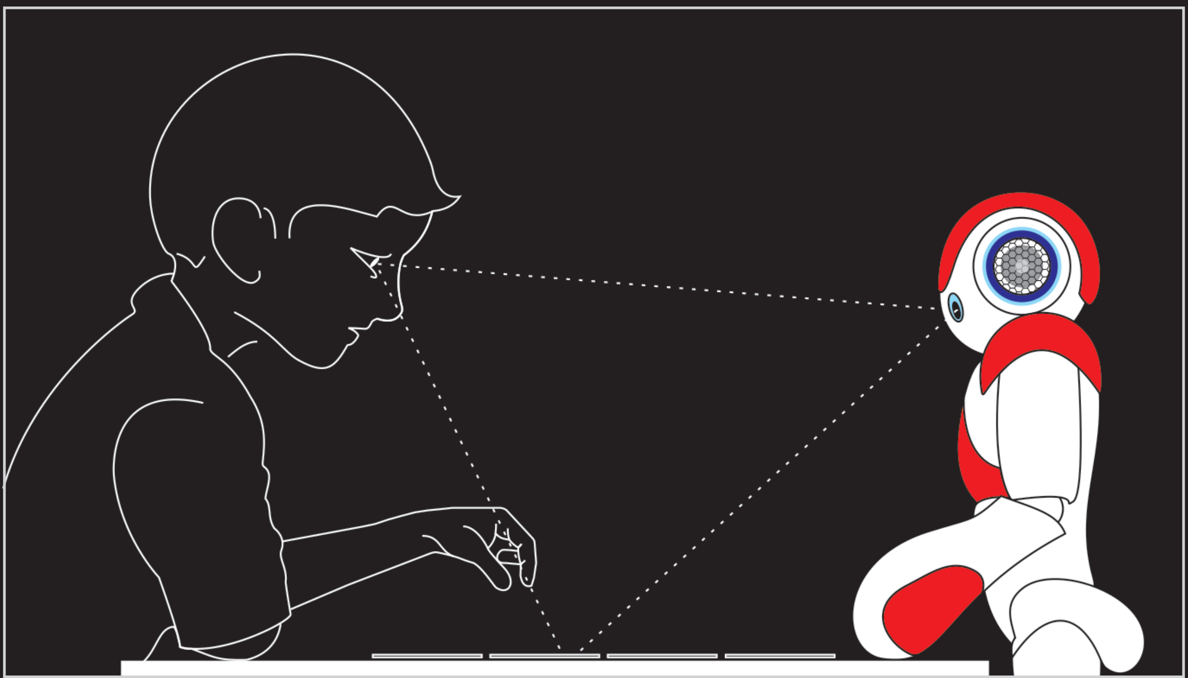
Gaze-based interaction for effective tutoring with social robots

GAZE-BASED INTERACTION FOR EFFECTIVE TUTORING WITH SOCIAL ROBOTS

EUNICE NJERI MWANGI



Gaze - Based Interaction for Effective Tutoring with Social Robots

Eunice Njeri Mwangi

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-90-386-5137-8

Keywords: Gaze - based interaction / Non -verbal behavior / Attentional cues / Directed attention / Human - robot interaction / Child - robot interaction / Social robotics / Education / Tutoring / Social training / Games for social robots

Cover Design: Ilse Modder | www.ilsemodder.nl

Cover Image: Shadi Kheirandish

Lay-out: Eunice N. Mwangi & Ilse Modder

Printed by: Gildeprint.nl, Enschede

Gaze - Based Interaction
for
Effective Tutoring with Social Robots

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus prof.dr.ir. F.P.T. Baaijens, voor een
commissie aangewezen door het College voor
Promoties, in het openbaar te verdedigen
op woensdag 28 oktober 2020 om 11:00 uur

door

Eunice Njeri Mwangi

geboren te Nyeri, Kenia

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. L. Chen
1e promotor:	prof.dr. G.W.M. Rauterberg
2e promotor:	prof.dr. M. Diaz-Boladeras (Universitat Politecnica de Catalunya)
copromotor(en):	dr.ir. E.I. Barakova
	prof.dr. A. Català-Mallofré (Universitat politècnica de Catalunya)
leden:	prof.dr. M.A. Neerincx (TUD)
	prof.dr. S. Kimani (JKUAT, Kenia)
	prof.dr.ir. L.M.G. Feijs (TU/e)

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

The work in this dissertation has been produced under the Erasmus Mundus Joint Doctorate Program in Interactive and Cognitive Environments. The research was conducted towards a joint PhD degree between the following partner universities:

Technische Universiteit Eindhoven

Universitat Politècnica de Catalunya

Acknowledgments

This PhD Thesis has been developed in the framework of, and according to, the rules of the Erasmus Mundus Joint Doctorate on Interactive and Cognitive Environments EMJD ICE [FPA n° 2010-0012] with the cooperation of the following Universities:

Alpen-Adria-Universität Klagenfurt – AAU

Queen Mary, University of London – QMUL

Technische Universiteit Eindhoven – TU/e

Università degli Studi di Genova – UNIGE

Universitat Politècnica de Catalunya – UPC

According to ICE regulations, the Italian PhD title has also been awarded by the Università degli Studi di Genova.

Acknowledgments

As my PhD journey ends, I would like to express my sincere and deepest gratitude to those who have supported and inspired me through this process.

First and foremost, I am thankful to God for the gift of good health and strength of mind to complete this work.

I would like to express my deepest gratitude to my supervisors, Prof Matthias Rauterberg, Prof Marta Díaz, Dr. Emilia Barakova, and Prof Andreu Català, for their tremendous support and helpful advice throughout my research work. Thank you for dedicating a lot of time to guide and mentor me to become a better researcher. Without your excellent training , the completion of this dissertation would not have been possible.

I would also like to extend my sincere gratitude to the teachers from DOE, Eindhoven, and El Margalló primary school in Vilanova I la Geltrú, Barcelona. Special thanks to teacher Mariken and teacher Jacqueline from DOE for your great interest in my research and for facilitating the children to participate in our study. This research would not have been possible without your contribution. I am extremely grateful to the teachers of El Margalló primary school for the warm welcome to your school and for graciously allowing us to set up our study within your institution. Your contribution to this work is invaluable.

My deepest thanks to the TU/e students, the children from DOE center and El Margalló School, and the children of my TU/e colleagues and friends who participated in our studies with a lot of enthusiasm. Thank you! It was such a delight to have you join our studies. You kept me inspired throughout my research work.

I'm also very grateful to the parents who voluntarily allowed their children to participate in our studies.

I would like to specially thank my colleagues and friends who have offered me incredible support during my research work. Many thanks to Merhnoosh, Kyra, Ruixin, and Joshua for your immense help during the preparation and running of my research experiments. You made this work possible.

I thank all my colleagues, lab mates, and friends at the department of industrial design, TU/e, and at the Technical Research Centre for Dependency Care and Autonomous Living (CETpD), UPC who have offered me incredible support during my studies. I had exciting conversations and stimulating discussions with most of you that inspired and kept me motivated through this process.

I would like to deeply thank my amazing parents, Mrs. Nelius Mwangi and Mr. John Mwangi, for their unwavering support and encouragement every step of my academic journey. Mum, thank you for your love and prayers. You have walked this journey with me, constantly reassuring me at every stage; there are no words to express my gratitude to you. Dad, although you are no longer with us, your words of wisdom and strength have kept me going. Thank you for teaching me to be resilient and persevering in challenging situations. I miss you!

My heartfelt thanks to my sisters Emma, Lillian, Lucy, and Betty, for cheering me on until the end. Many thanks to my brother Samuel, for your encouraging words and support. Loving thanks to my nieces and nephews. You all have made this achievement possible, and I am forever grateful.

Finally, I would like to express my deepest gratitude to all my friends for their encouragement and support along the way.

Summary

Gaze - Based Interaction for Effective Tutoring with Social Robots

The central thesis of this work is that effective gaze behavior can help build a shared understanding and mutual awareness between humans and robots, leading to positive outcomes in a tutoring interaction. Gaze behavior is an essential cue for social engagement and coordinated action, principally for tasks that imply human-robot collaboration, such as tutoring. The work presented in this dissertation is a compilation of findings from three empirical studies designed to explore the design space of gaze-based interaction to enrich human-robot interaction in educational settings where robots assume tutor or trainer roles.

In the first study, we examined how people perceive and interpret social cues from gaze provided by either a human or a robot tutor during a collaborative tutoring activity. The objective was to investigate whether people can notice and accurately interpret gaze-based cues from a tutor and whether they can accept the cues as help during learning interactions. We incorporated eye-tracking to examine gaze interaction during human-human and human-robot communications. We found that participants noticed the gaze cues from the robot tutor significantly more often than those of the human tutor. Consequently, we found that participants performed better with the robot tutor compared to the human tutor. These initial findings provide design recommendations for gaze-based communications to improve learning performance during human-robot tutoring.

Based on the results from the first study, we investigated how to implement gaze-based communication as an efficient help mechanism for robot-child tutoring. The objective was to examine child-robot gaze mechanisms to inform the robot's behavior design as a facilitator of children's task-solving. We carried out simultaneous observations of the child's gaze and the robot to examine the events of mutual gaze and gaze following patterns during the tutoring activity and to assess the impact of different child-robot coordinated gaze patterns on children's behavior and performance. We found that if a robot tutor provides gaze-based support, children perform better during the tutoring activity than when a tutor does not offer such cues. We also found that more events of mutual gazing patterns between the child and the robot tutor improve children's awareness of the tutor's intention during the activity leading to better performance. Therefore, we concluded that increasing gaze coordination between the child and the robot can improve performance and build mutual awareness during robot-based educative interventions.

In the last study, we investigated the nature and dynamics of gaze-based human-robot interaction (HRI) in tutoring. The objective was to examine intricate patterns of gaze interchanges between a child and a robot during the tutoring activity and to assess the impact of child-robot coordinated gaze on children's behavior and performance. We combined both observational and sequential lag methods to examine the relevant gaze sequences during a collaborative tutoring activity. We found that appropriate sequences and timing of the dyad's gaze behaviors between a child and a robot can lead to effective interactions between a child and a robot tutor. Based on these findings, we concluded that a robot tutor could positively influence the flow of the child's actions if the child interprets the social cues appropriately, improving the task execution and the play experience. This new understanding of the dynamic nature of gaze behavior during child-robot interaction contributes to the design of robot gaze behavior, to build better robot-based interventions in education and therapy settings. Overall, the findings from the user studies contribute to new design guidelines for gaze-based communications to improve learning performances and promote positive human-robot tutoring interactions.

In addition to the findings of the user studies, the main contributions of this dissertation include; First, an experimental framework for studying how gaze-based cues of robots can be applied to improve performance and quality of tutoring interaction. The experimental setting allows for simultaneous analyses of humans (adult-child) and the robot's gaze during a collaborative tutoring activity. Second, a coding scheme developed to measure the dynamics of child-robot interaction with an emphasis on coordinated and sequential gaze patterns between children and robots. The third is the use of simultaneous observational and lag-based methods to examine coordinated and interaction sequences of gaze between children and robots, helping unravel the dynamics of child-robot interaction in a tutoring setting. The lag-based methods provide an opportunity to investigate complex gaze sequences that—to our best knowledge—have not been previously explored in robot-based educational backgrounds or other human-robot interactions. The lag methods can be extended to analyze, in-depth, other interactive behaviors during human-robot interaction (HRI).

Contents

List of Figures

List of Tables

Chapter 1. **Introduction**

Advances in robotic technology have enabled robot integration into settings where they can interact collaboratively with human users. Accordingly, robots are gaining prominence, including within education (Belpaeme, Kennedy, Ramachandran, Scassellati, and Tanaka, 2018) therapeutic facilities for children (Albo-Canals et al., 2018; Henkemans et al., 2013; Scassellati et al., 2018) and in nursing homes for elderly (Fasola and Mataric, 2013; Feng et al., 2019; Perugia, Diaz-Boladeras, Catala-Mallofre, Barakova, and Rauterberg, 2020). A promising use of social robots is in tutoring, where robots take the role of tutors or trainers, for example, in rehabilitation centers for children with special needs. Besides, robots taking the tutor's role is inevitable, with schools facing teacher scarcity and the projected increase in student numbers.[*]

In the last few years, research has demonstrated the prospect for social robots to support tutoring (Belpaeme et al., 2018; Johal, Castellano, Tanaka, and Okita, 2018; Vogt et al., 2019). For instance, robot tutors can provide self-regulated learning (Jones and Castellano, 2018) and personalized interactions to children's affective and cognitive needs (Gordon et al., 2016; Ramachandran, Huang, and Scassellati, 2017). Robots can also elicit curiosity and foster engagement in children, which makes them useful learning tools (Gordon, Breazeal, and Engel, 2015). In therapy (social training), robots can help train fundamental social skills to children with autism (Huskens, Verschuur, Gillesen, Didden, and Barakova, 2013; Marino et al., 2020; Scassellati et al., 2018).

Although an increasing number of studies confirm the potential of robots in educational settings, much theoretical and practical work is required to make it a reality. Accordingly, to improve interaction outcomes in tutoring and other persuasion tasks, the design of intuitive interactions between robots and users is crucial. Human interactions are an excellent foundation on which to examine the design space for social robotic behaviors. For example, previous work shows that adding social communicative cues

[*] This chapter is partly based on selected paragraphs from the following publications (Mwangi et al., 2016; Mwangi, Barakova, Díaz-Boladeras, et al., 2018a; Mwangi, Barakova, Díaz, et al., 2018b; Mwangi, Barakova, Diaz, Mallofre, et al., 2017a; Mwangi et al., 2017b)

to robot behavior, both verbal or nonverbal, can improve task performance and learning (Admoni, Weng, Hayes, and Scassellati, 2016). Additionally, such cues can provide positive experiences; including engagement (Tapus et al., 2012; van Straten et al., 2018), emotional bonding (Diaz-Boladeras,2018), enhance compliance and persuasiveness (Ham, Cuijpers, and Cabibihan, 2015; Ghazali et al., 2018), and regulate human-robot collaboration (Huang and Mutlu, 2016).

While there is considerable progress towards natural interactions between robots and humans, how to design effective robot behavior to support human-robot tutoring interactions is still an open problem. This dissertation is aimed at understanding human-robot gaze-based interactions to inform the design of robots with credible and effective gaze behavior in tutoring and social training settings. To do so, we describe three user studies designed to investigate the concepts of mutual coordination — gaze-following and complex sequences of gaze behavior within robot tutoring interactions. These are well-controlled laboratory and field-based studies — with human participants (adults or children) interacting with a tutor (human or robot) in a board game tutoring activity. The dissertation adopts a multi-modal measurement and analysis approach that combines observational analysis and objective measurement techniques, including eye-tracking and lag methods, to analyze the interactive gaze behavior embedded in realistic tutoring settings. The findings from the user studies provide new guidelines to inform the design of gaze behavior to create effective and interactive human-robot tutoring interactions. The new knowledge of gaze-based communication can contribute to building relationships between robots and humans, to facilitate positive interaction outcomes, particularly concerning child-robot tutoring.

The remainder of this chapter describes the research background and the motivation for this research (Section 1.1) and outlines the research questions (Section 1.2). Section 1.3 describes the methodology followed in conducting this research. Section 1.4 describes the robotic platform, and Section 1.5 details the design of the board game activity used to perform the studies described in chapters 3 through 5. Lastly, Section 1.6 outlines the structure of the rest of the dissertation.

1.1. Motivation

In interactions between people, non-verbal cues such as gestures, facial expressions, body movements, and gaze help to establish understanding, mutual awareness, and a common ground for interaction (Frith, 2009). Gaze, in particular, is an essential part of nonverbal communication. According to Kleinke’s (1986) review of human-human interaction, gaze provides information, regulates interaction, and expresses an emotional state. Besides, the gaze is fundamental in the learning and development of children and is the basis of critical cognitive abilities such as the theory of mind and perspective-taking (Astington and Jenkins, 1995; Frischen, Bayliss, and Tipper, 2007; Hunnius, 2007). Therefore, gaze plays an essential role in enabling positive human social interactions.

Consequently, designing social gaze behaviors for a robot that are effective and credible is crucial. Prior work in human-robot interaction has shown that gaze can help build effective interactions between humans and robots (Admoni and Scassellati, 2017; Broz, Lehmann, Nakano, and Mutlu, 2012; Ruhland et al., 2015). In turn, this can result in positive outcomes, such as better recall (Mutlu, Forlizzi, and Hodgins, 2006), improved quality of communication (Kompatsiari, Ciardo, Tikhanoff, Metta, and Wykowska, 2019), and enhanced task performance (Boucher et al., 2012). There is also evidence that gaze is a powerful communicative signal as an implicit cue during collaborative activities (Mutlu et al., 2009) or in turn-taking (Palinko et al., 2015). However, how a social robot can apply its gaze behavior to achieve positive outcomes — in a tutoring scenario is still an open question — particularly regarding the role of deliberate — explicit attention-directing gaze cues and their relevance in learning scenarios. For example, is a robot tutor more effective when it provides (more) evident gaze cues or without cues or covert cues?

Research in child-robot tutoring and social training — autism therapy — has primarily investigated gaze as a communication signal with one-directional interactivity, paying little attention to the coordination and the interaction sequences between the child and robot. However, gaze interaction involves sequences of intertwined and coordinated looking behaviors. Moreover, the sequencing and timing of the gaze interactions, in particular, can be a reason for unnatural and, thus, ineffective human-robot interaction. Therefore, synchronized simultaneous observations of both the robot and the human's behavior can reveal dynamic patterns of gazing behavior that could impact communication and interaction outcomes. For instance, Kompatsiari et al. (2017, 2019) suggest that people feel more engaged when they establish mutual gaze interaction with the robot. However still, this and related studies do not explore the intricate gaze patterns of mutual interaction. Therefore, further investigation of these coordinated and sequential gaze patterns, particularly during child-robot tutoring, is needed.

This dissertation aims to investigate gaze-based communication in the space of creating effective tutoring interactions between humans and robots, and especially in the context of child-robot tutoring.

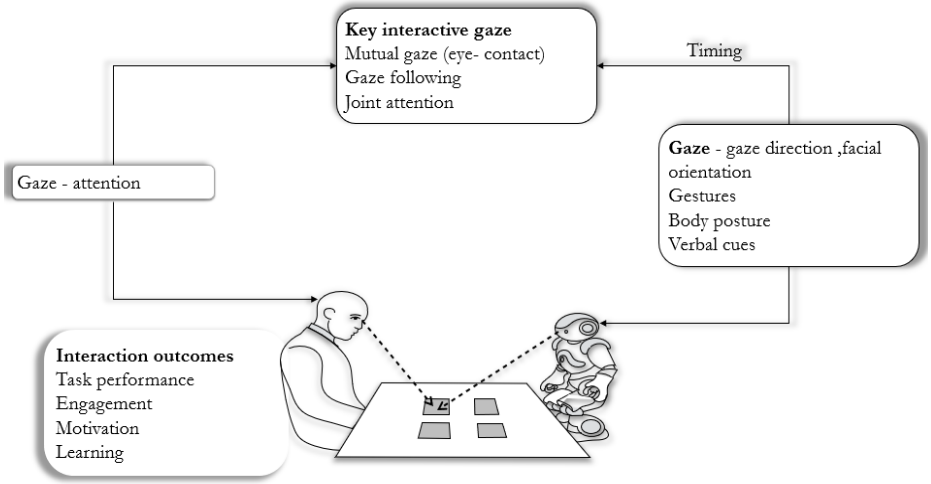


Figure 1-1 Research context: this dissertation investigates gaze-based communication to create effective tutoring interactions between humans and robots, and especially in the context of child-robot tutoring.

1.2. Research Questions

The main question addressed in this dissertation is how we can design effective and credible gaze behavior between people and robots, to facilitate a shared understanding and mutual awareness in a tutoring interaction?

The dissertation poses several sub-questions to answer the above question, addressed in chapters 3, 4, and 5. These questions focus on the following topics: exploring the influence of gaze-based cues in the context of tutoring, comparing the development of human-tutored and robot-tutored activities, and investigating the dynamic patterns of gaze-based interaction during human-robot tutoring.

The research questions are:

RQ1 — Do tutors' social cues of gaze influence performance in the context of a board game tutoring activity (either with a robot or with a human tutor)? To what extent?

(a) Do participants notice and properly interpret the gaze-based hints (i.e., tutor pointing with the eyes to a particular card) exhibited by the tutor while performing the card-matching task?
  i. Do participants notice the tutors' gaze behavior during the game?
  ii. Do participants attribute the tutor the intent to help with the gaze?

(b) Does the gaze behavior influence the choices of players in the card game?

RQ2 — Who is a better tutor: the human or the robot tutor?

(a) Is there any difference in task execution between the two conditions?

 i. Is there any difference in the time required to complete the game?

 ii. Is there any difference in the number of tries?

(b) Is there any difference in the participants' awareness of the tutors' behavior between the two conditions?

(c) Is there any difference in participants' judgment about the tutor (e.g., likeability, perceived presence) between the two conditions?

RQ3 — What are the dynamics of dyadic gaze-based interaction (mutual gaze, gaze-following, and joint attention) in the context of tutoring?

(a) How do we describe and identify the key patterns of coordinated gaze behavior in the context of tutoring?

(b) Which are the contextual and individual variables that affect their occurrence?

(c) How do we model the sequence of coordinated behavior according to individual and situational variables of interest, such as the flow of the game (i.e., a sequence of failed and successful tries)?

RQ4 — Does the coordinated gaze (mutual gaze, gaze-following) influence performance?

(a) Do dyadic patterns influence participants' awareness of the robot's behavior?

(b) Does the coordinated gaze influence task execution performance?

 i. Is there any difference in the time required to complete the game?

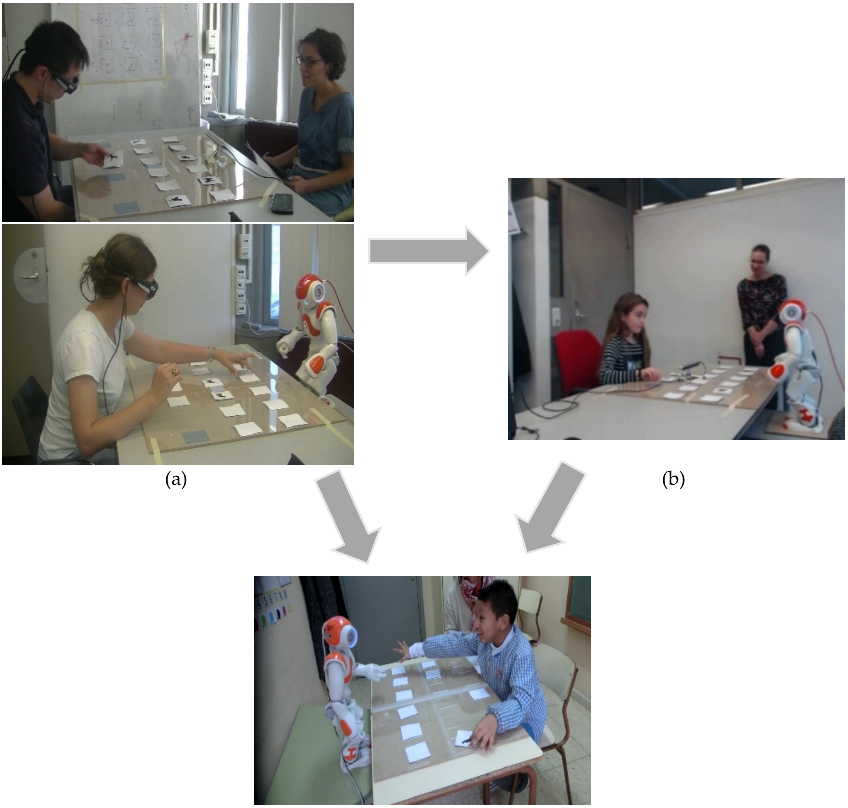
 ii. Is there any difference in the number of tries?

(c) Do dyadic patterns influence participants' social engagement?

1.3. Methodology

This dissertation presents three empirical studies designed to answer a combination of the research questions in section 1.2. The systematic studies follow a human-centered process that involves gaining a theoretical understanding from the literature on gaze interaction as it occurs in human interactions and through preliminary observations of people interacting in the same scenario in laboratory settings *(see* section 1.5*)*. After gaining this knowledge, the next step involves conceptualizing a theoretical framework, including writing hypothetical statements that guide the study's design. The next step involves preparing the platform for testing, including programming the robot's behavior, as well as a series of preliminary tests in the lab, to observe how people perceive the gaze motions, as well as the timing and the experimental task. The last step involves testing the research hypotheses in human-robot interaction evaluations to assess whether gaze influences human behavior, as is hypothesized.

We designed an experimental setting based on a board game scenario — memory game (Figure 1-4) to provide opportunities for gaze-based interactions. The interactivity of the robot behavior was gradually increased to make it possible for more complex interactive patterns to be created, as shown in Figure 1-2. The aim is to gather empirical evidence on how the participants respond to different manipulations in the robot's gaze behavior. Study 1 and Study 2, detailed in chapters 3 and 4, respectively, are laboratory-based, and Study 3 — the final experiment, is a field study conducted at an elementary school with children.



(a) (b) (c)

Figure 1-2 (a) Study 1: comparing gaze-based interaction between a human and a robot tutor (b) Study 2: concepts of dyadic gaze behavior — mutual gaze — gaze following in child - robot tutoring (c) Study 3: complex coordinated sequences of the child - robot gaze during a collaborative tutoring game interaction.

This dissertation adopts a multi-modal measurement and analysis approach that combines observational analysis, objective measurement techniques, including eye-tracking, self-reports, and lag sequential methods. The objective measures used in this dissertation cover the dimensions of task performance (e.g., the time to complete the game and number of tries to finish the game) and observable behavior (e.g., duration of gazes to the robot's face or the cards on the board). In this dissertation, the term performance is used as an overarching concept referring to participants' observed behavior during the trials, including execution (time and number of tries) and frequency and duration of gaze behavior. Therefore, performance refers to all the participant's observable behavior, including but not limited to task execution. Subjective measures include users' judgments of the tutor (e.g., likeability, social presence, intent) and a report on the subjective experience (e.g., engagement, enjoyment).

Study 1: In addition to the objective measures (task performance) and subjective self-reports, we incorporated eye-tracking to investigate participants' gaze behavior during human-human and human-robot communications (Figure 1-2 (a)). The benefit of using eye-tracking is that it provides data with high temporal resolution and can reveal detailed patterns of gaze behavior (Blascheck et al., 2014).

Study 2: We used observational analysis to examine child-robot dyad's gaze behavior during the game (including patterns of mutual gaze and gaze-following) (Figure 1-2 (b)). Observational methods are an effective way to assess child-robot interactions as children can sometimes not give direct answers to interview questions or when completing a questionnaire. Besides, prior studies have used observation methods to analyze child - robot interaction with success. For example, Díaz-Boladeras (2017) used observational and ethographic techniques to study emotional bonding in child - robot interactions. De Haas et al. (2017) used observational methods to evaluate gameplay interactions between children and robots.

Study 3: We combined observational and lag-based methods (Bakeman and Quera, 1995; Pohl, Wallner, and Kriglstein, 2016) to make a detailed analysis of the dynamic gaze interactions between the robot and the child (Figure 1-2 (c)). The lag-based methods provide an opportunity to look at more complex gaze sequences that have never been previously explored in robot-child tutoring, which is a necessary part of developing successful robotic interventions in tutoring and therapy settings.

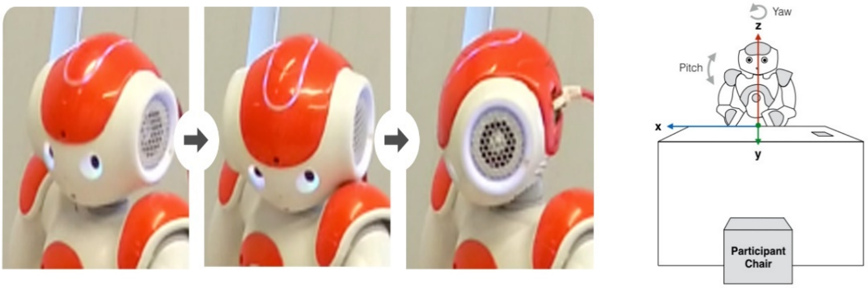
Table 1-1 A summary of research questions, experimental designs, robot interactivity, and evaluation methods for all three studies.

	Robot behavior and interactivity	Study design and evaluation
Study1 (RQ1; RQ2)	Simple attention directing gaze behavior Initial verbal instructions Closing verbal remarks	Mixed factorial design experiment Two-by-two (Between-subject; Tutor_Type: Human or Robot) and (Within-subject; Tutoring_Style: Gaze vs. No_Gaze) Objective measurements—performance Eye-tracking—gaze behavior Subjective (self-reports)—participants' perceptions
Study2 (RQ1; RQ3)	Simple attention directing gaze behavior Initial verbal instructions Closing verbal remarks	One-factor (Tutoring_Style) Within-subjects design with two conditions (Help (Gaze) vs. No_ Help (No gaze cues) Objective measurements performance Observational analysis—child and robot gaze units, dyadic gaze patterns (mutual gaze and gaze following) Subjective (self-reports)—children's perceptions
Study3 (RQ3; RQ4)	Complex joint attention gaze—improved timing and sequencing of the head/gaze movement of the robot based on Study 1 and study2 NAO face tracking—initial gaze interaction Initial verbal instructions Closing verbal remarks Verbal feedback throughout the game Expressive explanation /instructional gestures Winning and losing gestures accompanying verbal feedback	One-factor (Tutoring_Style) Within-subjects design with two conditions (Gaze with feedback vs. No gaze with feedback) Objective measurements—performance Observational analysis—child and robot gaze units and dyadic gaze patterns (mutual gaze and gaze following) Lag sequential analysis—patterns of joint attention gaze, the timing of dyad gaze behavior Subjective (self-reports)—children's perceptions

1.4. Research Platform

The robot platform used to perform the studies in this research is the programmable humanoid NAO from Softbank Robotics (Softbank, 2013). NAO is 58 cm tall, with 21 degrees of freedom. It can walk, speak, gesture, pan, and tilt its head and has a multitude of sensors. NAO has minimal facial features with a static mouth and eyes, and its face bears a resemblance to that of a child. NAO has been adopted widely for research focused on interactions with children with autism spectrum disorders, either for therapeutic (Gillesen et al., 2011) or for general educational—pedagogical purposes (Belpaeme et al., 2018). Plausible reasons could be its availability and minimalistic appearance, which make it suitable for young children, and its sensing and acting capabilities.

Figure 1-3 (b) shows a coordinate system that can help us describe its behavior in terms of the movement of different actuators from the perspective of the observer. The HeadYawAngleVal of NAO's gaze direction is the angle between the positive y-axis and a line drawn from the center to a fixed position. The yaw angle of the y-axis is 0, and a positive head yaw angle value is on the left side of NAO. NAO head yaw angles range from -119 to 119 degrees. The HeadPitchAngle of NAO's gaze direction is the angle between the *XY* plane and a line drawn from NAO's head location to a target square. The pitch (head joint front and back) angle increases from 0 to 29.5.



(a) (b)

Figure 1-3 (a) Research platform: design of the head movements for NAO robot (b) NAO coordinate system: the positioning of the robot during the experiments and the coordinate system that will be determining its head/gaze orientation.

Robot capabilities differ significantly from those of humans, and hence their ability to communicate gaze information. Therefore, to increase the effectiveness of robot gaze behaviors, it is crucial to establish how people perceive gaze cues while interacting with a robot in a specific context. Cheaper and more accessible robots are the feasible choice for use in education, as is illustrated in the review by Belpaeme et al. (2018). However, a notable challenge of designing gaze for a robot platform like NAO is that it lacks articulated eyes and, therefore, has to turn its entire head to direct its gaze. Even so, previous studies have shown that robots with no movable eyes can also direct gaze with head orientation (Cuijpers and Van Der Pol, 2013). For example, Cuijpers and Van Der Pol (2013) measured the region of eye contact with NAO and concluded that perception of gaze direction with NAO is similar to that of a human.

1.5. Interaction Setting — Board Game Design

We contextualized the user studies detailed in chapters 3, 4, and 5 in a play situation based on a board game activity, the 'Memory' game. In this task, participants played a card-matching game—finding matching pairs of cards—collaboratively either with a human or a robot tutor.

Figure 1-4 shows the design of the interaction setup used to conduct the studies detailed in chapters 3, 4, and 5. The participant and NAO sat at opposite sides of a table facing each other. A board grid with the card positions resembling a memory game was fixed on a table. The layout had 18 squares (8*8cm) organized in six (6) columns and three (3) rows. The 18 squares corresponded to 18 card positions. The squares were 10 cm apart in depth (y-axis) and 6 cm apart in width (x-axis).

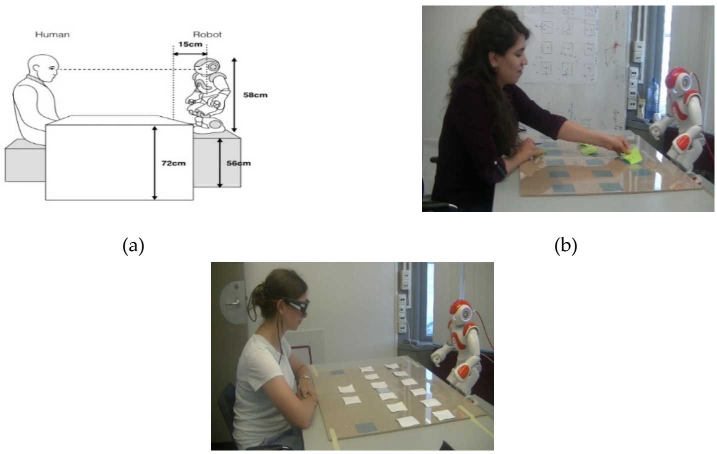
To measure NAO head angles to different card locations on the board, we placed NAO on a small desk 56 cm in height at the "Stand-init" pose (0, 0). The design grid had six squares in the x-direction, which was from the left to the right side of NAO, and three squares in the y-direction, which was the depth direction of NAO. The distance between NAO and the closest square position was approximately 20 cm away, and the furthest at the corner was about 60 cm. We attached a laser beam on the mid-section of the NAO head and adjusted it to point at the middle of the layout, using the "look at" module in Choregraphe program. We estimated the head pitch and head yaw angles for all the target positions using the motion screen on Choregraphe.

Card Position= {HeadYawAngleVal, HeadPitchAngleVal}

In our preliminary study (Mwangi et al., 2016) — "See where I'm looking at – Perceiving Gaze cues with a NAO Robot": we examined whether people accurately perceive NAO gaze cues and head angles directed towards different target positions on the board layout (Figure 1-4 (b)). Our motivation was that many settings used for therapeutic training and educational purposes are in the form of board games and other settings when the tutor and tutee are on the opposite sides of a table — as is the case in this research. Therefore, it is necessary to establish how people perceive gaze cues while interacting with the robot, and whether NAO can communicate gaze based on head movements, independent of eye movements.

From the results, we found that the NAO robot can provide gaze cues through appropriate head movements - facial orientation. Specifically, we found that when the head pitch angle is higher (24±2), and the depth is less, approximately 20 cm from the robot, participants detect the positions with good accuracy (Mwangi et al., 2016). These initial results informed the design of our interactive setting –

board game design and especially the card placement on the board, as shown in Figure 1-4 (c) (see *a complete description of the preliminary study and the board game design in* Appendix A*).*



(a) (b)

(c)

Figure 1-4 (a) Schematic illustration of the interaction setup (b) Participant marking the card positions NAO was looking at on the board layout (c) The design of the board game based on gaze perception results for NAO: the game consists of an arrangement of cards (14 cards), designed such that the head angles of NAO (facial orientation) can direct attention at different card locations when the participants and the robot are facing the board.

1.6. Thesis Overview

Chapter 2 - provides a review of related work on non-verbal communicative skills in human-social robot interaction. The emphasis is on both the design perspective of implementing gaze cues for social robots and on understanding how humans perceive robots' gaze behaviors.

Chapter 3 - investigates how people perceive and interpret social hints from gaze exhibited by either a robot or a human tutor during a tutoring interaction.

Chapter 4 - investigates dyadic gaze patterns during child-robot collaborative tutoring interaction regarding children's awareness of the tutor's intention, performance, and children's perceptions in a tutoring activity.

Chapter 5 - investigates the nature and dynamics of gaze-based human-robot interaction (HRI) and specifically, the dynamics of successful gaze sequences in the context of the child - robot tutoring.

Chapter 6 - provides a comprehensive discussion of the main findings and limitations of this research, including open challenges that inform future research directions.

Chapter 7 - provides conclusions based on the significant findings of the research and outlines the contributions of this dissertation.

Chapter 2. **State of the Art**

This chapter begins with a broader background reviewing non-verbal communicative cues in social human-robot interaction. Among these abilities, the emphasis is on studying the literature in human-robot interaction (HRI), which includes both the design perspective of implementing gaze cues for social robots and how humans perceive the robot's gaze behavior. In this context, the literature review focuses on the role-related specific skills needed to perform effective tutoring with an emphasis on intuitive non-verbal communicative abilities. Finally, this chapter highlights the implications from the literature, identifies gaps in the corps of knowledge, and develops the conceptual framework for empirical research.†

2.1. Non - Verbal Communication in HRI

Nonverbal cues, such as gestures, facial expressions, body posture, and body movements, form an essential part of human social interactions (Frith, 2009). Accordingly, non-verbal cues contribute significantly to the meaning exchanged in communication. Human-robot interaction has made considerable progress in mapping non-verbal human behavior to robots in an attempt to reach natural interaction. These studies show non-verbal cues can help build effective interactions between humans and robots in diverse domains (Admoni, 2016; Breazeal, Kidd, Thomaz, Hoffman, and Berlin, 2005; Kirchner, Alempijevic, and Dissanayake, 2011). For example, embedding nonverbal cues in robots can enhance task performance during human-robot interaction in diverse applications (Admoni, Weng, Hayes, and Scassellati, 2016; Boucher et al., 2012; Mutlu, Forlizzi, and Hodgins, 2006). Breazeal et al. (2005), for example, provides evidence that non-verbal communication, including gestures, can enable effective human-robot teamwork. Admoni et al. (2016) demonstrated how to achieve effective interactions with robots using nonverbal in various domains, for example, robot tutors, robot therapists,

† This chapter is partly based on selected paragraphs from the following publications (Mwangi et al., 2016; Mwangi, Barakova, Díaz-Boladeras, et al., 2018a; Mwangi, Barakova, Díaz, et al., 2018b; Mwangi, Barakova, Diaz, Mallofre, et al., 2017a; Mwangi et al., 2017b)

and robot coaches. Huang and Mutlu (2013) developed a toolkit to generate useful social behavior for robots to achieve positive outcomes in an educational setting.

Adding social cues to a robot's behavior can provide positive experiences, including enhanced engagement (Kompatsiari et al., 2019), improved compliance (Ghazali et al., 2018), emotional bonding (Diaz-Boladeras, 2018), and a common ground for interaction (Huang et al., 2015).

Humans naturally use many social cues to synchronize their operations in collaborative tasks. As a result, nonverbal cues, including gaze and gestures, have been applied to support fluent turn-taking during human-robot collaboration (Chao and Thomaz, 2012; Palinko et al., 2015). Hart et al. (2014) also found that non-verbal cues can help coordinate human-robot cooperative tasks in manufacturing settings.

Implications: These research efforts on mapping non-verbal human behavior to robots signify substantial progress towards natural and intuitive interactions with robots in diverse domains. The studies collectively support the notion that nonverbal communication patterns are essential aspects of human-robot interactions and may influence interaction outcomes in the context of tutoring. Thus, intuitively providing nonverbal cues is crucial for the successful application of robots in settings like education, coaching, or therapy. Thus, nonverbal communication is beneficial for understanding important design variables to improve human-robot relationships. For example, facial expressions — gaze — body posture — can provide information about attention and engagement. Therefore, examining non-verbal cues can help inform robot design guidelines for human-robot tutoring. While the above researchers have found potential benefits of enriching human-robot interaction with nonverbal cues, few methods available provide specific guidelines for designing them effectively, specifically in robot tutoring interactions. The next segment focuses on gaze behavior, which is an integral part of nonverbal communication and, therefore, in human-robot interaction.

2.2. Gaze in Human Communications

The gaze is a crucial non-verbal cue in human social interaction. Gaze facilitates a range of social functions during human-human interactions, including signifying attention (Frischen et al., 2007; Langton, Watt, and Bruce, 2000), coordinating turn-taking (Kendon, 1967), and communicating emotions and mental states (Argyle, Ingham, Alkema, and McCALLIN, 1973; Kleinke, 1986). The earliest social psychological studies and reports on the role of gaze in human behavior were by Argyle, Cook, and Cramer (1994), Kendon (1967), and Kleinke (1986). In this review, we focus mainly on the social function of gaze to the direction of attention to areas of information in the environment and to facilitate learning during social interactions.

Gaze and Attention: Attention is a crucial component of learning interactions. The gaze is particularly useful in directing another person's attention to areas of information in an environment during human social interactions (Emery, 2000; Frischen et al., 2007). Besides, humans are highly sensitive to the gaze behavior of the interacting partner. The sensitivity begins from a young age with newborns (three months) showing sophisticated gaze-following while interacting with their parents and caregivers (Brooks and Meltzoff, 2014; Brooks and Meltzoff, 2002; Meltzoff and Brooks, 2007). Such gaze-based interaction (gaze-following) facilitates the formation of joint attention, which is the basis of early childhood learning and the development of critical social cognition abilities such as the theory of mind and perspective-taking (Baron-Cohen, 2008; Frischen et al., 2007). Besides, deficits in gaze interaction in either dyadic or triadic joint attention abilities are some of the core communicative deficits associated with an autism spectrum disorder (Baron-Cohen, Leslie, and Frith, 1985; Baron-Cohen, 2008; Frischen et al., 2007).

Gaze and Learning: Neuropsychological evidence shows that perceiving and recognizing the gaze of the other person enables one to understand the intentions of the interacting partner (Emery, 2000). Therefore, seeing where another person is looking can guide attention to objects of interest that could be beneficial to the decision-making and reasoning during an activity (Emery, 2000; Langton et al., 2000). In turn, this could have a significant impact on task solving performance during activities. Such is particularly true for a tutor/coach who is always using gaze to keep the students engaged and to direct them to areas of information, which can influence their task solving. For example, in a classroom setting, teachers use their gaze to guide students' attention to points of information during learning — either on the board or elsewhere. Moreover, experienced tutors manipulate their gaze, either using more eye contact or directing their gaze towards students to keep the students engaged and to show attention.

Dyadic Gaze: Dyadic gaze behavior like mutual gaze, gaze-following, and joint attention indicate both engagement and the quality of social interaction. Farroni, Csibra, Simion, and Johnson (2002), argued that eye - gaze behavior helps to establish mutual coordination and a common ground for interaction. Coordinated gaze also influences how humans perceive and interpret interactions. For instance, mutual gaze indicates social engagement, gaze-following shows an understanding of others' attention, and gaze alternation is used to assess joint attention (Kaplan and Hafner, 2004; Kendon, 1967; Kleinke, 1986).

Several studies demonstrate the importance of mutual gaze in social interactions. Mutual gaze is the earliest form of interaction between a parent -caregiver dyad; therefore, social exchanges between mother and infant using mutual gaze create the first dyadic system in which the two individuals have similar control (Brooks and Meltzoff, 2002). Furthermore, eye contact is a very significant behavior for the caregiver, especially in the early phase of development, because it signals that the infant is participating in the interactive exchange (Brooks and Meltzoff, 2002; Meltzoff and Brooks, 2007).

Implications: The above sections highlight the significant role of gaze in communicating the direction of attention during social interactions and that persons are susceptible to the interacting partner's gaze behavior. Thus, at a minimum, seeing where another person is looking helps obtain cues that can act both implicitly and explicitly to direct attention and shape thoughts and decisions. In turn, such cues can be beneficial to the cognitive processes of the interacting partner, thereby improving their task solving and their experiences. Therefore, this characteristic of gaze is important to consider in the space of creating effective human-robot interaction. Besides, the ability of the gaze to facilitate such mutual coordination and awareness facilitates positive outcomes during interactions.

2.2.1. Social Gaze Definitions and Eye -Tracking Concepts

We have adopted the following definitions of interaction concepts of social gaze and task-solving eye-tracking concepts that are relevant to this research from different literature perspectives:

Table 2-1 Social gaze definitions and eye-tracking concepts

Social Gaze	Definition
Gaze - eye-gaze – looking at - look at	Looking at the other in the eyes, or the upper half of the face (Argyle, Cook, and Cramer, 1994).
Mutual gaze	The reciprocal gaze directed at the face region from one agent to the other (Argyle, Cook, and Cramer, 1994).
Gaze-following	Following the gaze direction of another and looking at the same target as the partner (Emery, 2000).
Joint attention	A sequence of coordinated gaze including mutual gaze and gaze-following that leads to sharing an attentional focus on an object of interest (Emery, 2000; Frischen et al., 2007).
Shared attention	Dyadic gaze whereby both persons attend the same object, as with joint attention, and are aware of each other's direction of attention (Emery, 2000).
Gaze cueing	"Use of perceived gaze direction to shift visual attention, that is, the seemingly automatic propensity to orient to the same object that other people are looking at" (Frischen et al., 2007).
Referential gaze or deictic gaze	Gaze at an object or location in space. Such a gaze sometimes occurs in combination with verbal references to an object, though it need not accompany speech (Emery, 2000).
Eye - Tracking Metric	**Definition**
Area of Interest (AOI)	Regions of a stimulus that are of interest to the researcher for a particular hypothesis. The researcher defines the AOIs either before or after an eye-tracking experiment. In our case, the area of interest is the face of a tutor (either the human or robot) (Blascheck et al., 2014; Holmqvist, 2011).
Fixation	The eyes' focus is on a particular scene lasting for approximately 200–300 milliseconds (Blascheck et al., 2014). Fixations are excellent measures of visual attention.
Fixation_count	Number of fixations within an area of interest
Fixation _time (ms)	Total fixation time on an area of interest. Fixation time is often associated with attention and visual processing. For example, more fixations on a particular AOI may indicate that the AOI is more important or more noticeable (draws attention) - of the participant than the other areas (Blascheck et al., 2014; Holmqvist, 2011).
Dwell_time (ms)	The time spent looking within a particular AOI, calculated by summing up the time the gaze coordinates were within the area of interest (Holmqvist, 2011). Dwell time can be used to answer the question of participant's interest - attention towards certain parts of the stimulus.
Revisits	The number of times the eye returns to an AOI that has already been visited (Farnsworth, 2018; Holmqvist, 2011). Revisits allow a researcher to study which areas of a stimulus repeatedly attract the attention of the participant.

2.3. Gaze in Human-Robot Interactions

Given the critical role of gaze in human communication, research into designing social gaze for robots has been extensive. There are several comprehensive reviews on gaze with social robots and designing effective gaze-based interactions for diverse human-robot interactions (Admoni and Scassellati, 2017; Broz et al., 2012; Mutlu, 2009; Ruhland et al., 2015).

Perceiving and Responding to a Robot's Gaze Behavior: Several studies have investigated how people perceive and respond to the robot's gaze behavior in diverse settings. For example, in a storytelling setting by Mutlu et al. (2006), participants recalled the story better when the robot looked longer at them. Previous works have also established that gaze is a powerful communicative signal without the observer's explicit awareness. For example, in object selection games, Mutlu et al. (2009) showed that participants could read and interpret unconscious leakage cues from the robot's gaze directed toward objects or locations in the environment. The authors claim that, in general, the gaze cue led to better performance in a guessing game task and even better with Robovie, which is more human-like than Geminoid. Palinko, Rea, Sandini, and Sciutti (2016a, 2016b) studied the impact of eye-gaze or head-based gaze estimation in a human-robot interaction experiment using the iCub robot. Yu, Schermerhorn, and Scheutz (2012) compared the timing of gaze behavior when interacting with either a robot or a human in a word-learning task. Their eye-tracking result revealed that people pay more attention to the robot's face than the human face during a word-learning task. Admoni, Hayes, Feil-Seifer, Ullman, and Scassellati (2013) examined the impact of frequency and the duration of gaze on the perception of attention during human-robot interaction, concluding that shorter, more frequent fixations are better for signifying attention than more prolonged and less frequent fixations. Andrist, Tan, Gleicher, and Mutlu (2014) combined three functionalities, including face-tracking and head detection, to design gaze aversion behaviors for conversational robots. Their findings showed that participants perceived the designed gaze as more

Gaze and Effective HRI: Prior work shows that gaze cues can enhance task performance and promote positive experiences during human-robot interaction. For example, Yoshikawa et al. (2006) explored both responsive and non-responsive gaze cues. They found that responsive gazes had a strong effect on the "feeling of being looked at" during the interaction. Moon et al. (2014) studied the effects of gaze behaviors in a handover task. They found that gaze cues can improve the handover timing and the subjective experiences in handover tasks. Boucher et al. (2012) examined the effects of gaze on the speed of communication in both human-human and human-robot interaction cooperative tasks. Their findings show that participants can use the gaze cues a robot to perform in an anticipatory manner and improve their performance in physical interaction tasks. Robots that establish mutual gaze with a human partner lead to positive evaluations. For example, Kompatsiari et al. (2017) suggest that people are sensitive to an artificial agent's mutual gaze, and they feel more engaged with the robot when they establish mutual gaze. Pfeiffer-Lessmann et al. (2012) examined the timing of gaze behavior in interactions between humans and a virtual human to build a joint operational model for artificial agents. Yonezawa, Yamazoe, Utsumi, and Abe (2007) showed that when a robot companion establishes a mutual gaze with human users, this leads to positive robot evaluations.

Gaze Interactions in Robot Learning — (robot coach -robot therapist-robot tutor)*:* Studies have also addressed the implicit role of gaze to support tutoring in collaborative settings. For example, Palinko et al. (2015) examined mutual gaze as a mechanism to enhance turn-taking in a collaborative teaching dictation scenario. Jokinen, Furukawa, Nishida, and Yamamoto (2013) demonstrated the importance of timing of gaze in turn-taking interactions during collaborative work, and as a mechanism to control turn-taking. Gaze behavior is important to examine in child tutoring, and therapy settings as the deficit in gaze interaction is argued to be a significant cause of autism. Previous work has shown that children with ASD show increased gaze behavior and shared attention during robot interactions than human interactions (Mavadati, Feng, Gutierrez, and Mahoor, 2015; Tapus et al., 2012). Therefore, for robots to be used in social training settings, especially for children with autism, the gaze is an important cue.

Implications: The above studies have found potential benefits of enriching HRI with gaze behavior with the unintentional and unconscious influence of gaze cues in a competitive -collaborative setting. However, how a social robot can utilize deliberate/explicit attention-directing gaze cues, particularly in learning settings, is still an open question. There is still no conclusive evidence whether robot tutors are more useful when providing (more) evident gaze cues or without cues or covert cues. Besides, as cheaper and accessible robots are a more feasible choice for use in educational settings, further investigation is required to figure out whether robots, such as NAO, can convey meaning through gaze cues based on head movements rather than on independent movements of the eyes.

In child-robot interaction (tutoring and social training), examining the coordinated and sequential gaze patterns can quantify, for example, whether the robot is gazing at the child mutually or whether the child is reflexively following the robot's gaze. Such data can inform the design of interventions to create effective robot tutors or for social training to improve learning outcomes. For instance, robot behavior could be designed in a manner such as to increase the level of mutual gaze in an interaction. A different intervention may be needed if the aim is to increase the child's' mutual gaze, such as increasing the robot's verbal cues or gestures or creating more prospects to encourage the child to look at the robot.

2.4. Designing Gaze Behavior for Social Robots

Table 2-2 recaps prior work on designing social gaze behavior for robots, including the type of robot used, the participants' details, methodologies and significant findings from the studies. The outlined studies follow different approaches to study gaze in various interactive robot applications, including educational settings, assistive robots, entertainment, and manufacturing. This dissertation draws from these previous research efforts to design gaze-based interactions for social robots in tutoring settings.

Table 2-2 Summaries of prior work on designing social gaze behavior for robots

Title	Robot — participants	Research objectives — questions	Methodological approach	Main results
(Admoni et al., 2013) Are You Looking at Me? Perception of Robot Attention is Mediated by Gaze Type and Group Size	myKeepon 53 participants (20 male, 33 female)	To examine which features of a robot's gaze make the robot appear to be attending to someone: frequency and gaze duration. Examined the influence of two types of gaze behaviors — short, frequent glances and long, less frequent stares.	Mixed 3 (group size) x 4 (gaze duration) between- and within subject's design	People feel the robot is looking at them when the robot uses shorter, more frequent glances than longer, less frequent glances.
(Boucher et al., 2012) I Reach Faster When I See You Look: Gaze Effects in Human and Human-Robot Face to Face Cooperation	I -Cub 5 participants	To determine whether the effects of the robot gaze can be generalized to performance improvements in a cooperative object selection task.	Object selection task Quantitative evaluation Qualitative evaluation	A robot can use gaze to support its speech in a cooperative object selection task. Humans can use the robot gaze to perform in an anticipatory manner in a cooperative task leading to better performance.
(Kinoshita et al., 2017) Trans gazer: Improving Impression by Switching Direct and Averted Gaze Using Optical Illusion	Trans gazer 12 (6 pairs) students (10 male; 2 female; Native Japanese) Age: 21 - 24	To evaluate the effects of Trans gazer's convex and hollow eyes, in one-to-many communication. To examine the impact of direct and averted gaze to multiple people simultaneously by changing the shape of the robot's eyes.	Seven-minute lecture about a cooking method to two listeners Within-participants design Likert method to quantify the participants' impressions of Trans gaze Video camera to ensure that the participants were looking at Trans	Hollow eyes can direct gaze more broadly, whereas convex eyes can convey gaze direction more correctly.

			gazer	
(Mutlu et al., 2006) A storytelling robot: Modeling and evaluation of human-like gaze behavior	Honda's ASIMO	To examine the impact of frequency of the robot gaze on recall during a storytelling task with two participants.	Storytelling activity with a robot and two participants ASIMO gaze was manipulated between the two participants during storytelling Pre and post questionnaires	Participants performed significantly better in recalling ASIMO's story when the robot looked at them more.
(Mutlu et al., 2009) Nonverbal Leakage in Robots: Communication of Intentions through Seemingly Unintentional Behavior	Geminoid and Robovie R-2 Geminoid—a near-human android Robovie—a humanoid with abstract, stylized human-like features 26 participants (17 male; 9 female; Native Japanese)	To examine whether people detect nonverbal leakages in robots. Do people attribute intentions to the robot? How do the physical characteristics of the robot affect these inferences?	Object selection task Performance evaluation: time Subjective evaluations: affective states	Participants could read and interpret unconscious leakage cues from the robot's gaze directed toward objects, which led to better performance in a guessing game task and even better with Robovie, which is more human-like than Geminoid.
(Yoshikawa et al., 2006) Responsive Robot Gaze to Interaction Partner	Robovie-R2, ATR Robotics) 39 Participants (20 male, 19 female) Ages: 18 - 24	To explore the effects of responsive gazes: Responsive: following gaze and aversive gaze Non-responsive: 100% staring gaze and independent gaze	Subjective evaluation Gaze detection tool	A robot that gazes responsively toward a human user is capable of eliciting a stronger feeling of being 'looked at' than a robot that uses non-responsive gaze behavior.

(Yu et al., 2012) Adaptive Eye - Gaze Patterns in Interactions with Human and Artificial Agents	Humanoid torso with a 2 DoF, Movable head and two 3 DoF arms. 48 students' participants	To investigate the micro-level behaviors of humans in interaction with human and artificial agents.	Word learning as an experimental task Comparative experimental design: Human confederate vs artificial agent Eye-tracking study	Participants—look more—to a robot partner's face than a human partner's face during a word-learning task.
(Palinko et al., 2015) Gaze Contingency In Turn-Taking For Human Robot Interaction	I Cub 8 participants (2 male; 6 female)	To examine whether mutual gaze can implicitly facilitate turn-taking in a dictation scenario. To compare a contingent behavior where the robot reacts to its partner glances to a purely rhythmic one.	Dictation scenario Objective and subjective evaluations of different strategies	Mutual gaze for a robot is an efficient means to seamlessly interact with human partners with different needs in a turn-taking task.

2.5. Summary

This review suggests that adding social cues to robots can lead to effective human-robot interactions. From this review, we draw the following: *First*, gaze and non-verbal communication, in general, are platform and context-dependent. Therefore, while there is extensive work focusing on designing gaze cues in human-robot interaction, how non-verbal cues and, especially, gaze can be manipulated to improve learning outcomes in human-robot tutoring is still an open question. The *second* relates to the importance of timing in gaze-based communication, which is crucial in designing effective interactions. Most of the prior studies have paid little attention to the interactive behavior and how gaze interactions unfold overtime during human-robot interaction. Therefore, further investigation of dyadic components -timing and sequences - of gaze interaction between humans and robots is needed to create more effective robot tutors. The following chapters (Chapters 3 through 5) will describe user studies designed to explore gaze interaction, mainly focusing on attention, mutual coordination, and intricately coordinated gaze patterns during human-robot collaborative tutoring. The next chapter focuses on the role of gaze in directing attention, which is a critical component of learning interactions.

Chapter 3. Gaze-Based Interaction in Human-Human and Human-Robot Tutoring

Attention is a crucial component of learning interactions. Gaze is particularly useful in directing another person's attention to objects of interest in an environment during social interactions (Emery, 2000; Frischen et al., 2007). Therefore, seeing the gaze of another person can guide attention to information of interest that could be beneficial to one's decision-making process and reasoning during an activity (Emery, 2000; Langton, Watt, and Bruce, 2000). In turn, such cues can be beneficial to the cognitive processes of the interacting partner, thereby improving their task solving. In classroom settings, for example, teachers manipulate their gaze to keep their students engaged and to direct their attention to points of interest during learning. Therefore, designing effective gaze behavior for robots is crucial to foster social engagement and coordinated interactions in human-robot tutoring.[3]

Prior research has demonstrated that gaze can help build better interactions with robots (Admoni and Scassellati, 2017; Broz et al., 2012). However, how a robot should apply its gaze behavior to improve learning performances in a tutoring context is still an open question. In this chapter, we examine whether gaze-based cues from a human or a robot can direct attention and influence participants' decisions and choices during a tutoring activity and whether humans can perceive and interpret such cues accurately. The underlying assumption is that gaze can help cue attention, influence decisions and thoughts, and improve performance.

To investigate the above, we conducted a well-controlled laboratory study in which participants played a matching board game in the presence of a tutor (either a human or a robot tutor), in two sets. In one of

[3] This chapter is largely based on the following publications (Mwangi, Barakova, Díaz-Boladeras, et al., 2018a; Mwangi, Barakova, Diaz, Catala, et al., 2017b)

the setups, the tutor provided gaze-based hints — looking towards the correct match — during the interaction. In the other case, the tutor did not provide help to the participants during the tutoring task.

We incorporated eye-tracking to examine participants' gaze behavior during the two tutoring styles and with the two different tutors. We assessed how the gaze-based tutoring style impacts participants' performance, gaze behavior, and the tutor's judgments compared to the no-gaze style. We identified two primary measures that are notably used to measure performance in memory games tasks. The duration the participant takes to complete the game and also the number of tries — moves — required to find all the matching cards.

We found that when a robot tutor provides gaze-based cues during tutoring, participants perform better using fewer tries to complete the task than when the robot tutor does not offer such cues. Further, we found that people identified the robot tutor's gaze hints more than the human tutor gaze hints and, as a result, performed significantly better using — fewer tries — with the robot tutor than with the human tutor. This chapter provides new recommendations for designing effective gaze-based communication to improve task performance in a robot-tutoring interaction.

The rest of this chapter is structured as follows: Section 3.1 provides background on prior works in human-robot gaze interaction, describes the objectives and the research questions. Section 3.2 describes the experiment design — the hypothesis, study conditions, measures, and participants' details. Section 3.3 details the experimental setup, the platform, and the design of the interaction scenario and the board game, "Memory." Section 3.4 outlines the results of the user study. Section 3.5 provides a comprehensive discussion of the main findings. Finally, Section 3.6 summarizes the main recommendations for designing a credible and effective gaze behavior robot-based educational setting from this chapter.

3.1. Background

The literature reviewed in Chapter 2 highlights the significant role of gaze in communicating the direction of attention during social interactions. Thus, at a minimum, seeing where another person is looking helps obtain cues that can act both implicitly and explicitly to direct attention and shape thoughts and decisions. Accordingly, following the gaze of another person can direct attention to various objects in the environment, allowing knowledge to be directly shared. Previous work has demonstrated that people are also sensitive to the robot's gaze directed at objects or locations in the environment. For example, the work of (Mutlu et al., 2009) shows that "leakage cues" — which are unintentional and unconscious both in their production and perception — can express intent during activity and improve task performance using the Robovie and Geminoid platforms. The above works focus on the implicit and unconscious influence of gaze cues in a collaborative setting.

This chapter investigates deliberate/explicit gaze-based interaction cues to build awareness and shared understanding to improve learning outcomes. Palinko et al. (2016) studied the impact of the eye- or head-based gaze estimation in a human-robot interaction experiment with the iCub robot. The robots used in these studies can move their eyes independently of the head. This poses the question of to what degree simpler and more available robots can perform gaze cueing effectively. Therefore, to increase the effectiveness of robot gaze behaviors, it is crucial to establish how people perceive gaze cues while interacting with a robot. As cheaper and better accessible robots are the more feasible choice for use in education and social training, we investigate whether robots such as NAO can convey such deliberate meaning through gaze cues.

The overall goal of this chapter is to examine whether gaze hints from a tutor (either human or robot) can direct attention and, therefore, influence the choices of human partners in a game-play tutoring task. We further examine whether humans can read such cues and accept help from a tutor, and, in turn, if these cues influence their decision-making and the performance of the interacting human.

In this regard, this chapter addresses the following questions:

RQ1 — Do tutors' social cues of gaze influence performance in the context of a board game tutoring activity (either with a robot or with a human tutor)? To what extent?

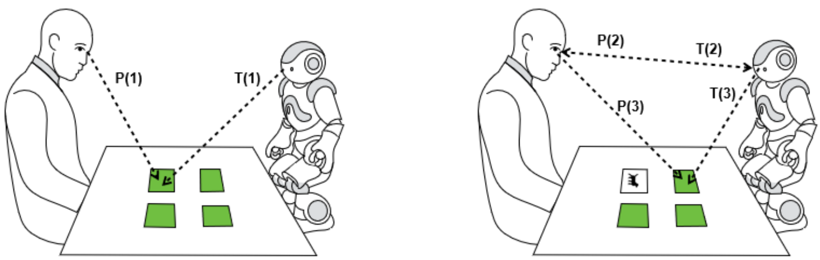
- (a) Do participants perceive and interpret the gaze-based hints (i.e. tutor pointing with the eyes to a particular card) exhibited by the tutor while performing the card-matching task?
    - i. Do participants notice the tutors' gaze behavior during the game?
    - ii. Do participants attribute the tutor the intent to help with the gaze?
- (b) Does the gaze behavior influence the choices of players in the card game?

RQ2 — Who is a better tutor: the human or the robot tutor?

- (a) Is there any difference in the participants' awareness of the tutors' behavior between the two conditions?
- (b) Is there any difference in task execution between the two conditions?
    - i. Is there any difference in the time required to complete the game?
    - ii. Is there any difference in the number of tries?
- (c) Is there any difference in participants' judgement about the tutor (e.g., likeability, perceived presence) between the two conditions?

To answer the above questions, we designed an experimental study to evaluate the effects of gaze hints in educational gameplay. We asked participants to play a matching card game with a human or a robotic tutor.

As depicted in Figure 3-1, we expected that the participants would notice the tutors' gaze while the tutor was looking at different cards on the board and follow the tutors' lead to the matching card. Thus, we hypothesized that gaze cues (facial orientation and gaze direction) from the tutor would draw the participants' attention to the matching card and, subsequently, influence their choices.



(a) (b)

Figure 3-1 The interaction flow: (a) P(1) - participant (left) turns over a card T(1) - tutor (right) looks at the selected card (b) T(2) - tutor gazes at the participant to draw attention P(2) - participant looks at the tutor T(3) - tutor moves on to look at a matching card P(3) - finally, the participant follows the tutor's gaze and looks to the matching card.

3.2. Experiment Design

In the following sections, we describe an experimental study investigating the influence of different gaze-based tutoring styles, as deployed by either a human or a robot tutor during a game tutoring activity, on participant's performance, gaze behavior and judgments of the tutor (presence and likeability). Figure 3-1 describes the interaction flow of the designed study.

3.2.1. Participants

The 20 participants who took part in the study (11 males and 9 females, between 19 and 33 years) were students recruited from TU/e university. The participants were from different cultural backgrounds, namely, Europe, Asia, and Africa. The game sessions took approximately thirty minutes. Participants received a coupon worth ten euro at the end of the experiment for their participation.

Ethics statement: Written consent was acquired from each participant prior to the experimental sessions. This was a non-clinical study without any harming procedure and all data were collected anonymously. Therefore, according to the Netherlands Code of Conduct for Scientific Practice (Principle 1.2 on page 5), ethical approval was not sought for the execution of this study.

3.2.2. Study Conditions

The study followed a two-by-two (Tutor_Type: Human or Robot) and (Tutoring_Style: Help vs No_Help) mixed factorial design experiment. The Tutor_Type variable was manipulated using a between-participants design and the Tutoring_Style variable was of a within-participants design. The two types of tutoring styles deployed by the tutor are as described below:

Help (presence of gaze cues): In this condition, the tutor (human or robot) provided gaze cues during the game. While introducing the game, the tutor informed the participant that they would help them. However, the tutor did not explicitly reveal the modality they would use to help. The tutor remained silent for the entire session. The tutor first looked at the card picked by the participant, then looked to the player, and, finally, to the matching card to attract and draw the participant's attention to the matching card.

No_Help (absence of gaze cues): In this condition, the tutor (human or robot) did not provide gaze cues during the game. The tutor only looked toward the participant and remained silent during the entire duration of the game.

Participants were assigned evenly to play in the presence of either the robot or the human tutor. Each participant interacted with the tutor in both the Help and No_Help conditions and the order of conditions was counterbalanced across trials.

Practical issues determined a between-subjects design between the two tutors due to availability and convenience for participants, who were requested to attend only one session. It was challenging to change the setting from one tutor to the other during the session. Therefore, one group was sent to the setting with the robot tutor and the other to the setting with a human tutor.

It is also important to note that we did not measure participants previous knowledge of the game or other individual variables such as the familiarity or attitude towards the robot, or psychological features like extroversion/introversion. Therefore, these individual variables can be considered confounding variables and they affect the strength of the main effect between the independent variables (Tutor_Type; Tutoring_Style) and the dependent variables: task execution, gaze, subjective experiences, and perceptions.

3.2.3. **Hypothesis**

We formulated the following hypotheses, as outlined below, regarding how the tutoring style (presence/absence of help gaze cues) and the type of tutor (human/robot) would affect participants' performance of the task and interactive behavior, and how the participants' gaze would differ between interactions with the human and the robot tutor.

(H1) The tutoring style (Help/No_Help) will influence task performance

H1.1: Participants will complete the task in less time in the Help condition than in the No_Help condition

H1.2: Participants will complete the task with fewer tries in the Help condition than in the No_Help condition

(H2) The type of tutor (human/robot) will influence task performance

H2.1: The type of tutor will influence the time of task completion

H2.2: The type of tutor will influence the number of tries

(H3) The tutoring style (Help/No_Help) will influence participants' gaze-based interaction during play

H3.1: Participants will look more into the tutors' face in the Help condition than in the No_Help condition

(H4) The type of tutor (human/robot) will influence participants' gaze-based interaction during play

H4.1: Participants will look more into the robot tutor's face than in the human tutor's face

H4.2: Participants will notice the gaze cues from the robot tutor more than from the human tutor.

(H5) The tutoring style (Help/No_Help) will influence participants perceptions

H5.1: Participants will evaluate the tutor more positively (perceived likeability and presence), during Help condition than in the No_Help condition

(H6) The type of tutor (human/robot) will influence participants perceptions

H6.1: Participants will evaluate the robot tutor more positively (perceived likeability and presence) compared to the human tutor

3.2.4. **Measures**

To evaluate the hypotheses detailed in 3.2.3, we employed the following measures:

Task performance: We identified two primary objective measures that are notably used to measure performance in memory games:

(a) Duration: the time it took participants to find all pairs of matching cards on the table.
(b) Number of tries: the total number of attempts required to find all the matching cards. A "try" consisted of choosing two cards.

Gaze behavior: Gaze direction was provided by the eye-tracking log. During the game, the participants were fitted with SMI eye-tracking glasses to capture their gaze direction. Eye-tracking provides data with high temporal resolution and can reveal detailed patterns of gaze behavior. We focused on the following eye-tracking metrics (*see definitions in* Table 2-1).

(a) Fixation_count: refers to the number of fixations within an area of interest (AOI_Face).

(b) Fixation_time: refers to the total fixation time within an area of interest (AOI_Face).

(c) Dwell_time: refers to the time spent looking at the area of interest, calculated by summing up the time the gaze coordinates were within the area of interest (AOI_Face).

(d) Revisits: The number of revisits provides information about how often participants returned their gaze to an area of interest (AOI_Face).

(e) Glances count: number of glances that occur during an interval of time, and specific to a target Area of Interest (AOI_Face).

Participants' perceptions: We used a questionnaire to evaluate participants' perceptions of the tutors' behavior, particularly perceived likeability, perceived presence of tutors, and judgments about the task. We included open-ended questions in the questionnaire asking the subjects to list the cues they observed or searched for in the tutors' behavior. In the end, we conducted semi-structured interviews to assess whether participants perceived any differences between the two conditions of the game *(see* Appendix B *for the questionnaire and a list of the post-experiment interview questions used to evaluate participants' perceptions).*

3.3. Methods

3.3.1. **Board Game Scenario**

The game used in this experiment was the card-matching game "Memory." The game was played with fourteen cards (seven matching pairs) with images of black dogs that varied slightly in shape. The dogs being the same color and not having much variance in shape increased the difficulty level of the game, as the number of pairs was relatively low. On the table, 14 cards were arranged in a rectangular layout.

The layout had six columns and three rows for a total of 18 cards or nine pairs. We placed the cards in the first two rows and two cards in the middle positions of the third row (Figure 3-2). This arrangement was informed by our preliminary experiment (*see* Appendix A *for a complete description of the board game design)*, where we examined whether the participants could accurately perceive the gaze direction of the NAO robot and the resolution needed for the head angles of the robot to direct at different card locations on the board.

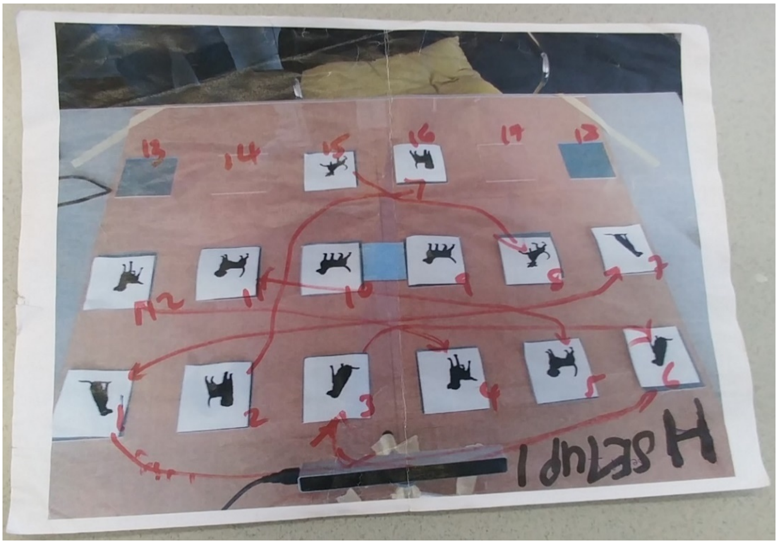


Figure 3-2 Board layout design: card arrangement configurations

The game began with the cards laid down on the board. The player then selected a card and tried to find a matching card. If the cards turned face up were similar (a pair of matching cards), the player took the couple and continued to match the cards; otherwise, the participant turned the cards face down and made a new try/move. The goal was to find all pairs in the smallest number of tries/attempts and in the shortest time possible. An attempt (try) consisted of choosing two cards. The game ended when the participant found all the matching pairs. Although the game is better and more enjoyable when two players play against each other, a single player could also play it. In this experiment, each participant played alone in the presence of either a human or a robot tutor.

3.3.2. **Experimental Setup**

Human Tutoring Setup

Figure 3-3 (a) depicts the human–human set-up. The tutor and the participant sat across from each other at the table, approximately 160 cm apart. The human tutor followed a pre-defined protocol of steps that detailed the rules of how to introduce the game and the sequence of how to shift her gaze during the game. Figure 3-4 captures the sequence of the tutors' gaze (human and robot tutor) from the eye-tracking videos in the Help condition. The tutor first looked at the chosen card, then looked to the face of the participant, and then looked to the matching card. This sequence of shifts in gaze direction was consistent for both tutors (human or robot) in the Help condition.

The same person acted as the tutor for all the participants and during all sessions. In front of the tutor was a printed photo of the card locations on the board layout. After each session with a participant, the cards were re-arranged. In this set-up, we logged the gaze of both the tutor and the participant. The Eye Tribe gaze direction tracker registered the gaze data of the tutor while the participants wore SMI eye-tracking glasses to capture their eye gaze behavior. The experiment was recorded using I-view ETG software and with a video camera. The reason for recording the gaze of the tutor was to examine if the tutor's gaze was consistent in all sessions and if varying behavior on the part of the tutor influenced the participants' gaze data in any way.

(a)

(b)

Figure 3-3 The experiment set-up: (a) human tutoring set-up and (b) the robot tutoring set-up

Robot Tutoring Setup

The human-robot setup included a humanoid robot, NAO, developed by Softbank robotics (Softbank, 2013), a personal computer, a webcam, and the same memory game (Figure 3-3; (b)). NAO is a 54 cm tall robot with a moveable head and facial features that bear resemblance to those of a child. To build the game, each card was placed in a fixed position on the board layout marked with a head pitch and yaw angle on the computer layout (Figure 3-2). The gaze cues were programmed such that each time a participant picked a card on the board, the robot's head angled to the position of the chosen card, then to the face of the participant (assumed at NAO's initial position), and, finally, to the location of the matching card. The cards on the board were matched using assigned card codes on the computer layout *(see* Appendix A *for a complete description of the board game design).* The design of the help condition gaze for the robot followed the concept of attention-directing gaze movement in human communication as shown in Figure 3-4.

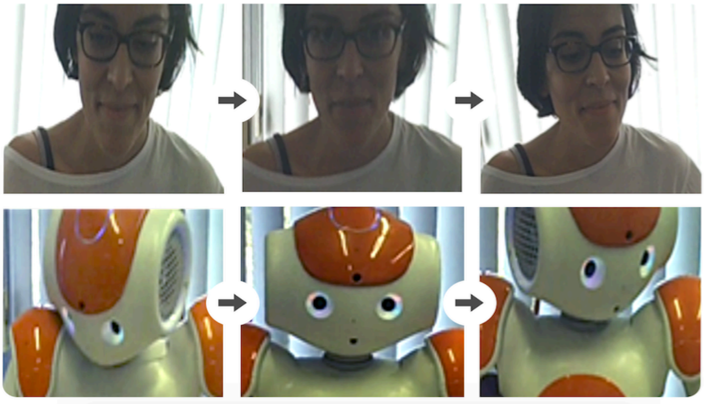


Figure 3-4 Tutors' gaze captured from eye-tracking videos: The tutor looks at the chosen card, looks to the participant, and then looks to the matching card.

3.3.3. **Procedure**

On receiving informed consent, the experimenter verbally provided the participants with details regarding the task and directions on how to play the game, including the use of the SMI eye-tracking glasses. At this point, the experimenter made sure not to disclose the study's objective. The SMI eye-tracking glasses were fitted on the participants. The experimenter then performed a calibration procedure for the eye-tracking system.

The tutor's (human or robot) instructions at the beginning of the game session Help_Condition:

> *"Hello, welcome to the game session. My name is Maka* ***(replaced with the name of the human tutor)*** *I am your tutor. I have a task for you. You are going to find pairs of matching cards on the table. You flip a first card and then flip a second card. If they do not match, turn both back and start over. If they match, leave them turned up. I am your tutor and I know the positions of all the cards on the table. I am going to help you! Please go ahead and flip the first card!"*

Notice that while introducing the game in the Help condition, the tutor (human or robot) informed the participants they would help him/her without revealing the modality it would use to help.

During the No_Help Condition, the tutors omitted the following line during the instructions:

> ***"I know the positions of all the cards on the table. I am going to help you!"***

The participants followed the tutors' instructions to complete the game. Once the game ended, the participants completed a post-experiment questionnaire *(see* Appendix B*)*. After each session, the study paused for around four minutes and the participants waited outside the room for the experimenter to rearrange the game for the second session. The same procedure was then repeated for the other condition. After interacting with the tutor in both conditions (the Help and No_Help conditions), the experimenter collected the demographic details and interviewed the participants to get more information on any differences they may have observed in the tutors' behavior between the two conditions.

3.4. Results

3.4.1. **Task Performance**

We analyzed the results of 20 participants (10 for the robot tutor condition and 10 for the human tutor condition, for a total of 20 tries in the Help condition and 20 tries in the No_Help condition).

We conducted a mixed-model ANOVA in SPSS, with the repeated measure Tutoring_Style (Help vs. No_Help) as the within-subjects factor and Tutor_Type (human or robot) as the between-subjects factor. We analyzed the effect of Tutoring_Style and Tutor_Type on the following two performance measures:

(a) Duration: We obtained the duration from video recordings, this being the period between the participant starting to play the game and completing it.

(b) Number of tries: We counted the number of tries/attempts that participants used from our video recordings.

The paragraphs below report on the results from performance, gaze behavior, and subjective measures. We first describe results from the performance measures to provide background information for the eye-tracking and subjective measures.

Table 3-1;Table 3-2 and Figure 3-5 provide results of the defined performance measures.

Table 3-1 Performance measures: duration (s) and number of tries

Tutor_Type	Help		No_Help		N
	Mean	SD	Mean	SD	
Durations (s)					
Human	118.00	38.31	118.70	45.02	10
Robot	124.40	27.77	145.10	83.52	10
Average	121.20	32.73	131.90	66.69	20
Number of tries					
Human	16.20	5.12	19.00	5.12	10
Robot	11.30	3.06	17.00	5.93	10
Average	13.75	4.81	18.00	5.49	20

Effect of Tutor_Type on Performance

Tests of between-subject effects showed a significant multivariate effect of Tutor_Type on performance measures (p=0.048*). There was no significant main effect of Tutor_Type on duration (F (1, 18) =0.913, p=0.352). However, we found a significant main effect of Tutor_Type on the number of tries (F (1, 18) =5.253, p=0.034*).

Effect of Tutoring_Style on Performance

Tests of within-subject effects showed a significant multivariate effect of Tutoring_Style (p=0.003*). There were no significant main effects on the duration between the Help and No_Help conditions (F (1, 18) =0.428, p=0.521). However, there was a significant main effect of Tutoring_Style on the number of tries (F (1, 18) =7.009, p=0.016*).

We found no significant Tutoring_Style by Tutor_Type interaction effects (p=0.654). Similarly, there was no significant Tutoring_Style by Tutor_Type interaction effects on the duration (F (1, 18) =0.374, p=0.549) and number of tries (F (1, 18) =816, p=0.378).

Table 3-2 Effect of Tutoring_Style at different levels of Tutor_Type

Tutor_Type	Tutoring_Style		MD	SE	Sig.
Duration (s)					
Human	Help	No_Help	-.7	23.13	0.976
Robot	Help	No_Help	-20.7	23.13	0.383
Number of tries					
Human	Help	No_Help	-2.8	2.27	0.233
Robot	Help	No_Help	-5.7	2.27	0.022*

*statistical significance, $p < 0.05$

Effect of Tutoring_Style on Performance at Different Levels of Tutor_Type

Duration: For the human condition, there was no significant difference in duration between the Help and No_Help conditions (p=0.976). Similarly, the mean difference in duration between the Help and No_Help conditions was not significant for the robot condition (p=0.383). (Table 3-2)

Number of tries: For the human condition, there was no significant difference in the number of tries between the Help and No_Help (p=0.233) conditions. However, for the robot condition, the mean difference in the number of tries between the Help and No_Help conditions (MD=5.7) was significant (p=0.022*), indicating participants performed significantly better with fewer tries with help from the robot tutor than without help.

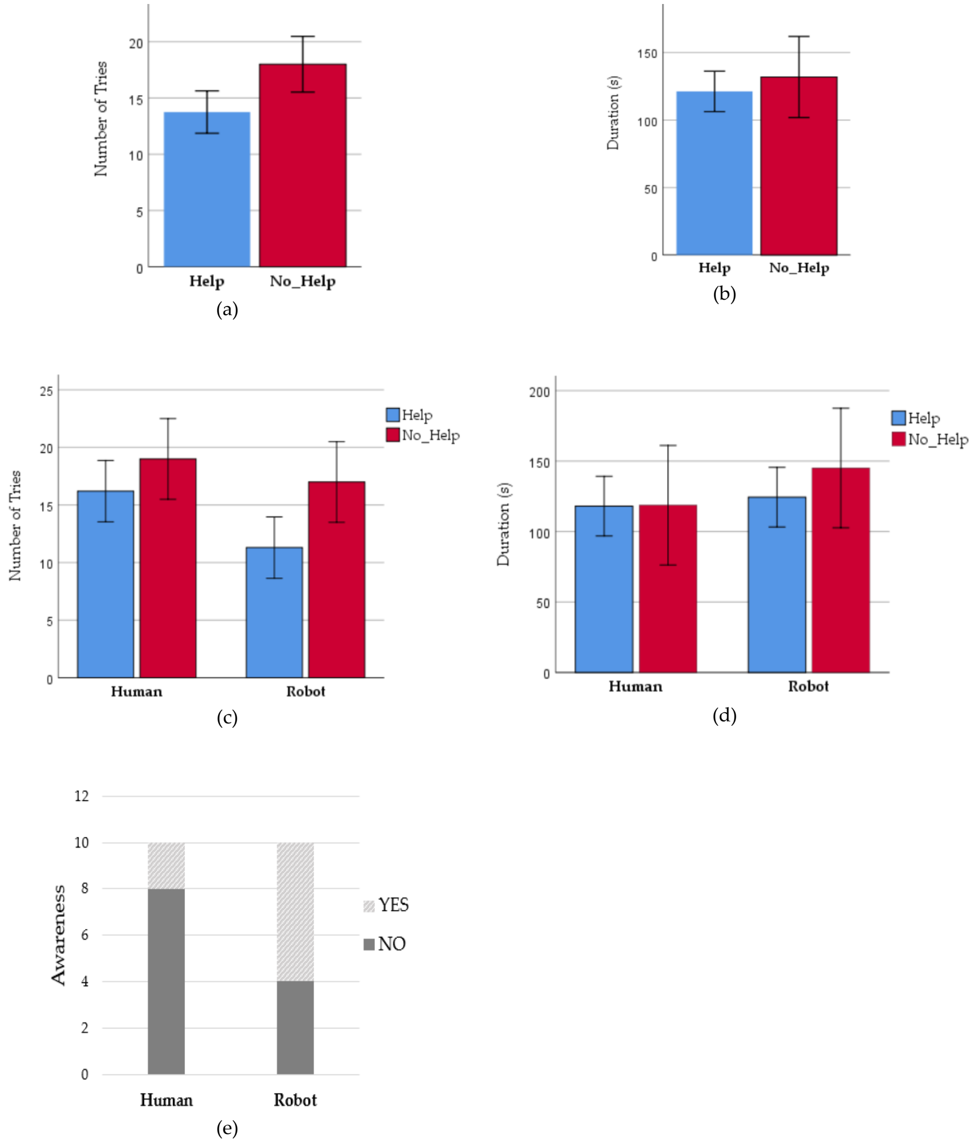


Figure 3-5 (a) Number of tries with and without help (b) the duration (s) it took participants to find matching cards with and without help (c) the number of tries with and without the help for the two tutors (d) the duration(s) it took to match all the cards with and without help from the two tutors (e) the number of participants who noticed the gaze cues during the Help condition for the two tutors.

Comparisons Between the Robot and the Human Conditions

Comparing both groups on duration, we found no significant differences in duration between the (Human condition *M*=118.00s *SD*=38.31) and (Robot condition (*M*=124.40s *SD*=27.77) (*p*=0.674, two-tailed) assuming equal variances during Help condition .Similarly there was no significant differences in duration between the (Human condition *M*=118.70s *SD*=45.02) and (Robot condition *M*=145.10s *SD*=83.52) (*p*=0.391, two-tailed) during the No_Help conditions assuming equal variances. We found significant differences in number of tries between the (Human condition *M*=16.20 *SD*=5.11) and (Robot condition *M*= 11.30 *SD*= 3.06) (*p*=0.018*, two-tailed) assuming equal variances during the Help tutoring style. However, there were no significant differences in the number of tries between (Human condition M= 19 SD=5.12) and (Robot condition *M*=17 SD=5.93) during the No_Help tutoring style (*p*=0.430, two-tailed), assuming equal variance.

Awareness of Tutor's Gaze Cues and Performance

From the post-experiment questionnaire, eight out of the twenty participants in both tutor conditions said they noticed help gaze cues from the tutor. Six out of the ten participants in the robot condition said they recognized gaze cues from the robot tutor while four did not see the help gaze cues .Two out of the ten participants in the human condition noticed help gaze cues from the tutor while the rest eight said they did not notice the gaze cues.

Table 3-3 shows the descriptive statistics regarding the duration(s) and number of tries alongside the participants' gaze awareness for both tutors.

Table 3-3 Awareness of tutor's gaze cues and performance

	Noticed Gaze Hints (YES group) N=8		Did Not Notice Gaze Hints (NO group) N=12	
	Mean	SD	Mean	SD
Duration (s)	122.13	31.62	120.58	34.83
Number of tries	9.75	1.98	16.42	4.25

We conducted an independent sample T-test using SPSS to compare the duration(s) and number of tries for participants who noticed gaze hints (reported as YES) and those who did not notice gaze hints (reported as NO) .There was no significant difference in the duration (s) —the time taken to complete the task — between those who noticed gaze (YES group *M*= 122.13 s, *SD*=31.62) and those who did not notice the gaze hints (NO group *M*= 120.58s, *SD*=34.83): *p*= 0.921 two-tailed, assuming equal variances). There was a significant difference in the number of tries between participants who noticed the gaze hints (YES group *M*=9.75, *SD*=1.98) and those who did not notice the gaze hints ((NO group *M*=16.42, *SD*=4.25): *p*=0.001*; two-tailed, assuming equal variances.)

Table 3-4 shows the descriptive statistics regarding the duration(s) and number of tries alongside the participants' gaze awareness for the human condition. Two out of the ten in the human condition noticed help gaze cues from the tutor while the rest eight said they did not notice the gaze cues

Table 3-4 Awareness of tutor's gaze cues and performance (duration(s) and number of tries) for the human condition

	Noticed Gaze Hints (YES group) N=2		Did Not Notice Gaze Hints (NO group) N=8	
	Mean	SD	Mean	SD
Duration (s)	111.50	2.12	119.63	43.25
Number of tries	11.50	3.54	17.38	4.89

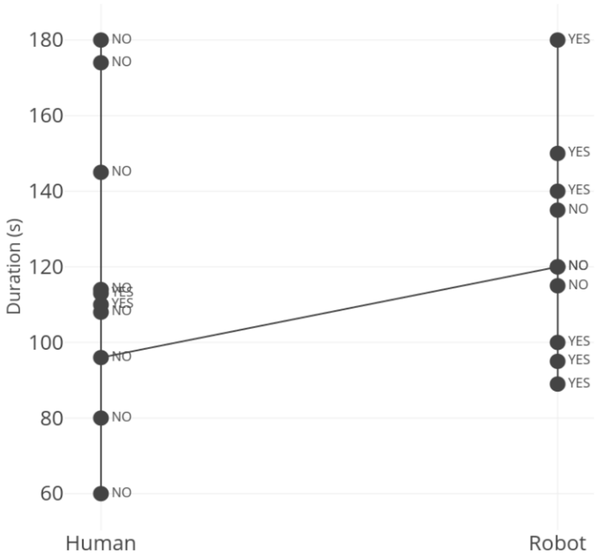
We conducted an independent sample T-test using SPSS to compare the duration(s) and the number of tries for participants who noticed gaze hints (reported as YES) and those who did not notice gaze hints (reported as NO) for the human condition. There was no significant difference in the duration (s) — the time taken to complete the task — between those who noticed gaze (YES group *M*= 111.50 s, *SD*=2.12) and those who did not notice the gaze hints (NO group *M*= 119.63s, *SD*=43.25): *p*=0.806 two-tailed, assuming equal variances). Similarly ,there was no significant difference in the number of tries between the participants who noticed the gaze hints (YES group *M*=11.50, *SD*=3.54) and those who did not notice the gaze hints ((NO group *M*=17.38, *SD*=4.89): *p*=0.156; two-tailed, assuming equal variances) in the human condition.

Table 3-5 shows the descriptive statistics regarding the duration(s) and number of tries alongside the participants' gaze awareness for the robot condition. Six out of the ten participants in the robot condition said they recognized gaze cues from the robot tutor while four did not see the help gaze cues.

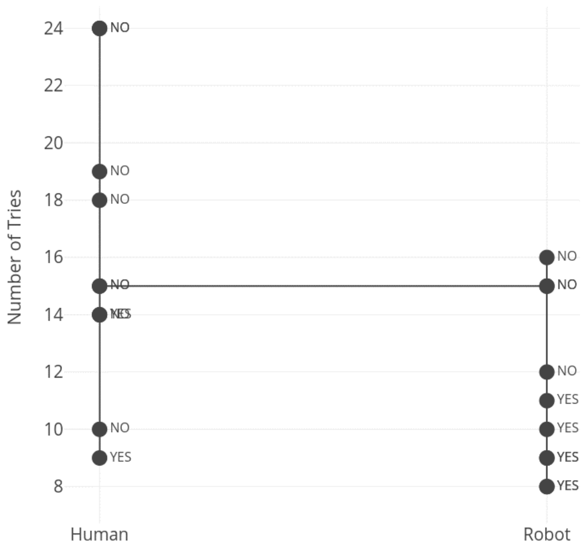
Table 3-5 Awareness of tutor's gaze cues and performance (duration(s) and number of tries) for the robot condition

	Noticed Gaze Hints (YES group) N=6		Did Not Notice Gaze Hints (NO group) N=4	
	Mean	SD	Mean	SD
Duration (s)	125.67	36.59	122.50	8.66
Number of tries	9.17	1.17	14.50	1.73

Similarly, we conducted an independent sample T-test using SPSS to compare the duration(s) and the number of tries for participants who noticed gaze hints (reported as YES) and those who did not notice gaze hints (reported as NO) for the robot condition. We found no significant difference in the duration (s) — the time taken to complete the task — between participants who noticed gaze (YES group M= 125.67 s, *SD*=36.59) and those who did not notice the gaze hints (NO group *M*= 122.50s, *SD*=8.66): *p*=0.872 two-tailed, assuming equal variances). There was a significant difference in the number of tries between those participants who noticed the gaze hints (YES group *M*=9.17, *SD*=1.17) and those who did not notice the gaze hints ((NO group *M*=14.50, *SD*=1.73): *p*=0.001*; two-tailed, assuming equal variances). Figure 3-6 shows how individuals who did not recognize the tutor was helping performed compared with participants who recognized the gaze cues. Figure 3-6 shows the duration (a) and number of attempts (b) in the help condition with text representing participants' awareness of the gaze cues (YES or NO) for both the human and robot tutors.



(a)



(b)

Figure 3-6 Awareness of tutor's gaze cues and performance (a) duration(s) and (b) number of tries

3.4.2. Gaze

In this subsection, we analyze the gaze behavior when the participants interacted with the human and the robot tutors. To analyze the recorded gaze data from the video, we used Begaze software to create custom trials of the video recordings, which included only the segment from the participant starting to play the game until game completion. From the trial images, we selected the face of the robot and that of the human tutor as the area of interest (AOI_Face), as shown in Figure 3-7 . We analyzed the results of 19 participants (10 for the robot tutor condition and 9 for the human tutor condition). There was a technical error in gaze recording for one participant in human condition during the No_Help condition.

(a) (b)

Figure 3-7 Area Of Interest (AOI) region: face of the (a) human tutor and (b) the robot tutor

We exported the metrics related to the trials and area of interest (AOI_Face for this case) to SPSS software. We conducted a mixed model analysis of variance in SPSS with the repeated measure Tutoring_Style (Help vs. No_Help) as the within-subjects factor and the Tutor_Type (human or robot) as the between-subjects factor. We analyzed the effect of Tutoring_Style and Tutor_Type on the following eye gaze measures: Fixation_count, Fixation_time, Dwell_time, Revisits and Glances *(see the definitions for these metrics in* Chapter 2; Table 2-1*).*

Effects of Tutor_Type on Eye - Gaze Measures

We found significant effects of the Tutor_Type on eye-gaze measures (p=0.018*). We found significant effects of Tutor_Type on Fixation_count (F (1, 17) =6.594, p=0.020*), Revisits (F (1, 17) =10.487, p=0.005*), Glances (F (1, 17) =12.066, p=0.033*) and dwell time (F (1, 17) =4.421, p=0.051). However, there was no significant effect of Tutor_Type on Fixation_time (F (1, 17) =4.112, p=0.059).

Table 3-6 Eye-gaze measures

Tutor_Type	Help		No_Help		N
	Mean	SD	Mean	SD	
Fixation_count					
Human	4.67	5.83	2.89	4.91	9
Robot	19.80	21.40	5.60	6.19	10
Average	12.63	17.44	4.32	5.64	19
Fixation_time					
Human	1097.97	1415.01	558.27	856.05	9
Robot	4205.72	5184.73	1001.52	1056.75	10
Average	2733.63	4107.58	791.56	967.35	19
Dwell_time					
Human	1245.78	1602.45	687.54	1044.21	9
Robot	5137.19	6309.69	1214.49	1330.52	10
Average	3293.89	5003.24	964.88	1201.18	19
Revisits					
Human	1.33	2.18	0.67	1.66	9
Robot	8.70	8.86	2.30	3.59	10
Average	5.21	7.46	1.53	2.89	19
Glances					
Human	2.11	2.37	1.00	2.00	9
Robot	9.60	8.97	3.20	3.68	10
Average	6.05	7.58	2.16	3.13	19

Effect of Tutoring_Style on Eye -Gaze Measures

There was no significant effect of Tutoring_Style on the eye gaze measures : Fixation_count (F (1, 17) =3.444, p=0.081) ; Fixation_time (F (1, 17) =3.753, p=0.070) ; Dwell_time (F (1, 17) =3.582, p=0.076); Revisits (F (1, 17) =3.520, p=0.078) and Glances (F (1, 17) =3.726, p=0.070).

Effect of Tutoring_Style on Gaze at Different Levels of Tutor_Type

Pairwise comparisons on the effect of Tutoring_Style on eye gaze measures at different levels of the Tutor_Type found significant differences between the Help and the No_Help conditions for Fixation_count (p=0.028*), Fixation_time (p=0.028*) and Dwell time (p=0.028*) for the robot condition. There were also significant differences for Revisits (p=0.024*) and Glances (p=0.029*) between the Help and the No_Help condition for the robot condition. For the human condition, there were no significant differences in Fixation_count (*p=0.779*), Fixation_time (p=0.705) and Dwell_time (p=0.749) between the Help and the No_Help conditions. There were also no significant differences between the Help and the No_Help condition for Revisits (p=0.810) and Glances (p=0.699) in the human condition.

Comparisons between the Robot and the Human Tutor on Eye Gaze Measures

Pairwise comparisons based on the estimated marginal means between the two tutors on the eye gaze measures found significant differences during Help for Revisits (p=0.027*) and Glances (p=0.027*) between the human and robot condition. There were significantly more revisits and more glances at the robot tutor's face than the human tutor's face. However, there were no significant differences in Fixation_count (p=0.056), Fixation_time (p=0.101), and Dwell_time (p=0.091) between the Human condition and Robot condition during the Help conditions. We found no significant differences between the human and the robot conditions during the No_Help conditions on all eye gaze measures : Fixation_count (p=0.309); Fixation_time (p=0.333); Dwell_time (p=0.354); Revisits (p=0.229) and Glances (p=0.130).

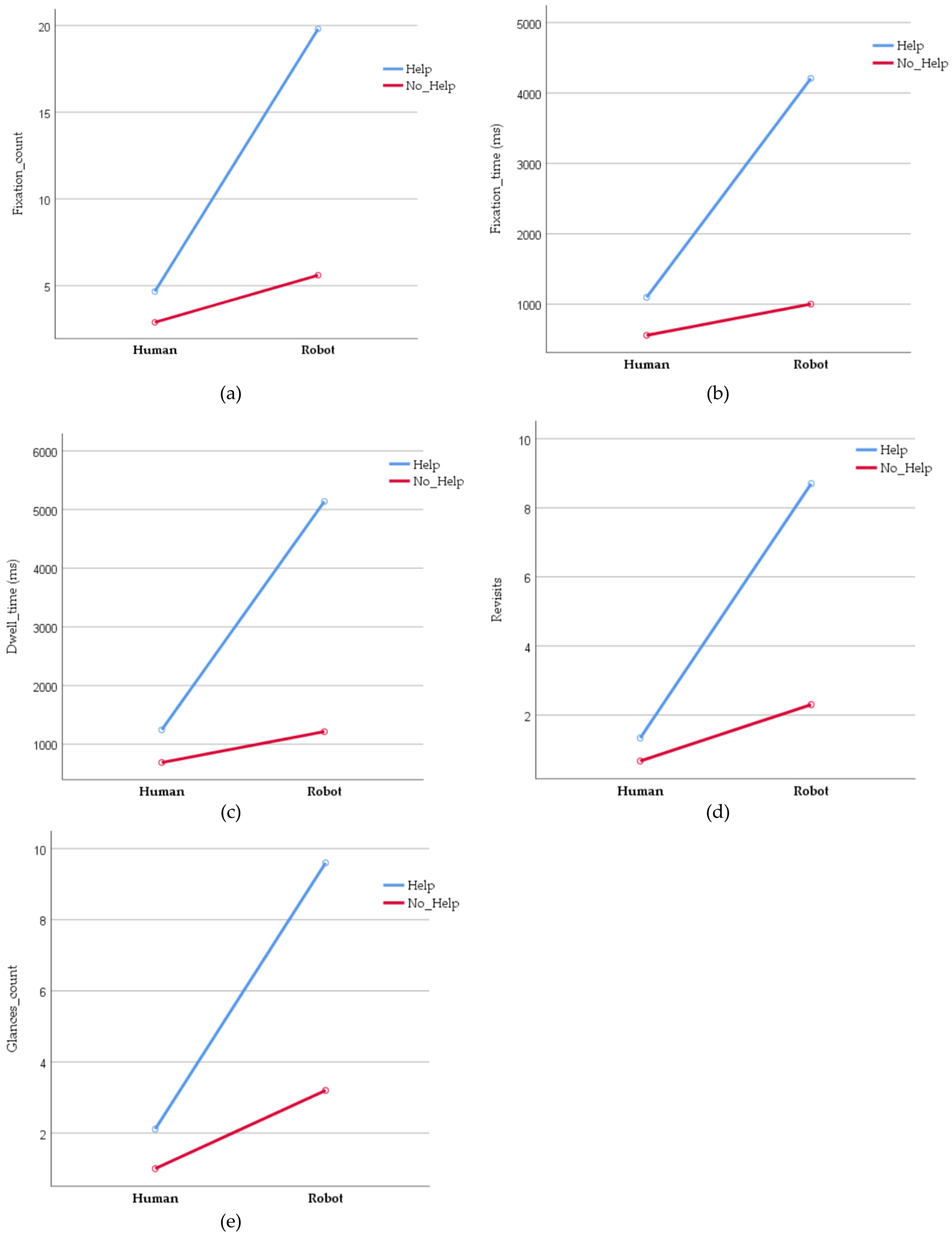


Figure 3-8 Comparing the effect of Tutoring_Style on eye-gaze measures at different levels of Tutor_Type

(a) Fixation_count (b) Fixation_time (c) Dwell_time (d) Revisits (e) Glances

3.4.3. **Participants' Perceptions**

For the subjective measures, we conducted a mixed model analysis of variance in SPSS with the repeated measure Tutoring_Style (Help vs. No_Help) as the within-subjects factor and the Tutor_Type (human or robot) as the between-subjects factor. We analyzed the results of 20 participants (10 for the robot tutor condition and 10 for the human tutor condition, for a total of 20 tries in the Help condition and 20 tries in the No_Help condition). Regarding the aspect of tutors' likeability and presence, participants rated the tutor on a five-point Likert scale: To facilitate the analysis of the Likert scale data, we coded the data as follows: 1: completely disagree; 2: disagree; 3: neutral; 4: agree, and 5: completely agree. We also included the 'not applicable' option, and these data were treated as missing values during the analysis. Figure 3-9 summarizes the results from our subjective measures, particularly the tutor's perceived likeability and presence.

Presence

A test of the within-subjects effect shows a significant multivariate effect of Tutoring_Style on presence measures (p=0.003*). There were significant main effects of Tutoring_Style on tutor's presence (F (1, 18) = 15.059, p=0.001*); 'tutor caught the participants' attention' (F (1, 18) = 9.529, p=0.006*); and tutor was attentive (F (1, 18) = 6.600, p=0.019*). A test of the between-subjects effect shows a significant multivariate effect of Tutor_Type on presence measures (p=0.001*). There was a significant main effect of Tutor_Type on the 'tutor caught the participant's attention' measure (F (1, 18) = 5.921, p=0.026*). However, the main effects of Tutor_Type on tutor presence (F (1, 18) = 3.009, p=0.100) and 'tutor was attentive' (F (*1*, 18) = 1.638, p=0.217) were not significant.

We found no significant multivariate interaction (Tutor_Type * Tutoring_Style) effects on presence measures (p=0.090). We found significant interaction effects on tutors' perceived presence (F (1, 18) = 4.366, p=0.051*) and 'tutors' behavior caught the attention of the participant' (F (1, 18) = 4.235, p=0.054*) but the interactions effects on tutors' perceived attentiveness (F (1, 18) = 0.492, p=0.491) were not significant.

Pairwise comparisons on the effect of Tutoring_Style on-presence measures at different Tutor_Type levels show that participants rated the robot tutor as more socially present (p=0.001*) and more attentive (p=0.033*) in the Help condition than in the No_Help condition. Moreover, they indicated that the robot tutor's behavior caught their attention more during the help condition than when there was no help (p=0.002*). However, we found no significant difference in how the participant rated the human tutor across all presence measures between the Help and No_Help conditions.

Pairwise comparisons based on the estimated marginal means between the two tutors on the presence measures show the robot tutor's behavior caught the attention of the participants significantly more compared to the human tutor (*p*=0.026*). However, there was no significant mean difference in how participants rated both tutors for the other two measures.

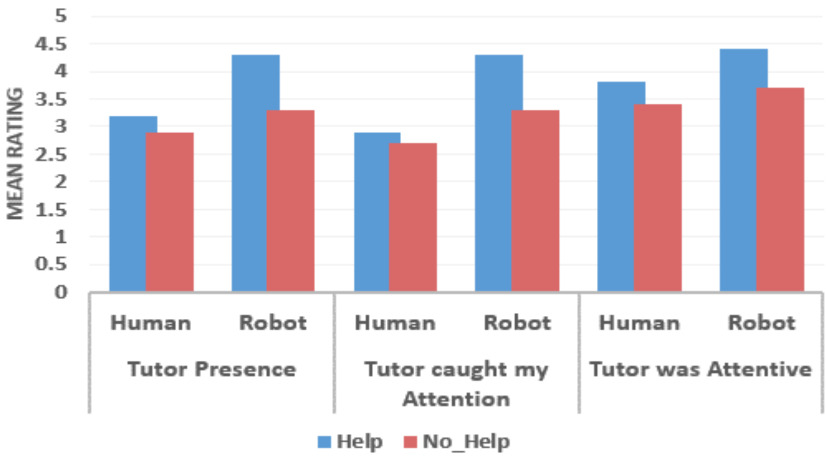
Likeability

A test of within subject effects found significant main effects of Tutoring_Style on the pleasant measure (F (1, 18) = 7.309, *p*=0.015*). However, there was no significant effect of Tutoring_Style on the participants' rating of tutors' kindness (F (1, 18) = 1.246, *p*=0.279) and friendliness (F (1, 18) = 2.000, *p*=0.174). There were no significant main effects of Tutor_Type on any of the likeability variables; tutors' kindness (*p*=0.167), pleasant (*p=0.*643) and friendliness (*p*=0.180).

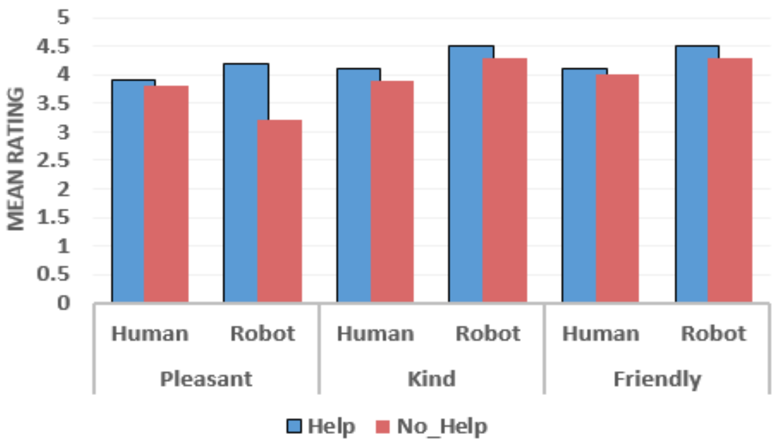
We found no significant multivariate interaction (Tutor_Type * Tutoring_Style) effects on likeability measures (*p*=0.390). We found significant interaction effects on the pleasant measure (F (1, 18) = 4.893, *p*=0.040*). However, there were no significant interaction effects on other measures; tutor's kindness (*p*=0.714) and tutor's friendliness (*p*=1.000).

Pairwise comparisons on the effect of Tutoring_Style on likeability measures at different levels of the Tutor_Type show that the participants rated the robot tutor as more socially pleasant (*p*=0.005*) in the Help condition than in the No_Help condition. However, there was no significant difference for robot tutor's kindness or friendliness ratings of the tutor between the two Tutoring styles. On the other hand, there was no significant difference on the rating of the human tutor in both the Help and No_Help conditions, across all the evaluated likeability measures.

Pairwise comparisons based on the estimated marginal means between the two tutors on the likeability measures found no significant difference regarding how participants rated both tutors for all likeability measures during both Help and No_Help conditions used.



(a)



(b)

1: completely disagree; 2: disagree; 3: neutral; 4: agree, and 5: completely agree.

Figure 3-9 Comparing the effect of Tutoring_Style on (a) presence measures and (b) likeability at different levels of the Tutor_Type

3.4.4. **Post-Experiment Interview**

In the end, we conducted a post-experiment interview in which we asked participants whether they noticed the helping cues from the tutor during the game and, if they did, whether that influenced their choice of cards. We also asked them what cues they expected the tutor would have used to help them in the game. Lastly, we examined how they perceived the robot's head movements, whether they regarded the movements as natural, too fast, or too slow, and if they considered the robot tutor's behavior to be automatic. *(See the list of the post-experiment interview questions at the end of* Appendix B*).*

Eight (8) participants in the Human condition said they did not notice the help, i.e., the tutors' gaze cues. Most of them indicated they focused on the game and did not look at the tutor. At least eighteen (18) of all the twenty (20) participants in both the Human and Robot conditions said they expected verbal —vocal—audio help from the tutor, as illustrated in the excerpts below:

> *"I expected verbal gaze cues from the tutor; I didn't pay attention to the tutor."*

> *"I noticed the gaze hint, but it took a while to get to the eyes. I expected speech; for example, when I get closer to the matching card, the tutor says something like warmer! Or hey, you have seen this card before. If there is no help, you are forced to remember, but now I relied on help."*

> *"I was focused on the game, so I did not look at the tutor; maybe in the future, you can put lights with different colors under the card or give some sound instructions. Audio feedback could be important".*

> *"I noticed and used the help as direction, but I prefer speech."*

Four (4) participants in the robot group did not notice the gaze cues. However, all of them reported seeing the head movements, but felt like the robot followed their moves rather than directing their attention to the matching cards as illustrated in the following excerpts:

> *"I noticed the robot was looking around, but it felt like it was following my moves, not showing me the positions. I felt as if we were both looking."*

> *"I did not get help! It seemed like he sees what I see! I was more focused on the task, and I thought he was watching what I was doing."*

> *"I was focusing on the cards and not looking at the robot. I thought the help would be vocal; the robot looked like it was moving with me, following me."*

> *"I did not feel like it was helping, just thought he was looking at the cards I was turning. I did not get the hint."*

We recorded mixed responses from participants regarding how they perceived the direction of gaze for the robot tutor.

> *"I got that the robot was helping; it was looking at the right card. Later, I felt the movements of the head were slower."*

> *"I got the tutor help—it looked at my card, then me, and then to the matching card. I just got one problem—to read the gaze direction from the angle. The head movements were pretty well-paced, not too slow, not too fast."*

> *"I got that the tutor was helping, was looking at the right card, but it was not easy to tell which one it was looking at. The robot felt natural, and the speed was nice. I felt like the robot was fully automatic."*

> *"Yes, I noticed the help, pretty natural, difficult to see the direction. I was not sure which card the robot was looking at when the cards were closer."*

3.5. Discussion

The work described in this chapter contributes to the design of gaze-based communication to improve learning performances in human-robot tutoring. The chapter presents a user study investigating how gaze-based hints from either a human or a robot tutor might influence performance (measured by the duration and number of tries for completing the task) and participants' judgments of a tutor during a card-matching game in a tutoring interaction. The study employed a two-by-two mixed design; Tutor_Type (Human or Robot) manipulated as a between-subject design and Tutoring style (Help vs. No_Help) manipulated as a within-subject design. The Help condition refers to the presence of gaze cues, and No_Help refers to the absence of the gaze hints during tutoring. In the following paragraphs, we outline our findings and discuss whether they support the formulated hypotheses in Section 3.2.3. We further give plausible explanations for our results.

In our first hypothesis *(H1)*, we projected that the tutoring style (Help/No_Help) would influence task performance measured using duration and number of tries. We found no significant difference in the duration between the two tutoring styles (Help vs. No-Help tutoring style) *(H1.1*). However, we found that participants used fewer tries to complete the task when the tutor helped with gaze hints during interaction compared to when the tutor did not help with gaze cues *(H1.2)*. We found that participants identified all the matching cards pairs with significantly fewer tries with help from the robot tutor than without help. However, there was no significant difference in the number of tries, with or without help, in the human tutor condition.

In our second hypothesis *(H2),* we predicted that the type of tutor (human/robot) would influence the participants' task performance regarding time and number of tries. The findings partially support our hypothesis: we found a significant effect of Tutor_Type on performance measures. There was no significant difference in the duration between the robot and the human tutor either during the Help or No_Help tutoring styles *(H2.1).* However, we found significant differences in the number of tries between the human and robot condition during the Help tutoring style *(H2.2).* We found no significant differences in the number of tries between the human and robot condition during the No_Help tutoring style.

In *(H3),* we predicted that the tutoring style (Help/No_Help) would influence participants' gaze-based interaction during play. There was no significant effect of tutoring style on eye gaze measures; we found no significant difference in participants' eye gaze measures during the Help condition compared to the No_Help condition *(H3.1)*. We found no significant differences in the eye gaze measures between the Help and No_Help conditions during the robot condition. We also found no significant differences in the eye gaze measures between the Help and No_Help condition during the human condition.

In *(H4),* we predicted that the type of tutor (human/robot) would influence participants' gaze-based interaction during play. The findings support our hypothesis; we found a significant impact of Tutor_Type on eye gaze measures. Results of Fixation_count, Revisits, and Glances on the AOI face of the tutor from the eye-tracking data set revealed that participants looked significantly more often at the robot tutor's face than the human tutor's face *(H4.1).* Moreover, we found substantial differences in the number of participants that noticed the robot and the human tutor's gaze cues. Twenty percent recognized the gaze cues from the human tutor, while sixty percent recognized the gaze cues from the robot tutor *(H4.2).*

In *(H5),* we projected that the tutoring style (Help/No_Help) would influence participants' perceptions of the tutor. The findings partially support our hypothesis; We found significant main effects of Tutoring_Style on presence measures: tutor's presence; 'tutor caught the participants' attention;' and the tutor was attentive. We found that the participants rated the robot tutor as more socially present; and more attentive in the Help condition than in the No_Help condition. Moreover, they indicated that the robot tutor's behavior caught their attention more during the help condition than when there was no help *(H5.1).* However, there was no significant difference in how the participant rated the human tutor across all presence measures between the Help and No_Help conditions

On likeability measures, we found that the participants rated the robot tutor as more socially pleasant in the Help condition than in the No_Help condition *(H5.1).* However, there was no significant difference in the tutor's kindness or friendliness ratings of the robot between the two Tutoring styles. On the other hand, there was no significant difference in the human tutor's rating in both the Help and No_Help conditions, across all the evaluated likeability measures.

In *(H6),* we projected that the type of tutor (human/robot) would influence participants' perceptions. The findings partially support our hypothesis: Regarding the presence measures, we found that participants evaluated the robot tutor more positively (tutor caught my attention) compared to the human tutor *(H6.1).* However, there were no significant differences in how participants rated both tutors for all likeability measures during both Help and No_Help conditions.

The significant difference between the number of participants who noticed the human gaze cues or the robot tutor might relate to diverse factors. First, the robot's novelty effect increased participants' attention, supported by the fact that participants with the robot tutor spent more time completing the task even with fewer tries. Moreover, the eye gaze data analysis indicated that participants spent more time looking at the robot; hence, they took more time to complete the task. Another explanation could be the robot's motors' sounds. While performing the gaze behavior by turning its head, it possibly attracted the participants' attention, making the robot's gaze behavior more salient than the human's.

Further, for the NAO robot, gaze behavior is achieved with large head movements, compared to the human tutor's gaze, which is subtler and is based more on the eye's movements than head motion. Therefore, we can consider that the robot's gaze behavior was overt. However, that of the human was a covert cue that affected its communicative effectiveness in assisting the players. Though covert cues can influence interaction even without one being aware of them, in this particular context and regarding the game flow, the gaze cues providing information about the matching card position have to be noticed by participants to be effective.

Moreover, the concept of intimacy regulation, where humans control their gazes to regulate the level of intimacy with their interaction partners, also applies (Argyle et al., 1994). We conjecture that participants restricted the gaze behavior according to social rules in the human condition, which was not present in the robot case. Additionally, in the human condition, the subjective reports indicated that most participants expected verbal help from the tutor, as the more natural modality of communication in the face-to-face situation, even more, provided the human tutor addressed the participant verbally at the introduction of the activity.

Results do not show significant effects on completion time (duration), even during the robot condition. There are several possible explanations for this fact. Firstly, as soon as the participant noticed the robot tutor was helping, they waited until the robot gazed at the matching card, even when they had an idea of where the matching card was. Secondly, as revealed in the interview responses, it took a while for some of the participants to read the robot's gaze direction. The difficulty in reading the robot's gaze may be attributed to the limitations of the robot used as the experimental platform for this study, as it lacks articulated eyes. The third probable reason is the duration of head motions during attention shifts from the flipped card to the participant's face and then to the matching card. Again, as shown in the qualitative responses, some participants indicated that the robot tutor's head movements were slow.

Further analysis shows that participants who reported seeing the tutor's gaze cues used fewer tries to complete the game than those who did not report noticing the gaze cues. However, there was no significant difference in duration for those who identified gaze cues in the robot condition and those who did not. For the human condition, there was no significant difference in performance between those who reported identifying the gaze cues and those who did not identify the gaze cues across both numbers of tries and duration measures. In this respect, our findings differ from Mutlu et al. 's (2009) or in collaborative scenarios by Palinko et al. (2015). These studies showed that gaze is a powerful communicative signal without explicit awareness from the observer. In the current study, the lack of awareness overrides the informative content of gazing. We believe it is because of the lack of subtlety of NAO's eyes, which do not have the impactful embodiment of human-like eyes as the robots used in the Mutlu et al. 's (2009) and Palinko et al. 's (2015) studies. We highlight the role of overt and covert gaze cues and their relevance in the learning scenarios in Chapter 6.

We recorded diverse responses from participants concerning how they perceived the direction, timing, and intent of the robot tutor's head movement (gaze). A few of the participants reported identifying the robot's head movements. However, they felt like it was following their tries rather than directing their attention to the matching card, missing the tutor's gaze informative content. These observations bring up the timing of gaze behavior and direction of head movements, highlighting two aspects of future work in Chapter 6.

Finally, we observed that participants in the robot condition appeared unperturbed; however, participants seemed a little uneasy in the human condition. One possible explanation is the unnaturalness of the human tutor's interactive behavior, shifting from casual, verbal-based communication during the introduction and the briefing phase to an absence of verbal communication during the game in both the Help and No_Help conditions.

The results from this chapter have positive implications for creating effective, human-robot gaze mechanisms to facilitate tutoring interactions. While there were no significant differences in the time taken to complete the task during the Help and No_Help conditions with the robot and human tutors, it is right to note that waiting for the robot's gaze cues increased the execution time but improved the accuracy of the selections. Therefore, for this case, the playful situation of following the smart cues of the robot should be very appealing as they provide a playful interactive situation. Furthermore, especially for a robot to be used in training social skills for children with development impairments (e.g., ASD), where the main goal is to be interactive and to take advantage of the help and interpret social cues and other's intent.

3.6. Summary

The findings in this chapter provide design recommendations that can help build effective gaze-based communication to improve performance in human-robot tutoring.

There are three significant conclusions from this chapter:

First, when a robot tutor provides gaze-based cues during a tutoring interaction, participants use fewer tries to complete the task than when a robot tutor does not give such cues. Therefore, providing gaze-based communication during learning can improve the robot's effectiveness as a tutoring agent. *Second,* participants notice the gaze cues (awareness of the tutor's intention to help) from a robot tutor more compared to a human tutor and, consequently, perform better with a robot tutor. *Third,* awareness of the tutor's intention to help (noticing the gaze hints) leads to better performance (fewer tries —increased accuracy of selections) during the tutoring game interaction.

In the future, researchers can further analyze the eye-tracking datasets to investigate finer- micro-level details of the interactive player-tutor-looking behavior, including intricate gaze patterns, to understand the differences between human-robot and human-human nonverbal communication. Such findings can further inform the design of computational gaze models in tutoring interactions.

In the next chapter, we investigate child-robot gaze mechanisms to inform the robot's behavior design as a facilitator of children's problem-solving competencies. We focus on gaze interaction concepts, including mutual gaze, gaze-following patterns, and coordinated sequences of joint attention to building capable robot tutoring agents for children.

Chapter 4. **Dyadic Gaze in Child-Robot Collaborative Tutoring Interaction**

The findings from the previous chapter have established design recommendations for how gaze-based communication can be manipulated to improve task performance during a human-robot tutoring interaction. In this chapter, we investigate how to implement gaze-based communication as an efficient help mechanism for robot-child tutoring. Specifically, we examine how children perceive and interpret gaze-based cues from robots and whether they can interpret such cues appropriately and under which conditions these social cues impact their performance and their cognitions about the robot. Beyond examining the influence of gaze cues from a robot on children's behavior, we adopt an observational approach to explore the dynamics of mutual gazing patterns and coordination of gaze direction during a tutoring game activity. Dyadic gaze behavior like mutual gaze, gaze-following, and joint attention indicate both engagement and the quality of social interaction. Therefore, examining the coordinated gaze behavior during child - robot interactions can promote more accurate interventions and, thus, further educational gains. Therefore, examining the coordinated gaze behavior during child - robot interactions can promote more accurate interventions and, thus, further educational gains.[4]

To achieve the above goal, we adopted the same setup as the one in the previous chapter to investigate the influence of the robot's gaze on children's performance and explore patterns of coordinated gaze between a child and a robot (NAO) during a card-matching game, "Memory." As in the previous chapter, children interacted with a robot tutor in a play tutoring situation in two sets. The robot tutor provided gaze clues in the first set—looking at the matching card—to assist children during the tutoring situation. In the other setting, the robot tutor did not provide such clues during the play. Consistent with previous findings, we found that children also improve their task performance when tutored by a robot, and this performance improves when the robot tutor uses more gaze cues. We found that more occurrences of mutual gaze and gaze-following patterns increase children's awareness of the robot's

[4] This chapter is largely based on the following publications (Mwangi, Barakova, Díaz, et al., 2018b; Mwangi, Barakova, Diaz, Mallofre, et al., 2017a;Mwangi et al., 2017c)

gaze hints and improve the efficacy of the robot tutor as a helping agent. These findings show that mutual coordination between the child and the robot can facilitate awareness of the tutor's intentions and, in turn, improve the performance of children during a tutoring task. The work in this chapter, therefore, demonstrates that robots can use gaze behavior effectively to enrich child-robot interaction in a tutoring context.

The rest of this chapter is structured as follows: Section 4.1 provides background on prior works in human-robot gaze interaction and describes the objectives and the research questions. Section 4.2 describes the experiment design — participants' details, study conditions, the hypothesis, and measures. Section 4.3 details the experimental setup, the platform, the interaction scenario, and the board game, "Memory," and describes the procedure followed during the experiment. Section 4.4 describes the behavioral system; Section 4.5 describes the coding process of the child-robot behaviors, and Section 4.6 shows the visualizations of the child -robot observations. Section 4.7 outlines the results of the user study; Section 4.8 discusses the main findings. Finally, Section 4.9 summarizes the main implications for designing credible and effective gaze behavior mechanisms in child-robot tutoring and education.

4.1. Background

In human social interactions, gaze facilitates mutual coordination and establishes a common ground for communication. Farroni et al. (2002) argued that gaze behavior is the most dominant mode of creating a communication link between interacting partners. This chapter focuses on gaze interaction concepts of mutual gaze, gaze-following, and joint attention, which are of substantial importance for natural human-human communication (Brooks and Meltzoff, 2005; Emery, 2000; Gernsbacher, Stevenson, Khandakar, and Goldsmith, 2008). Coordinated gaze behavior, including mutual gaze, indicates social engagement and the quality of interaction. Subsequent gaze-following shows an understanding of others' attention, and gaze alternation helps to assess joint attention. Consequently, as detailed in Chapter 2, Section 2.2, such gaze coordination is the foundation of children's learning, including the development of critical cognitive abilities such as the theory of mind and perspective-taking (Baron-Cohen et al., 1985; Brooks and Meltzoff, 2014; Brooks and Meltzoff, 2002). Therefore, examining the coordinated child -robot gaze behavior can promote more accurate robotic interventions and further educational gains.

Previous work shows that mutual gaze interaction can foster performance and promote positive experiences during human-robot interactions. For example, Kompatsiari et al. (2019) and Kompatsiari et al. (2017) suggest that people are sensitive to an artificial agent's mutual gaze and feel more engaged when a robot establishes mutual gaze. Yonezawa et al. (2007) show that when a robot companion establishes mutual gaze with human users, this leads to positive robot evaluations. A robot also seems more intentional when it provides a "responsive" gaze with the interacting partner compared to providing a fixed gaze (Yoshikawa et al., 2006). Admoni and Scassellati (2017) provide a comprehensive review of the design of robots' social gaze cues. Although a few studies have examined mutual gaze interaction in human-robot interaction research in diverse domains, the importance of dyadic gaze in child-robot collaborative tutoring and social training—autism therapies—is still an open question.

For instance, in the context of child-robot tutoring and autism therapies, research has mainly focused on whether children can read and follow the gaze cues exhibited by a robot (Mavadati et al., 2015; Robins, Dickerson, Stribling, and Dautenhahn, 2004). However, to date, the methodologies used to evaluate child-robot tutoring interactions only record the child's actions without noting the robot's concurrent behaviors. The drawback of such assessments is that they do not capture the interaction dynamics between the child and robot, which could impact the interaction outcomes. We argue that as dyadic gaze interaction involves sequences of intertwined and coordinated looking behaviors between child and robot, further investigation into these dyadic patterns is needed.

In this chapter, we developed a behavioral system to measure child-robot interaction regarding gaze behavior, engagement, and emotions. The emphasis is on coordinated and sequential gaze patterns—and the dynamics of—mutual gazing patterns between a child and a robot in a tutoring task. The aim is to identify the occurrences of mutual gaze and coordination of gaze direction patterns and assess how child-robot coordinated gaze behavior can improve the performance and participants' experiences during a tutoring game activity. Looking at these patterns, we aim to investigate whether children can read, interpret, and attribute intentions to gaze hints exhibited by a robot in a play situation.

In this regard, this chapter addresses the following questions:

RQ1 — Do tutors' social cues of gaze influence children's performance in the context of a board; To what extent?

(a) Do children perceive and interpret the gaze-based hints (i.e., the tutor pointing with the eyes to a particular card) exhibited by the tutor while performing the card-matching task?

i. Do children notice the tutors' gaze behavior during the game?

ii. Do children attribute the tutors' intention to help with the gaze?

RQ4 — Does the coordinated gaze (mutual gaze, gaze-following) influence performance?

(a) Do dyadic patterns influence children's interactive social engagement?

(b) Do dyadic patterns influence children's awareness of the robot's behavior?

(c) Do dyadic patterns influence task execution?

i. Is there any difference in the time required to complete the game?

ii. Is there any difference in the number of tries?

To examine the abovementioned questions, we used the designed board game task described in Chapter 3 in which a child plays "Memory" (a card-matching game) in the presence of a robot tutor (NAO). We propose that occurrences of appropriate sequences of the dyad's gaze behaviors (i.e., mutual gaze, gaze-following, and joint attention) would facilitate the effectiveness of the robot tutor as a helping agent and, consequently, improve the performance of the child fulfilling the task.

4.2. Experimental Design

In the following sections, we describe an experimental study investigating the influence of different gaze-based tutoring styles, as deployed by a robot during a game tutoring activity, on children's performance, gaze behavior, and children's judgments of the tutor.

4.2.1. Participants

Eighteen typically developing children (Age: 4 - 11 yrs.; Gender: Male: 10, Female: 8) took part in the study. The children were from a day-care center in the Netherlands and children of staff at the university. Most of the children were of European-Dutch and Asian backgrounds. The game sessions took approximately thirty minutes. We excluded three (3) children from further analysis for the following reasons. One of the children declined to participate in one of the sessions, and the other two were very young (below age 5) and needed a lot of help from the facilitator to play the game. All children received game gifts for their participation.

Ethics statement: This study was voluntary, and written consent was acquired from each child signed by their parents or guardians before the experimental sessions. This was a non-clinical study without any harming procedure, and all data were collected anonymously. Therefore, according to the Netherlands Code of Conduct for Scientific Practice (Principle 1.2 on page 5), ethical approval was not sought for the execution of this study.

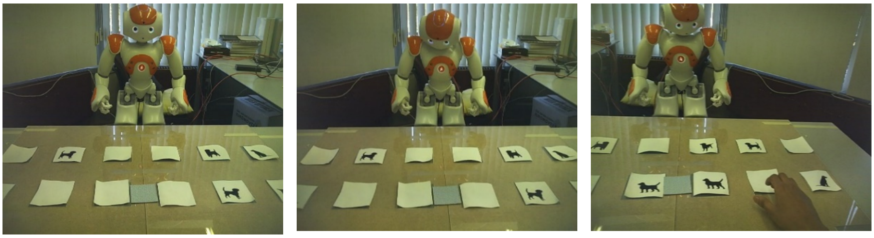


Figure 4-1 NAO gaze behavior: The NAO robot looking at different cards on the board

4.2.2. **Study Conditions**

The experiment followed a one-factor (Tutoring_Style) within-subjects design with two conditions (Help and No_ Help): The two types of tutoring styles deployed by the tutor are as described below:

Help: In this condition, the tutor provided gaze cues during the game. The robot's gaze cues were designed to point with the head — eye's direction to the matching card. The sequence of actions was similar to the previous chapter (Figure 3-1). The robot tutor would look at the card picked by the children, then look to the child's face, and then to the matching card to attract and draw the child's attention to the matching card.

No_Help: In this condition, the tutor (robot) did not provide gaze cues during the game. The tutor only looked at the child and remained silent during the entire duration of the game.

In both the Help and No_Help conditions, the tutor remained silent for the entire game duration.

4.2.3. **Hypothesis**

Based on the results of the previous study and the nonverbal theories of gaze in human-human communication (Emery, 2000; Frischen et al., 2007; Kleinke, 1986), we projected that higher levels—occurrences of mutual gaze and gaze-following patterns between the child and the robot would help draw the child's attention to the matching card, facilitate the robot tutor's effectiveness as a helping agent and, in turn, improve the performance of the child fulfilling the task.

In this regard, we formulated the following hypotheses: whether and how the help tutoring style influences children's gaze behavior and the influence on performance as a result of a richer gaze-based interaction.

(H1): The tutoring style (Help/No_Help) will influence task performance

H1.1: Children will complete the task in less time in the Help condition than in the No_Help condition

H1.2: Children will complete the task with fewer tries in the Help condition than in the No_Help condition

(H2): The tutoring style (Help/No_Help) will influence children's gaze-based interaction during play

H2.1: Children will look more into the tutors' face in the Help condition than in the No_Help condition

H2.2: Children will establish more eye contact with the robot tutor in the Help condition than in the No_Help condition

(H3): Child-robot coordinated gaze will influence the children's performance

H3.1: Child-robot coordinated gaze will influence children's awareness of the robot's behavior/engagement with the tutor)

H3.2: Child-robot coordinated gaze will influence execution in either execution time or the number of tries

4.2.4. **Measures**

To test the hypotheses in section 4.2.3, we employed the following measures:

Task performance: As in Chapter 3, we used the following two objective measures that are notably used to measure performance in memory games:

(a) Duration: the time it took the children to find all pairs of matching cards on the table.
(b) Tries: the number of tries required to find all matching cards. A try consisted of picking up two cards.

The goal of the game was to get all the cards flipped face-up (i.e., find all the matching card pairs) in the least number of tries and as quickly as possible.

Next, we developed a *behavioral system* to measure child - robot interaction regarding gaze behavior, engagement, and emotions during the play tutoring interaction.

Gaze behavior: Observations of gaze directions to indicate the child's attention to the robot, child - robot interaction, and child behaviors, such as searching for cards on the board and interactions with the facilitator. The behavioral system emphasizes the coordinated gaze — mutual gaze — gaze-following and sequential gaze patterns—and the dynamics of — mutual gazing patterns between a child and a robot in a tutoring task *(see* Table 4-1 *and* Table 4-2 *for the definitions of gaze behaviors)*

Physiological state:

(a) Engagement: We defined two categories as states: (1) disengaged and (2) The engaged state, which is complementary to the "disengaged" state. The assumption was that if the child was not disengaged, i.e., did not exhibit either of the disengaged cues detailed in Table 4-4, then she/he was engaged.

(b) Emotion: We defined three categories (positive, negative, and neutral) to be assessed by the researchers from facial expressions, body movements, and verbal behavior (*more details on the definitions of these measures are in Table 4-4).*

Perceptions: We used a post-test interview to evaluate children's perceptions of the tutor's behaviors. Particularly we asked the children whether they were aware of the tutor's help and the modality the tutor used to help them, their judgments about the tutor, and their perceptions of the task in the two conditions. All sessions were video-recorded to facilitate the offline analysis of these measures.

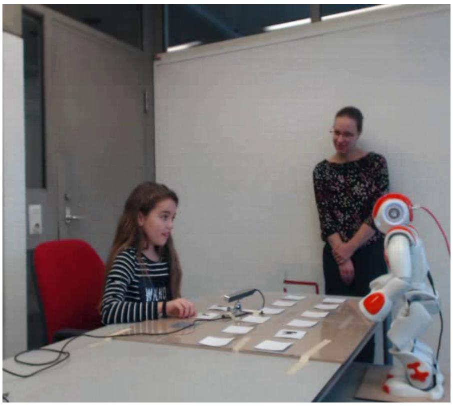


Figure 4-2 Experimental setting: child interacting with the NAO robot in a card-matching task. The facilitator is present at the session to support the child, if necessary. This setting is typical for child-robot interaction sessions.

4.3. Materials and Methods

4.3.1. **Experiment Setup**

The experimental setup included the NAO robot, a memory game, a webcam, and a personal computer (Figure 4-2). NAO is a 58cm tall robot from Softbank robotics (Softbank, 2013) with a moveable head and facial features that resemble those of a child. The robot tutor and the child sat across the table approximately 160 cm apart; there were fourteen (14) cards arranged in a rectangular board layout placed on a table *(see* Appendix A *for details regarding the board game design).* The algorithm is such that the robot automatically executed a sequence of head movements, as follows: the head angles shifted to the position of the chosen card (referred in this dissertation as card1), then to the face of the child, and then to the location of the matching card. Each card on the board was identified using a code, e.g., 1, 2, 3, and was placed in a specific position (head yaw and pitch).

4.3.2. **Procedure**

The child entered the experimental room (accompanied by the teacher, the parent, or the guardian). The robot was in a standing position on one side of a table, and the child sat on the opposite side of the table. A researcher who controlled the robot's behaviors was also present in the experimental room. The researcher sat in the corner of the room and followed the game using a webcam connected to her machine, to not be in the child's visual field so she would not interfere with the flow of the game. A facilitator was also present in the room to guide the children if they had any difficulties or answer any questions. The facilitator provided the child with a brief introduction of the robot and the game. The game began with the robot introducing himself as a tutor and giving the child instructions on how to play the game. The robot spoke either in Dutch or in English, depending on the language the child was comfortable using.

The tutor's instructions at the beginning of the game session Help_Condition

> *"Hello, welcome to the game session. My name is Maka. I am your tutor. I have a task for you. You are going to find pairs of matching cards on the table. You flip the first card and then flip a second card. If they do not match, turn both back and start over. If they match, leave them turned up. I am your tutor, and I know the positions of all the cards on the table. I am going to help you! Please go ahead and flip the first card!"*

Notice that while introducing the game in the Help condition, the robot tutor informed the child that it would help him/her without revealing the modality it would use to help.

During the No_Help Condition, the tutor omitted the following line during the instructions:

> *"I know the positions of all the cards on the table. I am going to help you!"*

After providing instructions, NAO remained silent and performed its gaze behavior as designed for either the Help or the No_Help condition, and the child played the first move by selecting one of the 14 cards (seven matching pairs) placed face down on the board and tried to find the matching one. If the cards turned face up were identical (a pair of matching cards), the child continued by making a new move; otherwise, the child turned the cards face down and made a new try/move. The game ended when the child found all the matching pairs.

Each child played the card game “Memory” in the presence of the robot tutor in both conditions of Help and No_Help, as described in Section 4.2.2. The order of conditions was counterbalanced. After playing the first game, the child waited for a few minutes for the experimenter to rearrange the card game. At the end of each game, the robot thanked and congratulated the child for playing the game. After both games, the facilitator asked the child the post-experiment questions on whether they noticed the tutor’s help cues and how they perceived the robot tutor in both conditions and the game.

4.4. Behavioral System

We developed a behavioral analysis system focused both on child and robot behaviors. There were four categories of a child's behavior: gaze (units and patterns), verbal, manipulation, and psychological state (engagement and emotions). For analyzing the robot's behavior, we registered both verbal and gaze behavior. Table 4-1 shows the child and the robot's gaze behavioural units, Table 4-2 shows the dyad's gaze patterns, including mutual gaze, gaze-following, joint attention, and the descriptions. The dyadic gaze analysis aims to examine the interaction between the child and the robot's gaze behavior. We focus on the following dyadic gaze patterns: mutual gaze, gaze-following, and joint attention, as illustrated in Table 4-2— *(see the definitions of social gaze behaviors in* Chapter 2; (Table 2-1*).*

Table 4-3 shows the verbal, behavioral units for both the child and the robot and the descriptions. Table 4-4 defines the behaviors used to assess the children's physiological state, i.e., engagement and emotions. Table 4-5 shows the behaviors of the children during card manipulation. Finally, Figure 4-3 shows snapshots of child and robot behaviors captured from the video observations depicting various robot and child gaze behaviors during the interaction.

Table 4-1 Child and robot's gaze units

Subject	Behavior	Categories	Description
Child	Gaze	Child_look_robot	Look at the facial area of the NAO tutor.
		Child_look_card1	Look at "Card1," which is the first card the child turns.
		Child_look_card2	Look at "Card2," which is the second card the child turns.
		Child_look_match	Look at "Matching," which is the correct card that matches Card1. If Card2 is the correct match, then it is coded as a match.
		Child_look_board	Looking generally at cards on the board or the board, for example, when the child is searching or thinking about which card to turn next.
		Child_look_facilitator	Look at the person helping during the experiment sessions.
		Child_look_else	Look at objects outside the game and its associated components (robot, board, and experimenter), for example, computers, windows, chairs, doors, or the roof. Also, an unknown gaze, i.e., when the coder cannot tell what object the child is looking at—the child is out of the screen or facing away from the camera or child is looking at another person other than the facilitator.
Robot		Robot_look_child	Look at the face/eyes of the child.
		Robot_look_card1	NAO tutor looks at Card1, i.e., the first card the child turns.
		Robot_look_match	NAO tutor looks at the matching card during the gaze condition

Table 4-2 Dyad's gaze patterns: in plain (transitions) — in blue italics (behavior continuity—not a transition): joint attention is depicted as a composite sequence of behaviors

	Dyad's behavioral units' sequence (timeline) →							
Mutual Gaze	**Child**	Look at robot						
	Robot	Look at child						
Gaze Following	**Child**	Look at robot	Look at X					
	Robot	Look at X	*Look at X*					
Joint Attention	**Child**	Look at robot	*Look at robot*	Look at matching	*Look at matching*	Look at robot	*Look at robot*	Look at matching
	Robot	Look at child	Look at matching	*Look at matching*	Look at child	*Look at child*	Look at matching	*Look at matching*

Table 4-3 Child and robot verbal behavioral units

Subject	Behavior	Categories	Description
Child	Verbal	Talk_to_robot	The child talks to the robot
		Talk_to_facilitator	The child talks to the person facilitating the experiment
		Talk_(expressive)	General verbal expressions during play
Robot		Opening_speech	NAO's instructions to the child on how to play the game.
		Positive_feedback	Statements when the child finds a matching card (fantastic, very good, well done)
		Negative_feedback	Statements such as "try again" and "not correct" if the child misses the correct match.
		Closing_speech	NAO congratulating and thanking the child at the end of each session.

Table 4-4 Physiological measures

Subject	Behavior	Categories	Description
Child	Engagement	Dis-engagement	(Closes eyes, looks down/head down; child covers face; a child is impatient; looks elsewhere)
		Engagement	If the child is not dis-engaged, then she/he is engaged
	Emotion	Positive	Enjoyment (clapping; winning sounds/gestures; Giggling) Happy (smiling/laughter/raising hands; positive verbal response (cool!))
		Negative	Frustration (negative verbal response; child impatient; uncomfortable) Fear, Confusion, Anger
		Neutral	When you cannot code any of the previous behavior (Positive or Negative)

Table 4-5 Child gameplay

Subject	Behavior	Behavior	Description
Child	Play	Touch_card1 Touch_card2 Turn-up_card1 Turn-up_card2 Turndown_card1 Turndown_card2	Playing the game (Turning cards)

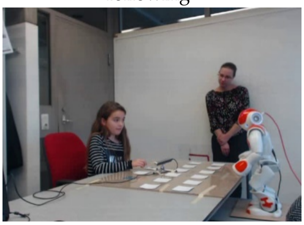
Robot_look_card1> Child_look_robot

Robot_look_match > Child_look_robot - child following the gaze of the tutor

Robot_look_match<>Child_look_match

Child_look_robot<> - Robot_look_child – Mutual gaze

Robot_look_match> Child_look_robot – gaze-following



Child_look_facilitator (person guiding and answering children questions)

Figure 4-3 Snapshots of child and robot behaviors captured from the video observations depicting various robot and child gaze behaviors during the interaction.

4.5. The Coding Process

The first step involved developing the coding scheme based on the established behavioral analysis system presented in Section 4.4. We used Noldus Observer XT software (Zimmerman, Bolhuis, Willemsen, Meyer, and Noldus, 2009) to conduct a behavioral analysis and to score the videos. First, we defined the two subjects—child and robot—i.e., that can initiate behavior during observations. Afterward, we identified the set of behaviors for each of the subjects. Figure 4-4 shows the coding scheme as captured from the observer software — the subjects on the left and the set of defined behaviors.

Gaze behaviors were defined as state behaviors, and behaviors by the same subject were mutually exclusive and exhaustive, meaning the behaviors excluded each other. And the subject state was known at any time. The set of gaze codes were both mutually exclusive and exhaustive (Bakeman and Gottman, 1997; Bakeman and Quera, 2011; Bakeman and Quera, 2012).

The coding process was done by having the researcher watch the videos and record the behaviors observed according to the coding scheme. During the experiment, we recorded thirty-five videos using a high-resolution webcam to capture the view of both the robot and the child. Thirty videos of 15 children in the Help and No_Help conditions were recorded and included in the analysis. There were 15 observations in the Help condition and 15 observations in the No_Help condition. The behaviors of the child and the robot were coded separately; that is, the researcher first coded the videos for child behaviors and then, in the second round, for robot behaviors. Reliability checks were performed throughout the coding process by randomly selecting and checking videos.

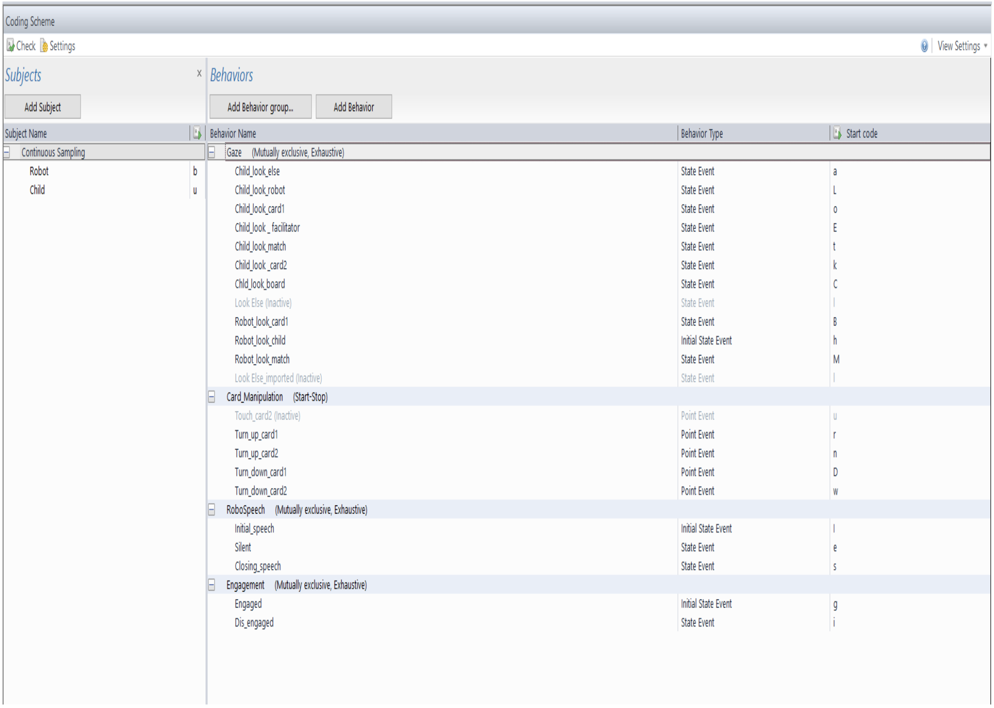


Figure 4-4 Initial coding scheme captured from the Noldus Observer XT 14

4.6. Visualizing Behaviors from the Observations

Figure 4-5 depicts a game situation in the Help condition of the coordinated interaction between the child and the robot and the corresponding sequences of mutual behaviors. The plot shows all the coded behaviors for the child in the top row and for the robot in the bottom row. In the top row is a timeline of the interaction in seconds. The rectangular bar indicates the duration of the occurrence of each event. In the figure, we can see the mutual relations between the behaviors. At the same time, in the second row, we see the child's behavior and the corresponding behavior of the robot in the bottom row. Each behavior appears on the visualization if it is coded at least once in the system. The length of the boxes indicates the duration, and the incidences of similar colored bar rectangles show the frequency of the occurrences of each behavior. A coordinated back-and-forth gaze alternation between the child and the robot's gaze and the cards (card1 or matching card) is a sign of a successful occurrence of joint attention, as illustrated in the visualization with a rectangle. Figure 4-6 shows mutual gaze occurrences, i.e., instances when the Child_look_robot and Robot_look_child behaviors co-occurs. In Figure 4-7, the highlighted instances show the moments when the behaviors Child_look_robot and Robot_look_match co-occur which is sign of successful occurrence of gaze following.

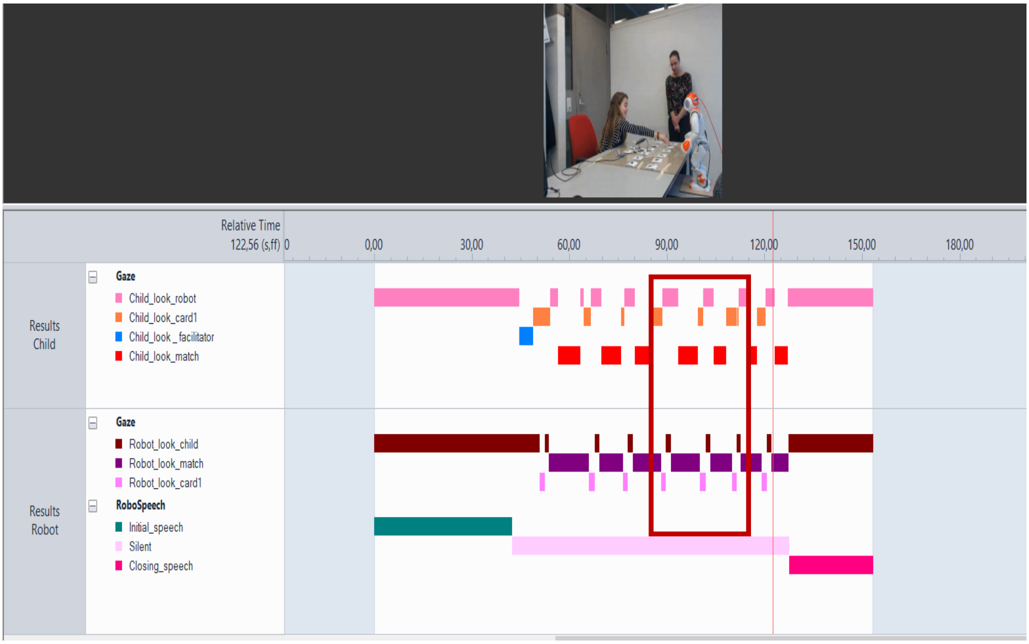


Figure 4-5 Video footage and visualization of events plotted horizontally against a time axis of coded child and NAO behavior from one play session in the Help condition. The red rectangular bar shows the instances of mutual gaze coordination between the child and the robot.

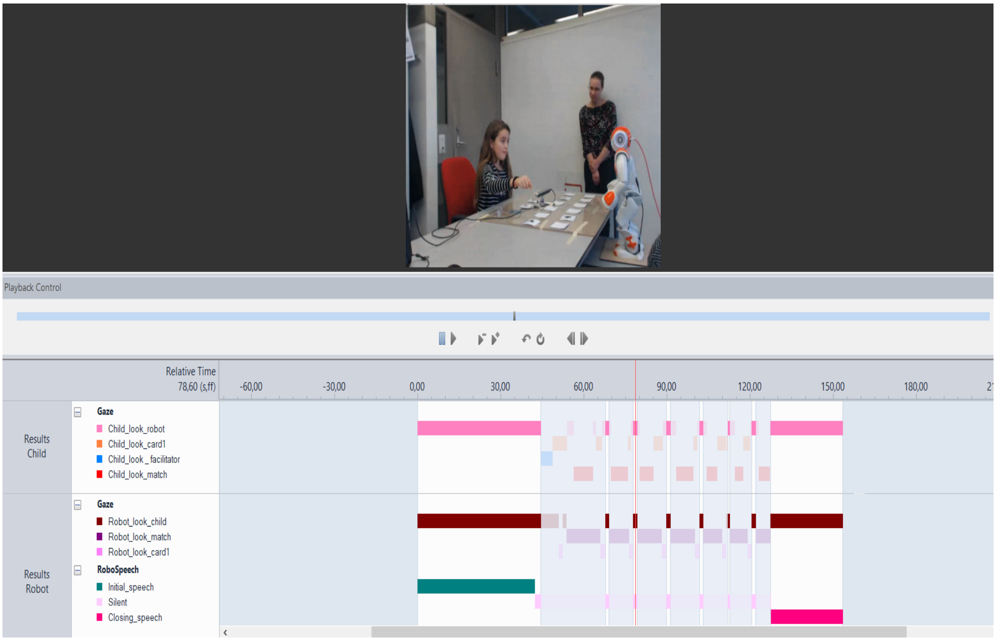


Figure 4-6 Higligting mutual gazing from child-robot observations: highlighted instances show the moments when mutual gazing occurred between the child and the robot, i.e., the behaviors Child_look_robot and Robot_look_child co-occur.

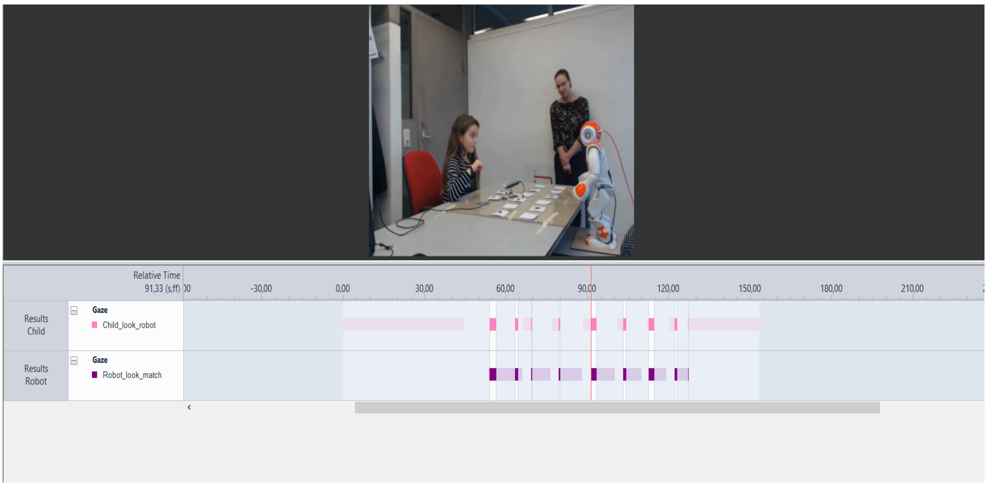


Figure 4-7 Higligting gaze following pattern from child-robot observations: highlighted instances show intervals when the following child and robot behaviors co-occurred: Child_look_robot and Robot_look_match. The pattern implies the child is looking at the robot while the robot is looking at the matching card.

4.7. Results

In this chapter, we analyzed the data of 15 children (age 6 -11) for 15 trials in the Help condition and 15 trials in the No_Help conditions. We excluded three (3) children from this analysis. The three children were exempted from the analysis due to the following reasons: one of the children failed to participate in one of the sessions of the game, and the other two were very young (below age 5) and could not play the game on their own.

4.7.1. Task Performance

To evaluate the effect of Tutoring_Style on children's performance - duration and number of tries, we conducted a repeated measure ANOVA with the Tutoring_Style (Help vs. No_Help) as the within-subjects factor using IBM SPSS statistics 25.

Table 4-6 Descriptive statistics: performance measures

	Help		No_Help		N
	Mean	SD	Mean	SD	
Duration (s)	167.87	75.70	169	79.18	15
Tries	14.07	4.90	17	3.65	15

Duration: There was no significant difference in the duration of the game between the Help and No_Help conditions (Help M=167.87s; No_Help M=169s; F (1, 14) =0.015, *p*=0.905).

Number of tries: There was a significant difference in the number of tries of the game between the Help and No_Help conditions (Help M= 14.07 tries; No_Help M=17 tries; *F* (1, 14) =5.331, *p*=0.037*).

4.7.2. Gaze

We conducted a repeated measure ANOVA with the Tutoring_Style (Help vs. No_Help) as the within-subjects factor using IBM SPSS statistics 25. The sections below report on the results. We analyzed thirty video observations of fifteen children (age: 6 -11 yrs.). There were fifteen trials in the Help condition and fifteen trials in the No_Help condition.

Child's Gaze Behavior

Duration(s): Based on quantified-coded data, we analyzed the durations children's gaze behavior for each condition (Help and No_Help conditions). The duration was defined as the time between the child picking the first card and taking the last one. The percentage shows the relative duration of gaze to the total time the child took to play the game, as duration varied with every game session.

Table 4-7 Children's gaze duration percentages (s) between Help and No_Help conditions (SE: standard error)

	Help		No_Help	
	Mean	SE	Mean	SE
Child_look_robot	28.41	5.53	7.21	1.87
Child_look_card1	21.91	2.05	32.88	1.51
Child_look_card2	42.62	3.62	47.98	1.64
Child_look_board	1.38	0.81	4.46	1.53
Child_look_facilitator	5.39	1.76	7.21	2.39
Child_look_else	0.29	0.21	0.17	0.15

We found a significant difference in gaze duration percentages for Child_look_robot between the Help and No_Help conditions (Help = 28.41% (observation duration(s)); No_Help =7.21% (observation duration(s)); F (1, 14) =14.06, p=0.002*). Children looked at the robot significantly longer during the Help condition compared to the No_Help Condition. Similarly, there was a significant difference in gaze durations for Child_look_card1 between the Help and No_Help conditions (Help = 21.91% (observation duration); No_Help =32.88% (observation duration(s)); F (1, 14) =18.25, p=0.001***)**. Children looked significantly longer towards card1 during the Help condition compared to the No_Help condition. However, there was no difference in looking duration at card2 in between the two conditions. A plausible explanation for this is given in the discussion section. In the percentage duration for Child_look_board (Help = 1.38%(observation duration(s)); No_Help = 4.46% (observation duration); duration; F (1, 14) =3.75, p=0.073), children looked longer at the board when there was no help than when there was help. However, the difference was not significant. There was no significant difference in percentage durations of Child_look_facilitator (p=0.56*)* and Child_look_else *(*p=0.65) between Help and the No_Help conditions.

Mutual Gaze (MG)

To highlight mutual gaze from the observations, we reduced the video observations to intervals when the behaviors—Child_look_robot and Robot_look_child—co-occurred, as shown in Figure 4-6, using the Observer XT. We examined the frequency and duration of mutual gaze (MG) patterns in both the Help and No_Help conditions. We conducted a repeated measure ANOVA in SPSS, with Tutoring_Style (Help vs. No_Help) as the within-subjects factor. We analyzed the results of 15 participants, for a total of 15 tries in the Help condition and 15 tries in the No_Help condition. We found a significant difference in the number of occurrences (frequency) of mutual gaze between the Help and No_Help conditions *(*p=0.001*). We also found significant differences in duration(s) of mutual gaze (MG) between the Help condition and No_Help condition *(p=0.004*).*

Table 4-8 Frequency and duration (s) of mutual gaze behavior in the Help and No _Help conditions

	Help Condition (N=15)		No_Help Condition (N=15)	
	Mean	SD	Mean	SD
Frequency	5.2	3.1	1.7	1.1
Duration(s)	32.4	1.1	13	12.2

Mutual Gaze and Awareness of the Tutor's Hints

Gaze awareness refers to whether the participant noticed the gaze hints of the robot tutor. From the post-experiment interview, eight children said they noticed the gaze hints from the tutor while the rest stated that they did not see the gaze hints.

Table 4-9 Frequency and duration (s) of mutual gaze behavior with and without awareness of the tutor's hints

	Noticed Gaze Hints (YES group) N=8				Did Not Notice Gaze Hints (NO group) N=7			
	Mean	SD	Range		Mean	SD	Range	
			Min	Max			Min	Max
Frequency	7.6	1.9	4	10	2.4	1.3	1	4
Duration (s)	40.6	17.3	24.5	78.9	20.3	18.5	1.0	56.3

Table 4-9 shows the descriptive statistics regarding the frequency and duration of mutual gaze patterns alongside the children's gaze awareness. We conducted an independent sample T-test using SPSS to compare the frequency and duration of mutual gaze for participants who noticed gaze hints (reported as YES) and those who did not notice gaze hints (reported as NO). There was a significant difference in the frequency of mutual gaze between children who noticed the gaze hints (YES group M=7.63, SD=1.92) and those who did not notice the gaze hints ((NO group M=2.43, SD=1.27): p=0.001*; two-tailed, assuming equal variances). We also found a significant difference in the duration of mutual gaze between children who noticed gaze (YES group M=40.62 s, SD=17.34) and those who did not notice the gaze hints (NO group M=20.27 s, SD=18.46): p=0.046*; two-tailed, assuming equal variances).

Gaze-Following and Awareness of Tutor's Hints

To highlight gaze-following from the video observations, we reduced the video observations to intervals when the following child and robot behaviors co-occurred: Child_look_robot and Robot_look_match. The pattern implies the child is looking at the robot while the robot is looking at the matching card. This co-occurrence is the key feature of joint attention, and a subsequent shared view pattern, Robot_look_match – Child_look_match, implies that both the robot and the child are looking at the correct card match at the same time.

Table 4-10 Frequency and duration (s) of gaze-following behavior with and without awareness of the tutor's gaze hints

	Noticed Gaze Hints (YES group) N=8				Did Not Notice Gaze Hints (NO group) N=7			
	Mean	SD	Range		Mean	SD	Range	
			Min	Max			Min	Max
Frequency	8.9	2.8	4	12	2.1	1.1	1	4
Duration (s)	22.4	15	9.6	55.7	4.2	3.6	1.0	10.3

We conducted an independent sample T-test in SPSS to compare the frequency and duration of gaze-following (Table 4-10) for children who noticed gaze hints as the robot was trying to help (reported as YES) and those who did not notice the robot hints (reported as NO). We found significant differences in the frequency of gaze-following patterns between children who noticed gaze hints (YES group M=8.88, SD=2.75) and those who did not recognize gaze hints ((NO group M=2.14, SD=1.07): p=0.001*; two-tailed, assuming equal variances). There were also significant differences in the duration of the gaze-following pattern between children who noticed the help through gaze hints (YES group M=22.43 s, SD=14.95) and those who did not notice the gaze hints ((NO group M=4.20 s, SD=3.63): p=0.008*; two-tailed, assuming equal variances). Figure 4-8 shows box plots showing the frequency and duration of mutual gaze and gaze-following pattern alongside the gaze awareness of the children.

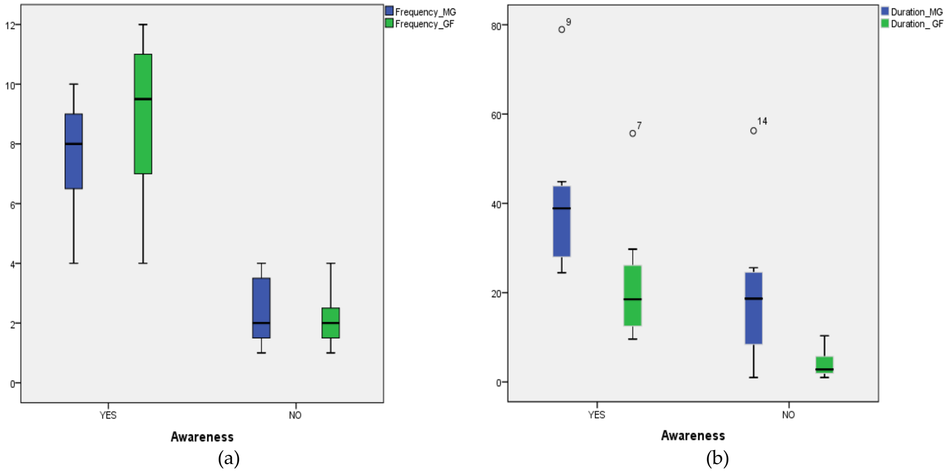


Figure 4-8 (a) Frequency of mutual gaze and gaze-following patterns (b) Duration (s) of mutual gaze and gaze-following patterns: Awareness — YES (noticed gaze hints) — NO (did not notice gaze hints)

4.7.3. Engagement

We examined the percentage levels of engagement during the Help and No_Help tutoring styles. From our findings, there was no significant difference in engagement during the game between the Help and No_Help conditions (Help = 99.38% (observational duration); No_Help = 97.89%; $p = 0.247$).

4.8. Discussion

This chapter continues the previous research line on how to design gaze-based communication to foster interaction outcomes during human-robot tutoring. We used a similar setting used in Chapter 3 to investigate whether gaze cues from a robot tutor can influence children's behavior during game tutoring tasks. Children interacted with a robot tutor in two conditions: one in which the robot provided gaze cues to help the children find matching cards on the table, and one in which the robot did not provide gaze help to the children. We analyzed execution performance, gaze behavior, and engagement level during the child-robot interaction to examine the impact of robot gaze cues. We also investigated the occurrences of repeating gaze patterns in the dyadic interaction between a child and the robot during gameplay.

In the following paragraphs, we outline whether the findings support the formulated hypothesis in Section 4.2.3:

In *(H1)*, we projected that the tutoring style (Help/No_Help) would influence task performance. The findings partially support this hypothesis. We found no significant difference in duration(s) —time taken to complete the game (*H1.1*); however, children used significantly fewer tries to finish the card-matching game with gaze hints from the robot tutor than without the gaze hints from the tutor (*H1.*2).

In *(H2)*, we projected that the tutoring style (Help/No_Help) would influence participants' gaze-based interaction during play. We found that children looked significantly longer to the robot tutor when the tutor provided gaze hints during the game than when the tutor did not give such hints *(H2.1)*. The findings support hypothesis *(H2.2)*; there were more occurrences and longer durations of mutual gaze patterns during the Help condition than in the No_Help condition.

On other child's gaze measures, we found that children looked at card1 significantly longer in the No_Help condition than the Help condition. The plausible reason could be because when the child flipped the first card (card1), they did not look at it, as they knew the robot would help with the clues. Again, interestingly, we found no significant difference in gaze duration on card2. The plausible explanation is that during the Help condition, the children looked at card2 longer to confirm it was the match. We found negligible percentages of durations of looking elsewhere in both conditions of Help and No_Help. These match our results on engagement, as we found no significant differences during the Help and No_Help conditions. There was a higher percentage of engagement during both sessions of the game. Only with two children did we notice some instances of disengagement.

In *(H3)*, we projected that the child-robot coordinated gaze would influence the children's performance. During the Help condition, eight children out of the fifteen children reported they noticed the gaze hints from the robot tutor—head movements of NAO to different cards on the table—while the rest did not see the gaze hints. The findings support *(H3.1):* We found significant differences in occurrences and durations of mutual gaze and gaze-following patterns for children that recognized the tutor's hints vs. those that did not notice it. These findings suggest that more occurrences and higher durations of coordinated mutual gaze and gaze-following pattern increased the children's awareness of the robot's helping cues. In turn, when children noticed that the robot tutor tried to help with the gaze hints, they looked more to the robot to take advantage of its cueing and performed significantly better *(H3.3).*

We found no significant effects on the duration to complete the task between the Help and No_Help conditions. There are several possible explanations for this. Firstly, we observed that when children noticed that the tutor was helping with gaze, they waited until the robot showed them the matching card, even when they had an idea of where the matching card was. Secondly, the robot's novelty effect, supported by the longer duration for the robot even when the number of tries was less, indicates that children who noticed gaze probably spent more time looking at the robot. Another probable reason is the duration of head motions during attention shifts from the flipped card to the participant's face and then to the matching card. This delay suggested the child was faster in picking the second card than the robot pointing with the eyes, with the subjective impression that the robot was following the child's gaze rather than trying to guide it. Lastly, a few of the children spent some time asking the experimenter questions during the game either because of confusion or when they saw the robot's movement and could not interpret it. Therefore, on a more conceptual level, the potential risk is that the robot tutor's hints might induce slacking, encouraging children to rely on his Help even when it would not be necessary. However, this might be the result of the use of explicit, overt cues. We discuss this further in Chapter 6 as an open question for human-robot interaction researchers to investigate the role of covert vs. overt gaze in learning scenarios and particularly in social training We examine the timing and sequencing of the gaze behavior between the child and the robot in Chapter 5.

In the post-experiment interview, eight out of the fifteen children said they noticed the robot's hints during the game while the rest said they did not see the hints. Gender and age are two factors to consider in future research regarding how children perceive and interpret robot cues, especially in tutoring interactions. However, it is good to note that from our observations, we noticed that age did influence the capacity of children to read Help from gaze with older children being more aware of the tutor's intent and following accordingly. In comparison, most of the younger children could not interpret the robot was helping them despite noticing the cues. This can be explained by social cognition developmental theories for children, including the theory of mind (Baron-Cohen et al., 1985). For example, the majority said they noted the robot tutor's head movements but could not interpret the head movements as gestures pointing to the matching card. From our research, we could not make such conclusions regarding age and gender, and we outline these as open questions in Chapter 6 that need further investigation.

This study compares with the findings from the previous chapter with adult participants using a similar setting on performance. In Chapter 3, the findings showed the participants performed better with the tutor's help gaze cues than without the help cues. Likewise, in both settings for the adults and children, most indicated they expected verbal cues from the tutor. We assume the design of the robot's behavior may have led participants to expect verbal help cues from the tutor because, in the beginning, the robot verbally introduced itself and gave instructions to the participant but remained silent during the entire interaction, only giving gaze clues. However, from our observation, we can highlight notable differences in how children perceive and interpret the robot's gaze cues. For example, in the child experiment, we observed a few children who proceeded to select different cards despite noticing the robot tutor was looking at a particular card, which was not the case for the adult participants. Again, these observations could be attributed to social cognitive theories of development—such as perspective-taking and agency attribution—which are fascinating to study with robots in the future (Baron-Cohen, 2008).

4.9. Summary

These results provide design recommendations that can be applied to create effective robot-child tutoring systems. There are two significant conclusions from this chapter:

First, when a robot tutor provides gaze cues during a tutoring interaction, children use fewer tries to complete the task than when a robot tutor does not give such cues. *Second*, more coordinated gaze interaction patterns (mutual gaze and gaze-following) increase children's awareness of the tutor's intention to help during the gameplay activity. Consequently, when children notice that the robot tutor tries to help with the gaze hints, they look more to the robot tutor, hence increasing the gaze interaction and performance.

In the following chapter (Chapter 5), we further our understanding of the dynamics of gaze interaction during child-robot tutoring. We combine observation analysis with lag-based methods to examine the interaction sequences of gaze between a child and a robot during a gameplay tutoring activity. We focus on identifying the sequential interaction patterns associated with joint attention (JA). Joint attention is a sequence of looking at behaviors (mutual gazing, gaze-following), leading to the sharing of attentional focus on an object of interest to ensure shared understanding. Such gaze sequences are known to be important in children's learning and development and are the basis of the theory of mind and perspective-taking (Baron-Cohen, 2008). As such, understanding how gaze coordination and interaction sequences of gaze unfold during child-robot interaction is of great importance to developmental robotics.

Chapter 5. **Identifying Dynamic Gaze Interactions during Child-Robot Tutoring**

This chapter investigates dynamic patterns of dyadic gaze behaviors between a child and a robot in a collaborative card-matching game (based on Memory game). The aim is to understand the effects of robot gaze behavior on child-robot interaction and attention during a tutoring task. Gaze is a social cue that indicates the children's attention to the robot, the mutual gaze, and joint attention to objects (cards in our case) (Frischen et al., 2007; Kleinke, 1986) that could be a basis for rapport, and the increase in the quality of child-robot interaction, and eventually result in better tutoring. Moreover, repeating sequences of behavior may encode typical and meaningful interaction patterns, which can help us provide guidelines for designing successful tutoring interactions.[5]

This chapter combines observational measures and lag methods to investigate the nature and dynamics of intricate patterns of gaze interchanges between the children and the robot. The lag-based method provides an opportunity to analyze dynamic and reciprocal gaze sequences in child-robot interaction rather than instances of (mutual) gaze or one-directional attention patterns that have never been previously explored in robot-child tutoring.

We conducted a field study in a classroom environment using the same interaction scenario as in the previous chapters, where a child plays a card-matching game in the presence of a robot tutor under two conditions. In one condition, the robot provides gaze-based interaction (e.g., glancing towards the correct match). In contrast, in the other condition, the robot tutor does not deploy any 'look at' behavior and keeps the gaze fixed on the child's face. However, in this chapter, the gaze sequence for the robot was specially designed to enhance its power to draw the child's attention to the matching card.

[5] Part of the work in this chapter is currently in submission

Moreover, we enriched the robot's behavior with verbal feedback after each move and expressive gestures of arms and hands to reinforce the robot's liveliness and eventual appeal.

We again coded the individual gaze behaviors of the child and robot from the video recordings of the game sessions using the extended coding scheme described in section 5.3.4. We included not only simple gaze behaviors (e.g., the child looks at the robot; robot looks at the matching card, and mutual gaze instance), as we did in Chapters 4, but also the complex coordinated sequences of joint attention. We have provided the definitions for the mentioned social gaze concepts in Chapter 2; Table 2-1. The descriptions of coordinated gaze patterns and sequences, as described in the behavioral system, are in Chapter 4; Table 4-2. Table 5-1 shows the gaze behavioral sequences examined in this study related to joint attention with the robot's gaze as the initial behavior and the child's gaze as the response behavior.

We analyzed the gaze data using sequential lag methods, both event-based and time-based, to examine if there are patterns of reappearing gaze sequences and the coupling of these patterns in helping the child pair the cards. Our particular focus was to investigate dyadic and sequences of gaze behavior—regarding efficacy and efficiency in task-solving and the quality of interaction with the robot—in a collaborative tutoring interaction.

We found that the robot’s gaze significantly impacts interactions between the children and robots during a collaborative tutoring task. Similarly, with the results from Chapter 4, we found more occurrences and longer durations of mutual gaze behavior with the presence of helping gaze cues from the robot tutor than when the robot did not provide such cues. We compared dyadic and sequential behavior pairs for two groups: The *Socially attentive group* — children with higher events of look at the robot behavior— and the *Non - socially attentive group* — children with lower events of looking at the robot. Looking at the robot, in this case, signals engagement with the tutor during the gameplay interactions. Children in the *Socially attentive group* performed significantly better than those in the *Non - socially attentive group*.

In line with findings from Chapter 4, we found more occurrences and longer durations of mutual gaze and gaze following patterns for the children in the *Socially attentive group* than the children in the *Non - socially attentive group.* Furthermore, we identified significantly more gaze sequences between the child and robot-related to joint attention in the *Socially attentive group*. However, there was no significant gaze sequences of joint attention in the *Non - socially attentive group.* These findings show that children who followed the robot tutor's gaze orientation performed better on the task than children who did not follow the tutor's gaze. The time-based lag analysis revealed that the timing of the child's response gaze occurred from zero up to six seconds after the initial gaze of the robot tutor. The interpretation is that any association between an initial and response gaze behavior for the *Socially attentive group* occurred from zero seconds up to six seconds and disappeared after six seconds.

These findings support the hypothesis that the occurrence of successful sequences of coordinated gaze behavior between the children and the robot facilitates children's awareness of the robot tutor's intentions in the interaction and the robot efficacy as a tutor.

The rest of this chapter is structured as follows: Section 5.1 provides the background on the importance of sequencing and coordinated gaze interaction and introduces lag-based analysis. Section 5.2 describes the experiment's design — participants' details, the hypothesis, study conditions, and the measures. Section 5.3 provides the experimental setup, the platform, and the design of the interaction scenario and the board game, and Section 5.4 outlines the results of the user study. Section 5.5 provides a comprehensive discussion of the results. Finally, Section 5.6 summarizes the significant contributions for designing effective gaze behavior mechanisms for robot interventions in tutoring and social training.

5.1. Background

The timing and the right sequences of interactions can make or break the quality of human communication, as shown in studies with newborns and children with autism spectrum disorder (Baron-Cohen, 1997; Brooks and Meltzoff, 2002; Meltzoff and Brooks, 2007). For example, turn-taking between infants and adults relies on sensitive timing and sequencing to the point that desynchronization during interaction results in the infants showing distress and avoidance, such as looking away (Brooks and Meltzoff, 2002,2014; Meltzoff and Brooks, 2007; Meltzoff, Brooks, Shon, and Rao, 2010).

Thus, the sequencing and timing of gaze interactions, in particular, can be a reason for unnatural and, therefore, ineffective human-robot interaction. However, there is limited evidence on how exactly the dynamic and interactive aspects of the social cues affect the interaction and the tutoring outcomes.

Gaze interaction involves sequences of intertwined and coordinated looking behaviors between interacting partners. Successful coordination of gaze leads to the transfer of information in a shared visual space. For instance, during conversational interactions, participants monitor their conversation partner's gaze to regulate their gaze behavior, control turn-taking, and indicate attention (Vertegaal et al., 2001). Although a large number of studies over the past decades have investigated gaze behavior and the crucial role it plays in communication (Admoni and Scassellati, 2017; Broz et al., 2012), how gaze interaction unfolds over time in child-robot interaction is still not understood. Besides, most of the present methods used to assess child-robot interactions mostly observe the child's behaviors without recording the robot's behaviors, and ignore reciprocal interactive behaviors. The disadvantage of such evaluations is that they do not capture the dynamics of the interaction, which can be seen in sequential intricacies of interactions between the child and the robot and could have an impact on the cognitive or affective learning outcomes. To address this gap, in the current chapter, we combine observational and lag sequential analysis methods to examine the sequential coordination of child and robot gazing behavior during a collaborative tutoring activity. The analysis of coordinated sequences of behavior during the child-robot interaction can provide a deeper understanding of how children interact with a robot during the task-solving and help design robots that are effective tutors.

Lag sequential analysis is a common method for examining sequential behavior pairs (Bakeman and Quera, 1995; Faraone and Dorfman, 1987; Pohl et al. ,2016) and has extensively been applied in behavioral research. For example, Montague and Asan (2014) and Montague et al. (2011) used the lag method to study and understand the dynamics of gaze interaction between patients and doctors within health systems. Borrero, England, Sarcia, and Woods (2016) and Woods, Borrero, Laud, and Borrero (2010) applied lag analysis to examine children's behaviors before and after parental attention during mealtime. Robinson, Anderson, Porter, Hart, and Wouden-Miller (2003) used lag analysis to study transition patterns of preschoolers' social play to identify precursors of problem behavior in individuals with autism.

However, to our knowledge, lag sequential analysis has not been applied in human-robot interaction to examine nonverbal communication and especially sequences of gaze interaction between the robot and the human. Previously in Chapter 4 (Mwangi, Barakova, Díaz, Mallofre, and Rauterberg, 2018), we found that if the robot is helping the children, i.e., uses its gaze to point at a matching card, simultaneous gaze between the child and the robot to each other and towards a card take place, and continues for the children who have noticed the gaze of the robot. We expect that the use of sequential analysis methods in human-robot interaction will help unravel the dynamics of child-robot interaction, which is a necessary part of developing successful robotic interventions in tutoring and therapy settings. In particular, the analysis of coordinated sequences and directions of gaze behavior provides a deeper understanding of whether, how, and under which conditions children can recognize and take advantage of robots' nonverbal prompts. Specifically, lag-sequential analysis helps examine whether one event contributes to the likely occurrence of another.

In this regard, this chapter addresses the following questions:

RQ3 — What are the dynamics of dyadic gaze-based interaction (mutual gaze, gaze-following, and joint attention) in the context of tutoring?

(a) How do we describe and identify the key patterns of coordinated gaze behavior in the context of?
(b) Which are the contextual and individual variables that affect the occurrence of key patterns of coordinated gaze?
(c) How do we model the sequence of coordinated behavior according to individual and situational variables of interest, such as the flow of the game (i.e., a sequence of failed and successful tries)?

RQ4 — Does the coordinated gaze (mutual gaze, gaze-following) influence performance?

(a) Do dyadic patterns influence the participants' awareness of the robot's behavior?
(b) Do dyadic patterns influence task execution?
   i. Is there any difference in the time required to complete the game?
   ii. Is there any difference in the number of tries?
(c) Do dyadic patterns influence participants' interactive behavior (i.e., social engagement)?

To explore the questions mentioned above, we used the same card game described in Chapter 3 and Chapter 4, where the child plays a card game in the presence of a robot tutor. For the study described in this chapter (see Table 1-1), we modified the timing and sequencing of the head/gaze movement of the robot based on earlier subjective reports where subjects (implicitly) assumed the tutor was performing gaze-following rather than gaze-cueing (*see gaze concepts definitions in* section 2.2.1*)*. Because of a particular delay—reaction time—between the child choosing the first card and the robot looking at the matching card which suggested the child was faster in picking the second card than the robot pointing with the eyes, with the subjective impression that the robot was following the child's gaze rather than trying to guide the child. In this study, the "reaction time" of the robot was reduced to overtake the child's move after the first choice.

The design of the gaze sequence for the robot tutor was modified as follows: (Figure 5-1)

The robot, from an initial state of looking at the child, gazed at the card the child picked—looked at the matching card—looked at the child's face—looked again at the matching card—and, finally, looked at the child's face again. This sequence repeated itself once the child picked a new card. Doing so, we expected that the child would follow the robot's gaze to the matching card and subsequently look back at the robot, i.e., perform joint attention (see definitions in Table 2-1; Table 4-2*), as it occurs in human-human interaction.*

Further, to enrich the interaction, in the present setup, the robot tutor deployed a friendly approach through verbal and nonverbal behavior. At the start, when the child entered the room, NAO tracked the child's face to create initial gaze interaction, to show interest and readiness of communication, and to increase its believability and friendliness. During the game, the robot gave verbal feedback to the child as to whether the card was correct or incorrect: when the child turned the right card, the robot said “very good” or “fantastic,” and when the child turned the wrong card, the robot gave statements such as “try again” or “not correct.”

Moreover, the robot deployed expressive gestures of the arms and hands, emphasizing verbal communication in the instruction section with pre-programmed explanation gestures and, during the game, it accompanied the verbal feedback with winning or losing gestures.

In the previous chapters, we have established that gaze behavior is an important nonverbal cue in human -robot interaction , as in human-human interaction, to convey information, attention, and awareness in a collaborative activity with a robot tutor. In this setting, therefore, gaze behavior indicates not only the child's attention to the robot but also the child's cognitive activity in the game, such as observing the images on the cards or scanning the board searching for the matching card. The assumption underlying this study is that the occurrences of appropriate sequences and coordination of the dyad’s gaze behaviors between the child and the robot would facilitate the effectiveness of the robot tutor as a helping agent and improve the performance of the child fulfilling the task. Thus, the robot could then positively influence the flow of the child’s actions—if the child takes the cues—improving the task execution and the play experience.

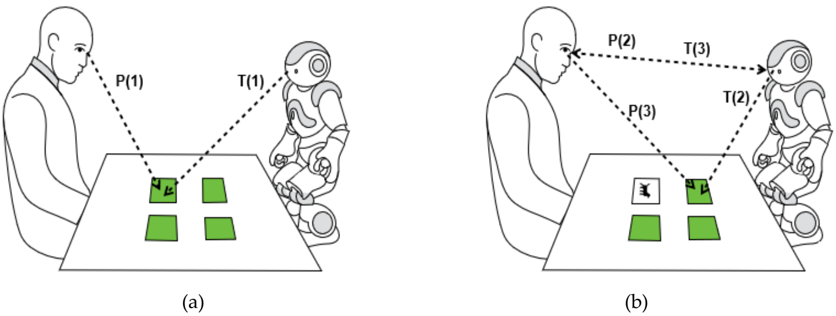


Figure 5-1 The improved interaction flow for the Gaze condition: (a) P(1) - child participant (left) turns over a card T(1) - tutor (right) looks at the selected card (b) T(2) - tutor gazes at the matching card and then towards the child T(3) to draw attention T(2) - tutor glances again at the matching card, and then to the child's face - T(3)

5.2. Experimental Design

5.2.1. Participants

Twenty-three children (Age: 7-8 years; Gender: Male: 11; Female: 12) from the primary school El Margalló in Vilanova i la Geltrú, Barcelona participated in the study. *El Margalló* is an inclusive education school where children with learning difficulties are in age-appropriate general education classes, fully participating in the classroom's regular activities, with the support of specialized staff, when required. The children with special needs attending the school include children with autism spectrum disorder and other psychological and physical conditions, such as cerebral palsy. The participants in this study were the whole class of grade 3. In this group, there were two children with autism spectrum disorder (ASD) and one with a congenital condition impairing motor and intellectual development. The teachers/parents/guardians of the children voluntarily completed written informed consent before the start of sessions *(see* Appendix D *for more details regarding the children's data).*

5.2.2. Study Conditions

The experiment followed a one-factor (Tutoring_Style) within-subject design with two conditions (the Gaze condition and No_Gaze condition). The two types of tutoring styles deployed by the tutor are as described below:

Gaze condition: In this condition, the robot provides gaze-based interaction (glancing towards the correct match). As mentioned in Section 5.1, the designed gaze sequence for the robot is an improvement over our previous studies to enhance its power to draw the child's attention to the matching card.

No_Gaze condition: In this condition, the robot tutor does not deploy any "look at" behavior and keeps the gaze fixed on the child's face.

In both conditions, the robot's behavior was enriched according to the results of the previous studies to reinforce the robot's appeal with verbal feedback after each move, with the more naturalistic and expressive movement of the arms and hands. Each participant interacted with the tutor in both conditions of Gaze and No_Gaze, and the order of conditions counterbalanced across trials.

5.2.3. Hypothesis

We formulated three hypotheses, as outlined below:

(H1): The tutoring style (Gaze /No_Gaze) will influence task performance

H1.1: Children will complete the task in less time in the Gaze condition than in the No_Gaze condition

H1.2: Children will complete the task with fewer tries in the Gaze condition than in the No_Gaze condition

(H2): The tutoring style (Gaze/No_Gaze) will influence participants' gaze-based interaction during play

H2.1: Children will look more and longer into the tutors' face in the Gaze condition than in the No_Gaze condition.

H2.2: Children will establish more eye contact with the robot tutor in the Gaze condition than in the No_Gaze condition

(H3): Child-robot coordinated gaze will influence the children's performance

H3.1: Child-robot coordinated gaze will influence participants' awareness of the robot's behavior/engagement with the tutor)

H3.2: Child-robot coordinated gaze will influence performance either in time or the number of tries

5.2.4. **Measures**

To evaluate the hypothesis mentioned above, we used the following measures:

Task Performance:

(a) Duration: The time it takes the participants to find all pairs of matching cards on the table;

(b) Tries: The total number of attempts required to find all matching cards. A "try" consists of choosing two cards.

Gaze behavior: To analyze the gaze behavior, we adopted the previously developed behavioral analysis scheme detailed in Chapter 4, Section 4.4. The structure of gaze analysis, as described in the previous chapter, is in three categories: child, robot, and dyad gaze (mutual gaze and joint attention) behavior. The additional gaze behavioral sequences - examined in this study are shown in Table 5-1. The highlighted and (in italic) are the specific sequences associated with mutual gaze, gaze-following, and joint attention.

Children Perceptions: At the end of the experiment, we performed semi-structured interviews to evaluate children's judgments of the tutor and task. To examine whether the child was aware of the robot's intent to help, the children were asked which cues they observed in the tutor's behavior. The questions also assessed whether children perceived any differences between the conditions of the game/task difficulty *(see* Appendix C *for more details on the experimental procedure and post-experiment interview).* All sessions were video-recorded to facilitate the analysis of performance and behavioral measures.

5.3. Materials and Methods

5.3.1. Interaction Scenario

The interaction set is the same card-matching game called "Memory," previously described in chapters 3 and 4. In this game, the child plays the card game alone, but in the presence of an expert (robot tutor). A detailed description of the board game design and arrangement of cards on the layout is in Appendix A.

5.3.2. Experimental Setup

This study was an “in situ” experience—at school with a whole class group, including children with special needs. The trials took place in a room familiar to the children, often used as a classroom. The robot was put in a standing position on one side of a small table where the game was placed (Figure 5-2), and the child sat on the opposite side. The camera was positioned to capture the behavior of both the robot and the child. A researcher who controlled the robot’s behavior sat in the corner of the room outside the child’s visual field, so she did not interfere with the flow of the game. Each card on the board was identified using a code, e.g., 1, 2, 3, and was placed in a specific position (head yaw and pitch).

Figure 5-2 A child playing the card memory game in the presence of the NAO tutor in a classroom. The child looks at the robot tutor while the robot is looking at the matching card.

In the Gaze condition, the experimenter controlled the robot's behavior: After entering the code for the chosen card, the robot executed a preprogrammed gaze sequence. First, the robot's head angled to the position of the first chosen card (referred here as card1), to the position of the matching card, towards the child again, back to the matching card, and then to the child's face. The automatic sequence of the robot's gaze behavior repeated each time the child picked a new card. In doing so, the robot alternated its gaze between the matching card and the child's face following a timed sequence and cycled back to mutual gaze to ensure a shared experience to elicit coordinated gaze behavior. This prompted the child to follow its gaze to the matching card and then similarly performed joint attention, as occurs in human-human interaction.

5.3.3.**Procedure**

At the beginning of each trial, the child entered the experimental room, accompanied by the facilitator. At this moment, the robot-tracking mode was on, and the robot was always looking to the door when the child entered. Then, the facilitator asked the child to take a seat beside him/her, provided the child with a brief introduction of the robot and the game, and handed over to the robot.

After the briefing, the robot tutor welcomed the child, introduced itself as a tutor, and instructed the child on the rules of the game in Catalan, as all children were conversant with the language.

In the Gaze condition, the text for the tutor's instructions was as follows:

> *"Hi, welcome to this play session. My name is Maca, and I am your tutor. I have a task for you. You have to find pairs of matching cards on the board.*
>
> *First, you take a card and turn it up on the board, and then you pick another one and turn it up.*
>
> *If they do not match, you turn both cards upside down so that the picture is hidden, and you start again. If the second card matches the first, you take the two cards and keep the pair next to you.*
>
> *You have to find all the pairs in the least number of possible attempts, choosing as few cards as possible.*
>
> ***I know the positions of all the cards on the table. I am going to help you!***
>
> *Please, let us start and choose the first card!"*

Notice that while introducing the game in the Gaze condition, the robot tutor informed the child that it would help him/her, without revealing the modality it would use to help.

In the No_Gaze condition, the briefing is the same except for the sentence referring to the help—in bold—that is substituted by the sentence **"You will have to play alone, I cannot help you now."**

After the briefing, the facilitator asked the child about the robot's role. If the child was not capable of referring to the tutor's role or did not understand NAO's instructions, the robot repeated the instructions, after which the child would begin playing the game.

During the flow of the game, the tutor provided feedback to the child using statements such as "very good" and "fantastic" when the child found a matching pair or "try again" and "you picked the wrong card" when the child made an incorrect move. At the end of every session, the tutor thanked and congratulated the child.

During the flow of the game, the facilitator answered questions from the children. After each session, the facilitator walked out with the child so that they could complete a post-experiment questionnaire regarding the awareness of the tutor's help, the game, and perceptions of the tutor's behavior in both conditions.

During this time, the investigator re-ordered the cards for the second session or the next participant. Each child played the card game in both conditions of Gaze and No_Gaze, with each session lasting approximately 20 minutes. This included the instruction period, the play, and the interview/questionnaire session. The order of presentation of conditions (Gaze/No_Gaze or No_Gaze/Gaze) was counterbalanced. All the sessions were videotaped to facilitate the analysis of performance and behavioral measures *(see* Appendix C *for the complete experiment protocol).*

5.3.4. **Coding Child-NAO behavior**

The coding of the video recordings was done using Noldus Observer XT 14 software (Zimmerman et al., 2009). There were three separate stages in each child-robot session: child entrance, the game session, and the closing part. All videos were coded for the duration between when the child sat down until the robot gave the closing remarks. Figure 5-3 shows the modified coding scheme, and the definitions for the behaviors used in this study are as defined in detail in the previous chapter (Chapter 4, section 4.4). The coder first coded the video in its entirety for child behavioral units and then again for robot behaviors using the Observer XT software. The process of coding took approximately 28 days. Forty videos were coded: 19 in the Gaze condition and 21 in the No_Gaze condition. The total duration for the videos in the Gaze condition was approximately 6082 sec, and the entire length of the videos in the No_Gaze condition was 4880.96 sec — the coder trained with five practice videos first before moving to the primary analysis to ensure reliability. Coding one video for all the behavior took approximately 30 minutes. Throughout the coding process, videos were randomly checked by members of the research team to assess reliability.

Figure 5-3 shows the modified coding scheme captured from the Noldus Observer XT 14, and Figure 5-4 shows video footage and a time event plot of the behavior visualization in the Gaze condition. The plot shows coded behaviors for the child in the top row and the robot in the bottom row. Each behavior appears on the visualization if it is coded at least once. On top is a timeline of the interaction in seconds. Below the timeline, we can see the child's behavior and, at the bottom, the concurrent robot behaviors. This visualization allows for identifying the timed relationships between both agents and figuring out particular patterns. The length of the boxes indicates the duration, and the number of rectangles of the same color shows the frequency of the occurrences of each behavior *(see* Appendix E *for all child - robot coordinated gaze visualizations— Gaze condition).*

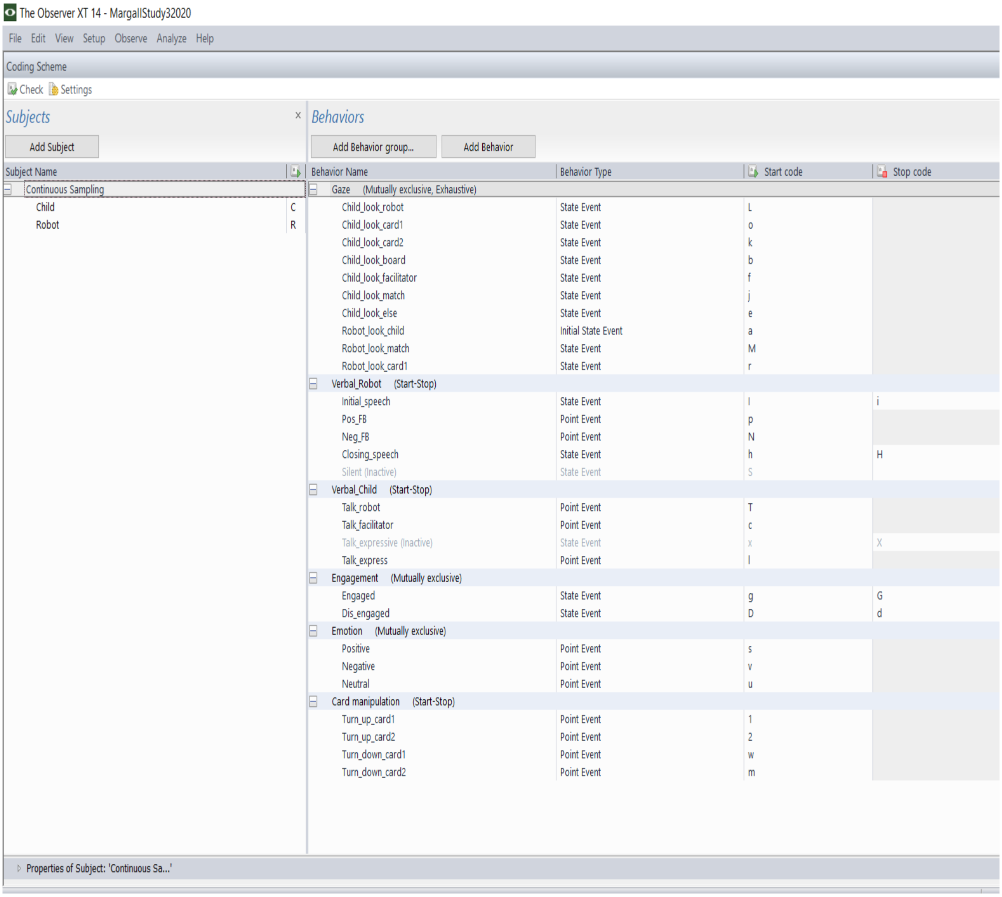


Figure 5-3 Modified coding scheme captured from the Noldus Observer XT 14

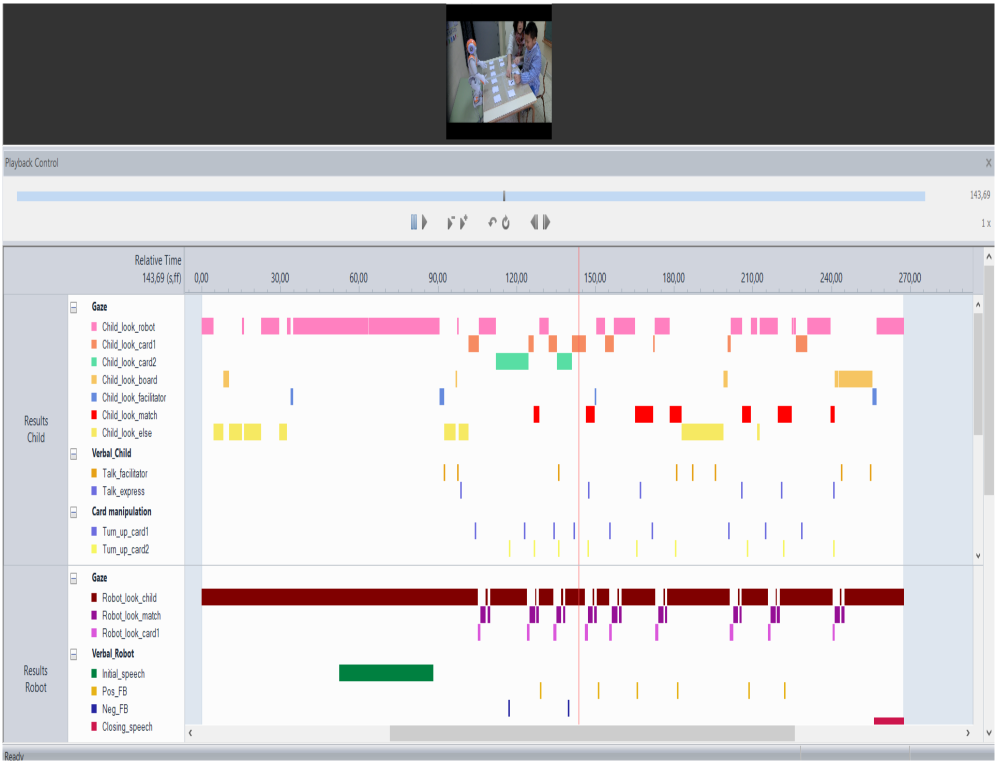


Figure 5-4 Video footage and visualization of events plotted horizontally against a time axis of coded child and NAO behavior units from one play session in the "Gaze" condition.

5.3.5. Gaze Behavior Patterns: Contingency Tables

In Chapter 4, Table 4-2, we presented the dyadic gaze behavior and their corresponding definitions. The joint attention gaze involves the coordination of gaze sequences between the child and the robot. These behavioral patterns include mutual gaze and gaze-following, and a subsequent sharing of attentional focus on an object of interest—shared gaze (Argyle et al., 1994; Emery, 2000; Frischen et al., 2007) *(see joint attention definition in* Table 2-1 *and* Table 4-2*)*. We examined the transitions from the robot's behavior to the child's behavior (i.e., the robot's gaze as the antecedent or initial behavior and the child's the consequent or response), as the robot's gaze was predetermined and followed a preprogrammed sequence. We considered the child and robot observations during the gaze condition, i.e., when the robot tutor provides gaze cues during the interactions.

The assumption was that during the Gaze condition, children would follow the social prompts from the robot tutor's gaze while the tutor was looking at different cards on the table and use these gaze cues from the tutor to select the matching card. Therefore, the projected pattern of coordinated behavior in the case of successful joint attention between the child and the robot was:

> *The child looks at the robot, and the robot looks at the child (mutual gaze) > the robot looks at the match > the child looks at the robot (gaze-following) > and the robot looks at the match and child looks at the match (shared gaze) (see Chapter 4; Table 4-2) for a complete description of dyadic gaze patterns — mutual gaze, gaze following and joint attention)*

The "and" means co-occurrence and ">" means the following events.

In this case, joint attention is highlighted as a complex sequence of gaze patterns as follows: *(see definitions for social gaze concepts in* Table 2-1*).*

Mutual gaze: refers to the instances when the Child_look_robot and Robot_look_child behavior co-occurs (Figure 5-6).

Gaze-following: The pattern — Robot_look_match and Child_look_robot — implies the child is looking at the robot while the robot is looking at the matching card and is noticing the robot's gaze direction to a particular card. And then a subsequent pattern — Robot_look_match > Child_look_match — indicates the orientation of the children's gaze to the robot's direction if the robot looks at the matching card, and the child turns to look at the matching card.

Shared gaze: The co-occurrence of — Robot_look_match and Child_look_match, behavior codes that depict shared attention on an object.

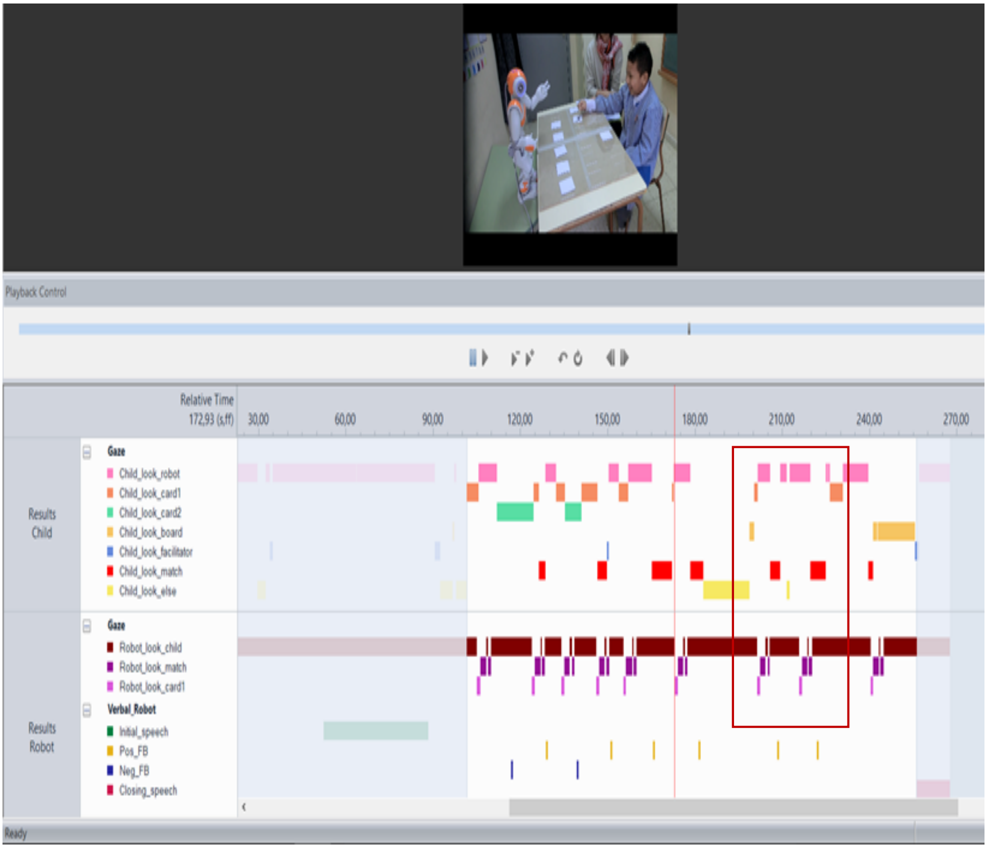


Figure 5-5 Child-NAO gaze coordination during the Gaze condition: A coordinated back-and-forth gaze alternation between the child and the robot gaze and the cards (card1 or matching card) signfies a successful occurrence of joint attention, as illustrated in the visualization with a rectangle.

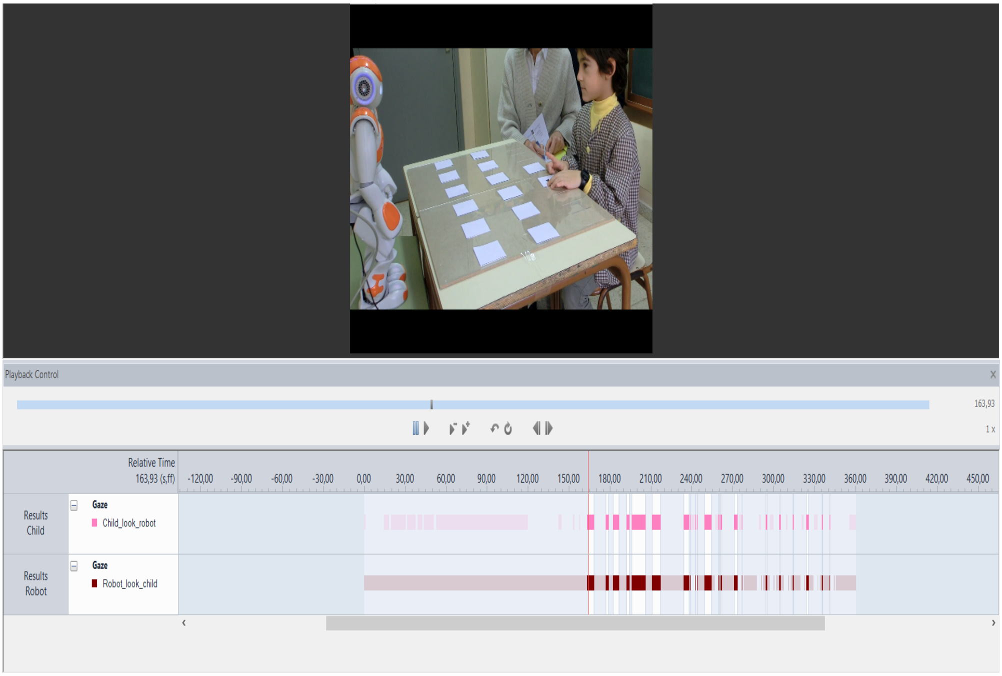


Figure 5-6 Highlighting mutual gazing: co-occurrences of 'Child_look_robot' and 'Robot_look_child' behavior

Table 5-1 Gaze behavioral sequences examined in this study — robot gaze as the initial behavior and the child gaze as the response behavior: highlighted patterns are joint attention gaze sequences. We use the short abbreviations to present the gaze behavior units for child and robot in the results tables.

	Sequential behavior pairs	
Robot Initiated Gaze	Initial Gaze	Response gaze
	Robot_look_child (**RLC)**	Child_look_card1 (**CLC1)**
	Robot_look_match (**RLM)**	Child_look_card1 (**CLC1)**
	Robot_look_card1 (**RLC1)**	Child_look_card1 (**CLC1)**
	Robot_look_child (RLC)	*Child_look_robot (CLR)*
	Robot_look_match (RLM)	*Child_look_robot (CLR)*
	Robot_look_card1 (RLC1)	*Child_look_robot (CLR)*
	Robot_look_child (**RLC)**	Child_look_match (**CLM)**
	Robot_look_match (RLM)	*Child_look_match (CLM)*
	Robot_look_card1 (**RLC1)**	Child_look_match (**CLM)**
	Robot_look_child (**RLC)**	Child_look_card2 (**CLC2)**
	Robot_look_match (**RLM)**	Child_look_card2 (**CLC2)**
	Robot_look_card1 (**RLC1)**	Child_look_card2 (**CLC2)**
	Robot_look_child (**RLC)**	Child_look_board (**CLB)**
	Robot_look_match (**RLM)**	Child_look_board (**CLB)**
	Robot_look_card1 (**RLC1)**	Child_look_board (**CLB)**
	Robot_look_child (**RLC)**	Child_look_facilitator (**CLF)**
	Robot_look_match (**RLM)**	Child_look_facilitator (**CLF)**
	Robot_look_card1 (**RLC1)**	Child_look_facilitator (**CLF)**
	Robot_look_child (**RLC)**	Child_look_else (**CLE)**
	Robot_look_match (**RLM)**	Child_look_else (**CLE)**
	Robot_look_card1 (**RLC1)**	Child_look_else (**CLE)**

5.3.6. **Lag-Based Analysis**

As previously stated, to examine the gaze behavior sequences between the child and the robot, we used the sequential lag method (Bakeman and Gottman, 1997; Bakeman and Quera, 1995, 2011). Lag sequential analysis is based on the contingency tables and is mainly used to determine how events follow each other. We applied the two methods associated with lag analysis to generate the contingency tables—event-based lag and time-based lag — described in the following subsections. The children were divided into two groups based on their frequency of looking at the robot. In the literature, the frequency of looking at the interacting partner is highly linked to engagement during social interaction (Kleinke, 1986). In our case, frequency represents the number of times the gaze occurred regardless of the gaze duration. Children with a "Child_look_robot" count > 10 were assigned to the *Socially attentive group* and those with a "Child_look_robot" count < 10 were assigned to the *Non - socially attentive group*. Twelve (12) children were categorized in *the Socially attentive group* and six (6) children in the *Non - socially attentive group*.

Event-Based Lag Sequential Analysis

For the event-based lag sequential analysis, we used a state lag order= +1 to generate the contingency tables, which means that, if the Robot_look_match was the initial gaze (Lag 0), the frequency of the Child_look_match as the response gaze (Lag 1) that happened right after Robot_look_match was recorded irrespective of time. To generate the contingency tables, we used the Noldus Observer XT 14 lag sequential module. The module calculates the frequency and transition probability. The transition probability refers to the number of transitions for a combination of criterion and target divided by the number of transitions from that criterion. Using the event-based lag, we generated two contingency tables for the *Socially attentive group* and the *Non - socially attentive group*, showing child gaze behaviors following the robot's gaze behaviors.

Time-Based Lag Sequential Analysis

For the time-based lag, we defined lag - based on a two-second interval. Lag zero (0) represents the moment the initial robot gaze behavior occurs, and Lag 1 [0 - 2 s] represents the first gaze behavior of the child that happened zero to two seconds later after lag zero (0). Lag 2 [2 - 4 s] constitutes the first behavior of the child that occurred two to four seconds after lag zero (0). Lag 3 [4- 6 s] and Lag 4 [6 - 8 s] represent intervals of four to six seconds and six to eight seconds, respectively. We generated two tables for each lag—for both the Socially attentive group and the Non - socially attentive group. Thus, there were eight contingency tables based on this method showing child gaze behaviors following the robot gaze behaviors.

Several statistical techniques are involved in lag sequential analysis. First is the likelihood ratio chi-square test by fitting a general log-linear model to the contingency tables to examine whether the cell values are independent. With a significant chi-square test, the next step is to calculate the expected count and adjusted standardized residuals. The adjusted residual indicates whether sequential behaviors are significant, assuming independence (Bakeman and Quera, 1995; Faraone and Dorfman, 1987; Pohl et al.,2016).

5.4. Results

We observed and analyzed thirty-six video recordings of child-robot interaction in two conditions (Gaze/No_Gaze). There was a total of 18 trials in the Gaze condition and 18 trials in the No_Gaze condition. Five (5) children were excluded from the primary analysis - (four) because of technical issues during the recording of the sessions and one of the children because s/he could not play the game.

We compared gaze behavior between two groups of children: The *Socially attentive group vs. Non - socially attentive group*. As highlighted earlier, the children were grouped based on the frequency of occurrence of Child_look_robot behavior during the game. Here, frequency represents the number of times the gaze (Child_look_robot) occurred regardless of the duration of gaze. Twelve (12) children were categorized in the *Socially attentive group* and six (6) children in the *Non - socially attentive group*.

Using the Observer XT Software, we selected the interval between the children picking the first card to the last card, i.e., the entire gameplay session. We excluded the instruction phase and the closing session when the child was saying goodbye and walking out of the room. The duration of the game session varied with different trials, depending on how long the child took to complete the game in both conditions (Gaze condition and No_Gaze condition).

5.4.1. **Task Performance**

Effect of Tutoring_Style on Performance

For this analysis, we conducted repeated tests ANOVA with the style of tutoring (Gaze vs. No_Gaze) as the within-subject factor. We excluded one participant who was considered as an outlier in this analysis. We found no significant difference in duration (s) (p=0.797) and tries (p=0.195) between the Gaze and No_Gaze conditions.

Table 5-2 Performance: descriptive statistics

	Gaze			No_Gaze		
	Mean	SD	N	Mean	SD	N
Tries	15.00	3.89	17	16.88	5.15	17
Duration(s)	185.78	58.88	17	180.38	73.64	17

Task Performance: Socially Attentive Group vs. Non - Socially Attentive Group

After performing an independent sample T-test using SPSS Statistics 25, we found a significant difference in the number of tries to complete the game between the *Socially attentive group* and the *Non - socially attentive group* (p=0.030*). Children in the *Socially attentive group* used significantly fewer attempts to find all the matching cards compared to the children in the *Non - socially attentive group*. However, there was no significant difference in duration between *Socially attentive group* and the *Non - socially attentive group* (p=0.240).

5.4.2. Gaze

Child's Gaze Behavior

We examined the effect of the Tutoring_Style (Gaze vs. No_Gaze conditions) on the children's gaze behavior frequency and durations. We analyzed both the frequency and the percentage duration of different gaze behavior units. We used the percentage duration as children took different durations to finish the game trials—the length varied with every game session. We performed a repeated measure ANOVA with the Tutoring_Style (Gaze vs. No_Gaze) as the within-subject factor using SPSS statistics 25. The sections below report on the results:

Table 5-3 Frequency of child gaze behavior between the Gaze and No_Gaze condition

	Gaze		No_Gaze	
	Mean	SD	Mean	SD
Child_look_robot	11.11	6.01	4.89	3.46
Child_look_card1	16.78	8.37	16.5	5.16
Child_look_card2	10.78	8.59	11.06	5.15
Child_look_else	0.39	0.85	0.83	1.2
Child_look_match	7.11	0.32	7.56	2.26
Child_look_board	4.78	4.88	4.06	3.73
Child_look_facilitator	4.67	3.77	2.11	1.97

Table 5-4 Percentage duration (s) of child gaze behavior between the Gaze and No_Gaze conditions

	Gaze		No _Gaze	
	Mean	SD	Mean	SD
Child_look_robot	17.37	11.06	5.11	4.18
Child_look_card1	24.17	5.35	28.46	5.2
Child_look_card2	27.56	13.11	36.34	7.43
Child_look_else	0.84	2.65	0.62	1.13
Child_look_match	17.23	6.62	21.24	7.48
Child_look_board	5.96	4.78	5.41	6.24
Child_look_facilitator	6.86	5.91	2.81	3.45

Frequency: There was a significant difference in the number of occurrences of the child looking at the robot during the game between the Gaze and No_Gaze conditions (p=0.001*). Children looked significantly more frequently to the robot tutor during the Gaze condition compared to the No_Gaze condition. There were significantly more occurrences of the child looking at the facilitator during the Gaze condition compared to the No_Gaze condition(p=0.007*). However, we found no significant differences in the frequency of occurrence for the rest of the categories of gaze behavior: Child_look_card1, Child_look_card2, Child_ look_board, and Child_look_else.

Duration: The interval duration is the time between the child picking the first card and taking the last one, i.e., the play session. The percentage shows the relative period of gaze to the total time the child used to play the game. There was a significant difference in percentage duration for looking at the robot between the Gaze and No_Gaze conditions (p=0.001*). Children looked at the robot significantly longer when the tutor provided gaze cues than without the gaze cues. Similarly, there was a significant difference in percentage durations of the child looking at card1 between the Gaze and No_Gaze conditions (p=0.009*). Children looked significantly longer at card1 when the tutor did not provide gaze cues than when the tutor provided gaze cues.

Equally, we found a significant difference in the percentage duration of the child looking at card2 between the two conditions (p=0.005*). Children looked significantly longer at card2 when the tutor was not helping than when the tutor was helping. A plausible explanation for this is given in the discussions section. We also found a significant difference in the percentage duration of the child looking at the facilitator between the Gaze and No_Gaze conditions (p=0.011*). Children looked significantly longer to the facilitator during the Gaze condition than in the No_Gaze condition. There was no significant difference in percentage durations of the Child_look_match, the Child_look_board, and the Child_look_else between the Gaze and No_Gaze conditions.

Effect of Tutoring_Style on Mutual Gaze

To highlight mutual gaze from the observations, we selected the intervals when the behaviors Child_look_robot and Robot_look_child behaviors co-occur, as shown in Figure 5-6 using Observer XT 14. Table 5-5 illustrates the descriptive statistics for the frequency and duration of mutual gaze (MG) patterns in both the Gaze and No_Gaze conditions. After performing a repeated measure ANOVA using SPSS Statistics 25, we found significant mean differences in frequency (p=0.001*) and duration (s) (p=0.002*) of mutual gaze between the Gaze and No_Gaze conditions *(see* Appendix D *for mutual gaze data for each child during Gaze and No_Gaze Conditions).*

Table 5-5 Frequency and duration (s) of mutual gaze between the Gaze and No_Gaze conditions

	Gaze Condition (N=18)				No - Gaze Condition (N=18)			
			Range				Range	
	Mean	SD	Min	Max	Mean	SD	Min	Max
Duration (s)	22.17	15.55	0.93	52.35	9.56	8.13	1.27	30.60
Frequency	11.11	6.94	2	26	5.44	3.70	1	12

In the next sections, we compare the mutual gaze, gaze-following patterns, and joint attention gaze sequences between the two groups of children.

Mutual Gaze: Socially Attentive Group vs. Non - Socially Attentive Group

Table 5-6 shows the descriptive statistics regarding the frequency and duration(s) of the mutual gaze for children in the *Socially attentive group and the Non - socially attentive group.*

Table 5-6 Frequency and duration (s) of mutual gaze for the *Socially attentive group* vs. *Non - socially attentive group*

	Socially attentive group (N=12)				Non - socially attentive group (N=6)			
			Range				Range	
	Mean	SD	Min	Max	Mean	SD	Min	Max
Duration(s)	28.50	15.17	0.93	52.35	9.52	5.28	2.54	15.18
Frequency	14.33	6.16	3	26	4.67	2.34	2	8

We conducted an independent sample T-test using SPSS Statistics 25 to compare the (frequency and duration of mutual gaze) for children in the *Socially attentive group* and those in the *Non - socially attentive group*. There was a significant difference in the frequency of mutual gaze between children in the *Socially attentive group* (*M*=14.33, *SD*= 6.16) and those in the *Non - socially attentive group* (*M*=4.67, *SD*=2.34); *p*=0.002*; two-tailed, assuming equal variances. We also found a significant difference in duration of mutual gaze between the children in the *Socially attentive group* (*M*=28.50 s, *SD*=15.17) and those children in the *Non - socially attentive group* (*M*=9.52 s, *SD*=5.28); *p*=0.010*; two-tailed, assuming equal variances).

Gaze-Following: Socially Attentive Group vs. Non - Socially Attentive Group

To highlight gaze-following from the video observations, we selected the intervals when the following child and robot behaviors co-occur: Child_look_robot and Robot_look_match. This pattern implies the child is looking at the robot while the robot is looking at the matching card and is noticing the robot's gaze direction to a particular card. This co-occurrence is the crucial feature in the pattern of joint attention and a subsequently shared view pattern formed by the co-occurrence of Robot_look_match and Child_look_match. This pattern implies the robot and the child are looking at the correct card match at the same time.

Table 5-7 Frequency and duration (s) of gaze-following for *Socially attentive group* vs. *Non – socially attentive group*

	Socially attentive group (N=12)				Non - socially attentive group (N=6)			
			Range				Range	
	Mean	SD	Min	Max	Mean	SD	Min	Max
Duration(s)	41.30	17.60	16.14	72.04	12.12	6.23	4.50	20.19
Frequency	13.83	5.38	7	23	5.67	2.42	2	9

We conducted an independent sample T-test in SPSS to compare the (frequency and duration of gaze-following) between the children in the *Socially attentive group* and those in the *Non - socially attentive group*. We found significant differences in the frequency of gaze-following-related patterns between children in the *Socially attentive group* (M=13.83, SD=5.38) and those in the *Non - socially attentive group* ((M=5.67, SD=2.42); p=0.003*; two-tailed, assuming equal variances). There were also significant differences in the duration(s) of gaze-following between children in the *Socially attentive group* (M=41.30 s, SD=17.60) and children in the *Non - socially attentive group* (M=12.12 s, SD=6.23 p=0.001*; two-tailed) assuming equal variances.

Patterns of Joint Attention Behavior: Gaze Sequences

We performed both event-based and time-based lag-sequential analyses to examine the gaze behavior sequences associated with joint attention occurrence in both the *Socially attentive* and Non - *socially attentive groups*. We focused on the gaze behavioral sequences in Table 5-1, with the robot gaze as the initial behavior and the child's gaze as the response behavior, and the highlighted patterns are joint attention gaze sequences. We use the ***short abbreviations*** as illustrated in Table 5-1 to present the gaze behavior units for the child and the robot in the contingency — results tables.

Event-Based Lag Sequential Analysis

Table 5-8 and Table 5-9 are the contingency tables for the *Socially attentive group* and the *Non - socially attentive group*, respectively.

Table 5-8 Observed counts: *Socially attentive group* (N=12)

Initial Gaze	Response Gaze						Total
	ClC1	CLR	CLM	CLC2	CLB	CLF	
RLC	55	50	14	28	15	16	178
RLM	6	36	33	29	15	13	132
RLC1	3	22	13	10	1	3	52
Total	64	108	60	67	31	32	362

Table 5-9 Observed counts: *Non - socially attentive gro*up (N= 6)

Initial_Gaze	Response_Gaze							Total
	ClC1	CLR	CLM	CLC2	CLB	CLF	CLE	
RLC	64	14	6	19	7	3	1	114
RLM	10	3	10	41	8	7	0	79
RLC1	5	3	11	10	1	1	0	31
Total	79	20	27	70	16	11	1	224

The likelihood ratio chi-square was 65.622 (df = 12) for the *Socially attentive group* table, and the likelihood ratio chi-square for the *Non - socially attentive group* table was 75.509 (df = 12). Therefore, cell values are significant and not distributed by chance.

Table 5-10 and Table 5-11 indicate the results of the event-lag sequential analysis for the *Socially attentive group* and the *Non - socially attentive group*, respectively, when the robot gaze is the initial behavior, respectively. The highlighted in italic indicate a significant association of the sequential patterns.

Table 5-10 Observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on event lag for the *Socially attentive group*

Initial_Gaze		Response_Gaze					
		ClC1	CLR	CLM	CLC2	CLB	CLF
RLC	Count	*55*	50	14	28	15	16
	Expected Count	*31.5*	53.1	29.5	32.9	15.2	15.7
	Adjusted Residual	*6.5*	-.7	-4.4	-1.3	-.1	.1
RLM	Count	6	36	*33*	29	15	13
	Expected Count	23.3	39.4	*21.9*	24.4	11.3	11.7
	Adjusted Residual	-5.0	-.8	*3.3*	1.3	1.4	.5
RLC1	Count	3	*22*	13	10	1	3
	Expected Count	9.2	*15.5*	8.6	9.6	4.5	4.6
	Adjusted Residual	-2.4	*2.1*	1.8	.1	-1.8	-.8

Table 5-11 Observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on event lag for the *Non - socially attentive group*

Initial_Gaze		Response_Gaze						
		ClC1	CLR	CLM	CLC2	CLB	CLF	CLE
RLC	Count	64	14	6	19	7	3	1
	Expected Count	40.2	10.2	13.7	35.6	8.1	5.6	.5
	Adjusted Residual	6.7	1.8	-3.2	-4.8	-.6	-1.6	1.0
RLM	Count	10	3	10	41	8	7	0
	Expected Count	27.9	7.1	9.5	24.7	5.6	3.9	.4
	Adjusted Residual	-5.2	-2.0	.2	4.9	1.3	2.0	-.7
RLC1	Count	5	3	11	10	1	1	0
	Expected Count	10.9	2.8	3.7	9.7	2.2	1.5	.1
	Adjusted Residual	-2.4	.2	4.3	.1	-.9	-.5	-.4

Time-Based Lag Sequential Analysis

Socially attentive group: The likelihood chi-square test showed significant dependence between the behavior pairs for Lag 1 [0 -2 s], Lag 2 [2 - 4 s], and Lag 3 [4 -6 s] tables. However, for Lag 4 [6 -8 s], the likelihood chi-square test was not significant. Thus, the children's responses occurred at between zero seconds up to six seconds after the tutor's initial behavior.

Table 5-12 Goodness-of-fit tests: likelihood ratio chi-square results for each time-based lag for the *Socially attentive group*

		Value	df	Sig.
Lag 1 [0 -2 s]	Likelihood Ratio	24.816	12	*.016*
	Pearson Chi-Square	23.376	12	*.025*
Lag 2 [2 - 4 s]	Likelihood Ratio	47.149	12	*.000*
	Pearson Chi-Square	46.899	12	*.000*
Lag 3 [4 -6 s]	Likelihood Ratio	41.709	12	*.000*
	Pearson Chi-Square	38.530	12	*.000*
Lag 4 [6 -8 s]	Likelihood Ratio	5.008	12	.958
	Pearson Chi-Square	3.977	12	.984

Table 5-13, Table 5-14, and Table 5-15 shows the observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on the time -based lag for the *Socially attentive group* for Lag 1 [0 -2 s], Lag 2 [2 - 4 s] and Lag 3 [4 -6 s], respectively. The significant behavior patterns observed are highlighted in italic.

Table 5-13 Observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on the time lag for the *Socially attentive group* — Lag 1 [0 -2 s]

Initial_Gaze		Response_Gaze						
		ClC1	CLR	CLM	CLC2	CLB	CLF	CLE
RLC	Count	*17*	59	17	37	16	16	0
	Expected Count	*9.9*	54.5	30.6	37.1	14.6	15.0	.3
	Adjusted Residual	*2.9*	.9	-3.4	.0	.5	.3	-.7
RLM	Count	9	60	45	43	21	19	1
	Expected Count	12.1	66.6	37.4	45.3	17.9	18.3	.4
	Adjusted Residual	-1.2	-1.3	1.8	-.5	1.0	.2	1.2
RLC1	Count	3	41	28	29	6	9	0
	Expected Count	7.1	39.0	21.9	26.6	10.5	10.7	.2
	Adjusted Residual	-1.8	.5	1.7	.6	-1.7	-.6	-.6

Table 5-14 Observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on the time lag for the *Socially attentive group* — Lag 2 [2 - 4 s]

Initial_Gaze		Response_Gaze						
		ClC1	CLR	CLM	CLC2	CLB	CLF	CLE
RLC	Count	*42*	49	12	30	13	14	0
	Expected Count	*28.6*	45.9	22.4	31.4	16.1	15.3	.4
	Adjusted Residual	*3.5*	.7	-3.0	-.4	-1.0	-.4	-.8
RLM	Count	27	47	17	31	14	17	0
	Expected Count	27.4	43.9	21.4	30.0	15.4	14.6	.4
	Adjusted Residual	-.1	.7	-1.3	.3	-.5	.8	-.8
RLC1	Count	4	21	*28*	19	14	8	1
	Expected Count	17.0	27.2	*13.3*	18.6	9.5	9.1	.2
	Adjusted Residual	-4.0	-1.6	*5.0*	.1	1.7	-.4	1.8

Table 5-15 Observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on the time lag for the *Socially attentive group* — Lag 3 [4 -6 s].

Initial_Gaze		Response_Gaze						
		ClC1	CLR	CLM	CLC2	CLB	CLF	CLE
RLC	Count	*62*	33	27	18	17	20	2
	Expected Count	*50.8*	45.7	21.4	27.8	14.5	17.9	.9
	Adjusted Residual	*2.4*	-2.9	1.7	-2.7	.9	.7	1.6
RLM	Count	50	42	16	28	12	15	0
	Expected Count	46.3	41.6	19.5	25.3	13.2	16.3	.8
	Adjusted Residual	.8	.1	-1.1	.8	-.5	-.4	-1.1
RLC1	Count	7	*32*	7	*19*	5	7	0
	Expected Count	21.9	*19.7*	9.2	*11.9*	6.2	7.7	.4
	Adjusted Residual	-4.2	*3.6*	-.9	*2.5*	-.6	-.3	-.7

Non - socially attentive group: The likelihood chi-square test showed significant dependence between the behavior pairs for the Lag 2 [2 - 4 s] only, and no other lags were significant.

Table 5-16 Goodness-of-fit tests — likelihood ratio chi-square results for each time-based lag for the *Non - socially attentive group*

		Value	df	Sig.
Lag 1 [0 -2 s]	Likelihood Ratio	15.147	12	.234
	Pearson Chi-Square	16.147	12	.185
Lag 2 [2 - 4 s]	Likelihood Ratio	33.620	12	*.001*
	Pearson Chi-Square	31.633	12	*.002*
Lag 3 [4 -6 s]	Likelihood Ratio	11.338	12	.500
	Pearson Chi-Square	9.813	12	.632
Lag 4 [6 -8 s]	Likelihood Ratio	6.359	12	.897
	Pearson Chi-Square	5.866	12	.923

Table 5-17 shows the observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on the time -based lag for the *Non - socially attentive group* for Lag 2 [2 - 4 s]. The highlighted values in italic show a significant association between sequential behavioral patterns.

Table 5-17 Observed counts, expected count, adjusted standardized residuals of the children's gaze in response to the robot's gaze based on the time lag for the *Non - socially attentive group* — Lag 2 [2 - 4 s]

Initial_Gaze		Response Gaze						
		ClC1	CLR	CLM	CLC2	CLB	CLF	CLE
RLC	Count	37	15	12	22	5	3	1
	Expected Count	27.3	11.2	10.4	34.1	5.6	5.6	.8
	Adjusted Residual	2.9	1.6	.7	-3.3	-.3	-1.5	.3
RLM	Count	25	9	7	32	4	3	1
	Expected Count	23.2	9.6	8.9	29.1	4.8	4.8	.7
	Adjusted Residual	.5	-.2	-.8	.8	-.5	-1.0	.5
RLC1	Count	6	4	7	31	5	8	0
	Expected Count	17.5	7.2	6.7	21.9	3.6	3.6	.5
	Adjusted Residual	-3.8	-1.5	.1	2.8	.9	2.8	-.8

5.4.3. **Children Perceptions**

In the end, the facilitator conducted a post-experiment interview with the children. The facilitator asked the children the following questions:

Q1: Was the game difficult? [Not at all difficult/A bit difficult/Not easy or difficult/Easy/Very easy]

Q2: Have you played the game before? YES/NO

Q3: Did the tutor help you? YES/NO; if yes, how did the tutor help you?

In the following sections, we present the children's answers to the above interview questions:

Q1: Based on the children's answers, there was no significant difference in regards to how the children perceived the task difficulty in the Gaze and No_Gaze conditions.

Q2: All children said they had played the game before, but with different images, for example, with other animals, cars, fruits, or houses.

Q3: Most of the children said they noticed the tutor's help. However, different children interpreted the cues from the robot differently, and most of the children answered that the robot tutor helped them with verbal feedback. Five (5) children, however, noted and followed the gaze cues from the robot tutor to find the matching cards on the board. In this section, we include some excerpts from the children regarding how they interpreted the tutor's verbal feedback and gaze cues.

Five children in the Gaze condition said they noticed the tutor's gaze cues while the tutor was looking at different cards on the game board. Below are the responses from children who saw the robot's gaze cues.

> *P02 [Gaze]: "He looked there, and I knew, he was looking over there because the card was there, it sometimes helps and sometimes not [points out] but with the gaze."*

> *P04 [Gaze]: "He moved his head to the sides — he moved his hands. I do not know, because I thought that if he looked, I thought 'it must be this one' But sometimes he fooled me and sometimes not."*

> *P08 [Gaze]: "He showed me the cards."*

> *P12 [Gaze]: "It helped me a lot, for example, he makes with the face where your pair was."*

> *P22 [Gaze]: "So, it looked at the right card — to give me a clue. With the face looking at the right cards."*

A number of children also noticed the movements of the robot's head. However, they said the robot was watching or looking at them while playing and were, therefore, not able to take advantage of the gaze cues from the robot tutor. Below are some excerpts from children's answers:

> *P05 [Gaze]: "He was looking at how I performed. When I took the two equals, he said to me, "fantastic, fantastic, fantastic."*

> *P06 [Gaze]: "Watch me and tell me things when I did not guess, 'try again,' and when I guessed, 'very well.'"*

> *P20 [Gaze]: "He was looking at how I was doing."*

As earlier mentioned, most children in the Gaze condition said the robot was using verbal feedback to help them. Below are some excerpts from children that interpreted the verbal feedback to be the modality the tutor was using to help.

> *P01 [Gaze]: "If I do not guess the right one, said: 'try again.'"*

> *P07 [Gaze]: "He said 'very well,' he gave me clues, when I was approaching [the right one] he said 'very well,' and that was a clue."*

P10 [Gaze]: "Telling me yes or no."

P11 [Gaze]: "He told me if I had done it right or not, as before."

P16 [Gaze]: "Telling me if I had done it right."

P17 [Gaze]: "He explains to me the instructions. And he told me if I did it right and if I didn't try again."

Remarkably, also in the No_Gaze condition, we found that children interpreted the verbal feedback from the robot as helpful clues.

P01 [No_Gaze]: "If I was right, he said 'great' or 'very good,' and if it did not guess he said 'try again.'"

P02 [No_Gaze]: "If I did it right, he said 'fantastic' or something like that."

P08 [No_Gaze]: "Tell me 'very well' or 'Try again.'"

P11 [No_Gaze]: "Telling me if I had done it well. If I had done right, he said 'very well.'"

P12 [No_Gaze]: "He was telling me when I was getting another: when it was not the right one, he said "try again" and when I got the right one 'Very good.'"

5.5. Discussion

The discussion presented in this chapter is a step forward in determining how sequential gaze coordination between the robot and the child affects awareness of the tutor's intention, social engagement performance, and the quality of interaction with the robot. To study the sequential occurrences of the gaze behavior pairs, we used both the event-based and the time-based lag methods. We compared sequential behavior pairs for two groups: The *Socially attentive group*—children with higher events of looking at the robot—and the *Non - socially attentive group*—children with lower events of looking at the robot behavior. The 'child_look_robot' behavior indicates instances when the child looks towards the robot's face. As is depicted in literature, looking at the interacting partner in this case, "look at the robot" measure is linked positively to engagement with the robot (Kleinke, 1986). Looking at the robot, in this case, signals social engagement during the tutoring interactions.

To conduct sequential analysis, we focused only on the transitions from the robot's behavior to a child's behavior as a following, as the other perspective—the robot gaze behavior is deterministic—always occurs in the same order. We generated two event-based contingency tables, one for the *Socially attentive* *group* and the other *Non - socially attentive group*, and eight time-based contingency tables, four for the *Socially attentive group* and four for the *Non - socially attentive group*: showing children's gaze behaviors following the robot's gaze behaviors.

In *(H1),* we projected the tutoring style (Gaze vs. No_Gaze) would influence task performance. We found no significant difference in the duration the children used to complete the task between the Gaze condition and the No_Gaze condition *(H1.1).* Similarly, we found no significant difference in the number of tries between the Gaze condition and the No_Gaze condition *(H1.2)* (Table 5-2). We identified the plausible reason for this from children's perceptions; most of the children interpreted the verbal cues as the helping behavior, and, in this case, the robot provided verbal feedback in both conditions. In Chapter 6, we provide further discussion concerning combining gaze with spoken interaction to build effective robotic tutors for children.

In *(H2)*, we projected that tutoring style (Gaze /No_Gaze) would influence the children's gaze-based interaction during play. The findings support the hypothesis. We found that children looked significantly more frequently and for longer durations at the robot tutor's face in the Gaze condition compared to the No_Gaze tutoring style *(H2.1)*. The mean frequencies and durations for the child's gaze behavior are in (Table 5-3) and (Table 5-4). Frequencies represent the number of times gaze occurred regardless of the duration of the gaze. Consequently, we found more occurrences and longer durations of mutual gaze during the Gaze condition than in the No_Gaze condition *(H2.2).* Children established eye contact more often and for longer durations with the robot tutor with the presence of gaze cues from the robot tutor than without the gaze cues from the tutor (Table 5-5).

In *(H3*), we projected that the child-robot coordinated gaze would influence the children's performance. The findings support our hypothesis; There were more occurrences and longer durations of mutual gaze and gaze-following patterns in the *Socially attentive group* compared to the *Non - socially attentive group* (Table 5-6 and Table 5-7) (*H3.1)*. Based on the event-based lag analysis, we identified significant gaze behavior sequences associated with joint attention (Table 5-10) in the *Socially attentive group (H3.1).* However, we found no significant gaze behavior patterns associated with joint attention for the *Non - socially attentive group* (Table 5-11). Consequently, we found that children in the *Socially attentive group* performed significantly better —using fewer tries to complete the game—when compared to children in the *Non - socially attentive group (H3.2).*

Table 5-10 indicates the results of the *event-based lag analysis* for the *Socially attentive group*. The highlighted values in italic indicate a significant association of the sequential patterns. We identified the following significant gaze sequences for the *Socially attentive group*:

Robot_look_child > Child_look_card1;

Robot_look_card1 > Child_look_robot;

Robot_look_match > Child_look_match.

These sequences can be interpreted as follows:

"Robot_look_child > Child_look_card1" —indicates the start of an interaction instance. The robot is looking at the child as if waiting for the child's move, and the child looks at the first card (referred here as card1).

"Robot_look_card1 > Child_look_robot"—suggests the child sees the gaze of the robot shift when the robot turns to look at the first card the child has picked, and the child looks towards the robot.

"Robot_look_match > Child_look_match" —indicates gaze-following to the gaze orientation of the robot tutor while the tutor shifts its gaze (head orientation) to the matching card, and the child follows the gaze to look at the matching card as well.

Thus, the potential interpretation of these findings is the children in the *Socially attentive group* responded to the robot's behaviors by looking at the robot when the robot looked at the first card they turned—referred to in this article as card1—and then followed the gaze orientation of the robot tutor to the matching card. Therefore, this sequence can be interpreted as a successful occurrence of joint attention or a shared gaze. It could be because of the child's implicit/explicit awareness concerning the robot's cues or the child implicitly or explicitly following the robot's gaze.

Table 5-11 shows the results of the *event-based lag sequential analysis* for the *Non - socially attentive group.* The highlighted values in italic indicate a significant association of the sequential patterns. We identified the following significant gaze sequences for the *Non - socially attentive group*:

Robot_look_child > Child_look_card1;

Robot_look_match > Child_look_card2;

Robot_look_card1 > Child_look_match.

These sequences can be interpreted as follows:

"Robot_look_child > Child_look_card1"—As earlier mentioned, this indicates the start of an interaction instance.

"*Robot_look_card1 > Child_look_match*" —indicates the child does not notice the robot's gaze shift when the robot turns to look at the first card the child has picked, and the child moves to find a matching card without looking at the robot.

"*Robot_look_match > Child_look_card2*"—shows unsuccessful gaze-following or lack of gaze-following, where the child misses the cues from the robot tutor while the tutor is looking at the matching card and, instead, selects a different card (card2 is when a child picks a second card that is incorrect).

The potential interpretation is that the children in the *Non - socially attentive group* were already gazing at the matching card when the robot looked at the first card they turned—referred to in this article as card1—and, therefore, did not follow the gaze orientation of the robot tutor to the matching card. There is not a coordinated exchange of non-verbal cues for supporting the right selection of the card. Therefore, this sequence can be interpreted as a lack of joint attention or a shared gaze. It could be because of the explicit unawareness of the child concerning the robot's cues or the failure in following the robot's gaze.

Table 5-13, Table 5-14, and Table 5-15 indicate the results of the *time-based lag sequential* analysis for the *Socially attentive group*. The highlighted values in italic indicate a significant association of the sequential patterns. For the *Socially attentive group*, the findings from the *time-based lag analysis* were as follows. First, Lag 1 [0 -2 s], Lag 2 [2 - 4 s], and Lag 3 [4 - 6 s] contingency tables showed significance in the likelihood ratio chi-square test; however, Lag 4 [6 - 8 s] was not significant. The interpretation is that any association between an initial and response gaze behavior for *the Socially attentive group* occurred from zero seconds up to six seconds and disappeared after 6 seconds.

Table 5-16 and Table 5-17 indicate the results of the *time-based lag sequential analys*is for the *Non - socially attentive group*. The highlighted values in italic indicate a significant association of the sequential patterns. For the *Non - socially attentive group*, only the Lag 2 [2 - 4 s] contingency table showed significance in the likelihood ratio chi-square test. Furthermore, the likelihood ratio chi-square test for all the other Lags were not significant. The interpretation is that any association between an initial and response gaze behavior in the *Non - socially attentive group* occurred from two seconds up to four seconds and disappeared after four seconds.

5.6. Summary

This chapter contributes new insights regarding how the lag-based method can be applied to analyze non-verbal communication in human-robot interaction. There are four significant findings from this chapter:

First, we found that more occurrences and higher durations of coordinated - mutual gaze and gaze following pattern - increased the children's social engagement with the robot tutor and their performance. *Second,* the event-based analysis shows that the robot's gaze significantly affects child gaze behavior during the interaction. *Third,* we found significant gaze sequences associated with joint attention in the *Socially attentive group*. Therefore, we have established that appropriate sequences of the dyad's gaze behaviors between a child and a robot during a tutoring interaction increase children's probability of successfully taking advantage of the robot's cues.

Fourth, proper timing of response of the child's gaze to the robot's gaze can lead to effective interactions between a child and a robot tutor. We found that the response gaze of children that were able to take advantage of the tutor's social cues effectively occurred from 0 to 6 seconds after the initial robot gaze. Therefore, understanding the dynamic nature of the robot-child gaze patterns adds to robot behavior designs, which can lead to more positive experiences during child–robot interaction.

For this analysis, twelve children - with higher events of look at robot behavior were pooled into the *Socially attentive grou*p, and six children - with lower events of look at robot behavior were pooled into the *Non - socially attentive group*. Thus, a limitation of these findings is that we could not examine whether the gaze sequences are consistent throughout each child's entire interaction. In the appendix section (Appendix E), we provide 18 gaze visualizations for all child- robot observations during the gaze condition showing how the child's gaze interacts with that of the robot in the entire interaction.

In Chapter 6, we discuss the main findings and the limitations of this work, including future research directions.

Chapter 6. **General Discussion**

In this dissertation, we present a body of work aimed at understanding human-robot gaze mechanisms to inform the robot's behavior design with effective gaze-based interaction to foster performance, social engagement, and mutual coordination in a tutoring setting. Towards this goal, we conducted three empirical studies (chapters 3, 4, and 5) to explore interactive gaze behaviors of mutual coordination with increasing complexity —from noticing of gaze to mutual gaze to gaze-following and complex sequences of gaze behavior within human-robot tutoring interactions. The empirical studies are well-controlled experiments in which participants interact with a tutor (human or robot) in the first study and robots with different levels of interactivity in the second and third studies in a designed board game task as a tutoring activity. The findings from this dissertation contribute new design implications for interactive and effective robot gaze-behaviors to improve learning performance and facilitate positive experiences in robot tutoring settings. This chapter presents the main findings and limitations of this work and open challenges that inform future research directions.

The rest of this chapter is structured as follows: Section 6.1 provides a discussion of the main findings of this work, Section 6.2 describes the research limitations, and Section 6.3 highlights open questions that inform future research directions on the topic of designing credible and effective gaze behavior mechanisms for robot interventions in tutoring and social training.

6.1. Main Findings

In the following sections, we discuss the main findings from the user studies described in chapters 3, 4, and 5. These include the experimental framework — interaction setting and board game; design guidelines for effective gaze-based interaction in robot tutoring; legibility of gaze behavior — comparing a human and robot tutor; sequential gaze analysis with lag-based methods; and laboratory vs. field user studies.

6.1.1. Experimental Framework

We contextualized the user studies detailed in chapters 3, 4, and 5 in a play situation based on a board game activity "Memory." The use of a board game that can be played by people with different developmental abilities and profiles is an important part of this work as it adds to the research line on social inclusion. Our motivation behind the design was that many settings used for therapeutic training and educational purposes are in the form of board games and other settings when the tutor and tutee are on the opposite sides of a table — as is the case in this research.

We validated the experimental setup with user studies involving adults and children (typically developing children) interacting with a robot in a tutoring context. However, this framework can be adapted and extended to engage children with autism disorder and older adults in activities through gaze interaction. Besides, the design of the board game targets a human and a robot's simultaneous gaze interaction to achieve a collaborative task, which is a common deficit with children with autism spectrum disorders.

Previous studies have also used games in social training of children with autism (Albo-Canals et al., 2018; Barakova, Bajracharya, Willemsen, Lourens, and Huskens, 2014). In such settings, games are designed to encourage children to engage in training practices better. Therefore, additional gaze-based interactions can increase this engagement. In elderly settings, Perugia et al. (2017) and Perugia et al. (2018) used cognitive games and robots to engage the elderly by mental stimulation on an emotional level. They found that the gaze is a crucial component of both cognitive and emotional engagement. Thus, gaze interaction can be included in the design of robots for the elderly, especially those with cognitive impairments.

Therefore, the interaction setting and the game used in the dissertation contribute a plausible testbed framework for designing gaze behaviors in general education — as a tool in tutoring and therapy.

6.1.2. Design Guidelines for Gaze-Based Interactions in Tutoring

The findings from the user studies we have conducted contribute new design recommendations on designing effective gaze behavior for robots, particularly in the context of tutoring and social training. We outline them below:

- *People (both adults and children) are sensitive to the gaze directions of a robot tutor. They can notice and interpret such cues accurately to improve their task performance — accuracy during robot tutoring activities.* The results in (Chapter 3 and Chapter 4) show that when a robot tutor provides gaze hints during tutoring interactions, people improve their task performance - use significantly fewer tries to complete the task.

- *People perform better when they are explicitly aware of robot tutor gaze cues in a tutoring interaction than when they don't notice the gaze cues.* In Chapter 3, the findings showed that persons that noticed the gaze hints of the tutor performed better during the interaction compared to those who did not recognize the cues (*see* Table 3-3; Table 3-5). In a follow-up study (Chapter 4), we also found that children that noticed the gaze cues of the tutor performed better than children who did not notice the helping cues. Further, in Chapter 5, we found that children in the *Socially attentive group* - with a higher frequency of look at the robot behavior, performed better compared to the children in the *Non-socially attentive group* -children with lower events of look at the robot behavior.
- *Coordinated child-robot gaze (mutual gaze and gaze following) increases children's knowledge of the helping cues from the robot during a tutoring interaction and the probability of taking advantage of the robot's cues successfully*. In chapters 4 and 5, we found that more occurrences and longer durations of mutual gaze and gaze following patterns for children in the *YES group* — children who noticed the helping cues from the robot tutor compared to children in the *NO group* — children who did not notice the helping cues. In turn, we found that when children are aware that the robot tutor tries to help with the gaze hints, they look more to the robot to take advantage of its cueing (*see* Table 4-9; Table 4-10; Figure 4-8; Table 5-6 and Table 5-7). In Chapter 5, we found that children in the *Socially attentive group* - had higher frequency and longer durations of mutual gaze and gaze following patterns compared to the children in the *Non-socially attentive group*.
- *More salient (Overt) gaze from the tutor may lead to better performance as the social cues are more noticeable compared to covert vs. no gaze in tutoring scenarios.* In Chapter 3, the results showed that participants noticed the gaze cues (awareness of the tutor's intention to help) more in the robot condition than in the human condition (*see* Figure 3-5 (e)). We identified that the NAO robot provides more salient gaze behavior compared to the gaze of a human tutor, which is subtle, and thus the robot gaze is much easier to notice due to several factors which we discuss in detail in Chapter 3; 3.5. A summary of these factors is in the following section 6.1.3. Thus, if a tutor provides (more) evident gaze cues, participants perform better than with no cues or covert cues during a tutoring activity.
- *Designing a robot with a dynamic gaze increases the effectiveness of the robot tutor as a helping agent*. In chapter 5, the results show that appropriate sequences of the dyad's gaze behaviors between a child and a robot can lead to effective interactions between a child and a robot tutor (see Table 5-10). Furthermore, proper timing of the response of the child's gaze to the robot's gaze can lead

to positive interactions between a child and a robot tutor (*see* Table 5-12). Thus, a robot tutor can influence the flow of the child's actions positively—if the child interprets the social cues appropriately during the interaction —and, in turn, improve the task execution and the play experience.

6.1.3. **Legibility of Gaze Behaviors**

Overall, the findings in this dissertation demonstrate that effective gaze interaction during human-robot tutoring can enhance performance and promote positive experiences for human partners (adults or children). These findings contribute to the growing body of work showing that gaze can help build effective interactions between humans and robots in diverse applications (Admoni and Scassellati, 2017; Boucher et al., 2012; Broz et al., 2012; Moon et al., 2014; Mutlu et al., 2006, 2009). An important finding from our work (Chapter 3) is that participants who noticed the robot tutor's gaze cues performed significantly better than those who did not notice the helping gaze cues. However, prior work has shown that the gaze is a powerful communicative cue without the observer's awareness. For example, in an object selection game, Mutlu et al. (2009) showed that participants could read and interpret unconscious leakage cues from the gaze of the robot directed toward objects or locations in the environment. Additionally, the authors showed that gaze cue led to better performance and even better performance with Robovie, which is more human-like than Geminoid. Palinko et al. (2015) found that mutual gaze can implicitly facilitate effective turn-taking in a dictation scenario using the I-cub robot. The plausible reason for this result is the use of explicit, overt cues from the NAO robot. As previously mentioned, the NAO platform used for this research has no articulated gaze. Therefore, it has to turn its entire head to communicate gaze while the robots used by Mutlu et al. (2009) and Palinko et al. (2015) have articulated eyes. Such robots can provide covert signals, with a gaze cueing that is effective even without reaching conscious awareness, while NAO achieves gaze interaction with an overt cue.

Furthering our discussion on the role of overt vs. covert cues in learning scenarios, results from Chapter 3 shows substantial differences between the human and the robot tutor concerning performance, participants' perceptions, and participants' awareness of the tutor's gaze cues. 20% of participants interacting with the human tutor noticed the helping gaze cues, whereas 60% of participants interacting with the robot recognized the tutor's gaze hints (Figure 3-5 (e)). As a result, people performed better when tutored by a robot (Table 3-1). Further, we found that persons that noticed the gaze hints of the tutor (either human or robot) performed better during the interaction compared to those who did not (Table 3-3; Table 3-5). Thus, we identified that the findings are not entirely dependent on the nature of the tutors (human or robot) but instead on the legibility of their gaze behaviors, which might depend on different factors. We outlined some of the factors in Chapter 3, section 3.5.

These include the need to rotate the robot's head to communicate gaze, owing to NAO's lack of articulated eyes, the sound produced during the motion of the head, and the curiosity the robot awakens (novelty effect). Other factors consist of the dynamics of the human tutor's gaze, which might look unnatural when scripted, and the intimacy regulation aspect (Argyle et al., 1973, 1994). This occurs during human social interactions. Therefore, the robot's gaze behavior for the NAO platform is achieved with large head movements. In contrast, the human tutor's gaze is much subtler, based more on the eye movements than on head motion.

Therefore, as mentioned in section 6.1.2, a potential interpretation is that the robot's gaze behavior is an overt cue. In contrast, the human's gaze is a covert cue that affects its communicative effectiveness in assisting the player during collaborative tutoring. We suggest that future work should delve more on this perspective to investigate the role of covert vs. overt signals and their relevance in learning scenarios. Future work should consider comparing our results, preferably with platforms with more articulated eye - gaze behavior, and examine how human perception and interpretation of gaze cues might differ.

6.1.4. **Sequential Gaze Analysis with Lag-based Methods**

In Chapter 5, we used the lag-based method to investigate complex sequences of gaze behavior between a child and a robot in a collaborative tutoring task. Lag sequential analysis is a common approach used to identify behavior pairs that follow each other (Bakeman and Quera, 1995; Faraone and Dorfman, 1987). The lag sequential method has previously been used to examine behavioral patterns in various research domains (Montague and Asan, 2014; Robinson et al., 2003; Woods et al., 2010). However, to our knowledge, there are still no studies that have used lag-based methods to analyze nonverbal communication in human-robot interaction research. Therefore, the work described in Chapter 5 is among the first studies in human-robot interaction to investigate the sequences of gaze interaction between humans (child) and robots during a tutoring interaction. We applied both event-based and time-based lag methods to examine gaze interaction between a child and a robot during the interaction. In our case, the event-based analysis answers the question of how robot and child gaze events follow each other during the interaction, irrespective of the time component. The time-based approach responds to the timing of the child's gaze in response to the robot's gaze behavior, distinguishing the association between gaze pairs in a different time window.

Using the lag-based methods to examine gaze interaction between the child and the robot during the tutoring activity, we found the following:

- *The robot gaze significantly impacts child gaze behavior during tutoring interactions. Thus, understanding the robot's gaze behavior and its impact on children's gaze during interaction can contribute to successful robotic interventions in tutoring settings.*
- *There are different gaze patterns — sequences for children in the Socially attentive group and children in the Non - socially attentive group.*
- *Appropriate sequences of gaze behaviors between a child and a robot can lead to effective child-robot tutoring interactions.*
- *Proper timing of the response of the child's gaze to the robot's gaze can lead to effective child-robot tutoring interactions.*
- *Understanding the dynamic nature of the robot-child gaze patterns adds to robot behaviors, leading to more positive experiences during child–robot tutoring.*

There is a growing body of research on nonverbal communication to improve human-robot interaction (Admoni et al., 2016; Kompatsiari et al., 2017). In the future, these studies can incorporate lag sequential analysis methods to evaluate robot-based interventions in different domains. Lag based methods would especially be useful for robots intended to act as trainers to children with autism spectrum disorder (ASD). For instance, Tapus et al. (2012) found that children with ASD show spontaneous social gaze interactions in response to robot behavior compared to human interactions. We suggest that lag-based methods can be applied to investigate gaze patterns for children with autism and compare with gaze behavior for typically developing children. Such gaze patterns evaluations can inform the development of computational models that can help diagnose and evaluate ASD. Therefore, future studies could go beyond measuring the frequency or duration of non-verbal behavior to understand intricate patterns during interactions. Understanding the significance of nonverbal communication in human-robot interactions may influence design guidelines, especially for the robot in the educational and social training context.

6.1.5. **Laboratory vs. Field Studies**

In this dissertation, we conducted both lab-based and field-based studies. The user studies described in chapters 3 and 4 are laboratory-based studies conducted in the social robotics lab at TU/e. The final experiment (Chapter 5) was a field study conducted at an elementary school in Barcelona with children of the same class in a room often used as a class by the children. We identified various benefits and limitations to these two different experiment locations:

It was possible to control the setting and potential confounding variables to a higher degree in the laboratory. Furthermore, it was easier to use more instrumentation; for example, in Chapter 3, we fitted the participants with eye trackers to register their gaze data. However, a limitation of the lab-based study is the restrictive nature of these environments, which decreases the validity of the research findings to real tutoring settings. Thus, the conclusions from these studies may not necessarily represent how people and robots or children and robots will behave in the real world. Therefore, the mutual benefit of the investigation in the school is that it more accurately revealed children's interactions with robots in the naturalistic setting—a class in which, eventually, the robot could be used in the role of a tutor.

However, conducting a field study is much more challenging compared to lab-based studies—for example, we experienced technical challenges during setup—and loss of data. There is also a need for additional resources—more people—and time required to carry out the investigations. While there are challenges as mentioned, a significant benefit of the field study at the school with a whole classroom of children contributes to the bigger goal of how a robot might work in a real tutoring environment.

6.2. Limitations

This work presented in this dissertation has several limitations that should be noted to improve future work. In the following subsections, we discuss the participant profiles (age, gender, and culture), the interaction scenario and robot platform (NAO), behavioral coding, and human-controlled experiments.

6.2.1.**Participants' Profiles**

There are three points to be noted about the profiles of the participants for the user studies discussed below:

Age: The user study in Chapter 3 involved adult participants between the ages of 18 and 33 years recruited from TU/e university. In Chapter 4, children participants were between the ages of 4 and 11 from a daycare center, including children of staff at the TU/e university. The study in Chapter 5 involved a homogenous group, a whole class, and children who were either 7 or 8 years. While we found that gaze cues could be applied to improve outcomes in both a child and adult context, we made various observations about how children and adults interact with a robot tutor and how they perceive the tutor's gaze. To some extent, we observed that age did influence the capacity of children to read help from gaze with older children being more aware of the tutor's intent and interpreting gaze accurately.

In comparison, some younger children were not able to accurately interpret the robot tutor cues despite noticing them. For example, in the children experiments (chapters 4 and 5), we observed a few children who proceeded to select different cards despite noticing that the robot tutor was looking at a particular card, which was not the case for the adult participants. These observations could be attributed to social cognitive theories of development—such as perspective-taking and agency attribution (Zwickel, 2009)—which are fascinating to study with robots in the future. This work provides initial findings on the ability of people (adults and children) to read and attribute intentions to a robot's gaze during a tutoring task. Future work should examine more in details regarding how adults and children interact with robots, mainly focusing on the subjective and participants' judgments of robots in a tutoring context.

Cultural - language*:* Gaze behavior is sensitive to cultural context and language (Argyle et al., 1973; Kleinke, 1986). We conducted the studies in different countries— the Netherlands and Spain —and, therefore, cultural differences associated with gaze behavior may have influenced the findings. The participants in Chapter 3 were mainly from Europe- Dutch, Asia, and Africa. Most of the children (Chapter 4) were of European-Dutch and Asian backgrounds. In Chapter 5, all the children were Catalan—Spanish native speakers. While we included participants from different backgrounds, we did not explore these cultural differences. Thus, future work should address this gap provided gaze is known to be highly culturally dependent.

Gender: Research in human communication suggests that gender has a significant impact on how people produce and respond to gaze cues (Argyle et al., 1994; Kleinke, 1986). In all our studies, we included participants of different genders; however, we did not balance the gender of participants or evaluate the impact of gender on our findings.

Future studies should consider investigations on how diverse populations, cultures, languages, gender, and age impact interactions with robots. In the final group of children (Chapter 5), there were two children with autism spectrum disorder (ASD). However, we did not compare the differences in behavior between these children with typically developing children. Future work should also consider examining differences in gaze interactions between these groups of children.

6.2.2. Interaction Scenario and the Robot Platform

All the studies described in chapters 3 through 5 involved a board game activity. Participants were required to find pairs of matching cards in the presence of either a human or a robot tutor. The task involved a face-to-face interaction, a board game on a table, primarily involving easy cognitive competencies and easy manipulative motor competences (manipulating the cards).

The NAO platform adopted for this research (Softbank, 2013) (see Figure 1-3) has no articulated eyes and is, therefore, incapable of eye movements. Instead, the NAO robot relies on head turns to indicate gaze direction. From our preliminary findings, we found that people can follow the orientation of NAO's head and its movements to judge gaze direction (see appendix A). In addition, in Chapter 3, we found that participants noticed the gaze cues of the robot tutor significantly more compared to those of the human tutor. Therefore, our findings support that robots such as NAO, without movable eyes, can provide gaze communication through appropriate head movements/facial orientation. These findings add to prior works that have found that robots without movable eyes can also communicate gaze direction. For example, Cuijpers and Van Der Pol (2013) measured the region of eye contact with NAO and concluded that perception of gaze direction with NAO is similar to that of a human.

However, the variability in appearance and capability of the robot's eyes is a significant aspect to examine in future studies. Previous research in human-robot interaction has used robots with varying levels of physical capabilities to study gaze interactions. These platforms range from simple cartoon-like robots to remarkably lifelike humanoids. For example, looking at the work of Mutlu et al. (2009), the authors compared the perception of "leakage cues" across two platforms: Robovie and Geminoid robots. Their findings show that participants perceived gaze cues more effectively with Robovie, which is more humanlike than Geminoid.

Consequently, future studies should explore whether the registered improvements, either in task performance and participants' experiences (perceived presence and likeability), can be achieved in different interaction scenarios and with different platforms. Future work should consider comparing our findings, preferably with platforms with more articulated eye gaze behavior, and examine how human's perception and interpretation of gaze cues might differ. For example, as earlier mentioned, robots with articulated gaze might be able to provide more covert signals compared to NAO, which provides more overt/explicit gaze cues. Based on social psychological literature, the gaze is highly linked to context (Argyle et al., 1973; Kleinke, 1986) and, therefore, how humans perceive gaze. It is, therefore, important to note that our findings are based on a specific interaction task—the board game scenario with a particular robotic platform (NAO). Therefore, whether the conclusions of these studies would generalize to other tasks, for example, remains unknown.

6.2.3. Behavioral Coding and Analysis

In chapters 4 and 5, we used observational methods to investigate simultaneous gaze-based interaction between a child and a robot in a collaborative tutoring activity. It is important to note that the behavioral system and coding scheme are context and platform-dependent and should be adapted to be useful for other HRI studies. However, the overarching categories more related to the theoretical framework of the gaze-based communicative behavior can be applied to a broader range of studies on human-robot gaze interaction. All coding in chapters 4 and 5 were carried out by a human coder using the observer XT 14 software (Zimmerman et al., 2009). While the observer XT helps to automate the coding process to a great extent and ease the coding process, there are still reliability concerns - errors associated with human coding.

One issue relates to the accuracy of the timestamps of the gaze events from the coding. For example, to examine mutual gazing - between the child and the robot *(see* Figure 4-6 *and* Figure 5-6 *for mutual gaze visualizations)*, we selected the instances when the Child_look_robot and Robot_look_child behaviors from child -robot observations co-occur. We suggest that future work can modify our scheme to define how long an overlap is needed to count as a mutual gaze. Depending on the accuracy of timestamps, frequency, and duration of mutual gaze, events can change. Besides, from the eye-tracking, the gaze often goes back and forth between the target and another object at a very high speed *(milliseconds),* which is not possible to achieve through human coding.

6.2.4. Human-Controlled Experiment

One more limitation of our work is the use of human control on some behaviors of the robot. While the sequencing and timing of the NAO head movements occurred automatically depending on the entered card code, the researchers had to control the robot behaviors, including the initial card reading and the verbal feedback behaviors. Whereas participants (Chapter 3) said they perceived the robot behaviors to occur naturally and autonomously, future studies can implement fully autonomous and intelligent applications to perform human-controlled behaviors in our experiments. For example, computer vision methods that can help detect when the card is correct or incorrect and gaze tracking to the locations of the human to provide more intuitive and responsive gaze behaviors.

6.3. Open Questions and Future Work

In this section, we discuss two open challenges that future work should take up towards designing effective gaze mechanisms for robots in the educational setting. While, in this dissertation, we have examined these issues to an extent, there are still open problems that need further investigation.

6.3.1. Timing of Gaze Behavior

In this dissertation, we have strived to investigate design guidelines for effective gaze behavior to build capable robot tutors. The gaze is a complex communicative cue involving many aspects that influence how it is perceived and interpreted. There are different facets to consider when designing gaze cues for a robot from the communication context to the particular robot's behaviors. Timing is a crucial aspect that impacts how people perceive gaze information. In our case, to design gaze behavior for NAO (Figure 3-4), we followed a systematic process that involved gaining a theoretical grounding on gaze interaction as it occurs in human interactions and preliminary observations of people interacting in the same scenario in the laboratory settings. We played the robot gaze motions to different cards on the board and asked participants how they perceived the head directions and timing of gaze behavior—natural, too fast, or too slow. We programmed the head movements of NAO based on these subjective reports and with the timing regarded as more natural.

However, from participants' perceptions (Chapter 3), we recorded diverse responses from participants concerning how they perceived the direction, timing, and intent of the robot tutor's head movements (gaze). A few of the participants reported they noticed the robot's head movements but felt as if it was following their tries rather than directing their attention to the matching card, missing the informative content of the tutors' gaze. Based on these reports, we identified that the readability of the gaze cues could have been compromised by wrong timing, letting the participants (implicitly) assume the tutor was performing gaze-following rather than gaze-cueing. To address these reports, in the last study (Chapter 5), we reduced the delay of the robot's head movements—reaction time—towards the first card (referred here as Card1). We also programmed the robot to give repeated short glances towards the matching card rather than a prolonged stare. In the previous studies, the robot was looking at the matching card for a long duration. However, in the final evaluation, we modified that so the NAO glanced twice at the matching card.

While we addressed the issue of timing to an extent, future work should examine the temporal aspects of gaze and movements in human-human interactions to build more realistic interactive robot gaze behaviors. Therefore, we recommend further analysis of the eye-tracking data sets (human-human and human-robot data) collected in Chapter 3 to investigate micro-level details of the interactive player-tutor-looking behavior. For example, to investigate responsive gaze behavior to either the robot or the human tutor during head turns or gaze shifts. The data can also reveal complex interaction sequences of gaze patterns to gain a better understanding of the differences between human-robot and human-human communication. Future studies should also investigate whether people perceive the robot's gaze timing in the same way as they do with humans, which we were not able to determine from our evaluations. Such works can utilize and extend our findings based on – time-based lag method to investigate the proper timing of human response behavior to the robot behavior to be able to use the social cues from the robot effectively. Additionally, for future improvements to our gaze behavior design and NAO communication capabilities, it would be important to include a gaze tracker to detect gaze directions of the human to create more intuitive, natural robot responses.

6.3.2. **Gaze and Other Social Cues**

In chapters 3 and 4, we examined gaze behavior in isolation from other social cues that form natural communicative interactions. At the beginning of the game, the tutor (human or robot) verbally provided instructions and the rules of the game to participants. During the entire play, the tutor silently gazed at different cards on the table in Help Setup. The tutor remained silent without providing any gaze cues in the No_Help case, only looking towards the participant until the game ended. While findings from these studies showed that participants—adults and children—improve their performance by utilizing the gaze help from the tutor, singling out gaze is a significant limitation for this research.

In social interactions, gaze occurs in coordination with other cues such as speech, gestures, body movements, touch, and facial expressions (smiling). Consequently, in Chapter 5, we enriched the robot tutor with a friendlier approach by combining verbal feedback with gaze cues. In this case, the robot tutor informed the children whether the card was wrong or correct after every try. Moreover, the robot deployed expressive gestures of the arms and hands, emphasizing verbal communication in the instruction section with pre-programmed explanation gestures and during the game. Also, the robot tutor accompanied the verbal feedback with winning or losing gestures. However, a limitation of our study design is we could not quantitatively establish the impact of these added social cues—verbal feedback—gestures in comparison to gaze, or as a combination gaze and spoken interaction on participants' performance and judgments. The reason is that the robot tutor provided verbal feedback in both game sessions (Gaze and No_Gaze condition). However, from our observations and children's perceptions (in Chapter 5), we found that most children would not make a new attempt in the game until the robot tutor gave them feedback on whether the card was correct or incorrect.

Similarly to the initial studies' (chapters 3 and 4) subjective reports, we found that most of the participants expected that the tutor would use verbal cues to help them in the game. Equally was the case with adult participants in Chapter 3—at least eighteen (18) of all the twenty (20) participants in both the human and robot conditions said they expected verbal/vocal/audio help from the tutor. We assume these expectations were grounded in verbal-based communication when the tutor introduced the activity at the beginning of the session. Therefore, we can establish that, possibly, the gaze in combination with feedback can improve the credibility and naturalness of the human-robot with robots, particularly in tutoring and social training. For example, Boucher et al. (2012) showed that a robot could use gaze to support its speech in a cooperative task. Thus, in the design of human-robot tutoring interaction, we recommend further research on a combination of gaze with other non-verbal cues to foster the efficacy of the robot either in tutoring or trainer roles.

Chapter 7. **Conclusion**

Robots will play a significant role in society, be it in health, entertainment, education, manufacturing. Amongst these domains, robots hold great potential in education and therapeutic training for children with special needs. In this dissertation, we aimed to investigate the human-robot gaze mechanisms to inform the design of robot behavior as a facilitator during tutoring interactions. Towards our goal, we conducted a series of user studies with children and adult participants both in a laboratory setting and in a regular classroom. First, we investigated gaze as a mechanism for directing attention during human-robot tutoring activities(chapters 3 and 4). Subsequently, we examined the concepts of mutual coordination—gaze-following and complex sequences of gaze behavior and their impact on children's performance and judgments within robot tutoring interactions(chapters 4 and 5).

The results of the user studies provide new insights and design recommendations for how to build effective, gaze behavior for robot tutors to impact learning outcomes positively. In Chapter 3, we found that participants perform significantly better when a robot tutor provides gaze-based hints during the tutoring activity than when the tutor does not offer such cues. We also found that people noticed the robot tutor's gaze hints more often compared to those of the human tutor. As a result, participants performed significantly better with the robot tutor compared to the human tutor. In Chapter 4, we found that children perform better and feel more engaged when the robot provides gaze cues and establishes mutual gaze during the interaction. Furthermore, we found that more mutual gaze interaction between the child and the robot increases children's awareness of robot tutors' intentions in a tutoring activity.

The study described in Chapter 5 is among the first studies in human-robot interaction that investigates the sequences of interaction between the robot and children in tutoring interactions. The study combines observational and lag sequential methods to capture the complexity of the dynamics of child-robot interaction, including complex gaze sequences between the child and the robot. The analysis of coordinated sequences and directions of gaze behavior provides a deeper understanding of whether, how, and under which conditions human partners—children—take advantage of robots' nonverbal gaze prompts. We found that successful sequences of gaze interaction between the child and robot tutors can facilitate better performance during the child- robot tutoring interactions.

We outline the contributions of this dissertation in Section 7.1 and a closing remark in Section 7.2.

7.1. Contributions

7.1.1. Methodological Contributions

This research adopts a multi-modal measurement and analysis approach. It combines observational analysis, objective measurement techniques, including eye-tracking, subjective (self-report), and lag sequential methods. Table 7-1 provides a detailed list of these contributions.

Table 7-1 Methodological contributions

Chapters 3, 4, and 5	An experimental board game design to study how gaze-based interaction can be applied to improve interaction outcomes both in performance (task execution) and participants' judgment (e.g., perceived presence and likeability) in the context of tutoring, coaching, and therapy.
Chapters 3 and 4	Simultaneous measurement of human and robot gaze, to improve the analysis of the HRI. Use of synchronized simultaneous observations of both the robot and the human's behavior to reveal precise patterns of gazing behavior and to describe and understand the complete phenomenon of human-robot communication.
Chapter 3	Eye-tracking data to investigate intricate gaze patterns - interactive player-tutor looking behavior (i.e., gaze-following, joint attention), to gain an understanding of the differences between human-robot and human-human nonverbal communication.
Chapter 4	Observational analyses based on the behavioral systems and coding schemes both for robot and child behavioral units. The observational analyses include the gaze pattern (as particular meaningful sequences of coordinated behaviors) as a unit of analysis.
Chapter 5	Implementation of sequential analysis methods in human-robot interaction to unravel the complex dynamics of child-robot interaction, which is a necessary part of developing successful robotic interventions in tutoring and therapy settings.

7.1.2. Theoretical Contributions

The theoretical contributions provide the basis for gaze-based interaction design to improve robot-based educational interventions. This work contributes three facets of new knowledge: designing gaze-based cues in the context of tutoring, comparisons between human tutors and robot tutors, and intricate patterns of gaze-based interaction during child -robot tutoring activities. Table 7-2 provides a detailed list of these contributions.

Table 7-2 Theoretical contributions

Chapter 3	Evidence that participants perform better with fewer tries to complete the card-matching activity when the tutor (either human or robot) provides gaze cues during gameplay than when the tutor does not give the gaze cues. Evidence that participants perform better with fewer tries to complete the game with the robot tutor than with the human tutor. Evidence that participants look more into the tutor's face when the tutor (human or robot) provides gaze cues during gameplay than when the tutor does not provide the gaze cues. Evidence that the participants look more into the robot tutor's face than to the human tutors' face during the gameplay interaction. Evidence that participants notice the gaze cues (awareness of the tutor's intention to help) more in the robot condition than in the human condition.
Chapter 4	Evidence that children perform better with fewer tries to complete the card-matching activity when the tutor provides gaze cues than without gaze cues from the robot tutor.
Chapters 4 and 5	Evidence that children look more into the robot tutor's face when the tutor provides gaze cues during gameplay than without gaze cues from the robot tutor. Evidence that children establish more eye contact with the robot tutor when the tutor provides gaze cues during gameplay than without gaze cues from the robot tutor. Evidence that more occurrences and higher durations of coordinated (mutual gaze and gaze-following) between the child and the robot tutor during gameplay increase the children's awareness of the robot's intent to assist, which, in turn, improves the task-solving performance regarding accuracy.
Chapter 5	Evidence that successful sequences of coordinated gaze behavior between the child and the robot during gameplay increases children's knowledge of the helping cues from the robot, which, in turn, improves the task-solving performance regarding accuracy.

7.1.3. **Practical Contributions**

The main practical contributions of this dissertation are the design and development of the board card game setting and the design of the robot's communicative behavior. Table 7-3 provides a list of these contributions.

Table 7-3 Practical contributions

Chapters 3, 4 and 5	The design and development of a board game as a tool in education and therapy.
Chapters 3, 4 and 5	The guidelines for gaze-based interactive behavior design –in diverse platforms.
Chapters 3, 4 and 5	The development of gaze-based communication for the NAO robot.

7.2. Closing Remarks

The findings from this dissertation provide recommendations for the design of credible and effective gaze behavior for robots in educational settings (tutoring and social training) to improve learning performance. Apart from the findings of the user studies, the main contributions of this dissertation include; First, an experimental framework to study simultaneous gaze interaction during human-robot tutoring activities. Second is a coding scheme developed to measure child-robot interaction with an emphasis on coordinated and sequential gaze sequences between children and robots. The third is the use of simultaneous observational and lag-based methods to examine coordinated and interaction sequences of gaze between children and robots, helping unravel the dynamics of child-robot interaction in a tutoring setting. The lag-based method, in particular, provides an opportunity to look at more complex gaze sequences, which is a necessary part of developing successful robotic interventions in tutoring and therapy settings. This dissertation also has several limitations, which we have outlined and open challenges that inform future research directions towards designing effective gaze behavior for robots in social training and tutoring settings.

Bibliography

Admoni, H. (2016). Nonverbal communication in socially assistive human-robot interaction. *AI Matters*, *2*(4), 9-10.

Admoni, H., Hayes, B., Feil-Seifer, D., Ullman, D., and Scassellati, B. (2013). Are you looking at me? Perception of robot attention is mediated by gaze type and group size. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 389-395). IEEE.

Admoni, H., and Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction*, *6*(1), 25-63.

Admoni, H., Weng, T., Hayes, B., and Scassellati, B. (2016). Robot nonverbal behavior improves task performance in difficult collaborations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 51-58). IEEE.

Albo-Canals, J., Martelo, A. B., Relkin, E., Hannon, D., Heerink, M., Heinemann, M., ... and Bers, M. U. (2018). A pilot study of the KIBO robot in children with severe ASD. *International Journal of Social Robotics*, *10*(3), 371-383.

Andrist, S., Tan, X. Z., Gleicher, M., and Mutlu, B. (2014). Conversational gaze aversion for humanlike robots. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 25-32). IEEE.

Argyle, M., Cook, M., and Cramer, D. (1994). Gaze and mutual gaze. *The British Journal of Psychiatry*, *165*(6), 848-850.

Argyle, M., Ingham, R., Alkema, F., and McCallan, M. (1973). The different functions of gaze. *Semiotica*, *7*(1), 19-32.

Astington, J. W., and Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition and Emotion, 9*(2-3), 151–165.

Bakeman, R., and Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge University Press.

Bakeman, R., and Quera, V. (1995). Log-linear approaches to lag-sequential analysis when consecutive codes may and cannot repeat. *Psychological Bulletin, 118*(2), 272–284.

Bakeman, R., and Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences.* Cambridge University Press.

Bakeman, R., and Quera, V. (2012). *Behavioral observation.* In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, and K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (p. 207–225). American Psychological Association.

Barakova, E. I., Bajracharya, P., Willemsen, M., Lourens, T., and Huskens, B. (2015). Long-term LEGO therapy with humanoid robot for children with ASD. *Expert Systems*, *32*(6), 698-709.

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

Baron-Cohen, S. (2008). Theories of the autistic mind. *The Psychologist, 21*(2), 112–116.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21).

Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., and Ertl, T. (2014). State-of-the-Art of Visualization for Eye Tracking Data. In *EuroVis (STARs)*.

Borrero, C. S., England, J. D., Sarcia, B., and Woods, J. N. (2016). A Comparison of Descriptive and Functional Analyses of Inappropriate Mealtime Behavior. *Behavior Analysis in Practice*, *9*(4), 364–379.

Boucher, J. D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., ... and Ventre-Dominey, J. (2012). I reach faster when I see you look: Gaze effects in human-human and human-robot face-to-face cooperation. *Frontiers in Neurorobotics*, *6*(3), 1-11.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., and Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 708-713). IEEE.

Brooks, R., and Meltzoff, A. N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental Psychology*, *38*(6), 958-966.

Brooks, R., and Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535-543.

Brooks, R., and Meltzoff, A. N. (2014). *Gaze following: A mechanism for building social connections between infants and adults.* In M. Mikulincer and P. R. Shaver (Eds.), *The Herzliya series on personality and social psychology. Mechanisms of social connection: From brain to group* (p. 167–183). American Psychological Association.

Broz, F., Lehmann, H., Nakano, Y., and Mutlu, B. (2012). Gaze in HRI: from modeling to communication. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction* (pp. 491-492). ACM.

Chao, C., and Thomaz, A. L. (2012). Timing in multimodal turn-taking interactions: Control and analysis using timed petri nets. *Journal of Human-Robot Interaction*, *1*(1), 4-25.

Cuijpers, R. H., and Van Der Pol, D. (2013). Region of eye contact of humanoid Nao robot is similar to that of a human. In *International Conference on Social Robotics* (pp. 280-289). Springer, Cham.

de Haas, M., Smeekens, I., Njeri, E., Haselager, P., Buitelaar, J., Lourens, T., ... and Barakova, E. (2017). Personalizing educational game play with a robot partner. In *Robotics in Education* (pp. 259-270). Springer, Cham.

Díaz-Boladeras, M. (2017). *Bonding with Robotic Pets. Children's Cognitions, Emotions and Behaviors towards Pet-Robots. Applications in a Robot Assisted Quality of Life Intervention in a Pediatric Hospital*. Doctoral dissertation, Universitat de Barcelona, Spain.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, *24*(6), 581-604.

Faraone, S. V., and Dorfman, D. D. (1987). Lag sequential analysis: Robust statistical methods. *Psychological Bulletin*, *101*(2), 312-323.

Farnsworth, D. (10). Most Used Eye Tracking Metrics and Terms. *IMotions. Luettavissa: https://imotions.com/blog/7-terms-metrics-eye-tracking/Luettu*, *1*, 2018.

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602-9605.

Fasola, J., and Matarić, M. J. (2013). A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, *2*(2), 3-32.

Feng, Y., Yu, S., van de Mortel, D., Barakova, E., Hu, J., and Rauterberg, M. (2019). LiveNature: Ambient Display and Social Robot-Facilitated Multi-Sensory Engagement for People with Dementia. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (pp. 1321-1333). ACM.

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences. *Psychological Bulletin*, *133*(4), 694-724.

Frith C. (2009). Role of facial expressions in social interactions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *364*(1535), 3453–3458.

Gernsbacher, M. A., Stevenson, J. L., Khandakar, S., and Goldsmith, H. H. (2008). Why does joint attention look atypical in autism?. *Child Development Perspectives*, *2*(1), 38-45.

Ghazali, A. S., Ham, J., Barakova, E., and Markopoulos, P. (2018). The influence of social cues in persuasive social robots on psychological reactance and compliance. *Computers in Human Behavior*, *87*, 58-65.

Gillesen, J. C., Barakova, E. I., Huskens, B. E., and Feijs, L. M. (2011). From training to robot behavior: Towards custom scenarios for robotics in training programs for ASD. In *2011 IEEE International Conference on Rehabilitation Robotics* (pp. 1-7). IEEE.

Gordon, G., Breazeal, C., and Engel, S. (2015). Can Children Catch Curiosity from a Social Robot?. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 91-98). ACM.

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., ... and Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (pp. 3951-3957). AAAI Press.

Ham, J., Cuijpers, R. H., and Cabibihan, J. J. (2015). Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *International Journal of Social Robotics*, *7*(4), 479-487.

Henkemans, O. A. B., Bierman, B. P., Janssen, J., Neerincx, M. A., Looije, R., van der Bosch, H., and van der Giessen, J. A. (2013). Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient Education and Counseling*, *92*(2), 174-181.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Huang, C. M., Andrist, S., Sauppé, A., and Mutlu, B. (2015). Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology*, *6*, 1049.

Huang, C. M., and Mutlu, B. (2013). The repertoire of robot behavior: Designing social behaviors to support human-robot joint activity. *Journal of HRI*, *2*(2), 80-102.

Huang, C. M., and Mutlu, B. (2016). Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 83-90). IEEE.

Hunnius, S. (2007). The early development of visual attention and its implications for social and cognitive development. *Progress in Brain Research*, *164*, 187-209.

Huskens, B., Verschuur, R., Gillesen, J., Didden, R., and Barakova, E. (2013). Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention. *Developmental Neurorehabilitation*, *16*(5), 345-356.

Johal, W., Castellano, G., Tanaka, F., and Okita, S. (2018). Robots for learning. *International Journal of Social Robotics*, *10*(3), 293-294.

Jokinen, K., Furukawa, H., Nishida, M., and Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *3*(2), 1-30.

Jones, A., and Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *International Journal of Social Robotics*, *10*(3), 357-370.

Kaplan, F., and Hafner, V. (2004). *The Challenges of Joint Attention*. 67–74.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, *26*, 22-63.

Kinoshita, Y., Yokoyama, M., Yoshida, S., Mochizuki, T., Yamada, T., Narumi, T., ... and Hirose, M. (2017). Transgazer: Improving Impression by Switching Direct and Averted Gaze Using Optical Illusion. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI* (pp. 53-62). IEEE.

Kirchner, N., Alempijevic, A., and Dissanayake, G. (2011). Nonverbal robot-group interaction using an imitated gaze cue. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 497-504). IEEE.

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychological Bulletin*, *100*(1), 78-100.

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., and Wykowska, A. (2018). On the role of eye contact in gaze cueing. *Scientific Reports*, *8*(1), 1-10.

Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., and Wykowska, A. (2019). It's in the Eyes: The engaging role of eye contact in HRI. *International Journal of Social Robotics*, 1-11.

Kompatsiari, K., Tikhanoff, V., Ciardo, F., Metta, G., and Wykowska, A. (2017). The importance of mutual gaze in human-robot interaction. In *International Conference on Social Robotics* (pp. 443-452). Springer, Cham.

Langton, S. R., Watt, R. J., and Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*(2), 50-59.

Marino, F., Chilà, P., Sfrazzetto, S. T., Carrozza, C., Crimi, I., Failla, C., ... and Ruta, L. (2020). Outcomes of a robot-assisted social-emotional understanding intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *50*(6), 1973-1987.

Mavadati, S. M., Feng, H., Gutierrez, A., and Mahoor, M. H. (2014). Comparing the gaze responses of children with autism and typically developed individuals in human-robot interaction. In *2014 IEEE-RAS International Conference on Humanoid Robots* (pp. 1128-1133). IEEE.

Meltzoff, A. M., and Brooks, R. (2007). *Eyes Wide Shut: The importance of eyes in infant gaze-following and understanding other minds.* In R. Flom, K. Lee, and D. Muir (Eds.), *Gaze-following: Its development and significance* (p. 217–241). Lawrence Erlbaum Associates Publishers.

Meltzoff, A. N., Brooks, R., Shon, A. P., and Rao, R. P. (2010). "Social" robots are psychological agents for infants: A test of gaze following. *Neural Networks*, *23*(8-9), 966-972.

Montague, E., and Asan, O. (2014). Dynamic modeling of patient and physician eye gaze to understand the effects of electronic health records on doctor–patient communication and attention. *International Journal of Medical Informatics*, *83*(3), 225-234.

Montague, E., Xu, J., Chen, P. Y., Asan, O., Barrett, B. P., and Chewning, B. (2011). Modeling eye gaze patterns in clinician–patient interaction with lag sequential analysis. *Human Factors*, *53*(5), 502-516.

Moon, A., Troniak, D. M., Gleeson, B., KXJ, P. M., Zheng, M., Blumer, B. A., ... and Crof, E. A. (2014). Meet Me where I'm Gazing: How Shared Attention Gaze Affects Human-Robot Handover Timing. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 334-341). IEEE.

Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *2006 6th IEEE-RAS International Conference on Humanoid Robots* (pp. 518-523). IEEE.

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. (2009). Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. In *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 69-76). IEEE.

Mwangi, E., Barakova, E. I., Díaz-Boladeras, M., Mallofré, A. C., and Rauterberg, M. (2018a). Directing Attention Through Gaze Hints Improves Task Solving in Human–Humanoid Interaction. *International Journal of Social Robotics*, *3*(10), 343-355.

Mwangi, E., Barakova, E. I., Diaz, M., Mallofre, A. C., and Rauterberg, M. (2017a). Gaze-based hints during child-robot gameplay. In *International Conference on Social Robotics* (pp. 413-422). Springer, Cham.

Mwangi, E. N., Barakova, E. I., Diaz, M., Mallofré, A. C., and Rauterberg, M. (2017b). Who is a Better Tutor? Gaze Hints with a Human or Humanoid Tutor in Game Play. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 219-220).

Mwangi, E., Barakova, E. I., Díaz, M., Mallofre, A. C., and Rauterberg, M. (2018b). Dyadic gaze patterns during child-robot collaborative gameplay in a tutoring interaction. In *27th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2018* (pp. 856-861). IEEE.

Mwangi, E., Barakova, E., Zhang, R., Diaz, M., Catala, A., and Rauterberg, M. (2016). See Where I am Looking at: Perceiving Gaze Cues With a NAO Robot. In *Proceedings of the Fourth International Conference on Human Agent Interaction* (pp. 329-332). ACM.

Mwangi, E., Díaz-Boladeras, M., Barakova, E., Català Mallofré, A., and Rauterberg, M. (2017c). Can children take advantage of Nao gaze-based hints during gameplay?. In *HAI 2017: 5th International Conference on Human-Agent Interaction* (pp. 421-424).ACM

Palinko, O., Rea, F., Sandini, G., and Sciutti, A. (2016a). Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5048-5054). IEEE.

Palinko, O., Rea, F., Sandini, G., and Sciutti, A. (2016b). Eye tracking for human robot interaction. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research and Applications* (pp. 327-328).ACM.

Palinko, O., Sciutti, A., Schillingmann, L., Rea, F., Nagai, Y., and Sandini, G. (2015). Gaze contingency in turn-taking for human robot interaction: Advantages and drawbacks. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 369-374). IEEE.

Perugia, G., Díaz-Boladeras, M., Català-Mallofré, A., Barakova, E. I., and Rauterberg, M. (2020). ENGAGE-DEM: a model of engagement of people with dementia. *IEEE Transactions on Affective Computing*.

Perugia, G., Rodríguez-Martín, D., Boladeras, M. D., Mallofré, A. C., Barakova, E., and Rauterberg, M. (2017). Electrodermal activity: explorations in the psychophysiology of engagement with social robots in dementia. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1248-1254). IEEE.

Pfeiffer-Lessmann, N., Pfeiffer, T., and Wachsmuth, I. (2012). An operational model of joint attention - timing of gaze patterns in interactions between humans and a virtual human. In N. Miyake, D. Peebles, and R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 851-856). Austin, TX: Cognitive Science Society.

Pohl, M., Wallner, G., and Kriglstein, S. (2016). Using lag-sequential analysis for understanding interaction sequences in visualizations. *International Journal of Human-Computer Studies*, *100*(96), 54-66.

Ramachandran, A., Huang, C. M., and Scassellati, B. (2017). Give me a break! Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 146-155).IEEE.

Robins, B., Dickerson, P., Stribling, P., and Dautenhahn, K. (2004). Robot-mediated joint attention in children with autism: A case study in robot-human interaction. *Interaction studies*, *5*(2), 161-198.

Robinson, C. C., Anderson, G. T., Porter, C. L., Hart, C. H., and Wouden-Miller, M. (2003). Sequential transition patterns of preschoolers' social interactions during child-initiated play: Is parallel-aware play a bidirectional bridge to other play states?. *Early Childhood Research Quarterly*, *18*(1), 3-21.

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., ... and McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and hci: Behaviour generation, user interaction and perception. In *Computer Graphics Forum,* 34 (6), 299-326.

Scassellati, B., Boccanfuso, L., Huang, C. M., Mademtzi, M., Qin, M., Salomons, N., ... and Shic, F. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, *3*(21).

Softbank. (2013). NAO the humanoid robot. *SoftBank Robotics*.

Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pintea, S., ... and David, D. O. (2012). Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interaction Studies*, *13*(3), 315-347.

van Straten, C. L., Smeekens, I., Barakova, E., Glennon, J., Buitelaar, J., and Chen, A. (2018). Effects of robots' intonation and bodily appearance on robot-mediated communicative treatment outcomes for children with autism spectrum disorder. *Personal and Ubiquitous Computing*, *22*(2), 379-390.

Vertegaal, R., Slagter, R., Van der Veer, G., and Nijholt, A. (2001). Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 301-308).

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., ... and Papadopoulos, F. (2019). Second language tutoring using social robots: a large-scale study. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 497-505). IEEE.

Woods, J. N., Borrero, J. C., Laud, R. B., and Borrero, C. S. (2010). Descriptive analyses of pediatric food refusal: The structure of parental attention. *Behavior Modification*, *34*(1), 35-56.

Yonezawa, T., Yamazoe, H., Utsumi, A., and Abe, S. (2007). Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In *Proceedings of the 9th International Conference on Multimodal Interfaces* (pp. 140-145). ACM.

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., and Miyamoto, T. (2006). Responsive robot gaze to interaction partner. In *Robotics: Science and Systems* (pp. 37-43).

Yu, C., Schermerhorn, P., and Scheutz, M. (2012). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *1*(2), 1-25.

Zimmerman, P. H., Bolhuis, J. E., Willemsen, A., Meyer, E. S., and Noldus, L. P. (2009). The Observer XT: A tool for the integration and synchronization of multimodal signals. *Behavior Research Methods*, *41*(3), 731-735.

Zwickel, J. (2009). Agency attribution and visuospatial perspective taking. *Psychonomic Bulletin and Review*, *16*(6), 1089-1093.

Appendix A. **Board Game Design**

This section presents a preliminary study designed to examine how people perceive gaze cues and head angles directed towards different target positions on a table when a human and a robot (NAO) are sitting against each other as in board game scenarios. The findings from this study informed the design of the board game layout and card arrangement and the overall design of interaction settings between the human and robot used to conduct the studies in Chapters 3 through 5.

Experimental setup

The experimental setup was as shown in Figure A-1(b). The participant and NAO sat at the two sides of a table facing each other. The table was approximately 80 cm in width, and the height of the table was 72 cm. A board grid with the card positions resembling a memory game was fixed on the table. The layout had 18 squares (8*8cm) organized in six (6) columns and three (3) rows. The 18 squares corresponded to the 18 card positions for the game. The squares were 10 cm apart in depth (y-axis) and 6 cm apart in width (x-axis). The layout was 600 mm in width and 900 mm long.

To measure the head angles, we placed NAO on a small desk 56 cm in height at "Stand-init" pose (0, 0). The design grid had six squares in the x-direction, which was from the left to the right side of NAO, and three squares in the y-direction, which was the depth direction of NAO. The distance between NAO and the closest square position was approximately 20 cm away, and the furthest at the corner was about 60 cm. We attached a laser beam on the mid-section of the NAO head and adjusted it to point at the middle of the layout, using the "look at" module in the Choregraphe program. We estimated the head pitch and head yaw angles for all the target positions using the motion screen on Choregraphe. [6]

Position= {HeadYawAngleVal, HeadPitchAngleVal}

NAO's coordinate system is as shown in (Figure A-1(a)). The HeadYawAngleVal of NAO's gaze direction is defined as the angle between the positive y-axis and a line drawn from the center to a fixed position. The yaw angle of the y-axis is 0, and a positive head yaw angle value is on the left side of NAO. NAO head yaw angles range from -119 to 119 degrees. The HeadPitchAngle of NAO's gaze direction is the angle between the xy plane and a line drawn from NAO's head location to a target square. The pitch (head joint front and back) angle increases from 0 to 29.5. For the setup, the highest

[6] This work is largely based on the following publication(Mwangi et al., 2016)

angle pitch value used was 24± degrees for the two middle positions in the first row. The highest yaw angle was 48± for position 1 and -48± for position 6. For the second row, the angle pitch decreases to 10±1 for the middle positions. The pitch angle decreased with an increase in the yaw angle for the positions on the sides.

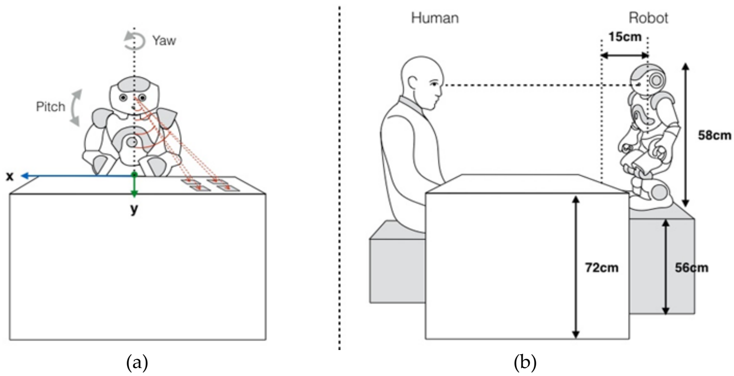


Figure A-1 (a) NAO coordinate system (b) schematic illustration of the setup

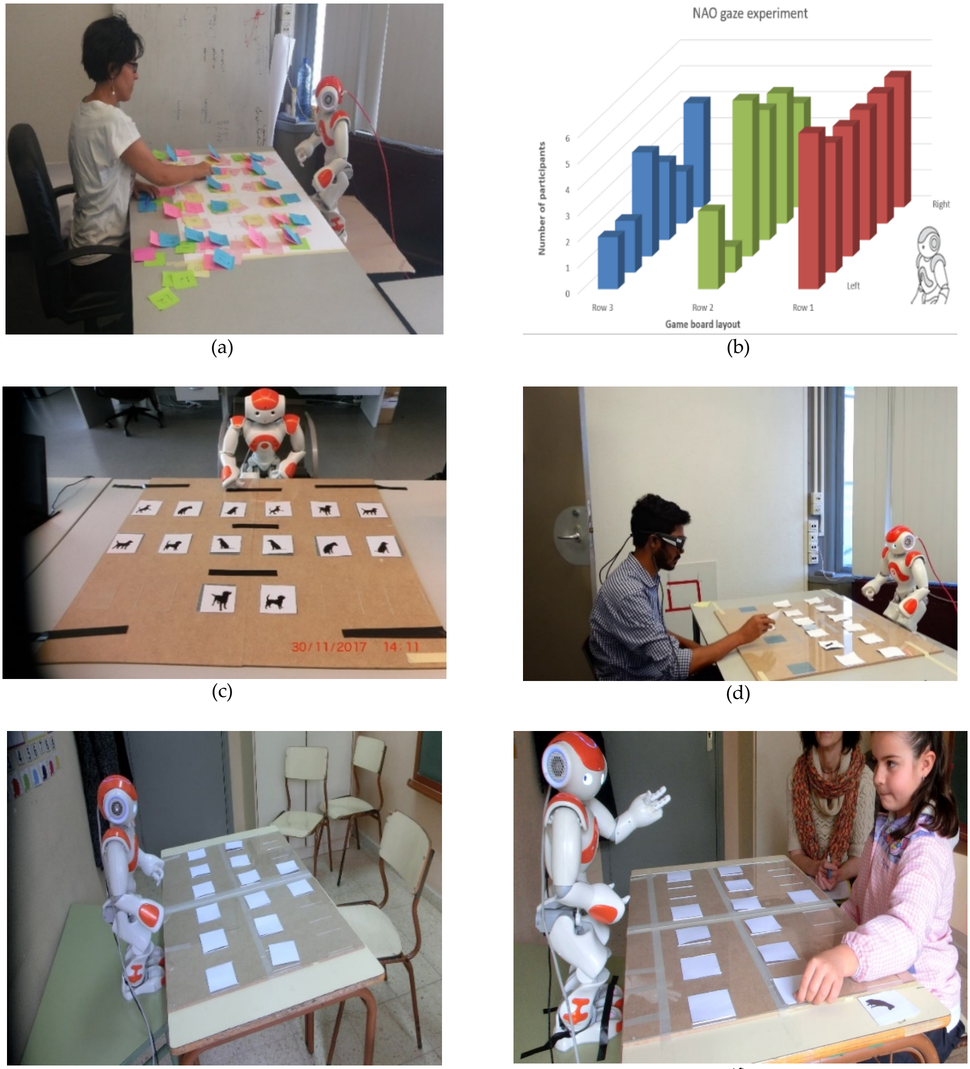
Procedure

Six students from the university participated in the preliminary study. Three were female, and three were male. The robot was placed on a small table 56 cm tall at (0, 0) with its face directed to the face of the person (Figure A-2(a)). The participant sat on a chair, which was adjusted to give an eye-height position with the robot. The distance between the robot and the participant was approximately 110 cm. The experimenter informed the participants of their role and gave them instructions regarding the experiment. We implemented a Java algorithm to turn NAO's head randomly to the 18 positions on the layout. Each participant interacted with the robot only in one trial. When the robot moved its head to a certain location, the participant wrote a number between 1 and 18 on a post-it and placed it where they perceived the robot was looking on the layout. Each trial lasted for about 5 minutes.

Implications for Design of the Board Game Layout

From the graph (Figure A-2 (b)) , we found that people are better able to perceive the square positions correctly more for the card locations that are in the row closest to NAO, i.e., when the head pitch angle is higher (24±2), and there is less depth in the y-direction), which is approximately 20 cm from NAO. The ability to perceive seems to lessen with the increase in depth. For example, the number of participants who perceived the gaze correctly decreased rapidly when the rows were far from NAO.

The number of correct perceptions was lower in Row 2 compared to Row 1 and continued to lessen for the third row, which is approximately 60 cm away from NAO for this layout. An interesting observation from the result is when NAO is looking at positions on its right. The participants were better able to perceive more than when the robot was gazing at the locations on its left. Observers were also able to correctly perceive the positions in the middle of the layout more. Observations also show that participants understood head yaw angles quite better as opposed to head pitch angles for this robot. However, with the increase in depth, the perception of yaw information seemed to lessen. The (Figure A-2(c)) shows the final gaze board design based on participants perceptions of the gaze cues of the NAO robot. We placed 12 cards on the first two rows (Row 1and Row 2) — 6 for each row — nearest to NAO and 2 cards in the middle positions of the third row farthest from NAO.



(a) (b) (c) (d) (e) (f)

Figure A-2 (a) Participant marking the card positions NAO was looking at on the board game (b) Gaze perception results for NAO: the number of participants who perceived the robot gaze correctly for each card location: Row 1; Row 2; Row 3. (c) The design of the board game based on the findings in (b) — 14 cards placed on the board layout, 6 cards placed in Row 1; 6 cards placed in Row 2, and 2 cards placed in the middle positions of Row3 (d) The final interaction setting design and the board game showing a human participant and a robot sitting across each other — (Chapter 3) (e) The robot setup in the classroom — (Chapter 5) (f) A child playing the matching card game and interacting with the robot tutor during the tutoring game activity in the classrom.

Appendix B. **Questionnaire**

In this section, we present the study questionnaire used to evaluate participants' perceptions regarding either the human or the robot tutor and the tutoring style (Chapters 3).

Dear Participant

I am undersigned, currently undertaking a Ph.D. Degree in Interactive and Cognitive Environments at the Eindhoven University of Technology. My research pertains to the use of social robots for the educational training. We are designing behaviors for social robots, and we would like you to assist in improving their designs.

Procedure

There are two sessions in this experiment, and you are kindly requested to answer the questions in the survey regarding each of the sessions. The whole survey will take approximately 8 minutes of your time. Please complete all of the items. Your participation in this study is voluntary, and all responses and data collected will be used only for the intended purpose of the research and will be treated in the strictest of confidence. Your name or any other personal identification will be not be collected or recorded at any time during the study.

Participant's Responsibilities

I voluntarily agree to participate in this study. I have the following responsibilities: perform experimental tasks, and answer the questionnaire to the best of my ability. I have read the conditions of the experiment, and I hereby acknowledge the above and give my voluntary consent:

Date_______________________________

Participant signature______________________________

Please direct any inquiries concerning this survey to:

PhD Student: Eunice N. Mwangi

Contact e-mail: e.n.mwangi@tue.nl /njeri.eunice@gmail.com

Supervisors: Prof. Emilia Barakova, Prof Matthias Rauterberg

* Required

Part 1: Demographic data

1. You are: *

Mark only one oval.

- Female
- Male

2. Your age: *

Part 2: Experiment Questions

3. To what extent do you agree/disagree with the following statements about the task you performed during the experiment: * Mark only one oval per row.

	Completely Disagree	Disagree	Neutral	Agree	Completely Agree	N/A
I enjoyed this task						
I found the task hard						
It was important for me to do a good job						
I would like to continue with this task						
I found this task easy to understand						
I have done a task like this before						
I needed help in this task						

Experiment: First Session

4. Please rate the following statements about the behavior of the tutor: * *Mark only one oval per row.*

	Completely disagree	Disagree	Neutral	Agree	Completely agree	N/A
The tutor was pleasant	◯	◯	◯	◯	◯	◯
The tutor was kind	◯	◯	◯	◯	◯	◯
The tutor was friendly	◯	◯	◯	◯	◯	◯
The tutor was likeable	◯	◯	◯	◯	◯	◯
The tutor was attentive	◯	◯	◯	◯	◯	◯

5. Please rate the following statements about your interaction with the tutor: *

Mark only one oval per row.

	Completely disagree	Disagree	Neutral	Agree	Completely agree	N/A
The tutor presence was obvious to me						
My presence was obvious to the tutor						
The tutor caught my attention						
I caught the tutors attention						
I was easily distracted from the tutor when other things were going on						
I remained focused on the tutor throughout our interaction						
The tutor did not receive my full attention						
The tutor behavior caught my attention						
I noticed the tutor 's help						

Experiment: Second Session

6. Please rate the following statements about the behavior of the tutor: * *Mark only one oval per row.*

	Completely disagree	Disagree	Neutral	Agree	Completely agree	N/A
The tutor was pleasant	◯	◯	◯	◯	◯	◯
The tutor was kind	◯	◯	◯	◯	◯	◯
The tutor was friendly	◯	◯	◯	◯	◯	◯
The tutor was likeable	◯	◯	◯	◯	◯	◯
The tutor was attentive	◯	◯	◯	◯	◯	◯

7. Please rate the following statements about your interaction with the tutor: * *Mark only one oval per row.*

	Completely disagree	Disagree	Neutral	Agree	Completely agree	N/A
The tutor presence was obvious to me	◯	◯	◯	◯	◯	◯
My presence was obvious to the tutor	◯	◯	◯	◯	◯	◯
The tutor caught my attention	◯	◯	◯	◯	◯	◯
I caught the tutors attention	◯	◯	◯	◯	◯	◯
I was easily distracted from the tutor when other things were going on	◯	◯	◯	◯	◯	◯
I remained focused on the tutor throughout our interaction	◯	◯	◯	◯	◯	◯
The tutor was not attentive to my problems during	◯	◯	◯	◯	◯	◯
I only focused on my task throughout the session	◯	◯	◯	◯	◯	◯
I noticed the tutor's help	◯	◯	◯	◯	◯	◯

Interactive Social Behaviors for the Tutor

8. Please list the kind of cues that you observed in the tutor behavior that helped you identify the matching card: *

9. Please list the kind of cues you think might be effective in improving the tutor's interactive behavior:

10. Any other comments about the study

Thank you for participating in this study

Post-Experiment Interview

The facilitator conducted a brief post-experiment interview with the participants. The facilitator asked the participants the following questions:

Q1: Did you notice the tutor's help? YES/NO; if yes, how did the tutor help you?

Q2: What behavior /cues did you expect the tutor would use to help you during the game?

Q3: How did you perceive the robot's head movements: were they natural, too fast, or too slow?

Q4: Did you consider the behavior of the robot tutor to be automatic?

Appendix C. **Experiment Protocol**

In this section, we present the experimental protocol and questionnaire for the child - robot study described in Chapter 5.

PROTOCOL NAO MARGALLÓ **NH / H**

First name: _______________

Age: ______

Consent: YES

Date: ___ / ___ / _______

Start time (getting in): _____ / ______ End time (getting out): ____ / _____

Facilitator: Eunice/ Andreu / Marta

Presentation (Facilitator)

"Hello, thank you for taking part in this activity.

We will play two rounds of this game. You must play on your own. You will have a tutor who will give you the instructions. After the first round, we will have some time outside of class, a couple of minutes, and then we'll enter the second part.

Here you have the board game and your tutor who will give instructions for the game."

Round_1 NO HELP

Instructions Robot:

"Hello, welcome to this session of play. My name is Maca, and I am your tutor.

I have a task for you. You have to find pairs of identical cards on the board. First, you pick up a card, and you turn it over, and then pick a second card. If the cards do not match, you turn them upside down so that the pictures are hidden and start again. If they match, you remove the two cards and keep them close to you.

You have to find all pairs with the smallest number of attempts, selecting as few letters as possible.

You have to play alone; I cannot help you now.

Please go ahead and choose the first card!"

Round_1 Questionnaire:

1- Was the game difficult?

Not at all / A bit difficult / Neither easy nor difficult / Easy / Very easy

2- Have you ever played before? YES/NO

3- What was the robot doing while you played? ________________________________

____

Round_2: HELP

Instructions Robot:

"Now we'll play another game of making pairs.

Remember, you must find all pairs with the least number of attempts by selecting as few letters as possible.

In this game, we will play together, we will make a team. I'm going to help you.

Please go ahead and choose the first card! "

Round_2 Questionnaire:

1- Was the game difficult?

Not at all / A bit difficult / Neither easy nor difficult / Easy / Very easy

2- What was the robot doing while you played? ________________________________

____

3- The robot, did it want to help you? YES / NO

HOW? ____________________________

Questionnaire at the End of the Game:

1- Have you ever played this game before? YES/NO

2- Did you like playing?

Not at all / Not much / Neither liked nor disliked playing / I liked / I liked it a lot

Goodbye:

Great. Thank you very much. We appreciate that you wanted to come and play. We need it to learn how to program the robot to make it more useful. Thank you so much!

PROTOCOL NAO Margalló **H / NH**

First name: _______________

Age: ______

Consent: YES

Date: ___ / ___ / _______

Start time (getting in): _____ / ______ End time (getting out): ____ / _____

Facilitator: Eunice/ Andreu / Marta

Presentation:

"Hello, thank you for taking part in this activity.

We will play two rounds of this game. You must play on your own. You will have a tutor who will give you the instructions. After the first round, there will be time outside of class, a couple of minutes, and then we will enter for the second part.

Here you have the board game and your tutor who will give instructions for the game."

Round_1: HELP

Instructions Robot:

"Hello, welcome to this session of play. My name is Maca, and I am your tutor.

I have a task for you. You have to find pairs of identical cards on the board. First, you take a card, and you turn it over and then a second card. If the cards do not match, you turn them upside down so that the pictures are hidden, and start again. If they match, you remove the pair and keep it close to you.

You have to find all pairs with the least number of attempts, selecting as few letters as possible.

In this game, we will play together, we will make a team. I know where the cards are placed. I'm going to help you.

Please go ahead and choose the first card! "

Round_1 Questionnaire:

1- Was the game difficult?

Not at all / A bit difficult / Neither easy nor difficult / Easy / Very easy

2- What was the robot doing while you played? ________________________________

____

3- The robot, did it want to help you? YES / NO

HOW? ____________________________

Round_2: NO_HELP

Instructions Robot:

"Now we'll play another game of making pairs.

Remember, you must find all pairs with the least number of attempts by selecting as few letters as possible.

You have to play alone; I cannot help you now.

Please go ahead and choose the first card! "

Round_2 Questionnaire:

1- Was the game difficult?

Not at all / A bit difficult / Neither easy nor difficult / Easy / Very easy

2- What was the robot doing while you played? ________________________________

____

3- The robot, did it want to help you? YES / NO

HOW? ____________________________

Questionnaire End Game:

3- Have you ever played this game before? YES/NO

4- Did you like playing?

Not at all / Not much / Neither liked nor disliked playing / I liked / I liked it a lot

Closing and Farewell:

Great. Thank you very much. We appreciate that you wanted to come and play. We need it to learn how to program the robot to make it more useful. Thank you so much!

Appendix D. **User Study Data**

Below we present each participant's details, including age, gender, awareness of tutor's gaze cues and performance (number of tries and duration(s)). We also provide gaze behavior data regarding frequency and duration of mutual gaze and gaze following patterns for the child - robot study described in Chapter 5.

Table D-1 Children details and performance (Chapter 5)

Participants	Gender	Age	Order	Noticed_ Gaze	Tries_ Gaze	Tries_ No_Gaze	Duration_ Gaze	Duration_ No_Gaze
P01	Girl	7	H_NOH	No	22	16	250.71	147.58
P02	Boy	7	H_NOH	Yes	10	15	194.36	114.21
P03	Girl	7	H_NOH	No	46*	21	453.24	214.51
P04	Girl	7	H_NOH	Yes	17	11	287.55	168.73
P05	Boy	7	H_NOH	No	15	12	128.03	106.17
P06	Boy	7	H_NOH	No	17	12	233.16	135.87
P07	Boy	7	H_NOH	No	15	11*	187.58	192.49
P08	Boy	8	NOH_H	Yes	9	12	154.52	140
P10	Boy	8	NOH_H	No	14	15	146.21	122.12
P11	Girl	7	NOH_H	No	11	26	101.1	316.88
P12	Girl	7	NOH_H	Yes	15	19	186.38	184.68
P13	Girl	7	NOH_H	No	17	15	227.16	185.25
P16	Girl	7	NOH_H	No	24	29	303.4	378.1
P17	Boy	8	NOH_H	No	15	18	142.91	173.84
P18	Girl	7	NOH_H	No	13	14	130.23	121.45
P20	Boy	*	H_NOH	No	14	24	132.2	259.25
P21	Boy	*	H_NOH	No	16	15	208.1	150.31
P22	Girl	8	H_NOH	Yes	11	19	144.74	169.5

Table D-2 Mutual gaze (MG) data (Chapter 5)

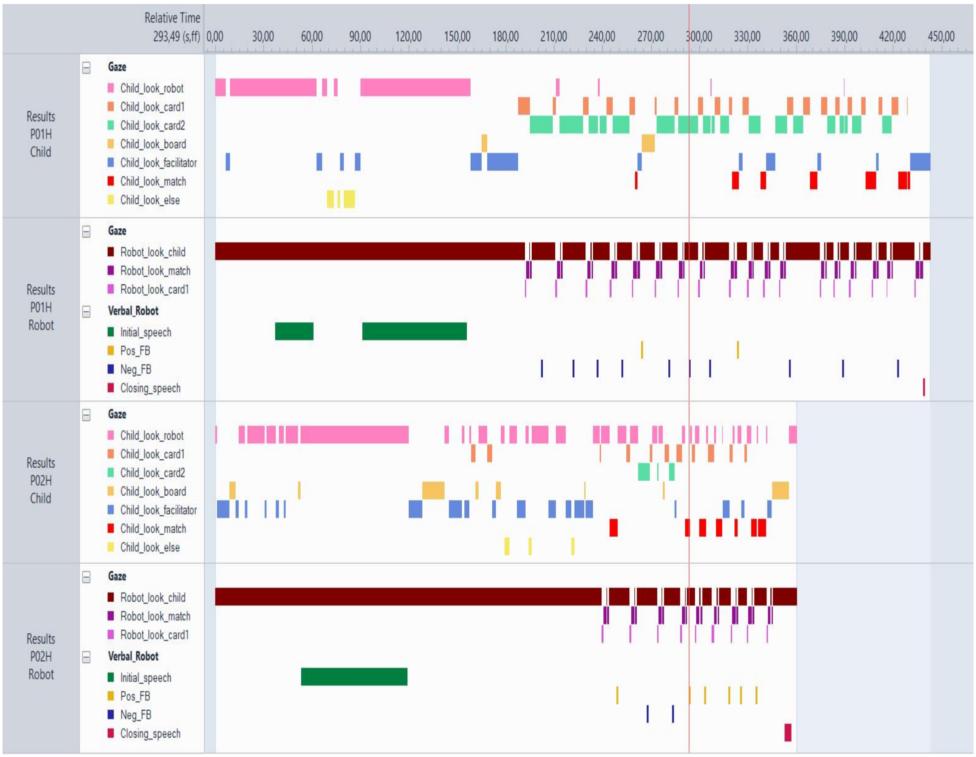
	Gaze Condition					No_Gaze Condition				
Participants	Frequency	Total Duration	Min Duration	Max Duration	% Duration	Frequency	Total Duration	Min Duration	Max Duration	% Duration
P01	3	2.54	0.37	1.47	0.57	1	1.27	1.27	1.27	0.62
P02	21	52.35	0.30	10.31	14.52	7	6.87	0.33	2.00	3.44
P03	4	5.24	0.23	2.37	1.09	1	1.94	1.94	1.94	0.76
P04	26	39.97	0.13	4.94	10.07	6	14.21	0.60	5.61	6.67
P05	7	15.18	0.67	4.10	4.44	2	1.84	0.53	1.30	1.06
P06	12	29.90	0.21	5.11	6.72	8	20.55	0.70	4.54	8.29
P07	13	25.92	0.32	4.08	8.13	12	30.60	0.30	5.17	8.66
P08	16	34.42	0.20	8.87	12.86	5	5.81	0.57	2.00	2.50
P10	8	17.46	0.40	5.77	7.82	3	4.47	0.03	2.84	0.85
P11	3	0.93	0.17	0.47	0.54	8	11.81	0.50	3.07	3.83
P12	19	16.02	0.10	1.87	6.18	5	6.44	0.57	1.90	1.65
P13	11	25.02	0.13	6.97	8.36	2	8.04	0.93	7.11	1.68
P16	8	15.15	0.43	6.41	4.17	2	5.34	1.07	4.27	1.73
P17	2	7.44	0.17	7.27	3.49	3	2.04	0.47	1.03	0.80
P18	4	11.56	1.27	6.34	6.06	2	1.40	0.50	0.90	0.77
P20	11	15.54	0.13	3.30	3.92	11	17.95	0.47	4.20	5.90
P21	14	52.02	0.60	11.04	10.43	9	15.01	0.60	5.07	7.88
P22	18	32.48	0.10	9.11	7.94	11	16.45	0.37	3.87	7.78

Table D-3 Gaze-following data (Chapter 5)

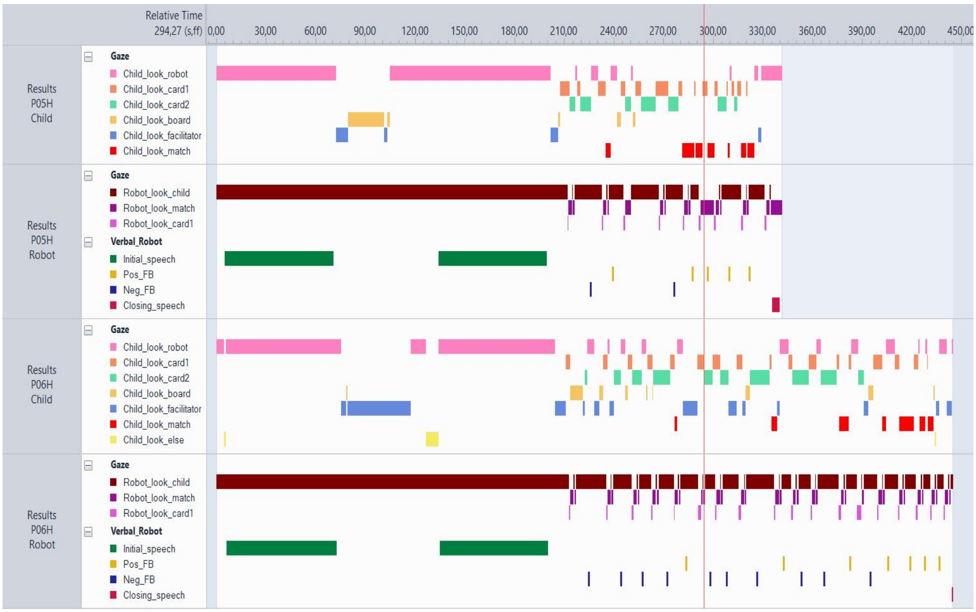
Participants	Frequency	Total duration	Minimum duration	Maximum duration	%observation duration	% Analyzed duration	Analyzed duration	Duration
P01	4	4.5	0.37	1.97	1.02	1.8	250.71	443.24
P02	23	72.04	0.67	10.31	19.98	37.06	194.36	360.49
P03	6	8.11	0.23	2.37	1.69	1.79	453.24	480.9
P04	21	56.09	0.41	9.34	14.14	19.51	287.55	397.68
P05	7	20.19	0.93	6.61	5.91	15.77	128.03	341.63
P06	12	37.75	0.9	5.55	8.49	16.19	233.16	444.79
P07	11	34.05	1.39	5.17	10.68	18.15	187.58	318.9
P08	11	51.14	0.73	8.87	19.11	33.09	154.52	267.66
P10	8	22.77	0.3	14.96	10.19	14.03	162.27	223.4
P11	7	16.14	0.01	6.42	9.29	15.95	101.15	173.63
P12	18	48.35	0.63	4.9	18.67	25.94	186.38	258.95
P13	9	29	0.67	6.97	9.68	12.76	227.16	299.49
P16	9	17.85	0.23	6.41	4.92	5.88	303.4	363.09
P17	2	7.87	0.43	7.44	3.69	5.51	142.91	213.14
P18	6	14.16	0.1	6.34	7.43	10.87	130.27	190.57
P20	12	18.32	0.4	3.3	4.62	13.86	132.2	396.85
P21	14	54.62	0.83	11.04	10.95	26.25	208.1	498.91
P22	20	55.36	0.7	9.11	13.54	38.25	144.74	408.96

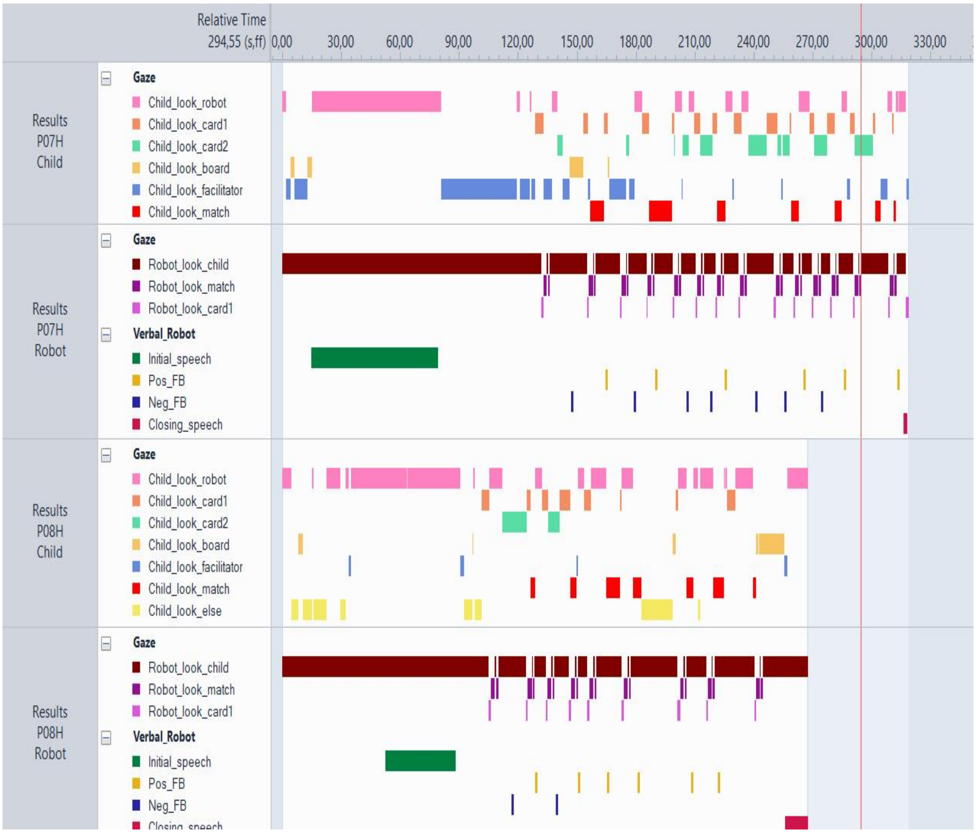
Appendix E. **Coordinated Gaze Patterns**

This section presents 18 visualizations of child – robot (NAO) observations in the gaze condition (Chapter 5). The time event plots show the interaction of the gaze behavior between the child and the robot during the Gaze condition, i.e., when the tutor provides gaze-based cues during the interaction. The upper row of each visualization shows the behaviors of the child, and the bottom row shows the gaze and the verbal behaviors for the robot tutor. The length of the boxes indicates the duration, and the occurrences of similar colored bar rectangles show the frequency of the events of each behavior. From the visualization, we can see the sequences and coordination of child and robot gaze during the interaction.

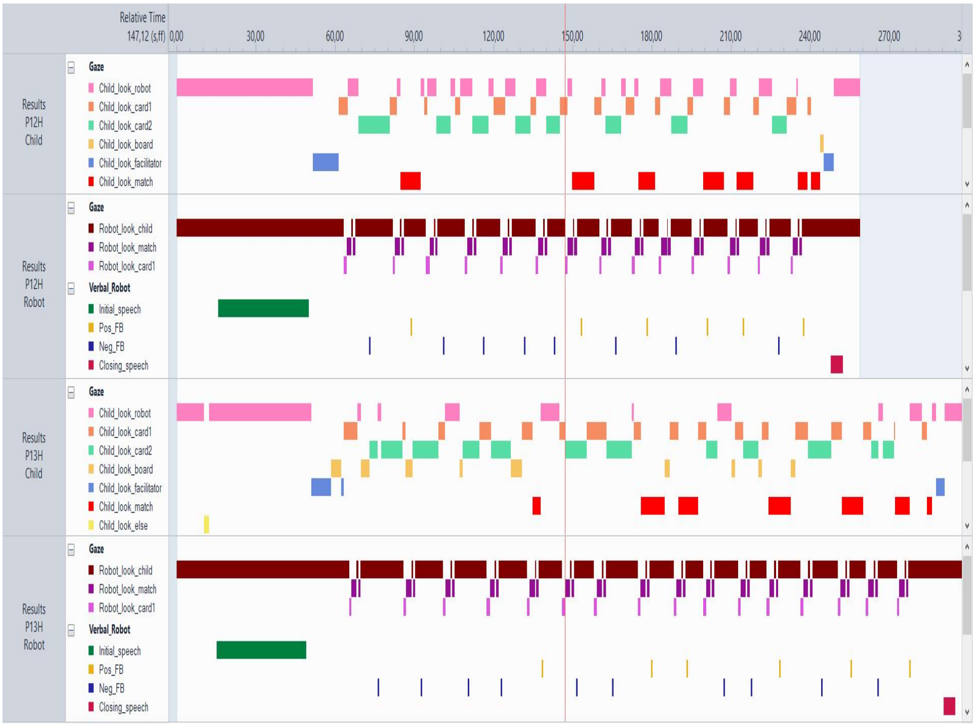


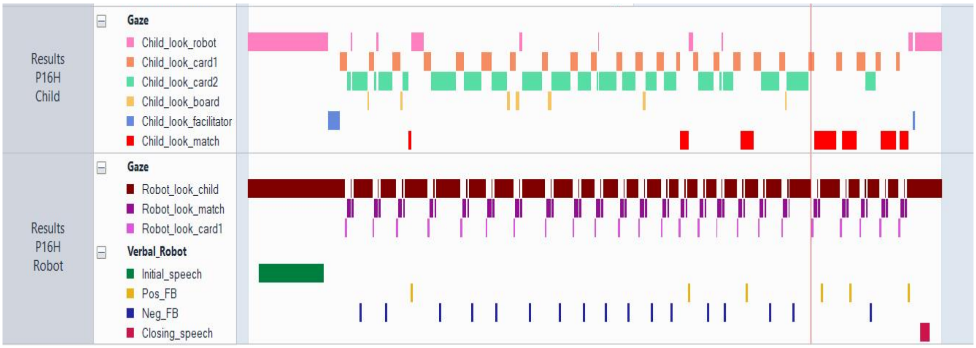

Relative Time
294,28 (s,ff)
0,00
30,00
60,00
90,00
120,00
150,00
180,00
210,00
240,00
270,00
300,00
330,00
360,00
390,00
420,00
450,00
480,00
Results P03H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Child_look_else
Results P03H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech
Results P04H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Child_look_else
Results P04H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech


Relative Time
294,27 (s,ff)
0,00
30,00
60,00
90,00
120,00
150,00
180,00
210,00
240,00
270,00
300,00
330,00
360,00
390,00
420,00
450,00
Results P05H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results P05H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech
Results P06H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Child_look_else
Results P06H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech


Relative Time
294,55 (s,ff)
0,00
30,00
60,00
90,00
120,00
150,00
180,00
210,00
240,00
270,00
300,00
330,00
Results P07H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results P07H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech
Results P08H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Child_look_else
Results P08H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech


Relative Time
147,13 (s,ff)
0,00
30,00
60,00
90,00
120,00
150,00
180,00
210,00
Results P10H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Child_look_else
Results P10H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech
Results P11H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results P11H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech


Relative Time
147,12 (s,ff)
0,00
30,00
60,00
90,00
120,00
150,00
180,00
210,00
240,00
270,00
Results P12H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results P12H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech
Results P13H Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Child_look_else
Results P13H Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech


Results
P16H
Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results
P16H
Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech


Relative Time
147,32 (s,ff)
0,00
30,00
60,00
90,00
120,00
150,00
180,00
210,00
Results
P17H
Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results
P17H
Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech
Results
P18H
Child
Gaze
Child_look_robot
Child_look_card1
Child_look_card2
Child_look_board
Child_look_facilitator
Child_look_match
Results
P18H
Robot
Gaze
Robot_look_child
Robot_look_match
Robot_look_card1
Verbal_Robot
Initial_speech
Pos_FB
Neg_FB
Closing_speech

Figure E-1 Child –Robot coordinated behavior visualizations for 18 children in Gaze condition

List of Publications

Journal Articles

- Identifying Dynamic Gaze Interactions during Child-Robot Tutoring (In submission)
- Mwangi, E., Barakova, E. I., Díaz-Boladeras, M., Mallofré, A. C., and Rauterberg, M. (2018). Directing attention through gaze hints improves task solving in human–humanoid interaction. *International Journal of Social Robotics*, *10*(3), 343-355.

Conference Proceedings

- Mwangi, E., Barakova, E. I., Díaz, M., Mallofré, A. C., and Rauterberg, M. (2018, August). Dyadic Gaze Patterns during Child-Robot Collaborative Gameplay in a Tutoring Interaction. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 856-861). IEEE.
- Mwangi, E., Barakova, E. I., Diaz, M., Mallofre, A. C., and Rauterberg, M. (2017, November). Gaze-based hints during child-robot gameplay. In *International Conference on Social Robotics* (pp. 413-422). Springer, Cham.
- de Haas, M., Smeekens, I., Njeri, E., Haselager, P., Buitelaar, J., Lourens, T., ... and Barakova, E. (2017). Personalizing educational game play with a robot partner. In *Robotics in Education* (pp. 259-270). Springer, Cham.
- Mwangi, E., Barakova, E., Zhang, R., Diaz, M., Catala, A., and Rauterberg, M. (2016, October). See where I am looking at: perceiving gaze cues with a NAO robot. In *Proceedings of the Fourth International Conference on Human Agent Interaction* (pp. 329-332). ACM.
- Mwangi, E., Diaz, M., Barakova, E., Catala, A., and Rauterberg, M. (2017, October). Can children take advantage of NAO gaze-based hints during gameplay? In *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 421-424). ACM.
- Mwangi, E. N., Barakova, E. I., Diaz, M., Mallofré, A. C., and Rauterberg, M. (2017, March). Who is a better tutor? : gaze hints with a human or humanoid tutor in game play. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 219-220). ACM.

Other Publications

- Kimani, S., Njeri, E., and Njue, J. (2013). Online Requirements and Portal Design for Female University Science and Technology Students in Kenya. In *IFIP Conference on Human-Computer Interaction* (pp. 403-410). Springer, Berlin, Heidelberg.
- Mwangi, E. N., Kimani, S., and Kimwele, M. (2014). Textual Emotion Communication with Non-verbal Symbols in Online Environments. In *International Conference on Human-Computer Interaction* (pp. 42-48). Springer, Cham.

Summer Schools

- EMJD ICE Thesis Contribution Presentation (Ph.D. third year) | University of Genoa, Italy 2017. *Title:* Gaze - Based Interaction for Effective Tutoring with Social Robots
- IEEE EURASIP S3P /EMJD ICE Summer School on Signal Processing | Cavalese, Italy 2016. *Title:* Gaze Hints with a Human or Humanoid Tutor in Game Play
- Summer School on Intelligent Sensing for Interactive and Cognitive Environments |Queen Mary University of London, Mile End Rd, London 2015. *Title:* An Eye - Contact Model for Social Robots

Biography

Eunice Njeri was born on July 24th, 1986, in Nyeri, Central Kenya. After finishing her Bachelor of Science degree in Computer Technology in 2009 at Jomo Kenyatta University of Agriculture and Technology (JKUAT) with a First-Class Honors, she studied Software Engineering (Master of Science) at the said university from 2011 and graduated in 2014. During her Master's research work, she designed an intelligent engine for determining users' emotions in online social environments. In January 2015, Eunice started her Ph.D. project under the Erasmus Mundus Joint Doctoral Program in Interactive and Cognitive Environments (EMJD - ICE). She commenced her research at the department of industrial design, Eindhoven University of Technology (TUE), Netherlands. In February 2017, she joined the Technical Research Centre for Dependency Care and Autonomous Living (CETpD) associated with Universitat Politècnica de Catalunya, Spain, to continue her Ph.D. research for one year. Her Ph.D. project focused on designing effective gaze-based interaction in educational settings (tutoring and social training), of which the findings are presented in this dissertation.

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UPC UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH