COMENIUS UNIVERSITY, BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

HUMAN-ROBOT INTERACTION: THE ROLE OF PRESENCE AND GAZE

MASTER'S THESIS

2022
BC. KASSANDRA FRIEBE

COMENIUS UNIVERSITY, BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

Human-Robot Interaction: The Role of Presence and Gaze

master's thesis

Study programme: Cognitive Science
Study field: Computer Science
Department: Department of Computer Science
Supervisor: RNDr. Kristína Malinovská, PhD.
Consultant: Mgr. Matěj Hoffmann, PhD.

Bratislava, 2022
Bc. Kassandra Friebe

28924551

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Kassandra Friebe
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Human-Robot Interaction: the Role of Presence and Gaze

Annotation: Studies in human-robot interaction (HRI) focus on the robot embodiment and the way it is presented to the human [1] adopting typical phenomena of social cognition. Social gaze [2] is a well-studied phenomenon of human-human interaction gaining interest in the HRI research [3]. However, attempts to establish connections between these two research domains have yet been sparse.

Aim: 1. Investigate whether humanoid robots that are presented to us in different ways, i.e. as co-present robots with us in the same environment or as telepresent robots via a computer screen, trigger gaze cueing effects in HRI, and whether such effect is presence-specific.
2. Design, implement and evaluate a behavioral experiment focused on a given goal.

Literature: 1. Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. International Journal of Human-Computer Studies, 77.
2. Frischen, A., Bayliss, A., Tipper, S. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. Psych. Bull., 133(4), 694.
3. Wiese, E., Weis, P., Lofaro, D. (2018). Embodied social robots trigger gaze following in real-time HRI. 15th Int. Conf. on Ubiquitous Robots. IEEE.

Supervisor:	RNDr. Kristína Malinovská, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	prof. Ing. Igor Farkaš, Dr.
Assigned:	01.10.2021
Approved:	01.03.2022

prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

....
Student

....
Supervisor

28924551

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Kassandra Friebe
Študijný program: kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Human-Robot Interaction: the Role of Presence and Gaze
Úloha prítomnosti a pohľadu v interakcii človeka a robota

Anotácia: Štúdie interakcie človek-robot (HRI) sa zameriavajú na stelesnenie robota a spôsob, akým je prezentovaný človeku [1], ktorý si osvojuje typické javy sociálneho poznania. Sociálny pohľad [2] je dobre študovaný fenomén interakcie človek-človek, o ktorý sa zaujíma aj výskum HRI [3]. Pokusy nadviazať spojenie medzi týmito dvoma výskumnými doménami sú však zatiaľ zriedkavé.

Cieľ: 1. Preskúmajte, či humanoidné roboty, ktoré sú nám prezentované rôznymi spôsobmi, t. j. ako roboty, ktoré sú s nami v rovnakom prostredí alebo ako teleprezentujúce roboty cez obrazovku počítača, spúšťajú v HRI efekty nabádania pohľadu a či je takýto efekt špecifický pre prítomnosť.
2. Navrhnite, realizujte a vyhodnoťte behaviorálny experiment zameraný na daný cieľ.

Literatúra: 1. Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. International Journal of Human-Computer Studies, 77.
2. Frischen, A., Bayliss, A., Tipper, S. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. Psych. Bull., 133(4), 694.
3. Wiese, E., Weis, P., Lofaro, D. (2018). Embodied social robots trigger gaze following in real-time HRI. 15th Int. Conf. on Ubiquitous Robots. IEEE.

Vedúci: RNDr. Kristína Malinovská, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: prof. Ing. Igor Farkaš, Dr.

Dátum zadania: 01.10.2021

Dátum schválenia: 01.03.2022

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

.... študent

.... vedúci práce

Acknowledgements: This thesis was supported by the Czech Science Foundation (GA CR), project no. 20-24186X. I would like to thank my supervisor Kristína Malinovská for her constant motivation to work together on this project, her confidence in my work and her willingness to discuss the topic in length. My research would have been impossible without the support of Sabína Samporová, who programmed the behavior of the robot and the experimental setup (see Samporová, 2022), and Matěj Hoffmann, who was up for collaboration on this project from the first minute and provided his insights and time. I would also like to thank Jakub Rozlivka and Lukáš Rustler for their help in programming the iCub and Adam Rojík for organizing the recruitment of participants. Finally, my thanks go to my family, friends, and boyfriend for their unwavering support and for putting up with my daily hardships. Without them, this achievement would not have been possible. Thank you.

Abstract

The way a robot is presented has several effects on human-robot interaction (HRI). In particular, research suggests that robots that are copresent in the same environment are evaluated more positively and provide better interaction outcomes than robots presented via a screen. However, how the physical presence of a robot affects simple social attention mechanisms has not been thoroughly investigated. Gaze cueing is a well-studied phenomenon in human-human interaction and is beginning to be of interest to the HRI community. The present work aims to establish the link between gaze cueing and physical presence in HRI and to contribute to filling the current research gap.

An experiment (N = 42) was conducted to investigate the influence of the physical presence of a robot and its gaze behavior on the reaction time of subjects in a gaze cueing paradigm. Participants were randomly assigned to one of two robot presence conditions (copresent robot: physically present iCub robot; virtual agent: screen version of the same robot) and were asked to locate the appearance of a target stimulus that was either congruent or incongruent to the location cued by the robot's gaze. After the experiment participants rated their perception of the robot by judging its anthropomorphism, animacy and likeability.

Participants showed a consistent gaze cueing effect irrespective of the robot condition they were assigned to, as indicated by slower reaction times in trials that were incongruent compared to the ones that were congruent. Against our hypothesis, the way the robot was presented had no effect on the strength of this effect. Additionally, in contrast to findings from previous studies, no differential effect of robot presence on ratings of the robot could be found. The results imply that gaze cueing as a basal phenomenon of human social cognition can also be found in interactions with humanoid robots. Against theoretical assumptions, the different ways of presenting the robot did not seem to alter the strength of the gaze cueing effect. Implications of the findings for the design of robots, as well as future research are discussed.

Keywords: human-robot interaction; social gaze; physical presence

Abstrakt

Spôsob, akým je robot prezentovaný, ovplyvňuje interakciu človeka a robota (human-robot interaction, HRI) rôznymi spôsobmi. Výskum v oblasti HRI naznačuje, že roboty, ktoré sú súčasne prítomné v tom istom prostredí ako človek, sú hodnotené pozitívnejšie a poskytujú lepšie výsledky interakcie ako roboty prezentované prostredníctvom obrazovky. Avšak to, ako fyzická prítomnosť robota ovplyvňuje jednoduché mechanizmy sociálnej pozornosti, nebolo dôkladne preskúmané. Pozorovanie pohľadom je dobre preskúmaný jav v interakcii človek-človek a začína byť predmetom záujmu komunity HRI. Cieľom tejto práce je zistiť súvislosť medzi naznačovaním pohľadom (gaze cueing) a fyzickou prítomnosťou v HRI.

Predmetom predkladanej diplomovej práce bol experiment s cieľom preskúmať vplyv fyzickej prítomnosti robota a komunikácie pohľadom na reakčný čas participantov v paradigme naznačovania pohľadom. Participanti boli náhodne zaradení do dvoch skupín. Každá z nich sa stretla s iným spôsobom prezentácie robota, a to s fyzicky prítomným robotom iCub a virtuálny agentom, čiže verziou toho istého robota na obrazovke. Úlohou participantov v experimente bolo lokalizovať výskyt cieľového podnetu, ktorý bol buď kongruentný, alebo inkongruentný s miestom, na ktoré upozornil robot svojím pohľadom. Po skončení experimentu účastníci hodnotili svoje vnímanie robota tak, že posudzovali jeho antropomorfizmus, animálnosť a sympatickosť.

U účastníkov sa prejavil konzistentný efekt naznačovania pohľadom v oboch experimentálnych skupinách, čo sa prejavilo pomalším reakčným časom v prípadoch, ktoré boli inkongruentné v porovnaní s prípadmi, ktoré boli kongruentné. V rozpore s našou hypotézou nemal spôsob prezentácie robota žiadny vplyv na silu tohto efektu. Okrem toho, sa na rozdiel od zistení z predchádzajúcich štúdií, neukázal vplyv prítomnosti robota na jeho vnímanie ľudmi v hodnotení po experimente. Z výsledkov vyplýva, že naznačovanie pohľadom ako bazálny jav ľudskej sociálnej kognície sa môže vyskytovať aj v interakciách s humanoidnými robotmi. V rozpore s teoretickými predpokladmi sa neukázalo, že by rôzne spôsoby prezentácie robota menili silu efektu naznačovania pohľadom. Diskusia o dôsledkoch zistení pre dizajn robotov, ako aj o budúcom výskume završuje túto diplomovú prácu.

Kľúčové slová: interakcia človek-robot; sociálny pohľad; fyzická prítomnosť

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Relevance of the Project

With technological progress, robots have become important partners, not only in isolation, but also in human living spaces. Human-robot interaction has therefore long been recognized as an important topic in human factors research. The development and design of robots that meet human needs and enable successful and enjoyable interactions is central to a field of research that ranges from psychology to robotics. When studying factors that shape human-robot interaction (HRI), one will quickly discover that there are many factors that shape social (human-robot) interaction that are worth studying. One factor that we have all become aware of in the Covid 19 pandemic is that physical presence can influence our interactions. It seems intuitively clear that being physically present in the same space is a fundamental state that affects how we interact, perceive the interaction, and our interaction partner. But how does physical presence really affect HRI? And how do new technologies like virtual reality fit into this scheme?

This thesis aims to provide an overview of the current state of research on the topic of presence in human-robot interaction and to expand our knowledge about it by empirically investigating how presenting a robot in different ways is shaping the way the robot is perceived and how social cues sent by it are used. After a thorough literature review, an experiment is conducted to investigate whether copresent robots, robots in the same environment, and virtual agents, simulated robots presented to the interaction partner via a computer screen, elicit an gaze cueing effect. Furthermore, it is explained how the experiment could be extended by presenting a robot in virtual reality. Thereby the thesis focuses on the proposal that presence on top of embodiment fundamentally shapes interactions and perception in human-robot interaction, and that virtual reality could offer new perspectives to research and development in the field of social cognitive robotics by bridging the gap between copresence and virtual presence.

In the remainder of this chapter, I will first give a recap of developments in human-robot interaction and factors shaping it. This includes a summary of studies that investigated the role of embodiment and presence, and will be followed by the introduction of virtual reality as a new tool in human-robot interaction research. Previous empirical findings addressing the comparability of virtual reality and other ways of presenting the robot will be reviewed and open questions in the field will be discussed. This will lead to the introduction of an important factor in human-human as human-robot interaction, social gaze, and finally lead to the formulation of the research question and hypotheses standing behind the empirical part of the thesis.

1.2 Previous Research

The research field of human-robot interaction (HRI) has evolved since the beginning of the 20th century with the advancement of technology and a new focus on interdisciplinary research (Goodrich and Schultz, 2008; Bartneck et al., 2020). New questions about the way humans and robots could work together and how best to afford this interaction were being asked. Recently, the question of how the embodiment of a robot and the way it is presented to the human shapes the interaction has concerned the scientific community. In particular, advances in virtual reality technology offer different possibilities than before and open up new spaces for HRI. In parallel, researchers started investigating how typical phenomena of social cognition relate to human-robot and human-avatar interaction. Social gaze is one well-studied phenomenon of human-human interaction and starts to be of more interest to the HRI research community. However, systematic attempts to establish connections between these two active research domains have yet been sparse. Before turning to them we shall look at the advances in HRI and the two topics of interest individually.

1.2.1 Human-Robot Interaction

While the idea of robots and robot-like behavior has been around for many years, technological progress has given it a sharp increase in popularity in the mid and late 20th century (Goodrich and Schultz, 2008; Bartneck et al., 2020). The earliest popular notion of the word robot was tellingly in Karel Čapek's 1920's play Rossum's Universal Robots (Čapek, 1923), which was soon to be followed by an abundance of science fiction literature (e.g. Asimov, 1950) and later, the development of the first robots. An important step away from science fiction took place in the 1950s, at the time of the industrial revolution, when the automotive industry was on the rise and industrial robots were produced to support production. The first known industrial robot was an industrial arm called Unimate, which removed parts from a dissecting machine (Gasparetto and Scalera, 2019). Since these developments, the general idea of what a robot is has changed and broadened. Robots changed in design and functionality, which is why, so far, researchers did not agree on one concise definition of the term robot. One often-cited definition is the one put forward by the Robot Institute of America in 1979, which describes a robot as “a reprogrammable multifunctional manipulator designed to move materials, parts, tools or specialized devices through variable programmed motions for the performance of a variety of tasks”. Other popular definitions stem from the IEEE, the Institute of Electrical and Electronics Engineers, which defines a robot as “an autonomous machine capable of sensing its environment, carrying out computations to make decisions, and performing actions in the real world” (Spectrum, 2020), and hence points towards some key elements that most definitions of robots highlight – robots’ ability to sense, compute, and act.

With the increasing popularity of multidisciplinary research approaches, a new way of studying and developing robots has emerged at the intersection of psychology, robotics, cognitive science, human factors, natural language, and human-computer interaction – human-robot interaction. While motivations to join the field of HRI may be broad, reaching from designing complex robotic systems with real-world applications to human centered use of robots as tools to understand social cognition to understanding how to develop complex intelligent systems – different researchers are united by a main theme. The field of human-robot interaction is concerned with the “understanding, designing, and evaluating robotic systems for use by or with humans.” (Goodrich and Schultz, 2008). Classical ways robots are applied in HRI are as telerobots, teleoperators, or automated vehicles.

Moreover, as technology continues to evolve and robot capabilities grow, these fields of work are being broadened and the possibility of humans and robots working together in teams is emerging (Hancock et al., 2011). Thus, the development of social robots (Sheridan, 2016) - i.e. robots that are designed to directly interact with humans, has become one of the key challenges in robotics research (Yang et al., 2018). With the growing possibility of robots working in all kinds of areas of social life, such as entertainment, teaching, or healthcare it becomes apparent, that research has to focus on the way humans perceive robots, specifically social robots, and the HRI itself. As a result, research has started to investigate typical phenomena of social cognition, such as gaze following (Wiese et al., 2018; Willemse and Wykowska, 2019), as well as factors that shape different facets of HRI. A meta-analysis conducted by Hancock et al. (2011) examined factors that influence trust in HRI and found that while human-related factors and environmental factors shape our perceptions of robots, factors that affect the robot, such as performance- and attribute-based factors, are particularly influential. These results highlight the importance of robot design for HRI and that future research needs to carefully consider the characteristics of the robot that enable meaningful interaction with humans.

1.2.2 Embodiment and Presence in Human-Robot Interaction

One robot-related factor that has recently gained popularity and has been shown to be important for HRI is the embodiment of the robot (Deng et al., 2019). Embodiment is a topic broadly discussed and studied in cognitive science. The questions of how we experience ourselves in a body that interacts with its environment and what kind of body is required for what kind of cognition are difficult ones, and the approaches to answering these questions are wide-ranging. Due to its popularity and broadness, it is seemingly hard to define the term embodiment in an overarching manner. Questions that are commonly asked with regards to embodiment are: How does having a body influence our thought processes? Our language? Our actions? What kind of body is needed for what kind of cognition? (Gibbs Jr, 2005). Researchers in the field of robotics and HRI are often referring to the definition by Pfeifer and Scheier (1999) who define embodiment as "[a] term used to refer to the fact that intelligence cannot merely exist in the form of an abstract algorithm but requires a physical instantiation, a body. In artificial systems, the term refers to the fact that a particular agent is realised as a physical robot or as a simulated agent." Following this definition, we will explore the effect of robot partners being embodied or having a body in HRI.

Embodiment in Human-Robot Interaction

While robots that are structurally coupled with their environment need to have a physical body to have an effect on their environment, many social interactions with robots can be accomplished in disembodied ways (Deng et al., 2019). The question of when it might still be relevant or of advantage to provide the robot with a physical body has been in the focus of HRI researchers over the past years.

Even though social interaction with robots can be accomplished in embodied and disembodied ways, studies investigating robots' physical embodiment indicate that physical embodiment influences HRI in a meaningful way. For example, in an early study conducted by Lee et al. (2006), in which participants were either introduced to a physical Aibo robot dog, or its screen equivalent, a positive effect of embodiment on ratings of the interaction with the robot and the robot's social presence could be found. After a first introduction, participants in both experimental groups were instructed to try out the robot's various sensors, which functioned the same in both versions, and to interact freely with the robot for 10 minutes. Ratings of the pleasantness of the interaction and the social presence of the robot were positively correlated with physical embodiment, indicating the importance of physical embodiment in HRI, even if it is not necessary to complete the interaction successfully.

In a similar study conducted by Kwak et al. (2013) children were asked to administer electric shocks to either a physically present robot or a simulated robot on a computer screen. Both versions of the robot showed colored bruises as a reaction to the electric shock to indicate pain. Afterwards, the children were asked to indicate how much they empathized with the particular robot. While both robot versions showed the same reaction to the stimulus ratings of empathy varied significantly across embodiment conditions. Children rated their empathy with the embodied robot as much higher than with the simulated robot.

Together, these results show that robot embodiment is a relevant factor to keep in mind when designing robots for HRI, however it has to be further explored what kind of embodiment exactly is needed for what kind of interaction and which other factors relating to embodiment might shape the interaction and the way we perceive our robot partners.

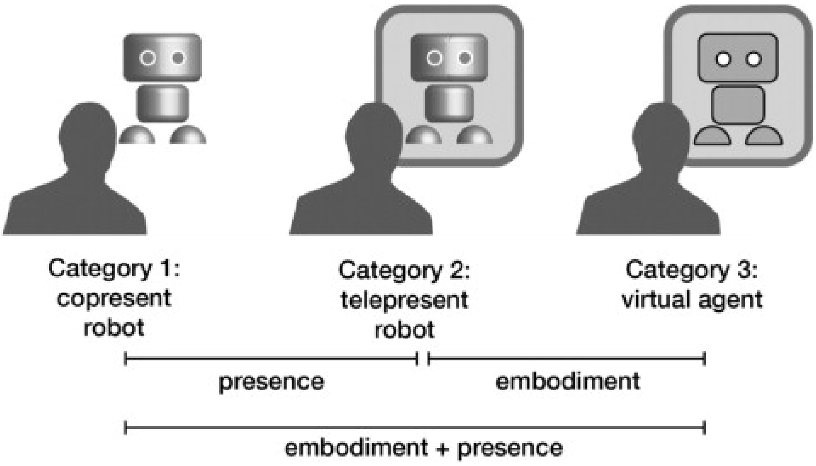
Physical Presence in Human-Robot Interaction

While these positive effects were for a long time attributed to embodiment of the robot, a meta-analysis conducted by Li (2015) indicates that it is physical presence rather than physical embodiment that accounts for those effects. Physical presence usually refers to the robots location in the same or a different environment than its interaction partner (e.g. Wainer et al., 2006; Bainbridge et al., 2008; Leyzberg et al., 2012), different environments hereby require the robot to be presented using some medium, e.g. a computer screen or virtual reality. The agents compared in the studies investigated in the scope of the meta-analysis were differentiated into three groups based on the way they were embodied and they way they were presented to their interaction partner (see Figure 1):

- *virtual agents* were defined as agents that are neither embodied nor present in the same environment as the participants
- *remote/telepresent agents* were defined as agents that are embodied but not present in the same environment as the participants
- *copresent agents* were defined as agents that are embodied and present in the same environment as the participants

A total number of 33 research works were included into the analysis. Individual analyses indicated that copresent robots are generally perceived more favorable and as more persuasive than telepresent robots. Multiple studies also reported better performance in HRI involving robots that are physically present in the same environment as the human interaction partner compared to screen based interaction scenarios. Interestingly, the authors did not identify differential results for the comparison of virtual characters and real-time video feeds of robots. Taken together these results indicate that while embodied robots generally elicit more positive evaluations, it is the effect of being physically present that is responsible for this effect.

Figure 1: Types of Robot Embodiment and Presence in HRI Literature



Note. From Li (2015); Copresent robots are physically embodied and present in the same environment as the observer, telepresent robots are physically embodied and presented via a medium, e.g. a computer screen, virtual agents are virtually embodied and presented via a medium.

For example, Wainer et al. (2006) carried out a study in which participants were (a) directly interacting with a physical robot; (b) indirectly interacting with a physical robot over a real-time video-conferencing link; or (c) interacting with a simulated robot on a screen. The task comprised of solving a Tower of Hanoi puzzle under the instruction of the robot and was followed by a set of questionnaires. Participants rated interacting with the co-located robot, i.e. the robot they were directly interacting with as more enjoyable and stated that they watched it more than remote-located (video-conference) robot or the simulated robot.

Physical Presence and Social Attributes

While some of the positive effects attributed to physical presence can be explained through task-affordances offered by the interaction partner being embodied in the same environment, some researchers argue that it might also shape the way we respond socially and understand robots as social relevant interaction partners (Leyzberg et al., 2012; Jung and Lee, 2004). In a study conducted by Bainbridge et al. (2011) using a Nico robot, the researchers investigated the effects of the robot interaction partner being physically present or presented on a screen on a series of behavioral and self-report measurements. While both ways of presenting the robot elicited similar responses with regards to task completion and simple interactions, such as greetings, participants were more likely to fulfil an unusual request voiced by the physically present robot and were granting it more personal space compared to the virtual robot.

One possible explanation might be that the way we present robots in HRI changes our perception of them with regard to their social standing and we thus treat them in a way we would treat other social actors. For example, Leyzberg et al. (2012) found positive effects on learning outcomes when participants were tutored by a physically present robot compared to one that was presented in a video or a disembodied voice, and concluded that “[p]erhaps robot’s physical presence increases its authority or social standing.” (Leyzberg et al., 2012). This is also consistent with the results of Jung and Lee (2004), according to which participants who interacted with a physical Aibo robot rated its social presence higher than that of the digitally presented version. It is important to note, however, that the authors were unable to replicate their results in a second experiment (2004, study 2).

In summary these findings indicate the importance of physical presence regarding social interactions. The reasons for possible positive effects can be manifold and can range from pure task physicality - as in work by Shinozawa et al. (2005) in which participants rated a copresent robot more favorably than a screen version of the same robot when interacting on a task in the “real” world (but not on a computer screen), its hightened social authority, as proposed by Leyzberg et al. (2012), or even effects on social presence evaluations, as indicated by findings of Jung and Lee (2004). Due to the novelty of this branch of research and the broad scope of phenomena that might be affected differently by the way a robot is presented, further investigation of the relationship of presence and different areas of social interaction will be required to better understand social human-robot interaction. Moreover, it is necessary to consider the use of new technologies such as virtual reality as they could compensate for situations in which it is impossible or not desirable for a robot to be physically embodied in the same space in an experimental environment. The reasons for this can be manifold and range from cost efficiency and safety to heightened naturalness of the experimental conditions.

Presence in Virtual Reality

Virtual reality (VR) is gaining popularity in HRI (Williams et al., 2018) as it could be a way to avoid problems researchers face in classical HRI studies and still create an immersive environment for HRI. Virtual reality (VR) can be described as a computer-based medium that perceives the user’s actions and interacts with their senses in such a way that the user feels mentally present in the simulation (Sherman and Craig, 2018). Historically, this kind of immersion was mainly possible using big screens in the so-called Cave Automatic Virtual Environment (CAVE) and was hence hardly attainable for the general public and rarely used by the research community. Nowadays, with the availability of more affordable and convenient Head-mounted displays (HMDs), it is more commonly used in research and can even be found in people's homes (Li et al., 2019).

In the field of HRI, the use of virtual reality technologies offers several new possibilities and advantages over standard laboratory studies or studies with robot simulators on conventional computer screens. Study environments can be created that are not only more immersive (Mestre, 2017) than their screen opponent, but also offer more possibilities to interact (Goedicke et al., 2018) and hence provide a more natural environment for research. In addition to that, they can offer more cost-efficient (Inamura and Tan, 2012) and safe (Shu et al., 2019) circumstances for research and development in the field of robotics, as tasks can be learned, optimized, or replicated in a controlled virtual environment instead of the real world. This would also open up new opportunities in the field of cognitive robotics, which is focusing on developing intelligent robots that can perceive the environment, act, and learn from experience to adapt their generated behavior accordingly to the interaction.

However, not only the development of robots in HRI can benefit from the use of virtual reality, but also the interaction itself and the robot perception could be positively shaped through the use of virtual reality. In a series of interaction studies, Liu et al. (2017) had participants collaborate with a robot in tasks that involved moving a plate so the robot can drop an item on it. The robot was presented either on screen or in virtual reality. Participants that were interacting with the robot that was VR-present were significantly quicker and more accurate in their decision on where to move the plate so that the robot could drop the items on it, suggesting that presenting the robot in VR may offer an advantage over presenting it on the screen in terms of task performance. While these results seem to indicate that VR could be a new tool for HRI, it is important to investigate how it affects different factors of the interaction and how it compares not only to telepresence but also to real-world copresence.

When Li et al. (2019) compared people's perception of a real robot and a virtual representation of the same robot in VR, as well as the way the robot presentation shaped the proxemics of the interaction, they found significant differences between the two conditions. Participants rated the real robot as more comforting and chose closer interaction distances with it compared to the robot presented in a virtual environment. Despite the fact that the results presented above indicate that presenting the robot partner in VR can influence specific outcome variables, the spectrum of variables to study is broad. This is why Wijnen et al. (2020) recommended replicating existing HRI research using VR to better understand how comparable results in VR are with well-studied real-world phenomena.

Summary: Embodiment and Presence in Human-Robot Interaction

Scientific investigation of factors shaping human-robot interaction have pointed to the importance of embodiment and presence of the robot. Both factors seem to positively influence perception of the robot, as indicated by higher ratings of trust, likeability and related factors. However, it is not always possible or preferable to introduce physically embodied robots into HRI. Therefore, new ways of presenting robots in HRI need to be explored. Advances in VR offer promising new spaces for HRI. To investigate effects of presence and embodiment, as well as the general transferability of HRI experiments to the real world, systematic studies are needed that compare the effects of different forms of presence on multiple outcome variables related to social interaction.

1.2.3 Social Gaze

Humans naturally engage in social interactions. These interactions involve complex exchanges of a variety of social signals, such as pointing, body posture, and movement fluidity. Of these social signals gaze particularly appears to play a crucial role as it is based on “hard-wired” neurocircuits in the brain, linking the superior temporal sulcus, amygdala and orbitofrontal cortex – regions of central importance to social cognition Emery (2000). In contrast to other social signals, gaze holds as specific role in perceiving information from and signaling information to others (Risko et al., 2016). For example, our gaze can cue another’s attention (Kuhn et al., 2009), indicate social interest (Stass and Willis, 1967) or physical attraction (Mason et al., 2005). At the same time, when we see another person’s eyes their gaze can give us information about their mental and emotional state (Baron-Cohen et al., 1997), as well as what their attention is focussed on (Frischen et al., 2007). This process of using someone else’s eye movement as information of what they are attending and shifting one’s own attention accordingly is called gaze cueing and is discussed as a prerequisite for joint attention (Emery et al., 1997), the case in which both persons visually attend the same object. Gaze cueing effects have been extensively studied and replicated in both humans and nonhuman primates using brain imaging as well as behavioral studies (for a review: Frischen et al., 2007).

Gaze Cueing Paradigms

Initial research on gaze cueing was conducted by Friesen and Kingstone (1998), who based their experiments on Posner's historical cueing paradigm (Posner, 1980; see Figure 2). Experiments using Posner's cueing paradigm usually proceed as follows: Participants are instructed to fixate on a cue in the centre of the screen and to respond to the appearance of a target stimulus on the left or right side of the cue by pressing the corresponding key on a keyboard as quickly as possible. Participants are usually better in detecting cued target stimuli compared to uncued target stimuli. These results indicate that people shift their attention towards the cued location.

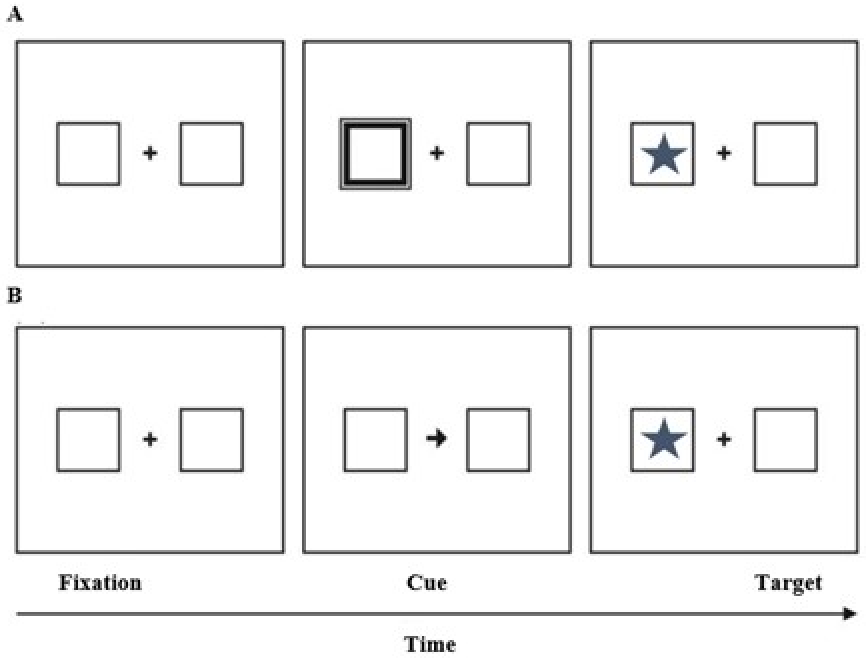
Friesen and Kingstone (1998) designed their experiment on the effect of attentional cues on attention shifting in adults in a similar fashion: Participants were asked to respond to target letters appearing on either the left or right side of the screen. The appearance of the stimulus was cued by a schematic face that either gazed at the corresponding side of the screen (valid trials; hereafter referred to as *congruent trials*), the opposite side of the screen (invalid trials; hereafter referred to as *incongruent trials*), or straight ahead (neutral trials) with varying cue-target stimulus onset asynchronies (SOA). Participants were informed that the gaze direction was not predictive of the appearance of the target stimulus. Still, results showed that participants reacted faster in congruent trials compared to incongruent and neutral trials. The appearance of this so-called gaze cueing effect showed for short cue-target SOAs and disappeared with longer cue-target SOAs (1,005 ms).

A later study conducted by Driver IV et al. (1999) replicated the results using photographs of faces that were either looking at the right or left side of the screen and worked as uninformative reoccurring cues. Reaction times for congruent trials were significantly faster than for incongruent trials, even when participants were informed that the gaze direction is uninformative of the appearance of the stimulus. Collectively these results suggest that observing another's gaze shift results in a corresponding shift of attention in the observer, even, when the gaze shift is of seemingly no informational value.

Processes Underlying the Gaze Cueing Effect

While studies suggest that gaze following occurs automatically in social interactions, there is yet no full understanding which types of stimuli and cognitive processes are exactly involved in processing the gaze of others. As a gaze cueing effect is consistently found in paradigms using non-social stimuli (e.g. Ristic et al., 2002) it is often explained in regards to salient stimulus-related information, based on *bottom-up processes*. In the case of social gaze cueing stimuli, the mechanism used to be explained with regards to

Figure 2: Posner's Cueing Paradigm



Note. Adapted from Frischen et al. (2007). Typical trial in a Posner cueing task. After fixation on a central cue, participants are asked to localize a target stimulus (here star) by pressing a corresponding key. RTs are expected to be shorter for congruent trials than incongruent trials. Example congruent trial on the top, neutral trial on the bottom.

purely physical and geometrical cues, such as the shape of the iris which deviates from the rest of the eye region or the brightness of the sclera compared to the darkness of the iris (Anstis et al., 1969; Ricciardelli et al., 2000). Such bottom-up processes are unconscious and hence not voluntarily controllable.

Top-down processes on the other hand, are driven by assumptions and are subject to voluntary control. They come into play when stimuli are ambiguous or symbolic and depend on higher-order cognitive evaluation. Recently, studies have investigated gaze cueing with a specific focus on possible top-down effects and found that modulation of social context information is related to changes in the gaze cueing effect (e.g. Dalmaso et al., 2020; Gobel and Giesbrecht, 2020). In addition, current research has pointed to the importance of beliefs about the gazer's mental state and thus its social relevance in gaze cueing paradigms. For example, a number of studies showed reduced gaze cueing effects after participants' beliefs about whether or not the observer could actually see the target were manipulated by informing them that the glasses they were wearing either allowed or blocked vision (Teufel et al., 2010; Morgan et al., 2018). Shorter reaction times for incongruent trials in the "blind" conditions suggest the importance of mental state attributions in interactions involving gaze cues.

Gaze Cueing in Human-Robot Interaction

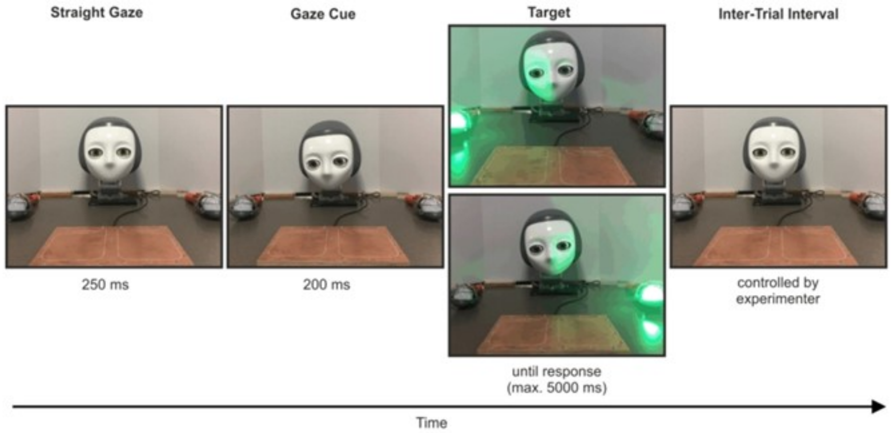
Because of its importance in social interaction, gaze was first introduced into robotic systems with the development of social robots. Famous examples of first gazing robots are Kismet (Breazeal and Scassellati, 1999) and Infanoid (Kozima and Ito, 1998), which were developed as conversational agents. The introduction of gaze was intended to support conversational flow by promoting engagement and directing attention (Admoni and Scassellati, 2017). However, research has shown that not every robotic stimulus is inducing gaze following effects in human observers (Admoni et al., 2011).

Martini et al. (2015) modulated a robots human-likeness by using a morphing paradigm in combination with a gaze cueing paradigm. The results indicate a nonlinear relationship of anthropomorphism and strength of the gaze cueing effect. Another important factor influencing gaze following behavior in HRI similar to results presented above seems to be mental state attribution and judgment of the robot's social relevance. In a study using photographs of human and robot faces, Wiese and colleagues found a gaze cueing effect specific to the attribution of mental states. While no gaze cueing effect was observed in a first experiment without manipulating participants' beliefs about the robot's mental states, a second and third experiment in which participants were told that the robot was operated by a human showed consistent gaze following behavior.

While most of gaze cueing studies only used static images of robots gazing, Wiese et al. (2018) employed a gaze cueing paradigm using an embodied humanoid robot. Participants were instructed to indicate the appearance of a light stimulus on their left or right side, which was cued by a Meka robot's gaze shift (Figure 3). Even though the researchers informed the participants that the robot's gaze is uninformative of the appearance of the light stimulus, participants seemed to follow the robot's gaze, as a congruency effect could still be found. Using a similar setup, Kompatsiari et al. (2018) were able to replicate these results with an iCub robot when the robot established mutual gaze with the observer before turning to the target stimuli - as in the previously reported experiment (Wiese et al., 2018) - but not when no mutual gaze was established.

Importantly, further controlled investigations will be needed, in which participants are faced with scenarios and robots that are constant except for their embodiment or the way they are presented, to better grasp the role of embodiment and presence on gaze cueing effects in HRI. To our knowledge, there is only one systematic study conducted to date that explores the relation of embodiment, presence and facial cueing.

Figure 3: Gaze Cueing Paradigm Using an Embodied Meka Robot



Note. From Wiese et al. (2018) "Embodied social robots trigger gaze following in real-time".

Mollahosseini et al. (2018) had participants interact with one of four types of agents, each different in terms of their embodiment and presence (virtual agent, copresent robot, telepresent robot and human) in interaction tasks that involved typical measures of communication - visual language, facial expressions and gaze. Results indicate that while visual speech does not seem to be influenced by the way the interaction partner is embodied and present, recognition of facial expressions, and especially eye gaze seem to be related to these factors. While further analysis of the data revealed a significant effect of embodiment, unlike other studies examining social interaction in HRI, no main effect of presence was found. This could be due to the nature of the task, which involved only still representations of gaze rather than actual movements, or the fact that it generally did not rely on social attributions that could relate to the presence of the robot, but on purely geometric cues. Similar to other studies investigating the effects of presence and embodiment in HRI, also this one did not include agents embodied in virtual reality. Since this new technology could have different effects than classical ways of embodying robots, the below presented research aims to provide further insights and fill the current gap in research on the effect of embodiment and presence on gaze following behavior in HRI.

Summary: Social Gaze

Social gaze is a fundamental component of human communication. It involves signaling and perception and is based on core mental networks, indicating its importance for human development and interaction. Gaze cueing - the effect of noticing another's gaze and directing one's own gaze in that direction - is discussed as a deeply embedded, automatic process and a prerequisite for higher social functions such as joint attention. While some researchers argue that it relies primarily on unconscious bottom-up processes, recent research has led to the suggestion that conscious top-down processes also play a central role in determining the strength of gaze cueing effects. In some studies using images of or actual embodied robots, gaze cueing effects were found to be consistent with those elicited by the sight of a human face. The role of mental state attributions, embodiment, and presence remains to be investigated to create a comprehensive picture of gaze cueing in HRI.

1.3 Scope of the Project

Taking the previously outlined results as a point of departure, we now turn to the empirical part of this work. As will become apparent, to better understand the phenomena of social cognition in human-robot interaction, it is fundamental to understand how different aspects of robot design affect the way we perceive and interact with our robot interaction partner. Crucially, as described in the previous section, robots are sometimes viewed and interacted with like social actors, however, different ways of presenting the robot can affect the extent to which we do so. While the study of gaze is a broad field, to our knowledge, there has not yet been a study that attempts to make connections between social gaze and the way we present robot interaction partners in HRI. Therefore, the present study aims to expand our knowledge of presence and gazes in human-robot interaction by investigating gaze cueing at different levels of robot presence in a systematic experimental paradigm.

1.3.1 Research Questions

Based on the assumption that physical presence influences robot perception and social interaction, with studies reporting differences between copresent robots and telepresent robots or virtual agents (Wainer et al., 2007), as well as first studies pointing towards differences between copresent robots and virtual reality-present robots (Li et al., 2019), we hypothesize that we will find differences in robot perception when comparing these three presence conditions.

Further, as robot presence has been shown to affect social HRI, we assume differential results in behavioral measures when comparing copresent robots with virtual reality present versions of the same robot and virtual or telepresent versions of the same robot. Crucially, based on findings that physical presence can positively influence a number of measures of social interaction like learning (Leyzberg et al., 2012), affording personal space (Bainbridge et al., 2011) or secret keeping behavior (Wijnen et al., 2020), we hypothesise participants to engage in more sociable interactions with robot showing heightened levels of physical presence, i.e. copresent robots $>$ virtual reality present robots $>$ virtually present robots (telepresent robots or screen agents).

With respect to social gaze, studies have pointed to typical gaze cueing effects when interacting with a physically present embodied robot (Wiese et al., 2018; Kompatsiari et al., 2018). However, the way in which gaze cueing is related to physical presence has not been sufficiently explored. Our goal is to investigate how the way we present a robot in social HRI influences typical gaze behavior. Based on this rationale it is asked whether we can replicate existing results on gaze cueing effects and whether the way a humanoid robot is presented to us influences the strength of this effect. Furthermore, it is explored to what extent the way we present a robot changes our perception of the same robot, as indicated by subscales of the Godspeed series (Bartneck et al., 2009).

1.3.2 Experimental approach

To approach these questions, an existing experimental paradigm that allows the study of attention cueing with humanoid robots (Wiese et al., 2018) was adopted to explore the effects of robot presence and gaze. The task comprised of localizing the change of light in a target stimulus, that was either congruent or incongruent to the location of a robot's gaze. As previous studies of the gaze cueing effect found a gaze cueing effect for embodied humanoid robots (Wiese et al., 2018), but reported inconsistent results for images of robots (Admoni et al., 2011; Wiese et al., 2014), we adapted the design to afford the exploration of the role of presence on this effect. The robot indicating the possible appearance of a target stimulus was either a copresent robot or a virtually present robot. This way possible differential effects of presence on gaze following behavior can be investigated.

Due to time constraints and the difficulty of implementing exactly the same conditions as the other conditions in VR, only the copresence and virtual presence conditions were implemented and studied for this work, while the inclusion of virtual reality offers a perspective for the future. However, because there has been no systematic research on the relationship between robot presence and gaze, this experiment is a novelty. Future research on how virtual reality fits into this relation may complete the picture.

1.3.3 Hypotheses and Predictions

Based on the findings by Wiese et al. (2018), it is hypothesized that robot eye movement will elicit a gaze cueing effect. This hypothesis gets reflected in the predictions that participants will respond quicker and more accurate in congruent trials compared to incongruent trials. Gaze cueing effect has consistently been demonstrated across different types of interaction partners (Driver IV et al., 1999) and even abstract symbols (Friesen and Kingstone, 1998). Thus, we expect to replicate the results of previous studies and contribute to a broader picture of gaze following behavior in humans.

Crucial for the stated research questions, it is further hypothesized that the robot’ presence affects this gaze cueing effect, as presence has been shown to influence multiple factors of HRI. Due to inconsistencies in research about presence and gaze following, we won’t make predictions about the direction of this effect. In an attempt to investigate the effect of presence on the gaze cueing effect, Lachat et al. (2012) studied gaze cueing in face-to-face human interaction and found a consistent gaze cueing effect, that was comparable to on-screen gaze cueing effects, while Admoni et al. (2011) found no gaze cueing effect for images of robots cueing a target. Wiese et al. (2014) were comparing gaze cueing effects in autistic individuals that were confronted with either a human or a robot interaction partner. Individuals in this experiment showed stronger gaze following behavior for the robot compared to the human. While these studies indicate that presence and embodiment can affect gaze cueing behavior, there has, to our knowledge, yet been no investigation of different levels of robot presence, on gaze cueing effects.

We expect to find no main effect of robot presence condition on RTs and accuracy, as no such has been reported to our knowledge.

Lastly, it is hypothesized that the way we present robots can influence robot perception. Based on the previously outlined results, we hypothesis participants to show differences in robot perception as indicated by the outcomes of the Godspeed subscales.

Chapter 2

Methodology

2.1 Experimental Setup

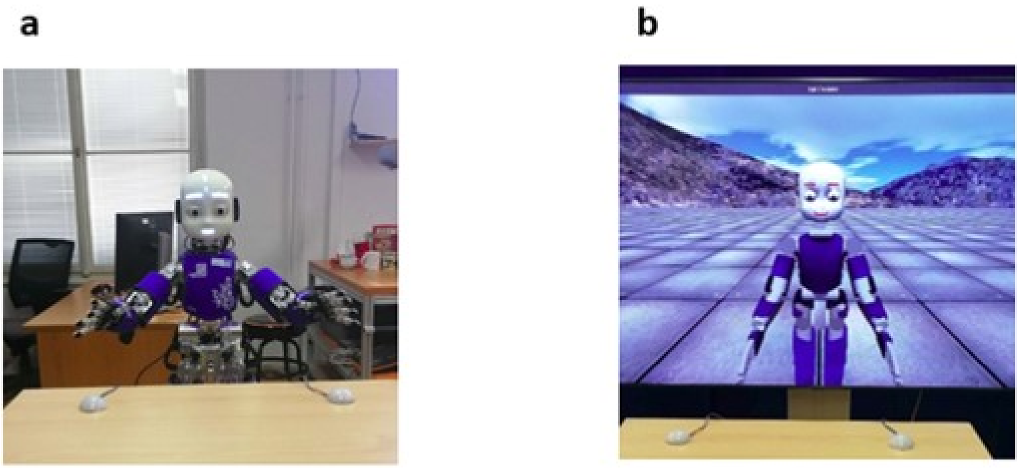
The present experiment was designed to examine both self-reported and behavioral effects of a robot's presence and gaze cues in a human-robot interaction task. During the experiment, participants were instructed to indicate the appearance of a target stimulus that was either congruent or incongruent with the position being gazed at by an iCub robot that was either physically present in the same room with the participants (copresence condition) or presented via a monitor (virtual agent condition) by pressing a corresponding key on a keyboard. After completion of the task, participants were asked to indicate the way they perceived the robot by completing the Czech translation of the three subscales “Anthropomorphism”, “Animacy” and “Likeability” of the Godspeed series (Bartneck et al., 2009). The experiment was conducted in April 2022 in the laboratories of the Department of Cybernetics of the Czech Technical University in Prague. The study was approved by the Committee for Research Ethics at the Czech Technical University in Prague under the reference number 00000-07/21/51903/EKCVUT.

2.2 Experimental Design

We used a 2 (copresence, virtual presence, between subjects) x 2 (congruent, incongruent, within subjects) mixed factorial design. The participants for this experiment were divided into two groups, each of which took part in one of two robot presence conditions. Conditions were run on separate days and participants were assigned to conditions based on their times available to participate in the experiment. Participants did not know beforehand what their task would be and which condition they would be assigned to.

In the copresence condition (Figure 4.a), participants performed the task in the same room as the iCub humanoid robot (Metta et al., 2010). The robot is approximately 104 cm high, weights about 24 kg, and has the shape of a 4-year-old child. Facial expressions were enabled. During the experiments, the robot always assumed the same neutral/positive expressions, to avoid confusing the participant or suggest that the participant's actions could arouse an eventual robot "emotional status". The robot behaviour was preprogrammed employing the Cartesian gaze controller described in Roncone et al. (2016).

Figure 4: Robot Presence Conditions



Note. Figure 4a shows the copresent iCub robot used in the experiment, Figure 4b shows the simulated version of the iCub robot (virtual agent) used in the experiment.

In the virtual agent condition (Figure 4.b), participants interacted with a simulated version of the iCub robot in the Gazebo simulator being presented in a face-forward pose, on a 65 inch Samsung LED monitor. The monitor and the presented image were adjusted so that height and head size of the robot matched the real iCub.

For both the copresence condition and the virtual agent condition, the experimental setup consisted of an 86.5 cm high table on which a keyboard and two custom-made light stimuli were placed 65cm apart from each other. Participants sat on a chair so that they could comfortably reach the keyboard with their hands. Depending on which condition the participants were assigned to, there was either a real iCub or a monitor showing the simulated iCub robot on the other side of the table. To summarize, the main difference between the two presence conditions was whether or not participants were interacting with a real iCub in the same room or a simulated iCub.

2.3 Participants

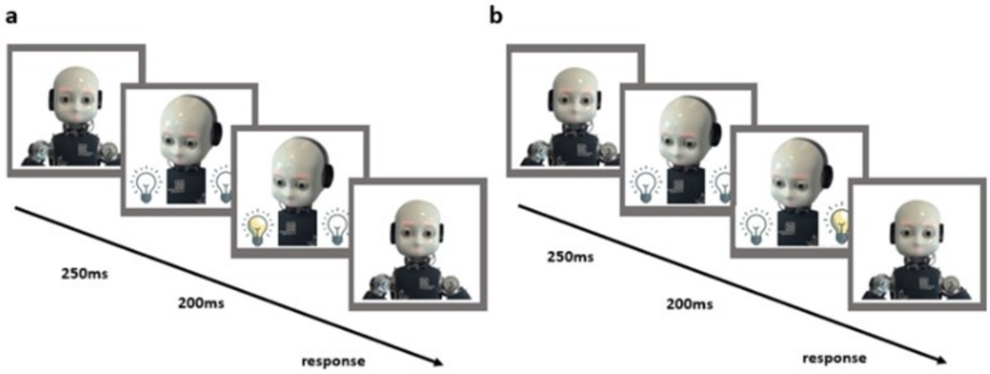
A total of 42 participants took part in the study, 3 of whom were excluded and retested due to interruptions or technical difficulties (for data of the three excluded subjects see Appendix). The final sample consisted of 42 participants, 21 each of whom were assigned to the copresence condition and the virtual agent condition. Participants were recruited via posters in the university building, Facebook, and email. Subjects conducting research on or working with robots were excluded from participation in the experiment. All participants provided written informed consent in line with the ethical approval of the study granted by the Committee for Research Ethics at the Czech Technical University in Prague. 26 participants were male, 16 were female. Participants were between 19 and 60 years old and the average age was 29.98 years (SD = 10.68). The mean age did not differ between experimental conditions (t=0.1, p=.9). When asked about their experience with robots on a scale from 1 (very poor) to 5 (very good), the mean score was 2.33 (SD = 1.2) and only one participant answered 5.

2.4 Procedure

Data collection for both conditions was conducted in the laboratories of the Department of Cybernetics at the Czech Technical University in Prague, Czech Republic. Upon arrival at the laboratory, participants provided written informed consent to participate in the experiment and learned about the experimental task by first reading the instructions and then having an experimenter describe the task to them. They were informed that the task was to locate the color change of one of two self-made light stimuli (for a reference see: Samporová (2022)) by pressing the appropriate key on a keypad, while the robot randomly changes its gaze direction. They then had the opportunity to ask questions about the task.

The experiment consisted of 80 trials, 40 congruent trials in which the robot looked at the lamp that was to change color and 40 incongruent trials in which the robot looked at the opposite lamp. In half of the conditions the robot looked at the left side and in the other half at the right side. All conditions were pseudo-randomly shuffled. Each trial began with iCub making eye contact by looking straight ahead in the direction of the observer. After 250 ms, iCub shifted his gaze either toward the lamp that was on his left side or toward the lamp that was on his right side. Subsequently, after 200 ms, one of the two lamps changed color, either on the corresponding side of the gaze cue or on the non-corresponding side of the gaze cue. When the target stimulus was presented, participants responded as quickly and as accurately as possible to the position of the target stimulus by pressing the “x” or “m” key on a standard keyboard. The target stimulus remained unchanged until a response was made or a time-out criterion (5000 ms) was reached. Then the light was turned off again and iCub looked straight ahead again to signal readiness to begin the next trial. Figure 5 shows an exemplary trial sequence.

Figure 5: Example Experimental Trial Sequence



Note. Example trial sequence with (a) congruent condition and (b) incongruent condition for the physically presence condition. Target stimuli are represented as schematic depiction (light bulb vs. self-made light stimuli in the actual experiment).

2.5 Measures

2.5.1 Godspeed Subscales

To assess perceptions of the robot, participants were asked after the experiment to indicate their perceptions of the robot by completing the Czech translation (for a reference see: Lehmann et al., 2020) of the three subscales “Anthropomorphism”, “Animacy”, and “Likability” of the Godspeed series (Bartneck et al., 2009). The Godspeed series is widely used in HRI research (Weiss and Bartneck, 2015), so its use can help provide results that are comparable across studies. It consists of a total of five subscales - Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety. In general, it is believed that high scores on these scales indicate that the robot is perceived positively, which in turn may lead to productive interactions. The anthropomorphism subscale consisted of 5 items (α = .757), the animacy subscale consisted of 6 items (α = .758), and the likeability subscale consisted of 5 items (α = .895). All items consisted of semantic distinctions, e.g., inauthentic/natural on a five-point Likert scale.

2.5.2 Reaction Time Measurements

Similar to previous studies on gaze cueing (Wiese et al., 2018; Driver IV et al., 1999), the influence of gaze cueing on participants' gaze following behavior was determined by measures of mean correct reaction time. A response was considered incorrect if it was made with the wrong key press, and considered correct if the correct key was pressed. Responses that were given in a response time that was more than 2.5 standard deviations away from the individual mean response time of a participant were excluded from further analyses. To further check for possible effects of a robot's physical presence on gaze following behavior, average correct response times (RTs) were calculated for each participant and each experimental condition (see next section for details of the statistical analyses).

2.6 Analysis

Statistical analyses were conducted using R version 4.1.2. The average correct RTs for congruent and incongruent trials were compared for each robot presence condition individually to check for consistency with previous studies regarding the strength of the gaze cueing effect. To test for the effect of cue-target congruence, for the virtual robot condition a t-test was calculated comparing the mean reactions times of congruent trials and incongruent trials. For the copresent robot condition a Wilcoxon test was calculated instead, to account for non-normally distributed data. Significant differences in reaction times between congruent and incongruent trials evidence the presence of a gaze cueing effect. Moreover, a mixed model analysis of variance (ANOVA) with a between-subject factor of robot presence (2: copresence vs. virtual presence), and within-subject factors of cue-target congruency (2: congruent vs. incongruent) was calculated. This analysis was used to assess the individual effect of cue-target congruency and the effect of presence specificity of gaze cueing. Robot presence specific gaze-cueing effects would be evidenced by a significant interaction between presence condition and cue-target congruency (over and above a main effect of congruency). By contrast, presence nonspecific gaze cueing would manifest in terms of a main effect of cue-target congruency (not accompanied by a presence x congruency interaction), with equal facilitation for all robot presence conditions of interest. All follow-up comparisons between conditions were carried out with paired t-tests.

Chapter 3

Results

The following section presents the results of the study. First off, different correlation tests checked for the association of demographic variables on the different outcome variables. Next, the first hypothesis was tested by conducting separate analyses for the two robot presence conditions, comparing the mean reaction times of congruent and incongruent trials. Afterwards, the presence specificity of a gaze cueing effect was tested by conducting a mixed Analysis of Variance with one within factor - cue-target congruency, and one between factor – robot presence. Further, to test the relation of robot presence on the way participants perceive the robot, Wilcoxon tests and T-tests were conducted. For all conducted analyses the necessary requirements have been tested and were met (see Tables 3 to 6 in Appendix B) if not indicated otherwise in the respective section.

3.1 Test for Control Variables

Firstly, to test for possible connections of the independent variables age, sex, and experience with robots on the outcome variables, mean reaction times in congruent and incongruent trials and ratings of anthropomorphism, animacy and likeability, correlation tests were conducted. Results of this can be seen in Table 1.

There were no significant correlations between the tested demographic variables and the mean reaction time for incongruent trials and ratings of the robot's animacy. Therefore, these variables were not considered as control variables in further analyses of associations on these two outcome variables. Importantly, the correlation analysis revealed a significant association of sex and ratings of the robot's anthropomorphism; $r=.32$, $p<.01.$, as well as sex and ratings of likeability; $r=.40$, $p<.01.$, indicating that overall women perceived the robot as more anthropomorph ($M = 13.6$, $SD = 3.39$) and more likeable ($M = 21.2$, $SD = 3.36$), than did men ($M = 11.3$, $SD = 3.18$; $M = 17.8$,

Table 1: Pearson Correlations Demographics

	1	2	3	4	5	6	7	8
1 sex	1							
2 age	-.50***	1						
3 knowledge of robots	0.26	-0.17	1					
4 anthropomorphism	-.32*	-0.08	0.05	1				
5 animacy	0.25	-0.1	-0.04	.74***	1			
6 likeability	-.40*	0.12	-0.08	.64***	.58***	1		
7 congruent RT	-0.19	.28*	-0.1	-0.28	-0.28	0.05	1	
8 incongruent RT	-0.21	0.31	-0.03	-.32*	-0.22	0.06	.93**	1

Note. *** $p<.001$, ** $p<.01$, * $p<.05$

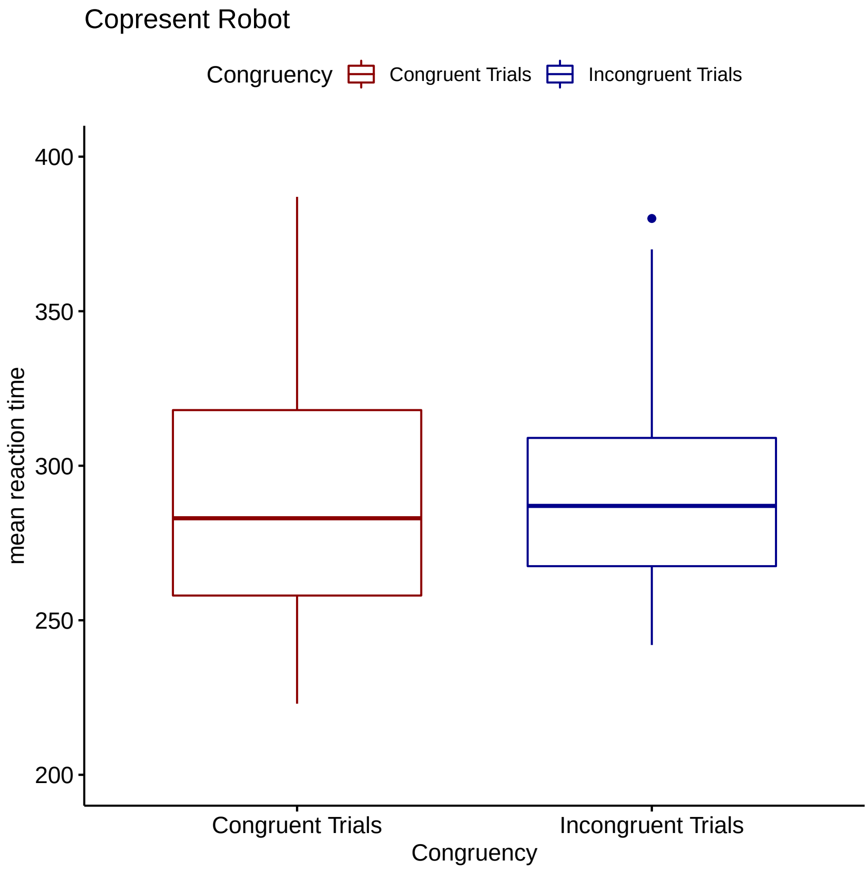
SD = 4.09).

Owing to theses significant correlations, additional test including sex as a covariate will be conducted to control for the relation of these variables. Further, a significant association of age and mean reaction times in congruent trials; r=.28, p<.01, indicates that in general younger participants responded quicker in congruent trial than did older participants. An additional analysis including age as a covariate will be conducted to control for possible effects on the investigated association of robot gaze and presence on mean reaction times.

3.2 Gaze Following Behavior by Groups

Mean reaction times were compared for congruent and incongruent trials, respectively for both conditions of robot presence. The results of the analysis of reaction time data are shown in Figures 6 and 7. For the comparison of mean reaction times in congruent and incongruent trials in the copresent robot group a Wilcoxon signed-rank test was calculated to account for non-normally distributed data as indicated by a significant Shapiro-Wilk test (W = .9, p = .009). On average, participants in the copresent robot condition responded faster on congruent trials (M = 291 ms) than on incongruent trials (M = 305 ms). Results of the Wilcoxon signed-rank test showed that this difference was statistically significant (p =.001), with a large effect size, r = 0.7. An additional t-test conducted on the mean reaction times of participants in the virtual agent group revealed a significant difference between congruent and incongruent trials (t (20) = -2, p = .03) with shorter reaction times for congruent (M = 284 ms) than incongruent trials (M = 291 ms). The effect size was at a moderate level, r = 0.5.

Figure 6: Differences in Average Reaction Times in Congruent and Incongruent Trials, Copresent Robot Condition

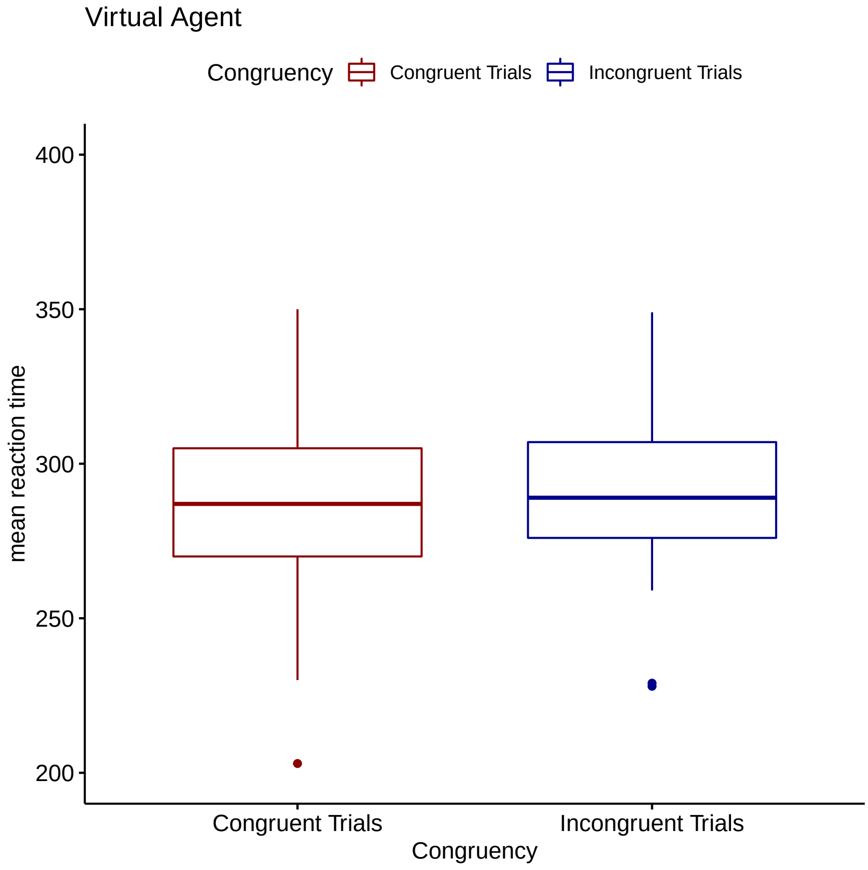


Note. Mean reaction times are reported in milliseconds.

3.3 The Effect of Presence and Gaze

In the following, mean reaction times were subjected to a two-way analysis of variance with two levels of robot presence (copresent robot, virtual agent) and two levels of cue-target congruence (congruent, incongruent) to test the effect of robot physical presence and gaze cueing behavior on participants' reaction times in a localization task (see Table 2). It is important to note that the data was not normally distributed in each group (see Appendix B). An ANOVA was conducted either way, as F-Tests have been reported to be relatively robust to violations of normality when homogeneity of variance is given (Blanca Mena et al., 2017). The main effect of gaze congruency yielded an F-ratio of F(1, 40) = 18.71, $p < .001$, indicating that mean reaction times differed significantly between congruent and incongruent trials. This result was followed by a simple comparison between the congruent and incongruent trials using a paired t-test. The results of the paired t-test showed that, on average, participants had faster reaction times on congruent cue-target trials (M = 288 ms, SD = 37.2 ms) than on incongruent cue-target trials (M = 298 ms, SD = 43.2 ms).

Figure 7: Differences in Average Reaction Times in Congruent and Incongruent Trials, Virtual Agent Condition



Note. Mean reaction times are reported in milliseconds.

This difference was significant t(80) = -1, $p < .01$, with a moderate effect size of r = 0.7. While participants in the virtual agent group (M = 287 ms) on average reacted faster than participants in the copresent group (M = 298 ms), the main effect of robot presence on participants' mean reaction times was non-significant F(1, 40) = 0.64, $p < .05$. Moreover, no significant interaction effect of gaze congruency and presence could be found, F(1, 40) = 2.22, $p < .05$.

Table 2: Gaze Congruence x Robot Presence Analysis of Variance

Source	Df	F	$\mu 2$	p
Robot Presence	1	0.64	0.02	0.43
Gaze Congruence	1	18.71	0.32	.001***
Presence x Gaze	1	2.22	0.05	0.14
Error	40			

Note. *** $p<.001$, ** $p<.01$, * $p<.05$

3.4 Robot Presence and Godspeed Indices

To test the hypothesis that robot presence influences the way participants perceive the robot in HRI, responses of the participants to the Godspeed subscales Anthropomorphism, Animacy and Likeability were taken into account. A standard t-test was used to examine the influence of robot presence on animacy ratings. Significant results in a Shapiro-Wilk test with mean ratings of anthropomorphism and likability as outcome variables indicated non-normally distributed data, so an unpaired two-samples Wilcoxon test was computed to test the influence of robot presence on perceived anthropomorphism and likability. Furthermore, since the correlation matrix revealed significant correlations between gender and the ratings of anthropomorphism and likability, linear regression models with sex as a covariate were conducted to control for possible influence of participants' sex. Mean ratings for all three Godspeed subscales by robot presence condition can be found in Figure 8.

3.4.1 Anthropomorphism

For ratings of the robot's perceived anthropomorphism, participants in the virtual agent condition assessed the robot's anthropomorphism slightly higher (M = 12.29, SD = 2.83) than people in the copresent robot condition (M = 12.05, SD = 4.03). Results of the independent samples Wilcoxon test, however, were not significant; W = 208.5, p = .77, r = -0.05. A linear model including sex as an additional predictor was tested due to the significant correlation of anthropomorphism ratings and sex. The model revealed no significant difference between presence groups, whereas ratings between sexes significantly differed ($p < .05$), with females (M = 13.56, SD = 3.44) rating the robot as more anthropomorphic on average than males (M = 11.31, SD = 3.21).

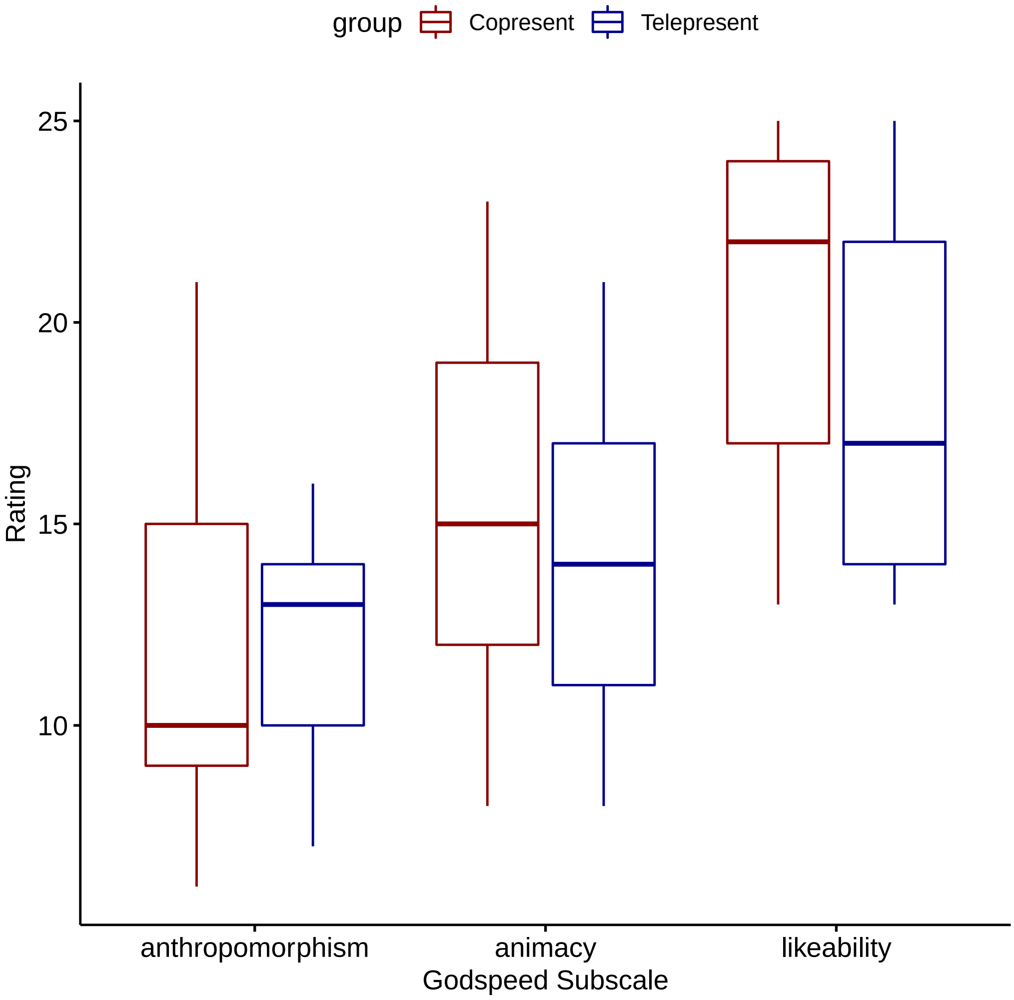
3.4.2 Animacy

On average, participants in the copresent robot condition rated the robot's perceived animacy higher (M = 15.19, SD= 4.50) than participants in the virtual agent group (M = 13.95, SD = 3.67). This difference was not significant; t(38.43) = 0.98, p = .33.

3.4.3 Likeability

Participants in the copresent robot condition assessed the robot's likeability slightly higher (M = 20.19, SD = 4.25) than people in the copresent robot condition (M = 18.10, SD = 3.94). Results of the independent samples Wilcoxon test indicate that this

Figure 8: Ratings of the Godspeed Indices by Group



Note. Scores in total numbers. The Anthropomorphism and Likeability scale consists of 5 sub-questions on a 5-point Likert scale with a maximum score of 25 points, and the Animacy scale consists of 6 sub-questions on a 5-point Likert scale with a maximum score of 30 points.

difference was not significant; W = 289, p = .09, r = 0.26. A linear model including sex as an additional predictor was tested due to the significant correlation of likeability ratings and sex. The model revealed no significant difference between presence groups, whereas ratings between sexes significantly differed ($p < .01$). On average, females (M = 21.23, SD = 3.41) rated the robot as more likeable than males (M = 17.84, SD = 4.13).

Chapter 4

Discussion

The present study investigated whether the way we present a robot in human-robot interaction influences basic social attention mechanisms, as measured by gaze-cueing effects as well as our perception of the robot, as measured by ratings of the Godspeed series (Bartneck et al., 2009) “Anthropomorphism”, “Animacy” and “Likeability”. Rather than confronting participants with only one way of presenting their robot interaction partner (Wiese et al., 2018; Kompatsiari et al., 2018), we varied the degree to which the robot is present with the participant, by having it be either copresent in the same room or virtually present via a computer screen, while keeping the stimuli and experimental setup constant. We hypothesized to replicate a gaze cueing effect similar to the one Wiese et al. (2018) found in their experiment using a Meka robot. Furthermore, we expected to find a robot presence-specific effect on the strength of this effect. Moreover, we hypothesized to find differential ratings of the robot’s anthropomorphism, animacy and likeability.

4.1 Interpretation of Statistical Analysis

Our results are consistent with previous research on gaze following with copresent robots (Wiese et al., 2018): Participants consistently exhibited gaze-following behavior, as evidenced by slower reaction times on trials in which the robot cued the wrong target compared to trials in which the cued location and target location matched. Hence, our results replicate the well-known finding that participants locate a target that is congruent with the cued direction more quickly than a target that is incongruent with the cued direction (Friesen and Kingstone, 1998; Driver IV et al., 1999). In particular, our results are in line with findings by Wiese et al. (2018) and Kompatsiari et al. (2018), that showed that a gaze cueing effect can be found when the target stimulus is directed by the gaze of an embodied robot. This finding is particularly valuable given the ongoing replication crisis in psychological research, which has highlighted the problem of replicating the results of many scientific studies (Open Science Collaboration, 2015) and allows for the generalization of the gaze cueing effect across different robotic platforms. Importantly, with an effect of 14 ms, the gaze cueing effect in our study is smaller than the 25 ms found by Wiese et al. (2018) but in line with findings of other studies using more controlled settings (Wiese et al., 2012; Wykowska et al., 2014). Differences in the extent of the effect might be explained by the design of the robot - possibly, the Meka robot used by Wiese et al. (2018) offers more social cues or other affordances than the iCub robot used in the present study or other robotic platforms. Further studies will be needed to better understand the relationship of robot design and gaze following behavior. Moreover, our results show that even the screen versions of our robot consistently triggered a gaze cueing effect as indicated by slower reaction times in incongruent compared to congruent trials, showing that gaze following in HRI is not limited to physically present versions of embodied robots.

Importantly, however, the same stimuli did not elicit varying degrees of gaze-cueing when comparing the different ways the robot was presented to the participants, as evidenced by a non-significant interaction of robot presence and cue-target congruency. Across conditions participants consistently followed the gaze, independently of whether they were confronted with a copresent iCub or a simulated iCub. As the novelty of this study lies in comparing the way the different ways of presenting the robot influence simple social attention mechanisms, as evidenced by the gaze cueing effect, there exists hardly any literature indicating similar or contradictory results. Importantly, however, these results add to findings of Mollahosseini et al. (2018) who were comparing the effect of embodiment and presence of four different types of agents on similar outcome variables. Results showed that embodiment but not physical presence was the factor that accounts for the significant difference in the participants' response, as indicated by no significant difference in results when presenting the participants with a copresent robot compared to a telepresent robot. However, the outcomes differed significantly for the comparisons of virtual agents and both forms of embodied robots. In contrast, our study found no difference in gaze following behavior between the presence conditions, as participants' reaction times were not significantly different when interacting with a virtual agent compared to a copresent robot (see section 3.3).

There are multiple possible reasons why physical presence might not additionally influence the gaze cueing effect. One could be that our results are based on the fact that robots in both conditions moved in a similar, human-like manner. Previous research has shown that (natural) movement is linked to mind attributions - famously for example in the Heider and Simmel illusion (Heider and Simmel, 1944), in which participants attribute mental states to three moving geometrical figures (two triangles that seem to "hunt" a circle). Studies examining how mental state attributions alter gaze following behavior in trials with photographs of robots have shown that this manipulation led to the occurrence of a gaze cueing effect that was not otherwise present (Wykowska et al., 2014). Differences between our results and those of studies using only photos of agents/robots (Admoni et al., 2011) or non-moving agents, as in the research of Mollahosseini et al. (2018), might be due to the association of motion and mind attributions leading to typical social interaction phenomena, such as gaze cueing.

Several researchers have found that copresent robots elicit more positive responses than simulated robots - they are generally rated as more likable, pleasant, and trustworthy (Li, 2015). In contrast to these results and our conjectures, the physical presence of the robot used in the present experiment did not seem to affect how participants rated it in terms of anthropomorphism, animacy, and likability. Importantly though, similar results were reported by Kiesler et al. (2008), who compared participants' ratings of four types of agents that differed in terms of their embodiment and physical presence after a brief conversational interaction. The results suggest that while embodiment seemed to play a role in the ratings of the anthropomorphism of the agents, the way they were presented did not make a significant difference.

The reason for these results may lie in the nature of the interactions presented. While studies involving touch and greater focus on spatial awareness might make the affordances of a robot's physical presence more apparent and hence lead to differential rating for copresent robots and simulated or telepresent robots (e.g. Lee et al., 2006), the interaction scenarios used in our experiment and in the experiment presented by Kiesler et al. (2008) do not really depend on the physical presence and might therefore produce different results. Moreover, to our knowledge, no study has yet examined how the physical presence of a robot affects the ratings of the three subscales of the Godspeed questionnaire we used. Therefore, conflicting results could be explained by differences in the scales used in different projects.

4.1.1 Summary: Interpretation of Statistical Analysis

Overall, the main findings of the present study can be summarized in the following points: 1) we replicated a gaze cueing effect as shown in several studies with images of human faces or physically present robots and extended our knowledge about the generalizability of this effect to an additional robotic platform. 2) Our study showed no robot presence-specific effect on gaze following, as well as 3) no presence-specific changes in ratings of anthropomorphism, animacy, and likability of the robot, suggesting that factors other than the way the robot is presented in the HRI may play a role here.

4.2 Limitations of the Study

There are some limitations to the current study that should be addressed in future research on real-time gaze following in human-robot interaction.

Although the sample size of our study was clearly large enough to detect an effect, it is important to note that sample size cannot be used as an indicator of generalizability. The population studied in our experiment was rather narrow, as we recruited mainly in the vicinity of the Czech Technical University. Further experiments should be conducted with other samples to further broaden the evidence.

While our results partially replicate those of researchers using other robotic platforms (Wiese et al., 2018; Kiesler et al., 2008), the interpretations and conclusions from our experiment are tied to the iCub robot. Since several studies have shown that differences in robot appearance strongly influence perceptions of and interactions with robots, it is important to follow up our results with different robots and their designs to understand what factors influence social attention mechanisms in human-robot interactions.

Further, our study combined behavioral measurements with self-assessments. This design helps to better understand processes and relate people's perceptions and actions. However, to better understand the underlying processes, the use of brain imaging techniques could be beneficial. MRI gaze cueing studies with human stimuli have shown that similar areas are activated when engaging in a theory of mind task or observing someone gaze in a certain direction (Calder et al., 2002). In the context of the presented research, such investigation might help better understand underlying processes such as mentalizing and how they are shaped by the way the robot is presented.

In addition, it is important to point out small differences between the two conditions in our experiment that future research could pay attention to. While the screen version of the robot moved silently, the joints of the copresent iCub made loud noises when it changed position. Further, the robots were presented in front of different backgrounds (see Figure 4). To rule out possible confounding effects, future studies should take special care in designing the different conditions so that factors irrelevant to the research question are stable across groups.

4.3 Implementations and Future Research

The study of social cognitive phenomena in the interaction between humans and robots is fundamental given how prevalent they will be in our future. Robots and humans will share not only work, but also living space. In order to interact with robots as social interaction partners, it is necessary to understand how typical features of human interaction can be generalized to robots. Our study adds to this understanding by investigating gaze following behavior in real time HRI with a copresent and a simulated robot.

The results of our study suggest that gaze cueing is consistent across different levels of physical presence of the robot. Consequently, it can be concluded that simple markers of social interactions with humanoid robots are similar to those in human-human interaction. This opens up a broad scheme of possible interaction scenarios in which participants communicate and interact with both on-screen and fully physically present versions of robots. With respect to the design of human-robot interaction scenarios, our observations suggest that it is possible to neglect the physical presence of the robot when the focus of the interaction is on basal mechanisms of social attention. This is particularly useful when considering the accessibility of physically embodied robots. Due to their cost and potential risks associated with using physical robots in HRI, they are still a luxury and can only be used in a limited scheme of tasks. The potential to use simulated versions of the same robot could therefore offer a number of advantages for both the development of and interaction with robots.

While these findings shed light on potential applications of physically present and simulated robots, to better understand the extent of these findings, future studies should consider the limitations of the present study in several ways: One way to further explore gaze tracking in HRI could be to add an additional presence condition by conducting the experiment in VR, as suggested in the introductory chapter. Virtual reality is a relatively new area of research and could be of great benefit to the HRI community as it offers the possibility of designing immersive but safe environments for interaction and development. Seeing the extent to which our results are applicable to virtual reality environments can guide future research in HRI. Further, future studies could aim at directly comparing virtual agents and telepresent robots. Based on the results reported by Li (2015) that interactions with virtual agents and telepresent robots were evaluated similarly and did not differ significantly in terms of behavioral outcomes, we do not expect a difference between these conditions, but a systematic investigation is nevertheless needed.

In addition, future research could aim to investigate the reproducibility of our results with different robotic platforms. This could help to better assess the generalizability of our results and investigate possible other factors that influence the strength of the gaze cueing effect, such as degrees of freedom of movement and anthropomorphic design. Extending the measurements to include imaging techniques and more detailed questionnaires that examine participants' perceptions of the robot could further explore the underlying mechanisms of social attention and add to future knowledge of social cognition in the context of human-robot interaction.

Chapter 5

Conclusion

The importance of understanding how social cognitive phenomena relate to human-robot interaction is becoming increasingly important with advances in robotics and related disciplines leading to robots becoming part of our daily lives and society. In this thesis, a specific phenomenon of social interactions - the social gaze - was studied with respect to human-robot interaction. Gaze is a fundamental aspect of human-human interaction due to its dual function as a signaler and perceiver and is based on specific neural circuits. Recent gaze cueing studies that used human faces to cue the appearance of a target stimulus consistently found a gaze cueing effect. Studies using images of robots or actual embodied robots yielded inconsistent results. In this work, we further investigated social gaze in HRI by using a gaze cueing paradigm with the humanoid iCub robot and a virtual version of that robot. As has been shown over the course of the Covid 19 pandemic, the physical presence of our interaction partners appears to be an important factor in interactions. While recent studies in the field of HRI have explored various positive effects of the physical presence of a robot, there has been no systematic investigation of how physical presence might affect the phenomenon of social gaze, with a specific focus on HRI. To investigate possible effects of a robot's presence on social gaze, participants were exposed to either a copresent or a simulated version of the robot.

Our results show that humanoid robots consistently elicit gaze-following behavior in human interaction partners, regardless of how they are presented during the interaction. Moreover, both copresent and simulated robots were rated similarly in terms of anthropomorphism, animacy, and likability. Although these results are partially inconsistent with our hypotheses, they contribute to further understanding of social HRI and point to important new avenues for future research. By replicating the gaze cueing effect with an iCub robot, we were able to extend our knowledge to another robotic platform to complement existing research. The introduction of physical presence as an additional factor opens the discussion on a possible generalization of results from screen or possibly VR-mediated interactions to real-world scenarios. This would help make HRI research more accessible and easier to replicate. While our results suggest that the way the robot is presented in HRI does not affect social phenomena such as gaze tracking, to fully understand the extent to which our results are generalizable to other environments and robots, it is important to additionally investigate the effects of robot design, motion, and interaction scenarios. While gaze tracking is an important feature of human social interactions that is now being investigated by the HRI community, other phenomena need to be explored to better understand social interactions between humans and robots.

Consequently, human-robot interaction research remains a field with unanswered questions and challenges. However, the growing recognition of the importance of the topic to our future and the interest of cognitive science and related disciplines provide hope that we will continue to move toward a comprehensive understanding of human-robot social interactions and the factors that influence them.

Bibliography

Admoni, H., C. Bank, J. Tan, M. Toneva, and B. Scassellati (2011). Robot gaze does not reflexively cue human attention. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Volume 33.

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

Anstis, S. M., J. W. Mayhew, and T. Morley (1969). The perception of where a face or television'portrait'is looking. *The American journal of psychology 82* (4), pp. 474–489.

Asimov, I. (1950). Runaround. i, robot. *New York: Bantam Dell* .

Bainbridge, W. A., J. Hart, E. S. Kim, and B. Scassellati (2008). The effect of presence on human-robot interaction. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 701–706. IEEE.

Bainbridge, W. A., J. W. Hart, E. S. Kim, and B. Scassellati (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics 3* (1), pp. 41–52.

Baron-Cohen, S., S. Wheelwright, and T. Jolliffe (1997). Is there a" language of the eyes"? evidence from normal adults, and adults with autism or asperger syndrome. *Visual cognition 4* (3), pp. 311–331.

Bartneck, C., T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Sabanovic (2020, 02). *Human-Robot Interaction: An Introduction*. Cambridge University Press.

Bartneck, C., D. Kulić, E. Croft, and S. Zoghbi (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics 1* (1), pp. 71–81.

Breazeal, C. and B. Scassellati (1999). How to build robots that make friends and influence people. In *Proceedings 1999 IEEE/RSJ international conference on intelligent* *robots and systems. Human and environment friendly robots with high intelligence and emotional quotients*, Volume 2, pp. 858–863. IEEE.

Calder, A. J., A. D. Lawrence, J. Keane, S. K. Scott, A. M. Owen, I. Christoffels, and A. W. Young (2002). Reading the mind from eye gaze. *Neuropsychologia 40* (8), pp. 1129–1138.

Čapek, K. (1923). *RUR (Rossum's Universal Robots): A Fantastic Melodrama in Three Acts and an Epilogue*. S. French.

Dalmaso, M., L. Castelli, and G. Galfano (2020). Social modulators of gaze-mediated orienting of attention: A review. *Psychonomic Bulletin & Review 27* (5), pp. 833–855.

Deng, E., B. Mutlu, and M. J. Mataric (2019). Embodiment in socially interactive robots. *Foundations and Trends in Robotics 7* (4), pp. 251–356.

Driver IV, J., G. Davis, P. Ricciardelli, P. Kidd, E. Maxwell, and S. Baron-Cohen (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual cognition 6* (5), pp. 509–540.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews 24* (6), pp. 581–604.

Emery, N. J., E. N. Lorincz, D. I. Perrett, M. W. Oram, and C. I. Baker (1997). Gaze following and joint attention in rhesus monkeys (macaca mulatta). *Journal of comparative psychology 111* (3), pp. 286.

Friesen, C. K. and A. Kingstone (1998). The eyes have it! reflexive orienting is triggered by nonpredictive gaze. *Psychonomic bulletin & review 5* (3), pp. 490–495.

Frischen, A., A. P. Bayliss, and S. P. Tipper (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological bulletin 133* (4), pp. 694.

Gasparetto, A. and L. Scalera (2019). From the unimate to the delta robot: the early decades of industrial robotics. In *Explorations in the history and heritage of machines and mechanisms*, pp. 284–295. Springer.

Gibbs Jr, R. W. (2005). *Embodiment and cognitive science*. Cambridge University Press.

Gobel, M. S. and B. Giesbrecht (2020). Social information rapidly prioritizes overt but not covert attention in a joint spatial cueing task. *Acta Psychologica 211*, pp. 103188.

Goedicke, D., J. Li, V. Evers, and W. Ju (2018). Vr-oom: Virtual reality on-road driving simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–11.

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

Hancock, P. A., D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors 53* (5), pp. 517–527.

Heider, F. and M. Simmel (1944). An experimental study of apparent behavior. *The American journal of psychology 57* (2), pp. 243–259.

iCub Tech IIT (2022, January). icub technical docs. `https://icub-tech-iit.github.io/documentation/`.

Inamura, T. and J. T. C. Tan (2012). Long-term large scale human-robot interaction platform through immersive vr system-development of robocup@ home simulator. In *2012 IEEE/SICE International Symposium on System Integration (SII)*, pp. 242–247. IEEE.

Jung, Y. and K. M. Lee (2004). Effects of physical embodiment on social presence of social robots. *Proceedings of PRESENCE* , pp. 80–87.

Kiesler, S., A. Powers, S. R. Fussell, and C. Torrey (2008). Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition 26* (2), pp. 169–181.

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

Kozima, H. and A. Ito (1998). Towards language acquisition by an attention-sharing robot. In *New methods in language processing and computational natural language learning*.

Krupke, D., S. Starke, L. Einig, J. Zhang, and F. Steinicke (2017, 10). Prototyping of immersive hri scenarios. pp. 537–544.

Kuhn, G., B. W. Tatler, and G. G. Cole (2009). You look where i look! effect of gaze cues on overt and covert attention in misdirection. *Visual Cognition 17* (6-7), pp. 925–944.

Kwak, S. S., Y. Kim, E. Kim, C. Shin, and K. Cho (2013). What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *2013 IEEE RO-MAN*, pp. 180–185. IEEE.

Lachat, F., L. Conty, L. Hugueville, and N. George (2012). Gaze cueing effect in a face-to-face situation. *Journal of Nonverbal Behavior 36* (3), pp. 177–190.

Lee, K. M., Y. Jung, J. Kim, and S. R. Kim (2006). Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies 64* (10), pp. 962–973.

Lehmann, H., A. Rojik, and M. Hoffmann (2020). Should a small robot have a small personal space? investigating personal spatial zones and proxemic behavior in human-robot interaction. *arXiv preprint arXiv:2009.01818* .

Lehmann, H., A. Roncone, U. Pattacini, and G. Metta (2016). Physiologically inspired blinking behavior for a humanoid robot. In *International Conference on Social Robotics*, pp. 83–93. Springer.

Leyzberg, D., S. Spaulding, M. Toneva, and B. Scassellati (2012). The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the annual meeting of the cognitive science society*, Volume 34.

Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies 77*, pp. 23–37.

Li, R., M. van Almkerk, S. van Waveren, E. Carter, and I. Leite (2019). Comparing human-robot proxemics between virtual reality and the real world. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 431–439. IEEE.

Liu, O., D. Rakita, B. Mutlu, and M. Gleicher (2017). Understanding human-robot interaction in virtual reality. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 751–757. IEEE.

Martini, M. C., G. A. Buzzell, and E. Wiese (2015). Agent appearance modulates mind attribution and social attention in human-robot interaction. In *International conference on social robotics*, pp. 431–439. Springer.

Mason, M. F., E. P. Tatkow, and C. N. Macrae (2005). The look of love: Gaze shifts and person perception. *Psychological science 16* (3), pp. 236–239.

Mestre, D. R. (2017). Cave versus head-mounted displays: ongoing thoughts. *Electronic Imaging 2017* (3), pp. 31–35.

Metta, G., L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, et al. (2010). The icub humanoid robot: An open-systems platform for research in cognitive development. *Neural networks 23*(8-9), pp. 1125–1134.

Mollahosseini, A., H. Abdollahi, T. D. Sweeny, R. Cole, and M. H. Mahoor (2018). Role of embodiment and presence in human perception of robots' facial cues. *International Journal of Human-Computer Studies 116*, pp. 25–39.

Morgan, E. J., M. Freeth, and D. T. Smith (2018). Mental state attributions mediate the gaze cueing effect. *Vision 2*(1), pp. 11.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science 349*(6251), pp. aac4716.

Pfeifer, R. and C. Scheier (1999). Understanding intelligence'the mit press. *Cambridge, MA* .

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology 32*(1), pp. 3–25.

Ricciardelli, P., G. Baylis, and J. Driver (2000). The positive and negative of human expertise in gaze perception. *Cognition 77*(1), pp. B1–B14.

Risko, E. F., D. C. Richardson, and A. Kingstone (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current Directions in Psychological Science 25*(1), pp. 70–74.

Ristic, J., C. K. Friesen, and A. Kingstone (2002). Are eyes special? it depends on how you look at it. *Psychonomic bulletin & review 9*(3), pp. 507–513.

Roncone, A., U. Pattacini, G. Metta, and L. Natale (2016). A Cartesian 6-DoF gaze controller for humanoid robots. In *Robotics: science and systems*, Volume 2016.

Samporová, S. (2022). Experimentálne prostredie pre interakciu človeka s humanoidným robotom icub. Master's thesis, Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislave.

Sheridan, T. B. (2016). Human–robot interaction: status and challenges. *Human factors 58*(4), pp. 525–532.

Sherman, W. R. and A. B. Craig (2018). *Understanding virtual reality: Interface, application, and design*. Morgan Kaufmann.

Shi, Z. and C. L. R. McGhan (2020). Affordable virtual reality setup for educational aerospace robotics simulation and testing. *Journal of Aerospace Information Systems 17*(1), pp. 66–69.

Shinozawa, K., F. Naya, J. Yamato, and K. Kogure (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International journal of human-computer studies 62*(2), pp. 267–279.

Shu, B., G. Sziebig, and R. Pieters (2019). Architecture for safe human-robot collaboration: Multi-modal communication in virtual reality for efficient task execution. In *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*, pp. 2297–2302. IEEE.

Spectrum, I. (2020, May). What is a robot?

Stass, J. W. and F. N. Willis (1967). Eye contact, pupil dilation, and personal preference. *Psychonomic science 7*(10), pp. 375–376.

Teufel, C., D. M. Alexis, N. S. Clayton, and G. Davis (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, & Psychophysics 72*(3), pp. 695–705.

Tikhanoff, V., A. Cangelosi, P. Fitzpatrick, G. Metta, L. Natale, and F. Nori (2008). An open-source simulator for cognitive robotics research: the prototype of the icub humanoid robot simulator. In *Proceedings of the 8th workshop on performance metrics for intelligent systems*, pp. 57–61.

Wainer, J., D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric (2006). The role of physical embodiment in human-robot interaction. In *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 117–122. IEEE.

Wainer, J., D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric (2007). Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 872–877. IEEE.

Weiss, A. and C. Bartneck (2015). Meta analysis of the usage of the godspeed questionnaire series. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 381–388. IEEE.

Wiese, E., H. J. Müller, and A. Wykowska (2014). Using a gaze-cueing paradigm to examine social cognitive mechanisms of individuals with autism observing robot and human faces. In *International Conference on Social Robotics*, pp. 370–379. Springer.

Wiese, E., P. P. Weis, and D. M. Lofaro (2018). Embodied social robots trigger gaze following in real-time hri. *2018 15th International Conference on Ubiquitous Robots (UR)* , pp. 477–482.

Wiese, E., A. Wykowska, J. Zwickel, and H. J. Müller (2012). I see what you mean: how attentional selection is shaped by ascribing intentions to others.

Wijnen, L., S. Lemaignan, and P. Bremner (2020). Towards using virtual reality for replicating hri studies. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 514–516.

Willemse, C. and A. Wykowska (2019). In natural interaction with embodied robots, we prefer it when they follow our gaze: A gaze-contingent mobile eyetracking study. *Philosophical Transactions of the Royal Society B 374* (1771), pp. 20180036.

Williams, T., N. Tran, J. Rands, and N. T. Dantam (2018). Augmented, mixed, and virtual reality enabling of robot deixis. In *International Conference on Virtual, Augmented and Mixed Reality*, pp. 257–275. Springer.

Wykowska, A., E. Wiese, A. Prosser, and H. J. Müller (2014). Beliefs about the minds of others influence how we process sensory information. *PloS one 9* (4), pp. e94339.

Yang, G.-Z., J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield, et al. (2018). The grand challenges of science robotics. *Science robotics 3* (14), pp. eaar7650.

Appendix A

Questionnaires

Demographic Survey

1. ID:
   ____
2. Věk/Age:
   ____
3. Pohlaví / Gender:
   O muž/male O ženy/female O ostatní/other
4. Jste pravák, oboustranný nebo levák? / Are you right-handed, ambidextrous or left-handed?
   O pravák/right-handed O oboustranný/ambidextrous O levák/left-handed
5. Znalost robotů / Knowledge of robots:
   Velmi špatná/very poor 1 2 3 4 5 Velmi dobrá/very good

Godspeed

napodobenina, "fake" / fake	1	2	3	4	5	přirozený / natural
strojový / machinelike	1	2	3	4	5	lidský / humanlike
nevědomý / unconscious	1	2	3	4	5	vědomý / conscious
umělý / artificial	1	2	3	4	5	jako živý / lifelike
pohybuje se strnule / Moving ridgidly	1	2	3	4	5	pohybuje se plynule, elegantně / Moving elegantly
mrtvý / dead	1	2	3	4	5	živý / alive
statický / stagnant	1	2	3	4	5	temperamentní / lively
mechanický / mechanic	1	2	3	4	5	organický, přírodní / organic
umělý / artificial	1	2	3	4	5	jako živý / lifelike
nečinný / inert	1	2	3	4	5	interaktivní / interactive
apatický / apathetic	1	2	3	4	5	vnímavý / responsive
Nebyl mi sympatický. / dislike	1	2	3	4	5	Byl mi sympatický. / like
nepřátelský / unfriendly	1	2	3	4	5	přátelský / friendly
nevlídný / unkind	1	2	3	4	5	hodný / kind
nepříjemný / unpleasant	1	2	3	4	5	příjemný / pleasant
strašný / awful	1	2	3	4	5	milý / nice

Appendix B

Additional Tables and Tests

A Shapiro-Wilk test was performed to test the assumption of normality of the t-tests (Table 4). Importantly, the data for the copresence group were not normal ($p < .05$).

Table 3: Shapiro-Wilk Test of Normality: RT Physical Presence (N = 42)

Group	W	p-value
Copresent	0.87	.009**
Telepresent	0.84	0.68

Note. *** $p<.001$, ** $p<.01$, * $p<.05$

To test the normality assumption of the analyses of variance, a Shapiro-Wilk test for normality was performed (Table 5). Importantly, the data in the copresence x incongruence group were not normal ($p < .05$).

Table 4: Shapiro-Wilk Test of Normality: RT ANOVA (N = 42)

Group	W	p-value
Copresent/Congruent	0.95	0.39
Copresent/Incongruent	0.88	.017*
Telepresent/Congruent	0.96	0.5
Telepresent/Incongruent	0.97	0.62

Note. *** $p<.001$, ** $p<.01$, * $p<.05$

A Levene's test was calculated to check for the assumption of homogeneity of variances.

Table 5: Levene's Test: Homogneity of Variance Assumption

Group	F	p-value
Congruent	1.73	0.2
Incongruent	2.38	0.13

Note. *** $p<.001$, ** $p<.01$, * $p<.05$

A Box's M test was calculated to check the assumption of homogeneity of covariances. Box's M is reported to be highly sensitive (Tabachnick & Fidell, 2001). Hence, p is regarded significant at p<.0001.

Table 6: Box's M-Test: Homogeneity of Covariances Assumption

Box's M	6.28
p-value	0.01

Note. * $p<.001$