego-Ego" condition (robot keeping the egocentric frame of reference) perform better in the last session compared to children in the "Ego-Add" condition (robot switching frame of reference).

The final hypothesis is based on the notion that engaging in perspective-taking activities with the robot can help children in related tasks. Hence, it looks at children's overall performance in the pretests and posttests.

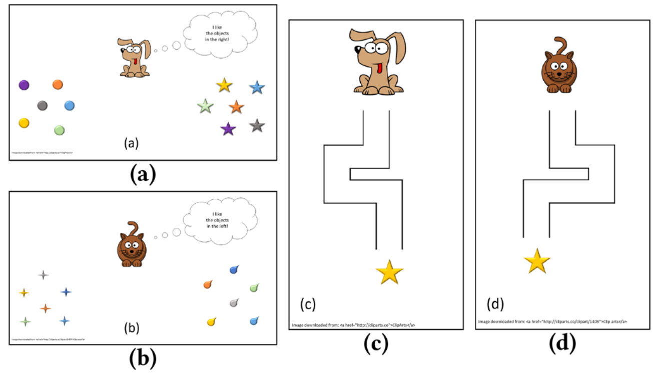

Figure 5: Examples of pretests and posttests: toys test (a) dog version and (b) cat version; path test (c) dog version and (d) cat version.

H3: children show better performance in the posttests compared to the pretests in taking the other character's perspective.

4.3 Perspective-Taking Tests

The activity was accompanied by two sets of pretests and posttests, which helped us evaluate if the task itself and interaction with the robot had any positive impact on their learning. The design of the tests was inspired by the experiments that commonly investigate level-2 spatial perspective-taking [21, 22, 47].

4.3.1 Left/Right Test or Toys Test. The toys test was designed to evaluate children's recognition of the perspective difference between themselves and another agent, i.e. the animal, that is facing them. We evaluated children's selection of the animal's favourite toy based on what the animal expressed in the test. If the child took the animal's perspective, we considered it a correct answer, otherwise, it was incorrect. Two similar versions of this test were designed with two different animals and different correct responses as shown in Figures 5a-5b. We alternated the tests as pretest and post-test between children. The instructions given to children are as follows: "The dog/cat thinks they like the right/left toy. Can you tell me which toy does the dog/cat like? (wait for the response)". In the dog version of the test, the dog says *"I like the right side toys"*, and in the cat version, it says *"I like the left side toys"*. The correct response to the dog version is *the balls* as they are on the right side of the dog and the correct response to the cat is *"the drops"*.

4.3.2 Test of Direction Sense or Path Test. The path test shown in Figures 5c-5d is a simplified version of "Money Standardized Test of Direction Test" developed by Money et al. and modified by Zacks et al. which has been adapted for children [30, 47]. Two versions of the test were designed with different animals and directions and were alternated as pretest and posttest. This test evaluates if children take the animal's perspective to guide them or not and how they guide the animal along a path. The instructions for this test are as follows: "The dog/cat wants to reach the star at the end of the road, can you describe the path that the dog/cat needs to take to reach the star? (help the child by saying 'move forward' and let them complete the instructions)". We gave children the freedom to describe the path as they were comfortable with. Any description within the following two approaches was acceptable. For example to guide the dog (in Figure 5c), a correct sequence for an approach that defines the path can be *"Forward-Right-Forward-Left-Forward"*. Another approach that is similar to walking with the animal in the path, can create the following sequence *"Front-Right-Front-Left-Front-Left-Front-Right-Front"*.

4.4 Procedure

The experiment took place in an empty room in an elementary school in Lisbon, Portugal, where children entered the room individually and met the two experimenters and the robot. Each child was seated in front of the robot with the Wacom tablet placed in between as shown in Figure 3. One of the experimenters was seated on the right side of the table and was responsible to ask the initial questions and explain the procedure to children in their native language, Portuguese. The second experimenter was seated on the other end of the table behind a laptop and was introduced to the child as the person who worked with the robot and was there in case the robot needed help. In reality, the second experimenter had access to a control panel that let them insert the child's instructions manually into the system. First, children responded to the pretests, then they went through one practice session and four sessions of the main game, and finally the posttests and a questionnaire about their impression of the robot.

5 RESULTS

5.1 Participants

A total of 35 participants (13 female, 22 male) between the ages of 7 to 9 took part in this study. They were selected from the 3rd and 4th grades of an elementary school to participate in the experiment. The study has received ethical approval from the university's ethics committee and parental consent from the parents of the participants before the main experiment. In addition to parental consent, all children provided verbal consent to participate and they were told they could withdraw from the experiment at any point they wished to. Moreover, the study was carried out after running a pilot with 7 children from 4 different age groups to test the system's functionality and to select the appropriate target age group [45]. The following analyses were carried out after excluding the data from 2 children, with 33 participants (11 female and 22 male) between the ages of 7 to 9 years old ($M = 8.22, SE = 0.12$). Among the 33 participants, 18 of them were starting their 3rd grade and 15 were starting their 4th-grade education. Among the 35 participants, 18 were in the Ego-Ego condition and 17 in the Ego-Add condition.

5.2 H1: Children's First Perspective Choice

To test this hypothesis we looked at children's first instruction in their instructor role, during the practice session. This instruction was before children received any feedback or became aware of the robot's perspective-taking abilities. In both conditions, children started as the instructor in the practice and session 1, as a result, we combined the participant data from both conditions to analyze this hypothesis. We have annotated children's instruction based on the frame of reference they used into *egocentric*, *egocentric-explicit*, *addressee*, and *addressee-explicit*. For simplicity, whenever we refer to egocentric or addressee-centric it means implicit unless it is stated otherwise. In the practice session, we noticed that none of the children used explicit utterances in their first instruction as presented in the first bar from the left in Figure 6. We used Chi-square goodness of fit to see if they significantly used more egocentric utterances compared to addressee-centric ones. On average children used more egocentric instructions compared to addressee-centric instructions, however, the test showed there is no significant difference between them ($\chi^2 = 3.6667, df = 1, p - value = 0.05551$). While the result rejects our hypothesis about significantly using egocentric utterances, it can be accepted when only considering the implicit utterances, since 100% of students used implicit instructions. Furthermore, the analyses show how children were more prone to start with implicit egocentric utterances before their information about their counterparts was updated. The rest of the bars in Figure 6 present the children's first two instructions in practice, session 1, and session 3, where they were the instructor. The figure shows how children adapted their instructions to accommodate the robot throughout the experiment.

5.3 H2a: Children's Frame of Reference Adaptation

In this part, we evaluated children's tendency to adapt their perspective to accommodate the robot. H2a tackles the frame of reference adaptation and H2b handles the perspective marking adaptation. To test H2a, we combined the "implicit egocentric" and "explicit egocentric" instructions into a variable called *egocentric** and "implicit addressee-centric" and "explicit addressee-centric" instructions into *addressee-centric** variable. Figure 7 shows the percentage of using egocentric versus addressee-centric for the first instruction in practice, session 1, and session 3. Children made more egocentric instructions in the practice session (66.66%) compared to session 1 (36.36%) and session 3 (36.36%). Since the robot was designed to have egocentric perception when receiving implicit instructions, the percentages show that children had adapted their instructions to the robot's frame of reference. We used Cochran's Q test as the

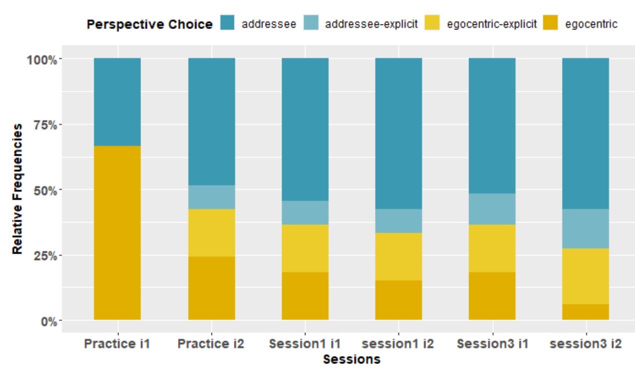

Figure 6: H1. Children's first two instructions in practice, session 1, session 3 categorized based on frame of reference and perspective marking. ("i1", "i2" refer to instruction 1, instruction 2).

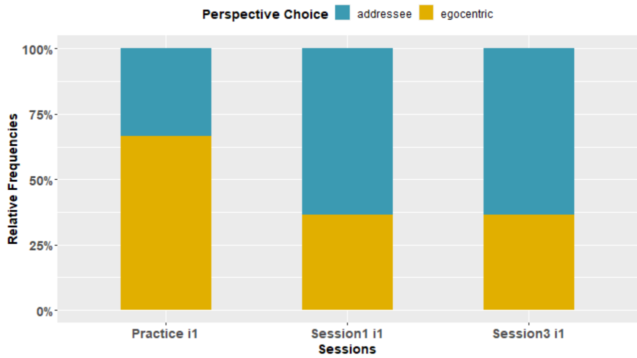

Figure 7: H2a. percentage of using egocentric vs. addressee-centric in the 1st instruction of the practice, sessions 1 and 3.

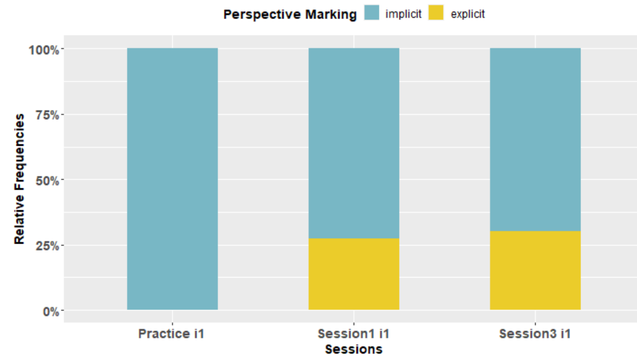

Figure 8: H2b. percentage of using implicit vs. explicit utterances in the first instruction of the practice, sessions 1 and 3.

variable is dichotomous with two levels and mutually exclusive and we need to compare them between three groups. Cochran's Q test determined that there was a statistically significant difference in the proportion of addressee-centric utterances over time $Q = 10.5263, df = 2, p - value = 0.005179$, hence we accept this hypothesis. We ran pairwise McNemar's Chi-squared test with Bonferroni adjustment which showed sessions 1 and 3 have significantly more addressee-centric utterances compared to practice. There is no difference between session 1 and session 3 as shown in Table $p - value = 1$. The result shows children made a significant change in their utterances after updating their mental model of the robot and then kept on instructing with that model.

5.4 H2b: Children's Perspective Marking Adaptation

To analyze H2b, we have combined the "implicit egocentric" and "implicit addressee-centric" instructions into a variable called *implicit** and "explicit egocentric" and "explicit addressee-centric" instructions into *explicit** variable. Figure 8 shows the percentage of using implicit versus explicit utterances for the first instruction in practice, session 1, and session 3. In the first instruction of the practice session, children only used implicit utterances, however, they started adopting explicit utterances in session 1 and session 3.

A brief look at the Figure 8 shows, that despite some children switching to explicit instructions in sessions 1 and 3 (27.28% and 30.31%), implicit instructions are still the dominant utterance. This shows a thought-provoking behaviour on the children's side, that they'd rather just switch left and right in their instruction to accommodate the robot rather than explicitly addressing the robot or themselves. This result can be partially due to the addition of extra cognitive processes simultaneously switching the frame of reference and marking the perspective. Similar to H2a, the data for this hypothesis is also dichotomous with two mutually exclusive levels, so we decided to use Cochran's Q test. Cochran's Q test determined that there was a statistically significant difference in the proportion of addressee-centric utterances over time $Q = 16.5455, df = 2, p - value = 0.0002554$, hence we accept the hypothesis. We ran pairwise McNemar's Chi-squared test with Bonferroni adjustment which showed sessions 1 and 3 have significantly more addressee-centric utterances compared to practice. There is no significant difference between session 1 and session 3.

5.5 H2c: Children's Performance vs. Conditions

In this hypothesis, we compared children's performance in sessions 2 and 4. Figure 9 shows children's performance i.e. reaching the goal (=1) and not (=0), in sessions 2 and 4 for Ego-Ego and Ego-Add conditions. The figure shows on average children performed better in session 4 compared to session 2. Considering the dichotomous and skewed data, we ran a McNemar Chi-squared test between sessions 4 and 2 for both conditions. For the Ego-Ego condition, McNemar's Chi-squared test with continuity correction showed ($\chi^2 = 4.9231, df = 1, p - value = 0.0265$), which means children significantly improved in winning the game in session 4 compared to session 2. On the other hand, McNemar's Chi-squared test with continuity correction ($\chi^2 = 0.36364, df = 1, p{-}value = 0.5465$) for Ego-Add condition showed no significant improvements between the two sessions. Unfortunately, despite the random allocation of children to each condition, we noticed that children in the Ego-Add condition showed a better initial performance in session 2 compared to the Ego-Ego condition. This might have contributed to the significant improvement of children in the Ego-Ego condition. As a result, we would not make any conclusive remark about this hypothesis until further experiments.

5.6 H3: Children's Performance in Pretest and Posttest

To check if children improved in responding to the posttest in comparison to the pretest, first, we looked at the type of data collected from each test. For each test, children either got 0 when they failed or 1 when they succeeded and the data for each test is dichotomous. We made a new variable by combining the result of both tests and looked at their overall performance in the tests. First, we analyzed the overall performance for normality using the Shapiro-Wilk test. A Shapiro-Wilk test for pretest and posttest showed a significant departure from normality, $W(32) = 0.78627, p = 1.786e - 05$ and

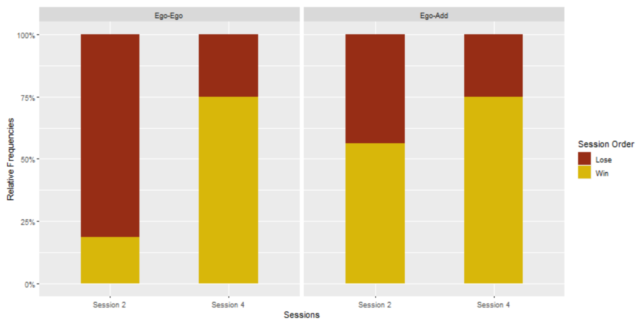

Figure 9: H2c. children's performance in session 2 and 4 for Ego-Ego and Ego-Add conditions.

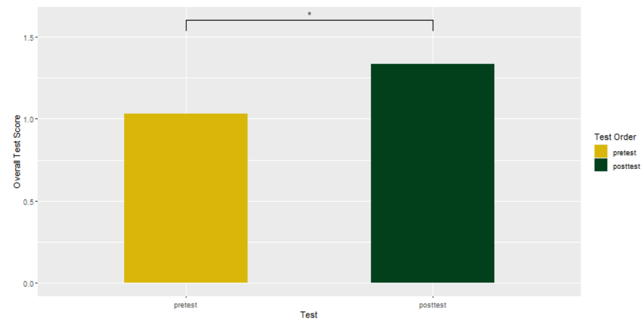

Figure 10: H3. Children's overall performance in pretest and posttest with combined score from toys and path tests.

$W(32) = 0.70321, p = 7.306e-07$, respectively. Considering that we expect children's performance to improve from pretest to posttest and the data is skewed, we performed a one-tailed Wilcoxon test. On average children performed better on the posttest ($Mdn = 2$) compared to the pretest ($Mdn = 1$). A One-tailed Wilcoxon signed-rank test with continuity correction indicated that this difference was statistically significant, $V = 29.5, p - value = 0.03826$, which lead us to accept this hypothesis. The overall test score is visualized in Figure 10.

6 DISCUSSION

In this work, we explored *children's first perspective choice toward the robot* and *children's tendency to accommodate the robot by adapting their perspective* in the context of completing a task with the robot. First, we designed a task that provided the ground for the child to take the robot's perspective in order to complete the task. Then, we configured the interaction using three elements that contributed to the way the instructions were composed or perceived. We implemented the interaction in the NAO robot and evaluated the children's instructions to the robot and their performance throughout the game. The interaction included a set of pretests, practice session, 4 main sessions, and posttests. Our initial findings demonstrate that children have the potential to change and adapt their perspective, i.e. frame of reference and perspective marking, to the robot's in the context of completing a task. They tend to create a mental model of their robot counterpart and the robot abruptly changing its frame of reference might affect children's performance or the flow of the interaction negatively. In the following paragraphs, we highlight the key findings of this study, how it can contribute to the future of designing robot's perspective-taking abilities toward children, and how children's understanding of the robot can guide the design of robots' mental models and perspective-taking behaviours.

Children tend to compose their first instruction by following the principle of least collaborative effort. This statement addresses children's behaviour in using more egocentric perspectives and not marking their instructions. The analyses show that children's first choice of perspective was *implicit egocentric*, however, they also had the tendency to correct their egocentric perspective to accommodate the robot after they failed to complete the task. These findings are aligned with the work of Epley et al., where they found that at the start of the interaction, both children and adults, first take their own perspective e.g. egocentric, and then through the course of interaction adjust it to another Epley et al.. As mentioned by Surtees and Apperly, taking owns perspective comes naturally, faster, and more accurately [37]. The implications of children's choice of frame of reference and perspective marking during interaction with a robot can be helpful to design the robot's behavioural model. This can further help the robot with creating a baseline of children's perspective-taking approaches in its first encounter with them.

Children significantly change their frame of reference and/or perspective marking to accommodate the robot's egocentric perception while instructing the robot. Further analyses show that children significantly changed their perspective choices particularly after the robot failed to reach the goal in the first session. Studies support that not only children but also adults tend to have automatic moments of egocentric perspective, however, adults tend to correct theirs faster than children [9]. While We observed children significantly changed to an addressee-centric frame of reference, they were not as prone to switch their perspective marking to an explicit one. These findings can be jointly explained by the iterative nature of perspective-taking [7] and the principles of least collaborative effort [44]. Children's tendency to stay with implicit instructions reveals that as long as a correct mental model of the counterpart is made, there is no need to complicate the instruction with additional information. The implications of children changing their perspective are highly valuable to us, particularly their choice of perspective when adapting to the robot's perspective.

The robot's change of perspective-taking and mental model should happen carefully and early in the interaction, otherwise, it can disturb the interaction. In the Ego-Ego condition, the instructions were explicit egocentric and in the Ego-Add condition, they were explicit addressee-centric. On one hand, we expected children to perform better in session 4 considering that the robot's frame of reference is explicitly expressed. On the other hand, we wanted to know if the robot switching its perspective from egocentric to addressee-centric would help children's performance, as they did not need to take the robot's perspective anymore. During the experiment, we observed that few children in the Ego-Add condition were confused by the robot's switch of frame of reference. Some of the children that had successfully completed session 2 failed in session 4, as they were rather confused by the robot's change of perspective or did not even realise that the robot's instructions were addressee-centric. We hypothesize the confusion is caused by the robot breaking the mental model that those children had already made. However, more experiments are needed to evaluate this hypothesis. It seems that in our case, changing the frame of reference from one session to another, with no particular necessity or prompting, harmed children's performance and hindered the flow of the activity.

6.1 Limitations and Future work

Our work has a few limitations that restrict the analyses and prevent us from making solid statements about some of our observations. Initially, we had planned a more complex experimental design where different scenarios were going to be evaluated. However, due to the limit in recruiting the minimum number of participants per condition, we opted for the current experimental design. More experimental conditions might have helped us discover the reasoning behind some of the children's choices. This limitation extends to the small sample size per condition, which goes back to the difficulties in recruiting children from schools. our next limitation corresponds to the uneven distribution of children between the two conditions. While most of the analyses were done based on the within-subject variable, we could not rely on the results from analyzing the between-subject variable. Despite that, children's performance showed an interesting pattern that can inspire future studies, in particular, to understand what type of behaviour from the robot can disturb the child's mental model of them. We designed the study in a semi-autonomous format, where the only time that the robot waits for human feedback is when the experimenter is inserting the child's instructions into the system. During the sessions, the robot asks for the child's feedback after every move it makes and the child can affirm the move is correct or not. However, we did not account for children asking for feedback from the robot and did not implement such steps into the design of the interaction. We observed that only 6 children out of 33 participants (18%) asked for the robot feedback after their first move in the second session. The decision was made to keep conditions uniform for all the participants. However, it can be considered as a design limitation, one that can be improved easily in the future developments of the interaction.

As for the future developments of the system, we expect to use the result from this study to develop a perspective-taking model for our robot. The model includes processes that decide if and when to accommodate children's perspective-taking abilities or challenge them depending on the goal of the interaction. The model is mainly inspired by the dual-process account of human judgments [8]. In the next steps, the model will be incorporated into the robot and evaluated based on how it accommodates children's perspectives or perceives them accurately. The platform can be used to challenge children to take different perspectives using more complicated and practical activities, such as the child and the robot collaborating to build something. The model can be further integrated with affective computing models to investigate how the robot's cognitive-affective states, such as the robot showing frustration, affects children's perspective-taking adaptation and perception of the robot [46].

7 CONCLUSIONS

This work introduced a platform that studied children's perspective choice and perspective adaptation to the robot's throughout the interaction. The components used for studying children's behaviour were the frame of reference and perspective marking. The interaction consisted of several sessions, with the child and robot taking turns to instruct each other and move objects around. In one condition, the robot was egocentric in all the sessions, and in the other condition, the robot switched to the child's perspective in the last session. It was observed that a considerable number of children, after realizing the robot's egocentric perspective did not switch to explicit expressions, instead, they switched to an implicit addressee-centric perspective, a behaviour aligned with the principles of least collaborative effort. Furthermore, the analyses hinted at the fact that when children assign a perspective model to a robot, they tend to keep making decisions using that model despite slight changes in the interaction. As a result, the robot's abrupt switch of frame of reference between sessions and with no prompting was found harmful in keeping the interaction transparent and it confused children. The study provided a set of measures for keeping the interaction with children natural, such as (1) switching between the frames of reference should happen when necessary or prompted and (2) being more explicit does not automatically make the interaction more understandable.

SELECTION AND PARTICIPATION OF CHILDREN

33 children aged 7-9 ($M = 8.22, SE = 0.12$) participated in the study. All children were attending the same school where the management staff and teacher showed interest in participating in the study. All parent were provided with a page describing general details of the study, the name of the researchers and a written consent form prior to the audio and video collection. Institutional recommendations were followed to insure data anonymization of all the logged data.

ACKNOWLEDGMENTS

Special thanks to the staff of Ideia o nosso sonho school for their participation and support. This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UIDB/50021/2020 and PD/BD/135150/2017, and CHILI laboratory at EPFL. This research was partially supported by TAILOR, a project funded by EU Horizon 2020 research and innovation program under GA No. 952215.

REFERENCES

[1] Aida Amirova, Nazerke Rakhymbayeva, Elmira Yadollahi, Anara Sandygulova, and Wafa Johal. 2021. 10 Years of Human-NAO Interaction Research: A Scoping Review. *Frontiers in Robotics and AI* 8 (2021).

[2] Yvonne Barnes-Holmes, Louise McHugh, and Dermot Barnes-Holmes. 2004. Perspective-taking and Theory of Mind: A relational frame account. *The Behavior Analyst Today* 5, 1 (2004), 15.

[3] Ellen Bialystok and Judith Codd. 1987. Children's interpretations of ambiguous spatial descriptions. *British Journal of Developmental Psychology* 5, 3 (1987), 205–211.

[4] Thomas Bolander. 2019. What do we loose when machines take the decisions? *Journal of Management and Governance* 23, 4 (2019), 849–867.

[5] Laura A Carlson-Radvansky and David E Irwin. 1993. Frames of reference in vision and language: Where is above? *Cognition* 46, 3 (1993), 223–244.
[6] Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22, 1 (1986), 1–39.
[7] Gideon Dishon and Yasmin B Kafai. 2020. Making more of games: Cultivating perspective-taking through game design. *Computers & Education* 148 (2020), 103810.
[8] Nicholas Epley, Boaz Keysar, Leaf Van Boven, and Thomas Gilovich. 2004. Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology* 87, 3 (2004), 327.
[9] Nicholas Epley, Carey K Morewedge, and Boaz Keysar. 2004. Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of experimental social psychology* 40, 6 (2004), 760–768.
[10] Kerstin Fischer. 2006. The role of users' concepts of the robot in human-robot spatial instruction. In *International Conference on Spatial Cognition*. Springer, 76–89.
[11] John H Flavell. 1977. The development of knowledge about visual perception.. In *Nebraska symposium on motivation*. University of Nebraska Press.
[12] John H Flavell, Eleanor R Flavell, Frances L Green, and Sharon A Wilcox. 1981. The development of three spatial perspective-taking rules. *Child Development* (1981), 356–358.
[13] Angela D Friederici and Willem JM Levelt. 1990. Spatial reference in weightlessness: Perceptual factors and mental representations. *Perception & Psychophysics* 47, 3 (1990), 253–266.
[14] Chris Frith and Uta Frith. 2005. Theory of mind. *Current biology* 15, 17 (2005), R644–R645.
[15] Adam D Galinsky, William W Maddux, Debra Gilin, and Judith B White. 2008. Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations. *Psychological science* 19, 4 (2008), 378–384.
[16] Julie A Hadwin, Patricia Howlin, and Simon Baron-Cohen. 2015. *Teaching children with autism to mind-read: The workbook.* John Wiley & Sons.
[17] Lauren Jay Harris. 1972. Discrimination of left and right, and development of the logic of relations. *Merrill-Palmer Quarterly of Behavior and Development* 18, 4 (1972), 307–320.
[18] Lauren Jay Harris and Ellen A Strommen. 1972. The role of front-back features in children's "front," "back," and "beside" placements of objects. *Merrill-Palmer Quarterly of Behavior and Development* 18, 3 (1972), 259–271.
[19] Adam Kendon. 1990. *Conducting interaction: Patterns of behavior in focused encounters.* Vol. 7. CUP Archive.
[20] William G Kennedy, Magdalena D Bugajska, Anthony M Harrison, and J Gregory Trafton. 2009. "like-me" simulation as an effective and cognitively plausible basis for social robotics. *International Journal of Social Robotics* 1, 2 (2009), 181–194.
[21] Klaus Kessler and Hannah Rutherford. 2010. The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology* 1 (2010), 213.
[22] Klaus Kessler and Hongfang Wang. 2012. Spatial perspective taking is an embodied process, but not for everyone in the same way: differences predicted by sex and social skills score. *Spatial Cognition & Computation* 12, 2-3 (2012), 133–158.
[23] Sören Krach, Frank Hegel, Britta Wrede, Gerhard Sagerer, Ferdinand Binkofski, and Tilo Kircher. 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PloS one* 3, 7 (2008), e2597.
[24] 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.
[25] Sau-lai Lee, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. 2005. Human mental models of humanoid robots. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2767–2772.
[26] Stephen C Levinson. 2003. *Space in language and cognition: Explorations in cognitive diversity.* Number 5. Cambridge University Press.
[27] Pascale Michelon and Jeffrey M Zacks. 2006. Two kinds of visual perspective taking. *Perception & psychophysics* 68, 2 (2006), 327–337.
[28] Henrike Moll and Andrew N Meltzoff. 2011. Foundation in Joint Attention. *Perception, causation, and objectivity* 286 (2011), 286–304.
[29] Henrike Moll and Michael Tomasello. 2006. Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology* 24, 3 (2006), 603–613.
[30] John Money, Duane Alexander, and HT Walker. 1965. *A standardized road-map test of direction sense: Manual.* Johns Hopkins Press.
[31] Reinhard Moratz, Kerstin Fischer, and Thora Tenbrink. 2001. Cognitive modeling of spatial reference for human-robot interaction. *International Journal on Artificial Intelligence Tools* 10, 04 (2001), 589–611.
[32] Jean Piaget, Bärbel Inhelder, and Alina Szeminska. 1960. The child's conception of geometry. (1960).
[33] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences* 1, 4 (1978), 515–526.
[34] Raquel Ros, Séverin Lemaignan, E Akin Sisbot, Rachid Alami, Jasmin Steinwender, Katharina Hamann, and Felix Warneken. 2010. Which one? grounding the referent based on efficient human-robot interaction. In *19th International Symposium in Robot and Human Interactive Communication*. IEEE, 570–575.
[35] Luc Steels and Martin Loetzsch. 2009. Perspective alignment in spatial language. (2009).
[36] Andrew Surtees, Ian Apperly, and Dana Samson. 2013. Similarities and differences in visual and spatial perspective-taking processes. *Cognition* 129, 2 (2013), 426–438.
[37] Andrew DR Surtees and Ian A Apperly. 2012. Egocentrism and automatic perspective taking in children and adults. *Child development* 83, 2 (2012), 452–460.
[38] Thora Tenbrink, Kerstin Fischer, and Reinhard Moratz. 2002. Spatial strategies in human-robot communication. *KI* 16, 4 (2002), 19–23.
[39] Andrew R Todd, C Daryl Cameron, and Austin J Simpson. 2017. Dissociating processes underlying level-1 visual perspective taking in adults. *Cognition* 159 (2017), 97–101.
[40] Michael Tomasello and Josep Call. 1994. Social cognition of monkeys and apes. *American Journal of Physical Anthropology* 37, S19 (1994), 273–305.
[41] J Gregory Trafton, Nicholas L Cassimatis, Magdalena D Bugajska, Derek P Brock, Farilee E Mintz, and Alan C Schultz. 2005. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 35, 4 (2005), 460–470.
[42] Karin M Vander Heyden, Mariette Huizinga, Maartje EJ Raijmakers, and Jelle Jolles. 2017. Children's representations of another person's spatial perspective: Different strategies for different viewpoints? *Journal of experimental child psychology* 153 (2017), 57–73.
[43] Henry M Wellman. 1992. *The child's theory of mind.* The MIT Press.
[44] Chengli Xiao, Liufei Xu, Yuqing Sui, and Renlai Zhou. 2021. Do people regard robots as human-like social partners? Evidence from perspective-taking in spatial descriptions. *Frontiers in Psychology* (2021), 4039.
[45] Elmira Yadollahi, Marta Couto, Wafa Johal, Pierre Dillenbourg, and Ana Paiva. 2020. Exploring the role of perspective taking in educational child-robot interaction. In *International Conference on Artificial Intelligence in Education*. Springer, 346–351.
[46] Elmira Yadollahi, Wafa Johal, Joao Dias, Pierre Dillenbourg, and Ana Paiva. 2019. Studying the Effect of Robot Frustration on Children's Change of Perspective. In *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, 381–387.
[47] Jeffrey M Zacks, JON Mires, Barbara Tversky, and Eliot Hazeltine. 2000. Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation* 2, 4 (2000), 315–332.
[48] Xuan Zhao, Corey Cusimano, and Bertram F Malle. 2016. Do people spontaneously take a robot's visual perspective?. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 335–342.