Environmental Physical-Virtual Interaction to Improve Social Presence with a Virtual Human in Mixed Reality

Kangsoo Kim
University of Central Florida

ENVIRONMENTAL PHYSICAL–VIRTUAL INTERACTION TO IMPROVE SOCIAL PRESENCE WITH A VIRTUAL HUMAN IN MIXED REALITY

by

KANGSOO KIM
B.S. Hanyang University, 2009
M.S. Hanyang University, 2011
M.S. University of Central Florida, 2015

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2018

Major Professor: Gregory F. Welch

ABSTRACT

Interactive Virtual Humans (VHs) are increasingly used to replace or assist real humans in various applications, e.g., military and medical training, education, or entertainment. In most VH research, the perceived social presence with a VH, which denotes the user's sense of being socially connected or co-located with the VH, is the decisive factor in evaluating the social influence of the VH—a phenomenon where human users' emotions, opinions, or behaviors are affected by the VH. The purpose of this dissertation is to develop new knowledge about how characteristics and behaviors of a VH in a Mixed Reality (MR) environment can affect the perception of and resulting behavior with the VH, and to find effective and efficient ways to improve the quality and performance of social interactions with VHs. Important issues and challenges in real–virtual human interactions in MR, e.g., lack of physical–virtual interaction, are identified and discussed through several user studies incorporating interactions with VH systems. In the studies, different features of VHs are prototyped and evaluated, such as a VH's ability to be aware of and influence the surrounding physical environment, while measuring objective behavioral data as well as collecting subjective responses from the participants. The results from the studies support the idea that the VH's awareness and influence of the physical environment can improve not only the perceived social presence with the VH, but also the trustworthiness of the VH within a social context. The findings will contribute towards designing more influential VHs that can benefit a wide range of simulation and training applications for which a high level of social realism is important, and that can be more easily incorporated into our daily lives as social companions, providing reliable relationships and convenience in assisting with daily tasks.

To my parents, family, and friends.

ACKNOWLEDGMENTS

First, I would like to thank my wonderful advisor, Prof. Greg Welch, for guiding and supporting me with full of inspiration, ideas, and patience on the journey of this dissertation research. Greg was not just my academic advisor, but a great mentor who opened my eyes to various aspects of research and cared for me and my family's well-being. Without his thoughtfulness, compassion, and respect, it would have been difficult for me to complete this work. It was a great pleasure being and working with him as one of his PhD students in an important part of my life.

I would also like to acknowledge all other committee members: Prof. Jeremy Bailenson, Prof. Gerd Bruder, Prof. Stephen Fiore, Prof. Avelino Gonzalez, and Prof. Gita Sukthankar. Of course all the inputs from the committees are invaluable, but especially I want to thank Jeremy and Gerd for sharing their insights and thoughts, and actually spending a lot of time together to discuss to make things right for my research.

Special thanks to all the current and former members of the SREAL lab: Prof. Charles Hughes, Dr. Leila Barmaki, Dr. Kate Ingraham, Dr. Sungchul Jung, Dr. Arjun Nagendran, Dr. Andrew Raij, Luke Bölling, Salam Daher, Manuel Gonzalez, Steffen Haesler, Jason Hochreiter, Eric Imperiale, Brian Kelly, Barbara Lee, Myungho Lee, Divine Maloney, Nahal Norouzi, Ryan Patrick, Remo Pillat, Ryan Schubert, and Greg Wiatroski. I acknowledge Prof. Eleazar Vasquez and Prof. Xin Yan for sharing their expertise in study design and statistical analysis with me. I also want to thank Dr. Peter Squire (Office of Naval Research) for the continuous funds for our lab and my work during the entire period of my PhD program.

I would like to thank Prof. Jong-Il Park (my BS and MS advisor at Hanyang University) and Prof. Woontack Woo (my undergraduate intern advisor at GIST) for their introduction and guidance to the research field of AR/MR. In addition, I want to thank all the research advisors, colleagues, and

friends in Korea including Prof. Sang-Hwa Lee, Dr. Byung-Kuk Seo, Dr. Junyeong Choi, Jungsik Park, Jae-Hyek Han, Youngju son, Prof. Ahyoung Choi, Prof. Changgu Kang, Dr. Youngkyoon Jang, Dr. Sejin Oh, Dr. Mingyu Park, and Dr. Choonsung Shin. From my college years to present, interacting with them always encouraged me to move forward. Special thanks to Byung-Kuk, Youngkyoon, and Choonsung for sharing their thoughts and discussing my research together.

I had three wonderful internship opportunities at Universal Creative, Facebook, and US Army Research Lab-West in the middle of my PhD program. Here I would like to thank all the people who helped and worked with me including my intern managers: Justin Schwartz, Dr. Sophie Kim, and Dr. Celso de Melo.

During the journey of my PhD, I met a lot of researchers and collaborators from different fields who have influenced me and my research directly or indirectly. I would thank all those people including Prof. Mark Billinghurst, Prof. Jonathan Gratch, Prof. Dongsik Jo, Prof. Sin-Hwa Kang, Prof. Si-Jung Kim, Prof. Gale Lucas, and Dr. Jongwook Kim. I learned a lot from them while reading about their research, discussing research ideas and working together with them. Particularly I want to thank Sangwoo Cho for sharing all the happiness and difficulty together as PhD students.

Finally, I would like to thank my family, the most important people in my life. My lovely wife and little kids, Jihyun Kim, Juli Kim, and Daniel Kim. Their endless support and cheer for me made this work possible. Thanks to my sister, Seoyoun Kim, who has always walked one step ahead and guided me since our childhood. Also, I thank my parents-in-law, Jinbae Kim and Kwangock Seo, for their considerate and warm heart supporting me and my family. The most special and sincere thanks go to my beloved parents, Jongo Kim and Jongsook Lee, who are the origin and the role model of my life. They have always shown faith in me and given me liberty to choose what I desired. I could have the strength and patience to work through all these years for their selfless dedication. Thank you so much from the bottom of my heart.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Virtual Humans (VHs) have been employed in various domains, such as entertainment, education, virtual tourism, telepresence, and medical or military training that benefit from technological surrogates (or stand-ins) for real humans, while representing or assisting them [1, 2, 3]. The potential of VHs is growing as the VHs can be pervasive and influential in our daily lives in the real world through Mixed Reality (MR) technology, along with the advent of cutting-edge Artificial Intelligence (AI) and Ubiquitous Computing like Internet of Things/Everything (IoT/IoE) [4].

In contrast to Virtual Reality (VR), which strives to dissociate users from the real world and immerse them in a computer-generated virtual environment, MR denotes a research field that aims to blend real and virtual worlds together. In particular, Augmented Reality (AR) technologies focus on superimposing virtual 3D objects on a user's predominant view of the surrounding real environment, thus generating the perceptual illusion that those virtual objects exist as part of the real world [5, 6]. Technological achievements in the field of consumer-level devices and displays such as the Oculus Rift, HTC VIVE, Microsoft HoloLens, and Magic Leap One, have rekindled public interest in MR/AR technologies with growing developer communities all over the world.

In many VR/AR/MR (VAMR) application scenarios, it is further desirable to create a sense of interaction and engagement not only with virtual *objects* but with interactive virtual *humans* or human-like representations. VHs, i.e., virtual *avatars* (user-controlled characters) or *agents* (characters controlled by a computer), have been popularly used as a substitute or surrogate for real humans while striving to maintain a similar sense of social interaction and engagement [7].

Most research related to VHs in VAMR has been focused on trying to increase the fidelity of social interactions with VHs while also trying to improve the *social influence* that VHs can exert over real humans. Social influence denotes the phenomenon that occurs when one's perception, emotions,

thoughts, or behaviors are affected by other real or virtual people [8].

Regarding social influence, different concepts and methods have been proposed to understand types of social influences [9], and different approaches have been introduced and evaluated that can measure the social influence exerted by a VH [10]. The most popular of these concepts are *social presence* or *co-presence*, which denote how much users feel that a VH is socially connected or co-located and present in the same space with them [11, 12]. A strong sense of social and co-presence can make users perceive a VH as more realistic and human-like [13], which usually can be expected to provide benefits related to the quality of social interaction and perceived engagement in the interaction with the VH [14, 15]. In some situations, such as VAMR training environments that use VH systems, such benefits to social and co-presence can translate to increased effectiveness and efficiency for the contextually relevant purpose (e.g., training) due to the improved quality of the interactions [16].

The general purpose of the research described in this dissertation is to improve our understanding of how human perception of a VH, in particular the sense of social and co-presence, varies with respect to the VH's characteristics, such as appearance, interactive capabilities, and behaviors, while users are directly interacting with the VH, e.g., via verbal or non-verbal communication. More specifically, the goal is to find effective and efficient ways to improve the quality of interaction and the performance of VHs in a social context in interactive MR systems.

This dissertation identifies a VH's physically plausible and coherent interactive behavior with the surrounding environment as a crucial factor in improving the VH's realism and presence. Human perception and behavior are analyzed through several human subject experiments examining the effects of different features in VH prototypes in real–virtual human interactions. The detailed study designs and findings will be illustrated and described throughout the dissertation while discussing the research implications for more influential real–virtual human interactions in a broader scope

of human–computer interaction paradigms. This chapter provides the overview of the dissertation by describing research motivations, the scope of the research, and thesis statements outlining the contributions.

1.1 Research Motivation and Objectives

Recent advances related to photo-realistic graphics [17], fast and reliable tracking [18], and intuitive interaction mechanisms [19], have shown much potential to improve the quality of VH systems in terms of the visual fidelity and interaction. However, despite the popularity of VHs and the continuous efforts to improve the social influence of VHs, e.g., social and co-presence, there is still a large difference in how VHs are perceived compared to real humans—the levels of social and co-presence with VHs are still relatively low compared to real humans [20].

VH systems in prior research have primarily focused on the VH's appearance [21] and direct (verbal) interaction with human users [22] to make the VH seem more realistic and human-like. However, VHs are still isolated from the surrounding *physical* environment in terms of the interaction—i.e., the inability to be aware of and influence the environment. In VR, where the purpose is to immerse users in a virtual world and isolate them from the real world, a VH's interaction with the physical environment might not be desirable, but in MR, where virtual sensations are superimposed on or composited with the real world [23, 24], the interaction of VHs with the physical world becomes a decisive factor [25].

For instance, let us consider a basic interaction, e.g., a verbal or non-verbal communication, with a VH while wearing a see-through head-mounted display (HMD) that provides a tracked (position and orientation) stereoscopic view of an augmented real world. The VH is visually presented via computer-generated graphics that are registered with the real world through the HMD as if the VH

is actually present in the physical surroundings. Now, if the VH does not exhibit awareness of or react to physical events, such as a water cup spilling or a fire alarm going off, the sense of social and co-presence with the VH would be immediately broken. Recent trends of MR technology incorporating other advanced technologies, such as AI and Ubiquitous Computing including IoT/IoE, have the potential to bridge this gap in current-state VH systems and enable highly interactive and intelligent virtual content [26], e.g., interactive VHs mimicking physical affordances [3].

During social interaction among humans, the surrounding environment is a critical factor to define and moderate one's perception of and behaviors with other humans or objects in it. Allwood [27] stated that the physical environment is one of the most important parameters that characterize social activities. Blascovich [28] defined social presence as "a psychological state in which the individual perceives himself or herself as existing within an *interpersonal environment*" (emphasis added). Intuitively, the physical environment where real and virtual humans are co-located in MR, and the VH's interaction with its physical surroundings will be important factors to understand and characterize the sense of social and co-presence with VHs in real–virtual human interactions.

Therefore, the objectives of this dissertation are to explore and understand how and in what ways the surrounding physical environment is contributing to human perception of natural interaction with VHs in MR, and to determine whether we can leverage any such knowledge to increase the corresponding sense of social and co-presence. To pursue these objectives, human subject studies are conducted incorporating different VH prototypes while varying the VH's visual embodiment and interactive behaviors with the surrounding physical environment. For example, the VH exhibits awareness of the surrounding physical objects by gaze or avoidance behaviors, and also influences the objects, e.g., turning on a floor lamp. Subjective responses and objective behaviors are collected and analyzed from the experiments. The experiments mainly focus on the sense of social and co-presence, but also cover different behavioral and perceptual aspects, such as the participant's conversational proactivity and perceived trustworthiness of the VH.

1.2 Exploration Areas and Scope of the Dissertation Research

Research on VHs, particularly in this dissertation, is inherently interdisciplinary, mainly positioned within the field of VAMR, but reaching out to the fields of AI, social and cognitive psychology, and ubiquitous computing including IoT/IoE, due to the nature of social interaction involving virtual and real humans and the shared (physical and virtual) environments that make up MR setups (Figure 1.1).

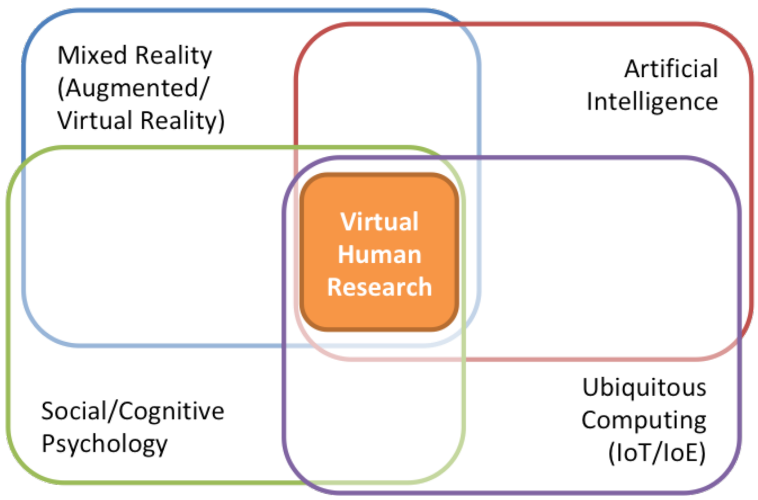


Figure 1.1: Interdisciplinary virtual human research involving mixed reality, artificial intelligence, social/cognitive psychology, and ubiquitous computing.

While interaction among real and virtual humans could also more broadly include interactions between multiple virtual avatars or agents, this dissertation primarily considers the direct interactions between real and virtual humans in MR (Figure 1.2); thus, the perception of the user's self-avatar body or VHs in the field of immersive VR are not in the scope of this dissertation.

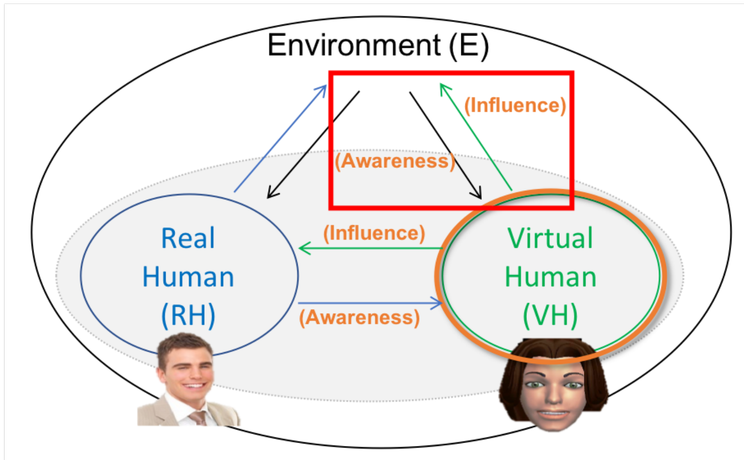


Figure 1.2: An illustration of real–virtual human interaction in a shared co-located physical environment.

There are two types of interaction for VHs in an MR environment: awareness (or sensing) and influence (or affecting/adjusting). There are four possible combinations of the physical space and the virtual space if we consider the "direction" of the interaction and the information being transferred (awareness) or action being performed (influence): physical to physical, physical to virtual, virtual to physical, and virtual to virtual (Figure 1.3).

Interactivity Direction	Within-Environment	Cross-Environment
Influence (Affect)	Virtual Environment (**VE**) influences **VE**	**VE** influences Real Environment (**RE**)
	RE influences **RE**	**RE** influences **VE**
Awareness (Sense)	**VE** is aware of **VE**	**VE** is aware of **RE**
	RE is aware of RE	**RE** is aware of **VE**

Figure 1.3: Physical–virtual interaction: influence and awareness within/across environments.

The research goal is to investigate the effects of a VH's environmental physical–virtual interaction on human perception and behavior; hence, the scope of this dissertation is primarily limited to interaction between physical and virtual environments, rather than interactions within the same environment.

Again, it should be noted that the research goal is to investigate the effects of interactive behaviors of VHs with the surrounding physical environment in real–virtual human interaction scenarios in MR. Hence, the overall approach is to prototype interactive VH systems and then evaluate the perception of the various VHs by human users. For the purpose of human subject studies, most VHs employed in this dissertation are controlled by a human experimenter behind the scenes following a Wizard-of-Oz paradigm [29], which can avoid known drawbacks of current-state AI based VHs. The focus of this dissertation is not on implementing AI based autonomous intelligent VH agents, although those agent systems, such as Amazon's Alexa, would benefit from the findings in this dissertation.

1.3 Thesis Statements and Contributions

Here, Thesis Statements (TS)—which are explored and supported by several experiments presented in this dissertation, regarding a VH's environmental physical–virtual interaction with the physically interactive visual embodiment, and plausible and appropriate behaviors—are set as follows:

- **TS 1**: A VH's increased physicality (or illusion of physicality) in the physical environment can increase social presence and natural behavior in real–virtual human interactions. (Chapter 4 and part of Chapter 5)

- **TS 2**: A VH's interaction with the physical environment, e.g., awareness of and influence over the environment, can increase social presence and co-presence with the VH. (Chapter 5 and Chapter 6)

- **TS 3**: A VH's visual embodiment and appropriate locomotion behavior interacting with the physical environment can increase the perceived trust and confidence in the VH. (Chapter 6)

To evaluate the thesis statements, several human subject studies incorporating different VH prototypes were conducted in MR environments. The specific studies supportive to each thesis statement are described in the chapters referred to in parentheses above.

Based on the results and findings from the human subject studies, this work provides several contributions:

- It provides a better understanding of the perception of virtual content, including VHs, and cognition in combined virtual and physical spaces.

- It presents insights for designing more realistic VHs that can interact with the physical environment, and which can be applied to VH systems for simulation and training applications.

- It introduces novel techniques and modalities for interaction between the virtual and physical worlds, which can be applied to a broad class of MR systems.

- Contemporary virtual assistant systems, such as Apple Siri, Google Assistant, Amazon Alexa, and Microsoft Cortana can benefit from the findings—e.g., by adopting a visual embodiment and appropriate behaviors related to the physical environment to improve users' perceptions of the virtual assistant.

Although most VHs used for the experiments in the dissertation are controlled by a human experimenter behind the scenes via the Wizard-of-Oz paradigm [29], the findings are informative for

a wide range of current-state and future VH implementations such as computer-controlled virtual agents and human-controlled virtual avatars. Also, while the scope of this dissertation was largely limited to designing and evaluating VHs with human-like appearances, some of the results and findings have more broad implications for non-human virtual content in MR as well.

1.4 Outline of the Dissertation

This dissertation consists of seven chapters including this introductory chapter. The remaining dissertation is structured as follows:

Chapter 2: BACKGROUND helps readers better understand VH research within the scope of this dissertation by addressing overall background knowledge such as the definition of MR, descriptions of various VHs that have been used previously in MR, and work related to user studies with VHs. The chapter introduces various prior research investigating human perception of and behavior with VHs in real–virtual human interactions, generally covering (but not limited to) the concept of social influence and social presence.

Chapter 3: PROPOSED ENVIRONMENTAL PHYSICAL–VIRTUAL INTERACTION IN MR introduces the environmental physical–virtual interaction techniques that are proposed to improve the sense of social presence with a VH in real–virtual human interactions. The chapter describes three important aspects to consider for physical–virtual interactions: the VH's visual embodiment; the VH's environmental awareness and influence; and the situational plausibility and coherence of the VH.

Chapter 4 through Chapter 6 present five human subject experiments that have been conducted for this dissertation—the details of those experiments and the findings were already published in various international conferences or journals, e.g., [30, 31, 32, 33, 34]. The experiments vary a VH's

visual embodiment and awareness/influence behaviors with the surrounding physical environment, and examine the effects on the perceived social presence with the VH. Each of the chapters describes one or more experiments in the topic that the chapter title represents.

Chapter 4: INFLUENCE OF PHYSICAL BODY IN HUMAN BEHAVIOR presents a large-scale exploratory study employing a humanoid robot to identify the effects of a VH's physicality on human behavior when interacting with the VH. The results from the study support the first thesis statement about the effects of a VH's physicality (TS 1 in Section 1.3) by observing stronger conversational proactivity and reactivity of human interlocutors with the physical robot compared to a live video stream of the robot displayed across multiple video screens. In other words, more people tended to initiate a conversation with a physically embodied robot than with a life-size video stream of the exact same robot, and people were also more likely to respond to the actual robot rather than the video representation when it attempted to start a conversation. Other behavioral changes with respect to the VH's physicality and the human user's personal profile (e.g., age and gender) are also discussed.

Chapter 5: SPATIAL AND BEHAVIORAL COHERENCE IN MR describes three human subject experiments related to the proposed environmental physical–virtual interaction in MR environments. These experiments varied a VH's interactive behavior with the surrounding physical environment to investigate its effects on participants' perception and behaviors (e.g., social presence and avoidance behavior). Overall, the results support the second thesis statement about social presence (TS 2 in Section 1.3) as well as the first statement about the illusion of physicality (TS 1 in Section 1.3).

Chapter 6: PHYSICALLY INTERACTIVE VISUAL EMBODIMENT AND IMPROVED TRUST presents a final experiment showing that the VH's visual embodiment and environmental physical–virtual interaction with locomotion behavior not only improves the sense of social pres-

ence with the VH, but also the perceived trustworthiness of the VH. The findings of the experiment support the second thesis statement about social presence (TS 2 in Section 1.3) and the last thesis statement regarding trustworthiness (TS 3 in Section 1.3), and provide insights for more practical implications about the social influence of VH interactions.

Chapter 7: CONCLUSIONS AND FUTURE WORK summarizes the dissertation by revisiting the thesis statements and the findings from the conducted experiments. An analysis of the contributions and implications of this research is finally followed by a discussion of promising future work in this area.

CHAPTER 2: BACKGROUND

In this chapter, background information will be provided in the scope of the goals and directions of this dissertation for the readers, and the current state of related work in VH research will be summarized. First, a definition of MR will be discussed with related concepts, such as AR and VR. Also, a concept of pervasive or ubiquitous MR that illustrates virtual entities are absorbed in our daily lives is described via a convergence with other cutting-edge technologies, such as Ubiquitous Computing (IoT/IoE) and AI. Second, the concept of *Social Influence* and the *Presence* terminology in VAMR or technology-mediated communication are described, while addressing important measures to evaluate the systems in real–virtual human interactions. Third, the concepts of avatars and agents are explained within a broader concept of technological human surrogates, and examples of VH systems in VAMR are discussed along with related forms such as robotic humans. Finally, related work on VH studies will be reviewed summarizing the main findings and results from the studies.

2.1 Pervasive Mixed Reality

This section will describe the general definitions of VR, AR, and MR along with other related concepts, and the efforts to combine the virtual world with the real or physical world perceptually. Ubiquitous Computing (IoT/IoE) and advanced AI that are merged with VAMR technologies are discussed as effective strategies to build highly immersive and interactive physical–virtual environments.

2.1.1 Mixed Reality in Reality–Virtuality Continuum

In the field of computer science, VR refers to a technology-mediated simulation that can create real or artificial experiences, where people can interact with a simulated computer-generated virtual environment. The term "Virtual Reality" was coined and popularized by Jaron Lanier in the 1980s at a time when the technology was becoming available to researchers and end users [35]. However, even before the 1950s, VR-related concepts and technologies have been anticipated and discussed in Science Fiction books. Moreover, research prototypes and applications with more than visual and aural sensory modalities have existed since the 1950s. For example, the first immersive multi-sensory simulator called "Sensorama" was introduced in 1962 [36], which provided sights, sounds, smell, wind, and vibration feedback. Anticipating future advances in this field, up to the point where human users may not be able to distinguish the virtual from the real, various terms such as simulated reality, hyperreality, and synthetic reality have been introduced and defined [37]. In such forms of reality, users may not be aware of whether the outside world is simulated or not through high-fidelity simulated multi-sensory feedback that is indistinguishable from natural sensations and perception [38].

Unlike VR, which strives toward complete immersion of users in a virtual environment, in AR/MR, users can experience virtual content blended with real objects and environments as if the content exists as part of the real world. MR has existed in many forms since Sutherland presented the first prototype AR head-mounted display and discussed the "Ultimate Display" in the 1960s [39, 40]. Although the concept of MR is evolving and there are different perspectives to define it, the typical definition of MR is traced back to a paper written by Milgram and Kishino in 1994 [6], which is the first academic paper to use the term "Mixed Reality" in the context of computer interfaces. They defined MR or the display as "a particular subset of VR related technologies that involve the merging of real and virtual worlds somewhere along the 'virtuality continuum' which con-

nects completely real environments to completely virtual ones." As shown in Figure 2.1, the "Reality–Virtuality Continuum" spans from completely real to completely virtual environments, and includes AR and Augmented Virtuality (AV). AR denotes experiences that superimpose virtual objects on the real environment, while AV can be thought of as the other way around when superimposing real objects on the virtual environment.

Figure 2.1: Reality–virtuality continuum by Milgram and Kishino [6].

Later, Mann [41] placed MR in a more generalized concept of Mediated Reality including VR and other forms of "modulated" reality like *diminished* reality, which tries to remove objects visually from the perceptible real world (Figure 2.2).

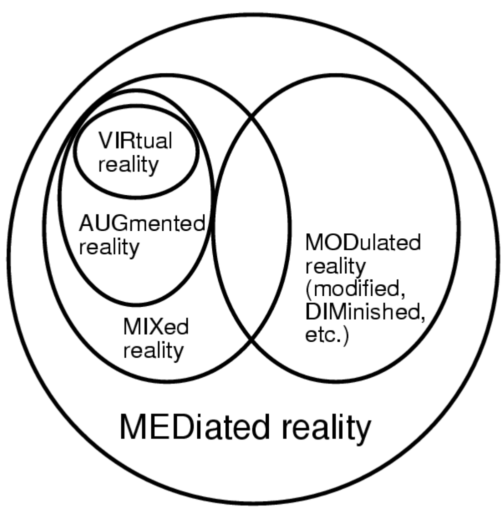


Figure 2.2: The concept of Mediated Reality by Mann [41].

Mann et al. [42] recently illustrated a variety of concepts related to "reality" including "Extended Reality (XR)" that tries to include the extreme ends of the virtuality continuum, i.e., reality and virtuality, which Milgram and Koshino did not include under the concept of MR. They presented a broader term, "Multimediated Reality (All R)", which is multidimensional, multisensory, multimodal, and multidisciplinary, by building a sophisticated taxonomy that covers different continuums related to virtuality and reality in a perspective of interactive multimedia (Figure 2.3).

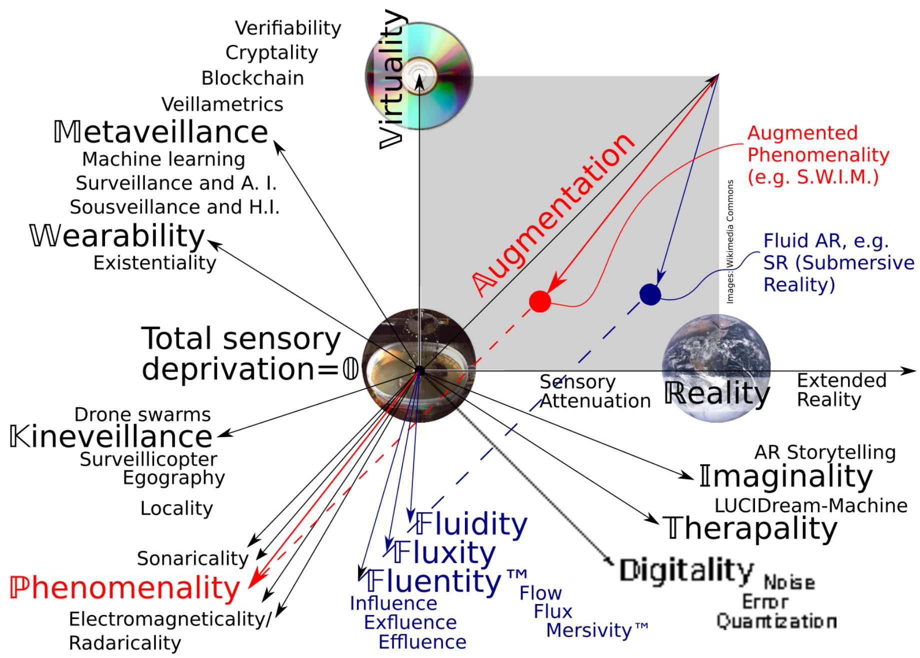


Figure 2.3: The Multimediated Reality Continuum by Mann et al. [42].

In this dissertation, the focus is on AR/MR while VR is considered separately from AR/MR due to the nature of VR trying to diminish the perception of the physical environment in VR experiences. The most popular definition of AR is from Azuma in 1997 [24], where he defined AR as three characteristics: (i) it combines real and virtual objects in a real environment; (ii) it runs interactively in real time; and (iii) it registers (aligns) real and virtual object with each other (in 3D). According to his recent paper in 2017, Azuma predicted a brighter future for AR compared to VR in commercial

markets because of the potential to improve the user's understanding of and interaction with the real world, while it could replace all other display form factors, such as smartphones and tablets, via wearable devices like head-worn AR displays [43]. Recently the term, MR, is often used to describe a highly interactive version or expansion of AR [44]. In that sense, MR involves not only the interaction between human users and computers—e.g., virtual entities—but also the surrounding environment in the context of interaction, which aligns with the present dissertation research (Figure 2.4).

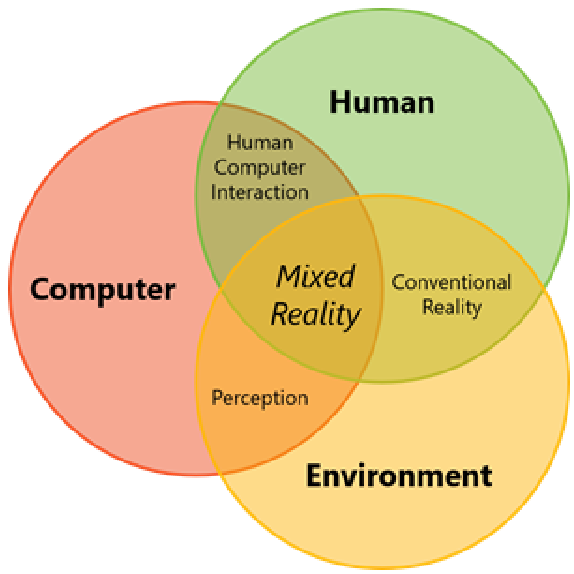


Figure 2.4: A diagram of environmental input and perception for Mixed Reality by Microsoft. [44].

AR/MR technology has recently developed significantly with powerful but affordable computing devices for compelling AR/MR experiences, and now encounters a golden age with unprecedented public interest over the history of MR research [45]. For example, a multi-user mobile AR game, Pokémon Go, was widely adopted around the world—25 million active users in the United States and 40 million worldwide; 500 million downloads worldwide in 2016 [46], and AR/VR startups raised over $3 billion investment in 2017 across 28 AR/VR categories [47], while major IT com-

panies, such as Apple, Facebook, Google, and Microsoft, have been investing and developing their own MR platforms and technologies.

2.1.2 Ubiquitous Computing and Artificial Intelligence

In previous section, we reviewed different "reality" concepts related to MR, and addressed current trends of MR in the industry and public domains. All those reality concepts addressed above try to describe real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables [42]. In the combined (/mixed) environments, we want to extend human abilities and secure the benefits for our society's convenience and comfort.

To achieve this, the convergence of MR technology with other related technologies, such as Ubiquitous Computing and AI, will be a strong driving factor. *Ubiquitous Computing*, which is also called *Pervasive Computing* or *Ambient Intelligence*, is a technological concept that describes the notion that all computing occurs anytime and everywhere so that users do not distinctly realize that it is happening. A pioneer of Ubiquitous Computing, Mark Weiser, stated that "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it." For example, a door intelligently recognizes the personnel who have authority to enter the room and opens automatically for them. Or, a refrigerator identifies missing food items in it and makes a delivery order to fill it up. Advanced AI technology in ubiquitous environments can enable not only more sophisticated interpretations of and semantic understanding in the given context, but also appropriate and useful actions performed by pervasive computing modules, such as IoT/IoE devices—Things are connected to the Internet and communicate with each other remotely.

Combined environments between the real and virtual worlds in the domain of Ubiquitous Computing and VR have been proposed and researched for more than a decade [48, 49]. Paradiso and

Landay presented a concept of *Cross Reality*— ubiquitous mixed reality environment that comes from the fusion of these two technologies (ubiquitous sensor/actuator networks and shared online virtual worlds)" (parentheses added) [48]. Kim et al. defined *Ubiquitous Virtual Reality (U-VR)* as "A concept of creating ubiquitous VR environments which make VR pervasive into our daily lives and ubiquitous by allowing VR to meet a new infrastructure, i.e. ubiquitous computing" [49].

Nevertheless, there are still many obstacles for the widespread adoption of AR/MR technologies in our everyday life in pervasive ways. Azuma addressed the important challenges to overcome for ubiquitous AR/MR, e.g., precise tracking anywhere and anytime, wide field of view (FoV) of optical see-through near-eye displays, innovative interfaces, and semantic understanding of real-world objects [50]. While AR/MR technology is advancing together with strong public interest and industrial investments, recent trends toward a merger of AR/MR with Ubiquitous Computing and advanced AI algorithms has the potential to make virtual content even more intelligent and interactive up to the point where they may be perceived as true social entities—VH avatars or agents socially influential to the users are examples of such virtual content [51, 52]. In such advanced AR/MR environments, we will expect natural and more seamless interaction between the virtual content and the physical environment. To meet these expectation, advanced AR/MR technology that enables more dynamic physical–virtual interaction has to be devised, and rigorous user studies are needed to understand how and in what ways the surrounding physical environment is contributing to human perception of natural interaction. This dissertation describes research towards the future of AR/MR content, particularly about the perception of intelligent virtual content in ubiquitous AR/MR environments.

2.2 Presence and Social Influence

In human society, how to be perceived by the people around oneself and how to influence them are critical to determine one's successful social relationships. An American writer, Dale Carnegie, published a book about how to influence people by changing one's behavior treating them in 1936, and the book has been sharing insightful lessons about social relationships and influence until today [53]. These days, Social Network Services (SNS), such as Facebook or Instagram, and other social media connecting people over the world via computing devices have increased the user's ability to influence other users and to be influenced by them. Some celebrities have more than 100 million followers on their SNS profile, and influencer marketing uses their strong social influence to create trends in commercials and changes in the followers' behavior and thoughts [54].

In social psychology, *Social Influence* refers to "the general category of influences between and among individual persons" [55] as one's attitudes, emotions, thoughts, or behaviors can be changed and affected by others. For example, the *social facilitation* theory states that one performs simple tasks better but complex tasks worse when one is in the presence of others [56, 57]. Likewise, within the more specific context of VHs, social influence indicates how interacting with a VH can affect or change a real human's attitudes, emotions or behaviors. Previous studies have indicated that VHs can exert social influence over real humans under different circumstances. For example, researchers found similar social facilitation effects with VHs as with real humans. Park and Catrambone [56] demonstrated the similar effects for both simple and complex tasks with VHs, while Zanbaka et al. [57] observed a decrease in performance in complex tasks when the VH was present although they did not see an improvement in simple tasks. The phenomenon could be explained by Reeves and Nass' "Media Equation" [58] which states that people have an inherent tendency to treat computers or new media as if they are real people. They noticed that many people interact with computers in a "fundamentally social and natural" way. Thus, a VH mimicking a

real human's appearance and behaviors should be able to promote a relatively high level of social influence with human users.

The phenomenon is intuitively correlated with one's sense of being together with or being connected with others, i.e., *co-presence* or *social presence* [59]. While there is no universal agreement on the precise definitions of social presence and co-presence, some researchers regard them as different but correlated concepts [16, 12], whereas in the VR community, these terms are sometimes used interchangeably [60]. Harms and Biocca consider co-presence as one of several subdimensions that embody social presence [61], and Blascovich et al. [28, 8] define social presence both as a "psychological state in which the individual perceives himself or herself as existing within an interpersonal environment" and "the degree to which one believes that he or she is in the presence of, and dynamically interacting with, other veritable human beings." Riva [62] defined social presence as "the non-mediated perception of an enacting other (I can recognize his/ her intentions) within an external world," and stated that "the highest level of Social Presence is the identification of intentional congruence and attunement in other selves (the self and the other share the same intention)." Considering the definitions addressed above, we can expect that the plausibility of the context and the surrounding environment where the social interaction takes place are important factors in the sense of social presence or co-presence, for example, due to enhanced mutual awareness [63] or a shared inter-personal environment [28, 8].

In a broader sense of *Presence*, i.e., the sense of "being there," Slater [15] introduced the refined concepts of *Place Illusion (PI)* and *Plausibility Illusion (Psi)*. PI is largely related to factors such as visual appearance and sensorimotor contingency and intuitively related to the sense of "being there." Psi refers to the sense that "the scenario being depicted is actually occurring" and that it requires a "credible scenario and plausible interactions between the participant and objects and virtual characters in the environment." When the plausibility fails to be consistent, that could cause a "break in presence" [62, 64]. Slater also emphasized that "when Psi breaks, it is unlikely

to recover" [15]. Due to the nature of Psi as it relates to interactions between real and virtual objects and humans, it could be highly related to the concepts of social presence and co-presence as well, and maintaining plausibility in the environmental context should be emphasized as important for the sense of social and co-presence.

Lombard and Ditton defined presence as the sense of *non-mediation*, which means that we can perceive presence via a technological medium if we can be totally oblivious to the existence of the medium [65]. MacIntyre et al. addressed technical problems and contextual inconsistencies as discouragements of achieving the sense of non-mediation [66]. Technical problems normally include lack of computational power or accuracies, such as low-performance computing devices or inaccurate tracking in AR. Contextual inconsistencies are about a break of the realism of a scenario, and sensory and behavioral aspects.

Contextual inconsistency in social context could be related to one's expectation based on their experience from life-long interactions with real people. A *negative expectancy violation* is an unfavorable response due to the discrepancy between expected behaviors and enacted behaviors [67]. One could say that such a negative expectancy violation in social interactions and the notion of an *"Uncanny Valley"* [68] could be related with the failure of the sense of non-mediation or contextual inconsistencies. Thus, for the higher sense of social or co-presence, this obliviousness to mediation should be emphasized as an important factor as well.

Measures of social and co-presence are an important part of understanding social influence of VHs. Thus, various measures of social influence including social or co-presence for VHs or any other forms of mediated interaction have been proposed over the years. Not only subjective ratings [69, 12, 61, 70, 71], but also objective behavioral and cognitive measures [72, 73, 74] have been employed for stable and accurate evaluation of human perception and social influence of VHs.

2.3 Virtual Humans in Mixed Reality

Humans have been representing other humans in the context of acting for over 2,500 years [75]. For at least 500 years, we have been using technology to create *human surrogates*. Traditionally, two terms have been used to denote manifestations of human surrogates: avatars and agents. The distinction is based on the controlling entity, which could be either a human (avatar) or a computer algorithm (agent). The word avatar first appeared in the science fiction novel Snow Crash [76], in which avatars were introduced as virtual entities controlled by human users. More rigorously, Bailenson and Blascovich [77] define an avatar as "a perceptible digital representation whose behaviors reflect those executed, typically in real time, by a specific human being". If a human surrogate is labeled as an agent, the common assumption is that its behavior is controlled by a computer program rather than a real human being. Analogous to the avatar definition, an agent is "a perceptible digital representation whose behaviors reflect a computational algorithm designed to accomplish a specific goal or set of goals" [77]. In the broader sense, "surrogate" captures various types of human representations, while not being encumbered by traditional distinctions between digital and physical form as well as the nature of the agency. For example, in the mid 1500s, Leonardo da Vinci was designing and building mechanical animatronic humans [78], and in the mid 1900s we began using electromechanical technology to represent other humans, including both *Robotic Humans* (RH) such as audio-animatronic humans pioneered by Walt Disney, and computer graphics *Virtual Humans* [1]. Nagendran et al. [79] explored human surrogate spaces and their different characteristics in terms of three different axes: appearance, shape, and intelligence, while including both RHs and VHs and even further reaching out to real humans and the mixed form of RHs and VHs.

The focus of the dissertation is on the VHs representing or replacing real humans in AR/MR, where real and virtual worlds are merged; however, RHs are often compared with VHs in VH research

due to the shared interests, e.g., studying human perception of and behaviors with them, yet there are clear characteristic differences from VHs, e.g., physicality and virtuality. Thus, here several types of VH systems in different physical settings are illustrated together with RH systems for clarifying the scope of VHs in the dissertation while providing a broader background from related literature.

2.3.1 Robotic Humans

RHs with human-like appearance have appeared in films such as “Sonny” in *I, Robot* (2004) or “Andrew” in *Bicentennial Man* (1999). In robotics, (humanoid) robots are replacing real humans in manufacturing environments, and exploring/navigating robots play important roles in some scenarios such as tasks related to space research [80]. Due to the use of the RHs for the interaction with real humans, not only the technical control mechanism [81] but also its social and emotional effects [82] have been actively researched.

These RHs have a physical body and can physically interact with real humans and their environments, and it has been shown that they could have social influence on human users. For example, Kiesler et al. [83] presented preliminary results indicating that eating habits could be influenced by the presence of robotic or virtual agents. Siegel et al. [84] studied the effect of a robotic human’s gender in persuading individuals to make a monetary donation. They found that men were more likely to donate money to the female robot. Kanda et al. [85] studied the behaviors of elementary students with an interactive humanoid, and found that the younger the participants were, the more time they spent with the robot. Ogawa et al. [86] developed a humanoid robot and found that people changed their negative feelings toward the robot to positive feelings once they hugged it. In the study, they revealed that the older group tended to have a good impression of the robot from the beginning of the interaction and talked rather than listened to the robot. Wiltshire et al. [87]

discussed how to assess social cues and interpret social signals in social robot interactions, and describe a taxonomy outlining the current state of the art in sensor systems and computational techniques for detecting social cues and extracting social signals. Fiore et al. [88] studied how the gaze and proxemic behavior of a mobile robot could be perceived as social signals in human–robot interactions and affect one's sense of social presence with the robot. Walters et al. [89] researched proxemics with a robot with respect to a participant's age, and found that children tended to stand further away from the robot, compared to adults. Warta et al. [90] identified and developed a set of research questions about human perceptions of robots, such as similarity, complementarity, and agency, to guide research directions for a better understanding of social cognitive constructs to enhance the collaboration performance between real and robotic humans.

2.3.2 Virtual Humans

VHs are characterized by computer-generated animated graphics that observers can visually perceive via display devices, such as monitors like TV screens, HMDs, or projectors. One critical advantage of VHs compared to other forms of real or animatronic humans is that they can easily change their appearance and behaviors. They are also easier to control and less likely to require physical maintenance and repair because they do not have physical components such as actuators. However, this virtuality without physicality can cause that human users do not feel that the VH is actually physically present in the space where the interaction is happening. If presented through a stereoscopic head-mounted display, VHs could elicit an illusion of physicality, in that they can be perceived to occupy space in the world in which the user is present. Nevertheless, for instance, if the VH is to shake hands with a user, the user may see the VH's hand outstretched, but still cannot feel it since it does not have any physical manifestation in the real world.

2.3.2.1 Virtual Humans in Monitors and Immersive Projection

The most common VH setting is a monitor-type display due to the wide accessibility and the easy of use. Various applications with VHs in this setting have been used for different social context scenarios ranging from military or medical simulation and training to education and entertainment applications.

Gratch and his colleagues have developed several VHs for military and medical applications, and explored many different aspects of the VHs including both technical improvements and human perceptions. Their "Simsensei Kiosk"—a VH interviewer for aiding in healthcare decisions—was a fully autonomous VH system that could recognize a human user's verbal and nonverbal behaviors, e.g., natural language understanding and face detection (including gaze and expression recognition), and showed potential in face-to-face interactions [91]. In educational training, Dieker et al. [92] employed a virtual environment system with VHs, called "TeachLivE" to train education students who planned to be teachers, and showed its usefulness in teacher training. Sagar et al. [93] introduced an autonomous animated VH face with high-fidelity graphics and a neurobehavioral model for its realistic behaviors. Hoque et al. [94] developed and used an interactive and expressive VH for interview training. Their VH used multimodal information from the real human partner, such as verbal and nonverbal behaviors, and could generate appropriate responses in context. Some are interested in making avatar models of real humans. For instance, DeMara et al. [95] and Lee et al. [96] presented the design of a Lifelike Responsive Avatar Framework (LRAF), which used a virtual representation of a real human.

In spite of the popularity of settings with monitors, there is an effective barrier between real humans and VHs, which could aggravate the perception of VHs, e.g., co-presence, because of the obvious boundary between the real space and the virtual space by the display form factor. To reduce this noticeable boundary between the real and virtual spaces, VHs are often displayed on wide

projection screens so that the users can be immersed by the visual stimuli. Same as for VHs displayed on monitors, VHs displayed on immersive projection screens are useful and popular in military simulations and training scenarios, where involving real humans and environments is challenging or impossible due to potential physical harm or damages [97], and the validity and effectiveness are shown by many human subject studies [98, 99, 13, 100]. Projections with wide screens and over entire walls have been popularly used with realistic rendering and lighting techniques for VHs and other virtual content in entertainment like the theme park industry to make the audience feel a high social presence or co-presence with the content [101]. Besides, such immersive projection technologies are used in mediated interaction, e.g., as a facilitator for tele-presence. Pejsa et al. [102] at Microsoft Research developed a life-sized tele-presence system called "Room2Room" (R2R). The R2R approach employs digital projectors to display a remote participant in a plausible physical location in the local room, for example in a chair. This approach leverages the available (appropriate) physical affordances of the rooms, and maps local and remote participants to physical locations using either a *predefined* approach where the mapping is specified a priori, or an *optimal facing* approach where the mapping is determined on an ad hoc basis, depending on the participant's locations, movements, etc.

Although the wide projection approach can improve the level of immersion in VH simulations, the VH's interaction with the surrounding environment is still limited and the physical settings can be burdensome to install and calibrate.

2.3.2.2 Virtual Humans in Immersive VR HMD

Apart from viewing the real environment, VHs can be rendered in immersive VR, where users wear a VR HMD blocking the real environment and being solely immersed within the VR environment. The VHs and the users are co-located within the shared VR environment. The users might have

their own virtual avatars in the space and interact with the VHs through the self-avatars. The users will experience the illusion of physicality of the VHs by the spatial occupancy and gestures; however, real humans are aware of the spatial disconnect between the VH (in a virtual space) and themselves (in a real space).

Many social science studies that require a highly immersive social context employed VHs using VR HMDs to make the simulation as realistic as possible. For example, Guadagno et al. [103] used an immsersive VR environment with an HMD to investigate the role of the VH's gender and behavioral realism in persuasion. Zibrek et al. [104] studied the user's proxemic behavior with VHs in immersive VR while varying the VH's appearance and rendering style. Smith and Neff [105] displayed an embodied VH via a VR HMD to present the user's non-verbal behavior, e.g. deitic gestures, in collaborative discussion scenarios. In the direction of social studies, Bailenson and colleagues have actively used immersive VR HMDs to place human participants in realistic social situations with VHs [106, 10, 107]. Also, in popular applications related to journalism films, VHs and VR HMDs are introduced to broadcast real-world events such as the tragic states in war fields shown in Project Syria[1].

2.3.2.3 Virtual Humans in AR HMD

While the VHs addressed above have been popularly used, and the validness and effectiveness have been proved for the purpose of certain scenarios, the interaction of the VHs with the surrounding physical environment is not available or very limited, which could aggravate the perception of the VHs in co-located real–virtual human interactions. For example, the VHs in monitors, projections, and VR HMDs are not spatially perceived to be freely walking around the real world where the users are living.

[1] `https://docubase.mit.edu/project/project-syria/`

Recently, AR/MR HMDs have shown their potential to elicit an illusion of a VH's presence and physicality in the real world. Unlike VHs in VR HMDs, because VHs in AR/MR HMDs are rendered in a real space, human users might perceive the VHs as sharing the real space. For instance, holograms shown in the film "Star Wars" offer examples of such cases. Torre et al. [108] superimposed an AR human that could play a checker game with real human users in a shared MR environment. Obaid et al. [109] conducted a study comparing the user's voice level while interacting with a VH in AR and VR. Jo et al. [110] developed an AR tele-presence framework using a VH in AR controlled by a remote user, and discussed how to maintain the VH's realism in the physical local place by adapting its motion to the surrounding physical objects. They discussed a problem of physical discrepancy between two locations in VH-based tele-presence systems, e.g., the chair a user is sitting on can be different in shape and size from another chair his/her avatar should be sitting on in a remote location, which could reduce the VH's naturalism and realism. They tried to resolve the problem by matching virtual objects with remote real objects and using VH's motion adaption techniques so that they can maintain the VH's environmental plausibility. Piumsomboon et al. [111] presented a remote collaborative system with a VH avatar. In their system, a miniature virtual avatar in AR represented the remote user who was wearing an immersive VR HMD for better quality of communication in MR collaboration scenarios. Holz et al. [3, 112] surveyed various forms of agents in a fully physical, a fully virtual, or an AR/MR environment in the context of social interaction, and detailed the advantages and issues with social interaction with AR/MR agents. With respect to applications of VHs in AR/MR, Magnenat-Thalmann et al. [2] summarized various fields that would benefit from employing interactive VHs, such as industrial training and cultural heritage guidance.

In AR environments, maintaining plausibility in the shared physical–virtual space via spatial and behavioral coherence could be intuitively important in human perception of the VHs. Microsoft introduced a game called "Fragments" [113] where people can see and interact with VHs in AR

through an AR HMD, called HoloLens, and emphasized the need for visual conflict-free real–virtual relationships and interactions, for objects and humans while presenting the virtual characters behaving coherently with the physical environment in the game. Previous work also illustrated the importance of physical–virtual interaction in AR with VHs, particularly emphasizing the environmental awareness (sensing) and influence (affecting) with the surroundings [25].

2.3.2.4 Virtual Humans without Visual Embodiment

Related, but different from VHs with a visual appearance, voice-controlled Intelligent Virtual Assistant (IVA) systems, such as Amazon Alexa, have been widely adopted in practical use cases. Thanks to the advance of AI with voice recognition and speech synthesis technology, such IVA platforms are becoming more and more popular in our society. The future of highly intelligent and interactive IVAs has been illustrated in science fiction media, such as the movie *Her* (2013)—a story about a man who falls in love with the disembodied voice of an IVA. As depicted in the film, IVAs will become a social entity mimicking human intelligence.

However, most current-state commercial IVAs mainly focus on voice commands and voice feedback, and lack the ability to provide non-verbal cues, which are an important part of human social interaction. Again, AR/MR has the potential to overcome this challenge by providing a visual embodiment for the IVA. A human-like visual representation could enrich the communicative channels that convey the agent's status and intentions by presenting gestures and motions as social behaviors. The visual representation and social behaviors of an IVA can help users understand what the IVA is currently doing or is going to do and what its intentions are in the given context. In this manner, the visual embodiment and social behaviors of an IVA accomplished by AR/MR have much potential to increase the sense of social presence.

To the best of my knowledge, there are few or no studies about the effects of VH's spatial and

behavioral coherence (i.e., natural occlusions and behaviors avoiding implausible physical–virtual conflict) on social influence of VHs, particularly in AR. Therefore this dissertation is trying to fill the gap in the research field.

2.4 User Studies with Virtual Humans

The potential of VHs to influence real humans has been studied in a wide range of contexts. Madary and Metzinger [114] described previous research supporting the claim that (social) experiences with VHs could affect one's behaviors during the interaction with a VH, and even after the interaction, i.e., changing one's behavior in the real world after the experience. The social influence of VHs on real humans has been studied with various measures—both subjective ratings and objective behavioral and cognitive measures [73]. Here, previous perception studies with VHs, particularly related to social presence with, behavioral influence of, and trustworthiness of VHs, are reviewed.

2.4.1 Social Presence and Co-Presence with Virtual Humans

The most common measures related to the perception of VHs are social presence and co-presence. Different characteristics of the real and virtual humans have been observed to influence a real human's sense of social and co-presence during the interaction with VHs, such as the perceived agency of VHs [7], the level of anthropomorphism [12], and the fidelity or realism of the VH's appearance and behaviors [115, 116, 117]. Fox et al. [7] investigated the relationship between the perceived agency of VHs and measures of social influence (e.g., presence, physiological measures, or interpersonal distance). They found that a VH was more influential when participants perceived that the VH was controlled by a real human (i.e., an avatar), than if it was perceived to be controlled by a computer algorithm (i.e., an agent). Nowak and Biocca [12] did not find any agency effects,

but found that a higher anthropomorphism of the VH resulted in a reduced sense of social and co-presence, which conflicted with their hypothesis. They explained this result by stating that a more anthropomorphic image might reinforce the participant's expectations about realistic behaviors of the VH which could not be entirely met in the experiment.

Bailenson et al. [118] studied one's level of co-presence in a multi-user shared immersive virtual environment while manipulating the VH avatar's non-verbal behaviors: i) human forms with head movements, ii) human forms without head movements, and iii) human voice only. They reported one's higher sense of co-presence in a condition with head movements compared to the other conditions. Bailenson et al. [119] also compared different forms and behaviors of realism using a VH. In their study, they used video-conference (high behavioral and form realism), voice-only (low behavioral and form realism), and "emotibox" avatar (high behavioral realism and low form realism) conditions, and showed that participants rated the lowest score for "emotibox" in the self-reported co-presence.

Huang et al. [120] developed a VH, which could more precisely predict the timings of back-channel feedback and end-of-turn in conversations. They used one's sense of social presence with the virtual agent as a measure of "rapport" with the virtual agent—the feeling of being "in sync" with a conversational partner. Wang and Gratch [121] emphasized that a VH exhibiting only mutual gaze (i.e., eye contact) could reduce the sense of rapport with the agent. They suggested that VHs should exhibit other gestures such as head movements and body postures along with attentive gaze to improve the quality of real–virtual human interaction. Jo et al. [110] conducted a study with an AR tele-presence framework using a VH, and found higher responses in user experience and presence with the VH system by resolving the matching problem between virtual objects and remote real objects. Chuah et al. [13] developed interactive VHs with a physical body (e.g., mannequin legs) in a medical application and concluded that increasing the physicality of VHs could encourage higher social presence.

Regarding the *physicality* or the *illusion of physicality*, which generally refers to the perceived degree to which a VH appears to the user as a physical entity with form and shape, Li [122] points out that physicality has two different dimensions in many previous studies: *physical embodiment*—whether the VH or RH has a real/physical body or not (e.g., a tele-present robot compared to a simulated virtual agent with a similar appearance); and *physical presence*—the fact that it is physically present close to the user (e.g., robots that are co-present compared to tele-present). Li summarized the results of 33 previous publications related to physicality and suggested that a positive perception was attributed to the physical presence of surrogates (e.g., RHs and VHs), not the physical embodiment.

2.4.2 Behavioral Influences of Virtual Humans

Social influence of VHs has been evaluated not only with subjective self-reported ratings but also behavioral observations, such as a user's gaze behavior (e.g., eye contact) and proxemic behavior (e.g., personal space). Bailenson et al. [118] showed evidence supporting the "equilibrium theory" with VHs, i.e., that mutual gaze and proxemic behavior are inversely related to each other. They found that people maintained more space around VHs than non-human-like virtual objects [123] and that one's perception of high co-presence with the VH associates with more eye contact. Also, they found that people maintained greater distance from VHs when approaching their fronts compared to their backs, and gave more personal space to the VHs who engaged them in mutual gaze [124].

Argelaguet et al. [125] investigated obstacle avoidance behavior while walking around real and virtual static obstacles inclucing a VH in a large immersive projection setup. The results showed a decreased walking speed along with an increase of the distance from the obstacles when facing virtual obstacles compared to real ones. Also, participants kept more distance when the obstacle was

anthropomorphic compared to an inanimate object and the orientation of the obstacle also influenced the avoidance behavior. Bönsch [126] presented a study showing that people avoided a VH in a CAVE environment, and also preferred the VH's collaborative collision avoidance behavior, e.g., the VH's behavior to step aside for them.

A human user's personality traits also are important factors that characterize their behaviors with VHs. Wieser et al. [127] investigated one's level of social anxiety and behavior with VHs. They suggested that people with high social anxiety showed a complex pattern of avoidance behavior, e.g., women with high social anxiety tended to avoid eye contact with male VHs exhibiting direct gaze behavior. Obaid et al. [109] conducted a study comparing the user's voice level while interacting with agents in AR and VR, and found that in both conditions the user's voice level compensated for the distance to the agent, although the effect was stronger in AR, i.e., users more strongly perceived the distance between themselves and the AR agent. The result might be explained by participants perceiving the distance between themselves and the AR agent with less perceptual error [128]. Garau et al. [129] evaluated human user's responses including presence, co-presence, and physiological signals, with respect to a VH's different degrees of responsiveness. Their results did not show a significant relationship between one's perceived co-presence and the VH's degree of responsiveness. However, they suggested that one's lower level of computer usage was linked to a higher level of co-presence. Further behavioral dynamics in locomotion between real humans and VHs or crowds were studied by Warren and Fajen et al. [130, 131, 132], and Neth et al. [133] leveraged the behavioral dynamics while walking with a VH to avoid collisions with physical objects in the real world using a redirected walking algorithm.

For different behavioral influences, Guadagno et al. [103] investigated the role of a VH's gender and behavioral realism in persuasion, and found that the VH was more persuasive when it had the same gender as the user, and exhibited greater behavioral realism. They also found in-group favoritism for female participants, i.e., women liked a female virtual character, while male partic-

ipants did not show in-group favoritism. Rosenberg et al. [134] found that people who helped a VH in distress by using a virtual "super power"—in this case the ability to fly in a virtual environment—were more likely to help people in the real world afterward.

In addition to behavioral observations, some tried to examine physiological responses with VHs. As mentioned earlier, Garau et al. [129] included physiological signals as the measures to investigate the effects of a VH's different degrees of responsiveness in their study. Also, Obaid et al. [135] used VHs in AR to evaluate the relationship between the physiological arousal of users and the VH's behavior associating with cultural differences (e.g., personal space and gaze), and suggested that mutual gaze had a higher impact on one's sense of arousal than the interpersonal distance.

2.4.3 Trust and Confidence in Virtual Humans

The perceived trust and confidence in VHs are often researched together with the level of social presence because of their potential correlation.

Bente et al. [136, 137] observed an increased interpersonal trust along with a strong social presence in network-based communications using embodied virtual representations with a high level of nonverbal activity, while Riegelsberger et al. [138] showed that an embodied virtual representation still elicits a lower level of trust than a video conference setting with real humans. Pan and Steed [139] compared three different forms of communication including embodied VH interaction with respect to the perceived trust in advice-seeking situations, and found that the virtual form was not preferred compared to the other forms, i.e., face-to-face and robotic embodied interactions. They suggested that the physical presence of the robot representation might have influenced the trust assessments positively.

Related to the trustworthiness of VHs, one's self-disclosure behaviors with VHs have been investigated. Lucas et al. [140] found supportive evidence that people tend to have more self-disclosure behaviors when VHs are perceived as an agent (i.e., an autonomous computer) than an avatar (i.e., a real human behind the scene) in stigmatized scenarios like clinic interviews. Mell et al. [141] also showed high self-disclosure behavior with a VH agent in a sensitive topic asking about one's financial status.

CHAPTER 3: PROPOSED ENVIRONMENTAL PHYSICAL–VIRTUAL INTERACTION IN MR

The previous chapter summarized background knowledge about current MR technology and some related work in VH research, emphasizing the importance of seamless interaction of a VH with the surrounding physical environment with respect to the social influence of the VH.

In this chapter we describe in detail the proposed environmental physical–virtual interaction, which was briefly introduced in Chapter 1, for improving the sense of social presence with a VH in pervasive MR environments. Specifically, we will discuss three important aspects to consider for a VH's environmental physical–virtual interaction: (1) the visual embodiment of the VH, (2) interaction that includes physical awareness and influence, and (3) the plausibility and coherence of the VH's appearance and behavior in the MR environment. In an attempt to improve the social presence of VHs in MR settings, these three aspects were carefully considered in the designing of the VH prototypes used in the experiments presented in this dissertation.

3.1 Visual Embodiment of Virtual Humans

Current-state commercial virtual assistant systems, such as Apple Siri or Amazon Alexa, do not normally have visually embodied appearances other than the physical device providing the audio input and output for the system. Such virtual assistants rely primarily on verbal discourse to communicate with the users with only a few simple visual indications on the device, such as different colors or types of animated LED lights, to convey the current state of the system. The inability for more complex or nuanced information to be conveyed visually could lead to possible confusion or misunderstandings about the expected responses, actions, or objectives of the assistant.

VAMR has the potential to overcome this challenge by providing a visual embodiment for the virtual assistant. A human-like visual representation could enrich the communicative channels that convey the assistant's status and intentions by presenting gestures and motions as social behaviors. Thus, researchers in Intelligent Virtual Agents (IVA) have proposed various types of visually embodied VHs with human-like appearances, and claimed that having a visual embodiment affords richer, more smooth and efficient real-virtual human communication by adding the visual interface modality [142, 143]. In particular, a VH with a human-like appearance can be much more suitable for providing social signals and cues that are generally present in human–human communication, e.g., sharing of emotional state or conveying action intent through non-verbal expressive behaviors [144, 145].

The proposed VH's interaction with the physical environment in this dissertation benefits from the VH's visual embodiment in MR, specifically by having a human-like appearance that can exhibit non-verbal behaviors.

3.2 Interaction: Awareness and Influence

The term "*look and feel*" has traditionally been used by software designers in computer science when discussing graphical user interfaces (GUI), to describe the interface design. In this context, the "look" is the visual appearance, such as the colors, shapes, and layout, whereas the "feel" refers to the dynamic behavior of the visual components like buttons and sliders.

In research looking at VHs as embodied interfaces, the "feel" can be thought of as corresponding to the VH's interaction while the visual embodiment of the VH, as discussed in the previous section, is the "look". This dissertation explores the relationship of the environment that is within an interactable range of the VH as well; consequently, the "feel" should not only cover the interaction

with any human users, but also the interaction with the surrounding MR environment. The VH's interaction, which is generally intended to mimic a real human's, can be characterized by the VH's apparent awareness of, and ability to influence, the environment through expressive behaviors, e.g., gaze, facial expressions, locomotion, and gestures. The behaviors related to awareness and influence are normally performed in a perception-action sequence or cycle. For example, when a person becomes aware of someone trying to enter the room, she may open the door, after which she may become aware of the other person's awareness of her, leading her to greet that person, and so on. A basic model of awareness and influence in human interactions is illustrated in Figure 3.1.

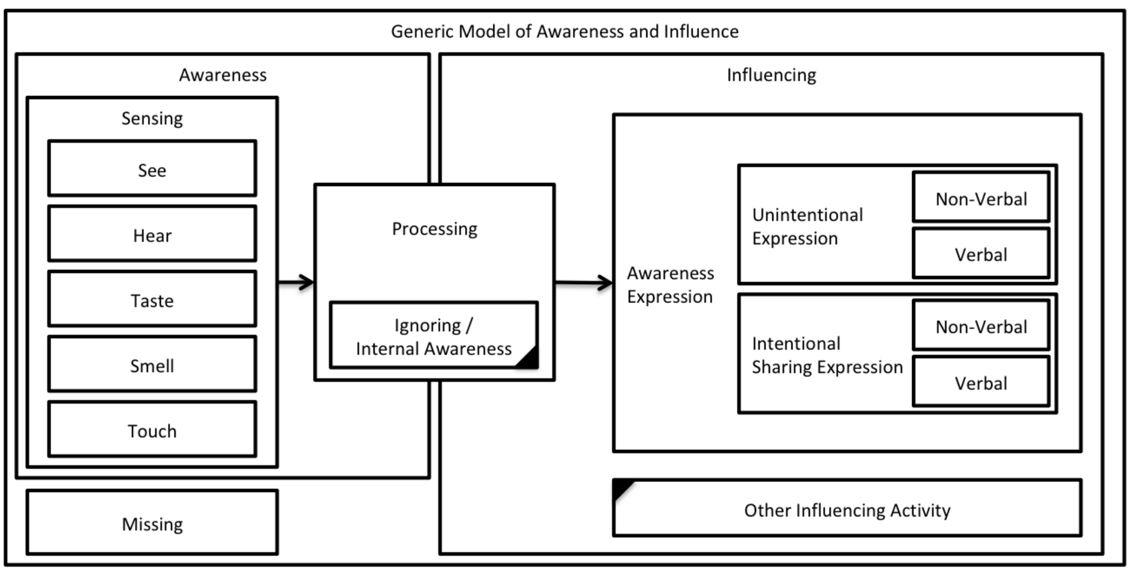


Figure 3.1: A basic model of awareness and influence in human interactions.

In a MR setting, where a VH exists in the real world with (dynamic) physical objects and real humans, achieving interaction between the VH and the physical environment is relatively more challenging as compared to robotic humans in the real world or VHs in a virtual world, due to a VH's inherent lack of physicality [25]. For example, consider a scenario in which a VH needs to sit on a real chair, currently pushed under a real desk, in the MR environment (refer to Figure 3.2).

People would normally expect that the VH should be able to recognize the physical chair (i.e., be aware of it) and physically move the chair to sit on it (i.e., be able to influence it). We may assume that various methods exist to achieve the recognition of the chair, such as through computer vision techniques or radio-frequency identification (RFID) tags. However, it is still challenging to physically move the chair without additional, unusual hardware to control the chair, such as an IoT-enabled actuator (refer to Section 2.1.2).

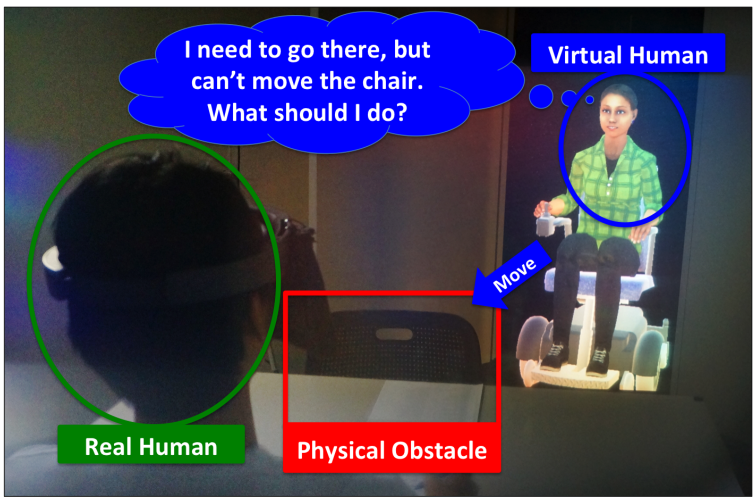


Figure 3.2: An example of challenging physical–virtual interaction in MR. The VH cannot physically move to the space where the real chair is already present.

The experiments looking at human perception and behavior as discussed in this dissertation do not focus on the implementation of hardware for mechanical actuation or control of physical objects in the environment for general purposes, as that is outside the scope of this research. Rather, the experiments specify the scenarios that involve the VH's interaction with the surrounding environment and make the VH be perceived interactive in those particular settings—i.e., an illusion of the VH's physical environmental awareness and influence.

3.3 Plausibility and Coherence

Once the VH has the ability to interact with the physical environment, by exhibiting awareness of and/or influencing the environment, the interactive behavior should be plausible and coherent within the context of the MR environment. The concepts of coherence and plausibility, as they relate to immersive VR experiences, are used to describe maintaining or not breaking the sense of spatial presence, i.e., the sense of being there within the VR environment (refer to Section 2.2). More broadly, the concepts of plausibility and coherence could be associated with the counterfactual theories of causation—the understanding of causality (the cause and effect relationship between events) influences judgments of the plausibility of events [146, 147]. In epidemiology research, Höfler defined *plausibility* and *coherence* stating "the observed association can be plausibly explained by substantive matter (e.g. biological) explanations" and "a causal conclusion should not fundamentally contradict present substantive knowledge" respectively [148]. The difference between plausibility and coherence is subtle. Plausibility is about whether one could imagine that one event would have caused the observed result if it had really happened, while coherence is about whether the relationship between the event and the observed result would fit in the existing theory, which is already assumed correct [149].

For the aforementioned chair example in the previous section, there are several approaches that could potentially deal with the problem of the VH's lack of physicality while considering the plausibility and coherence in the VH's behavior, as described below (also refer to Figure 3.3):

- **IGNORE**: The VH could just accept the limitation (any physical-virtual mismatch) and ignore the physical object, e.g., the VH passes through the chair. However, this can cause a "break in presence" or negative expectancy violation and harm the sense of social presence with the VH (Section 2.2).

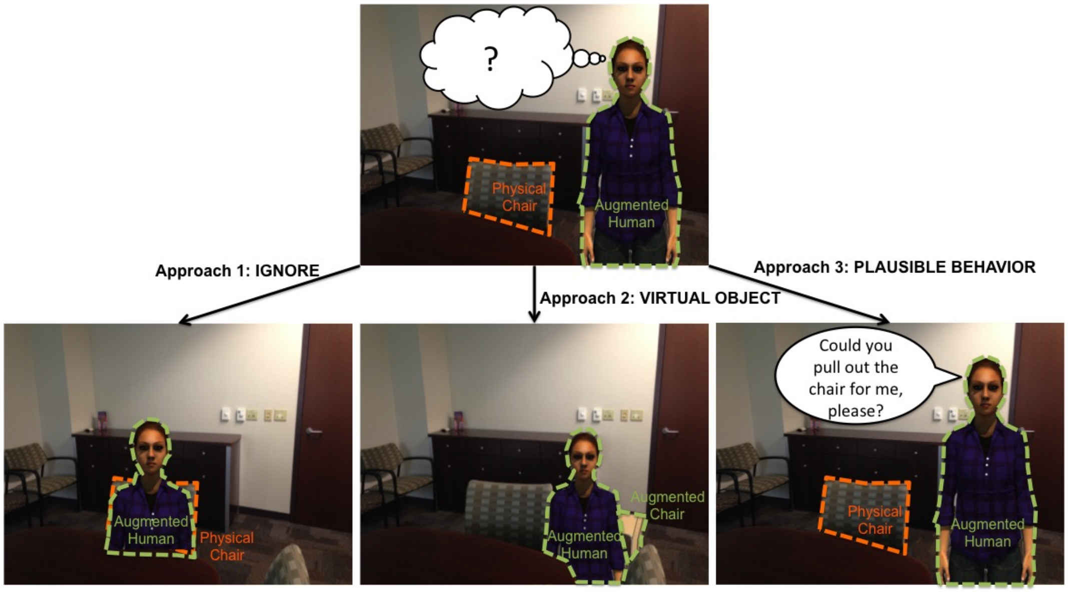


Figure 3.3: Examples of potential approaches to deal with the issue of physical–virtual interaction. Top: the VH cannot physically move the chair to sit on. Bottom-left: the VH can ignore its limitation and just be overlaid on (co-located with) the current physical setup. Bottom-center: An augmented object (chair) can be used to maintain the plausibility. Bottom-right: the VH can ask a real human's help to move the chair, not only maintaining plausibility, but also perhaps enhancing the perceived social and co-presence with the VH.

- **VIRTUAL OBJECT**: The VH can use a virtual chair in the MR environment, so that she can can pull out and sit on the virtual chair. If this is not possible because of some physical limitations, such as a lack of unused space in which to render a virtual chair, the physical chair could be visually erased and a virtual chair displayed at the same location (i.e., via diminished and augmented reality). In this way, plausibility might be maintained.

- **PLAUSIBLE BEHAVIOR**: The VH can ask for help from a real human to alter the physical environment, without revealing its functional deficiency, e.g., the VH might ask a nearby real human to pull out the chair. In this way, the limitation due to the lack of physical influence by the VH might actually present an opportunity to enhance the overall perception of the VH's engagement in the interaction without harming the contextual plausibility.

For the proposed environmental physical–virtual interaction of a VH in MR, the VH's behavior should be plausible and coherent within the MR environment and within the social context with the real human users. To achieve this, the VH should first obey the physical rules in the real world unless it is framed as an entity that has "super powers" like passing through obstacles or flying in the air. In addition, the VH should also regulate social norms to be more realistic and influential to the human users while interacting with them and the environment.

CHAPTER 4: INFLUENCE OF PHYSICAL BODY IN HUMAN BEHAVIOR

In this chapter, a large-scale exploratory study in a public space [30] is introduced and described to understand how human behaviors interacting with technological human surrogates can be differed by the surrogate's physicality, e.g., robotic or virtual humans (refer to Section 2.3). In this way, we can first verify the importance of virtual human's (illusion of) physicality in human perception and behavior in real–virtual human interactions, before we focus on the experiments that actually evaluate the effects of the proposed environmental physical–virtual interaction for real–virtual human interactions in Chapter 5 and Chapter 6.

For the study, human behavior interacting with a robotic human is compared with the behavior with its live video stream. By using the video stream as an extremely realistic virtual form of the robot, the confounding impact by the visual difference between the robotic and virtual humans can be ruled out, but the effect of physicality can be focused. Prior research has investigated the effects of the human surrogate's physicality and gesturing in human perceptions and social influence of the surrogate. However, those studies have been carried out in research laboratories, where the participants were aware it was an experiment, and the participant demographics were typically relatively narrow—e.g., college students. Here, the present study involves 7685 people in a public space, where they were unaware of the experimental nature of the setting. Their behaviors interacting with different human surrogate settings, in terms of its physicality and gesturing, are collected and statistically analyzed. The behaviors are evaluated using several variables, such as proactivity, reactivity, and proximity. Several interesting phenomena are identified, and those could lead to hypotheses developed as part of future hypothesis-based studies. Based on the measurements of the variables, people are more likely to be engaged in a human–surrogate interaction when the

surrogate is physically present, i.e., a robotic human, but movements and gesturing with its body parts have not shown the expected benefits for the interaction engagement. In addition to the main effects by the surrogate's physicality, statistically significant differences are found regarding the demographics of the people in the study, such as higher engagement for females than males, and higher reactivity for younger than older people. Implications of the findings will be also discussed for practitioners aiming to design a technological surrogate that directly interact with real humans, while emphasizing the effects of virtual human's physicality in human behavior.

4.1 Material

4.1.1 Technological Human Surrogate

As mentioned above, a physical robotic human surrogate is utilized in the study, which we call the *RoboThespian*—a life-size robotic human manufactured by Engineered Arts Limited (Figure 4.1). For the experimental setting, the RoboThespian was controlled (inhabited) dynamically by a real human through Wizard of Oz paradigm [1] [150].

4.1.1.1 RoboThespian (Robotic Human Surrogate)

The RoboThespian includes a projector and short-throw optics in its head, allowing us to change its facial appearance and expressions dynamically via computer-rendered graphics and animations. To support gesturing, the RoboThespian uses a combination of electric and pneumatic actuation (fluidic muscles). The RoboThespian is fitted with ten fluidic muscles that control the following joints on each arm: shoulders (roll, pitch, yaw), elbow (pitch), and wrist (pitch). Six independent

[1] In a *Wizard of Oz* paradigm a human subject is made to believe that a human surrogate behaves autonomously when it is controlled by a human.

servo motors control the head (roll, pitch and yaw) and the torso (roll, pitch and yaw). On each hand, the thumb is fixed while each of the four remaining fingers are actuated in a binary manner (extended or curled) using directional control valves. The finger actuation is intended to be purely for gesturing, pointing, or other types of non-verbal communication—the lack of thumb actuation and low force exerted when the fingers are closed makes the hands unsuitable for gripping or interacting with objects. As described, the upper-torso of the RoboThespian has a total of 24 independently controllable degrees of freedom. While the legs of the RoboThespian *can* be actuated (allowing the RoboThespian to squat down), they were fixed in a rigid standing configuration for this study. The natural low impedance characteristics of the pneumatic actuators make the RoboThespian relatively safe for use in an environment where other humans will be nearby during an interaction. Custom clothing was fashioned to fit over the metal and plastic frame of the RoboThespian so that it could provide a more plausible human-like appearance while not overly encumbering the motion. In addition to shoes, pants, and a long-sleeve shirt, as wig was also fastened to the head of the RoboThespian to give it hair and to hide the parts of the plastic head shell that did not have projected imagery (Figure 4.1).

4.1.1.2 Human in the Loop

We say that a surrogate has *agency* when a computer algorithm is used to generate autonomous responses during interpersonal communication (including both verbal and nonverbal behaviors). While still having made significant strides, enabling complete agency in technological human surrogates is not yet possible. The current state-of-the-art research in AI cannot yet replicate the intelligence level and natural behavior of humans in social interaction. Thus, many previous studies involving social interactions with technically sophisticated surrogate systems have used a human-in-the-loop to control the surrogates (i.e., a Wizard of Oz paradigm with a human controller operating the surrogate behind the scenes).

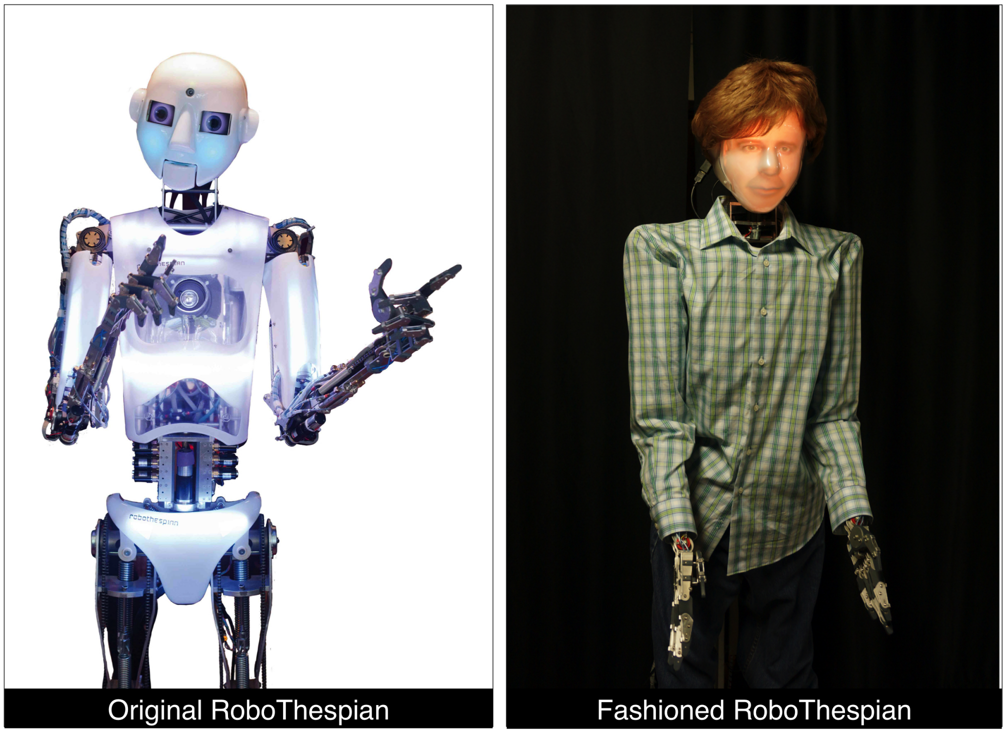


Figure 4.1: The RoboThespian™ used in the study. To provide a more plausible human-like appearance compared to the system as shipped, custom clothing was fashioned to fit over the metal and plastic frame, and a wig was fastened to the head (right photo).

For this study, a male professional actor was hired and trained to control the surrogate, the RoboThespian, from a remote location. The actor used an interface with an infra-red camera (TrackIR) and a magnetic tracking device (Razer Hydra) to affect the surrogate's facial animations, head movements, and upper-torso gesturing. The actor was also able to view the environment around the surrogate's location via a commercial video-chat program (Skype) and a camera set up in the interaction space. The camera was positioned to provide the actor with an approximation of the surrogate's viewpoint. In addition, the actor wore a microphone headset, which allowed verbal communication with remote people. In this way, the actor could speak naturally while controlling the surrogate and exhibit appropriate verbal and nonverbal responses in context. The details

of the control mechanism (AMITIES) is described in [151]. For the study, two different modalities of the surrogate (i.e., surrogate's physicality and gesturing) were varied to see their effects in human–surrogate interactions (described in the section below); however, the human controller (human-in-the-loop) was not informed of the surrogate condition of the day and could not see the surrogate through the video-chat camera view to the people in the remote place considering the confounding effect of the controller's surrogate condition awareness.

4.1.2 Experimental Setup

4.1.2.1 Surrogate and Environment Settings

The study was conducted in the lobby of Partnership 3 building at the Institute for Simulation and Training, University of Central Florida, where a more diverse population sample can be involved. Since there were two different states of physicality (the physical RoboThespian, or the video stream on three 65" wide-screens), either the RoboThespian or the screens must be moved in/out of the lobby according to the physicality condition of the day. For the physical robot conditions, the RoboThespian and its peripheral devices (e.g., an air compressor and a PC to control the RoboThespian) were placed in the lobby. For the screen conditions, the RoboThespian and its peripheral devices remained in the lab space, and three live HD video streams of the RoboThespian from three HD camcorders were run through the building and fed to the three screens in the lobby. Because of the long distance between the lab space and the lobby, HDMI-to-Ethernet converters were used to transmit the signal through the Ethernet ports in the lab/lobby, as opposed to direct HDMI connections. The physical settings needed to be practically deliverable and easy to set up because some forms of equipment had to be moved in/out of the lobby for each session. To ease the transition processes most of the equipment was attached on rolling tables and moved together. Two speakers for conveying the surrogate's voice (the controlling actor's voice) were placed on

the table behind/near the surrogate (RoboThespian or screens) so that people would perceive the voice as coming from the surrogate. For collecting their behavioral data during the interaction, a Microsoft Kinect sensor was installed on top of the black curtain rod above the surrogate, and a microphone was placed on the floor next to the surrogate's feet. A webcam was placed near the Kinect sensor on the curtain rod for the controlling actor's view. All other devices on the tables were hidden behind black curtains. Since there was a task that the surrogate requested the people in the lobby for taking a photo of itself in the interaction scenario, a camera that people could use was placed on a chair about five meters away from the surrogate setup. In this way, the surrogate could point to it during the gesturing conditions. The overview of the surrogate and environment settings are shown in Figure 4.2.

4.1.2.2 Surrogate Control and Communication Connections

The human controller (a trained professional actor) controlled the surrogate from a remote room separated from both the lobby and the lab space while viewing the surrogate environment (lobby) through a commercial video-chat program (Skype). A webcam for the Skype call was configured in the lobby not to see the surrogate so that the controller was not aware of the physicality and gesturing conditions.

A human observer was hired and trained to observe the interaction between the surrogate and people in the lobby. The observer was tagging interesting moments of the interaction next to the controller. The details about the human observer will be explained in Section 4.4.1. Multiple client-server software connection frameworks were used among the human controller, human observer, and the RoboThespian. The framework allowed the controller to manipulate the RoboThespian's behaviors through controlling devices (refer to Section 4.1.1.2) and the observer to create pre-defined/custom tags in-situ while the interaction was happening. The controller wore a headset to

communicate with people in the lobby through the video call. As the human observer also needed to see and hear the surrogate environment for appropriate tagging, a mirroring monitor and audio splitter were used for duplicating the human controller's feeds. Overall diagrams for the settings are shown in Figure 4.3.

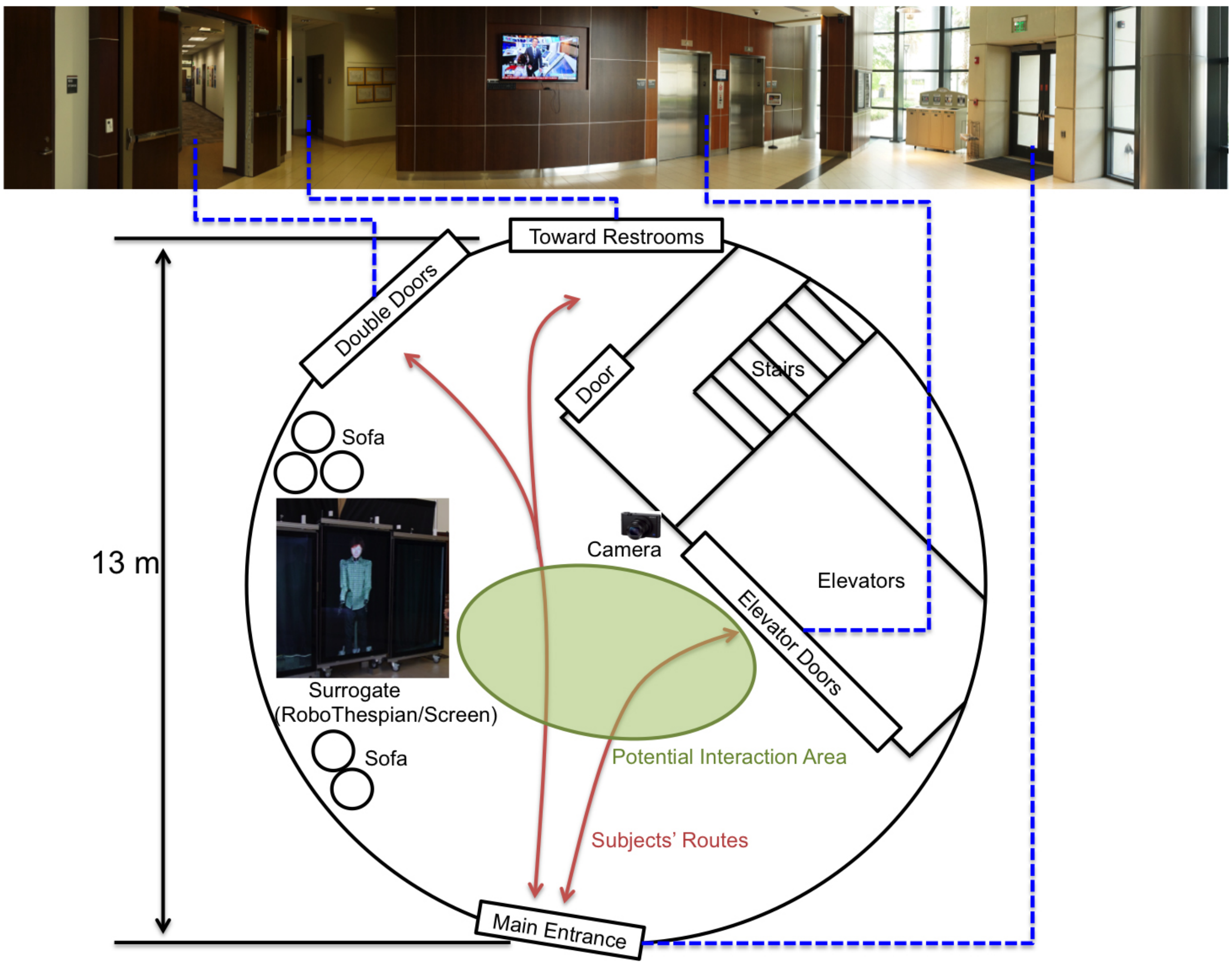


Figure 4.2: The physical experimental layout. Top: a panoramic image of the lobby from the surrogate's perspective. Bottom: the floor plan of the lobby, with the potential interaction area and typical routes indicated.

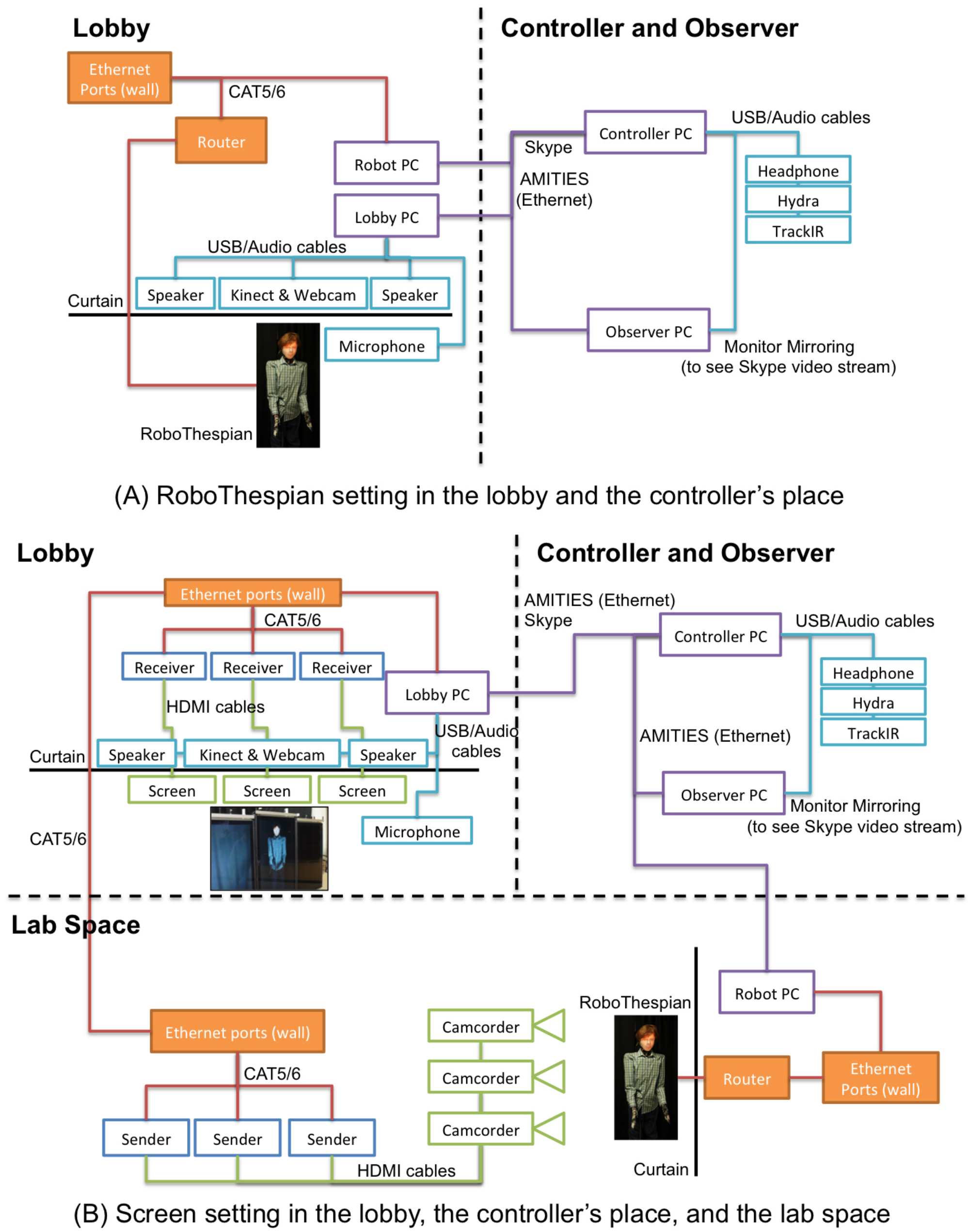


Figure 4.3: Diagrams indicating the components and connections corresponding to the two physicality conditions. (A) The condition where the RoboThespian is physically present in the *lobby*. (B) The condition where the RoboThespian is physically present in the *laboratory*, and viewed by the people in the lobby via HD video feeds displayed on the wide-screen display setup.

4.2 Methods

A 2×2 factorial design was prepared to explore the effects of surrogate physicality and gesturing (Figure 4.4). The independent variable *physicality* had two levels: (i) the RoboThespian was physically present in the local environment, or (ii) three large screens displayed a real-time video stream of the RoboThespian. The real-time video stream was used to minimize differences between the two levels in physical/behavioral authenticity and the visual fidelity of the human surrogate. The independent variable *gesturing* also had two possible states: (i) the RoboThespian exhibited gestures with the upper torso (including arms and hands) or (ii) the RoboThespian exhibited no upper torso gestures. Note that, independent of the physicality and gesturing level, the RoboThespian could move its head freely (under the control of the remote actor). This allowed the RoboThespian to show attention and interest by turning its head toward interlocutors while speaking to them. Figure 4.4 depicts the 2×2 factorial design visually, and the four corresponding experimental groups are described below.

- Group I (RoboThespian *with* Gesturing): People encounter the ***RoboThespian***, which is physically present in the local environment, and the RoboThespian ***can perform the upper-torso gesturing***. The upper-torso gesturing mostly includes arm movement such as opening arms, hand shaking, and pointing.
- Group II (RoboThespian *without* Gesturing): People encounter the ***RoboThespian***, which is physically present in the local environment, and the RoboThespian only moves its head, but ***cannot move the upper torso*** (including arms and hands) at all.
- Group III (Screen *with* Gesturing): People encounter the video stream of the remotely-located RoboThespian through the wide ***Screen*** consisting of three aligned large TV displays, and the RoboThespian on the screen ***can perform the upper-torso gesturing***.

- Group IV (Screen *without* Gesturing): People encounter the video stream of the remotely-located RoboThespian through the wide ***Screen*** consisting of three aligned large TV displays, and the RoboThespian on the display can only move its head, but ***cannot move the upper torso*** (including arms and hands) at all.

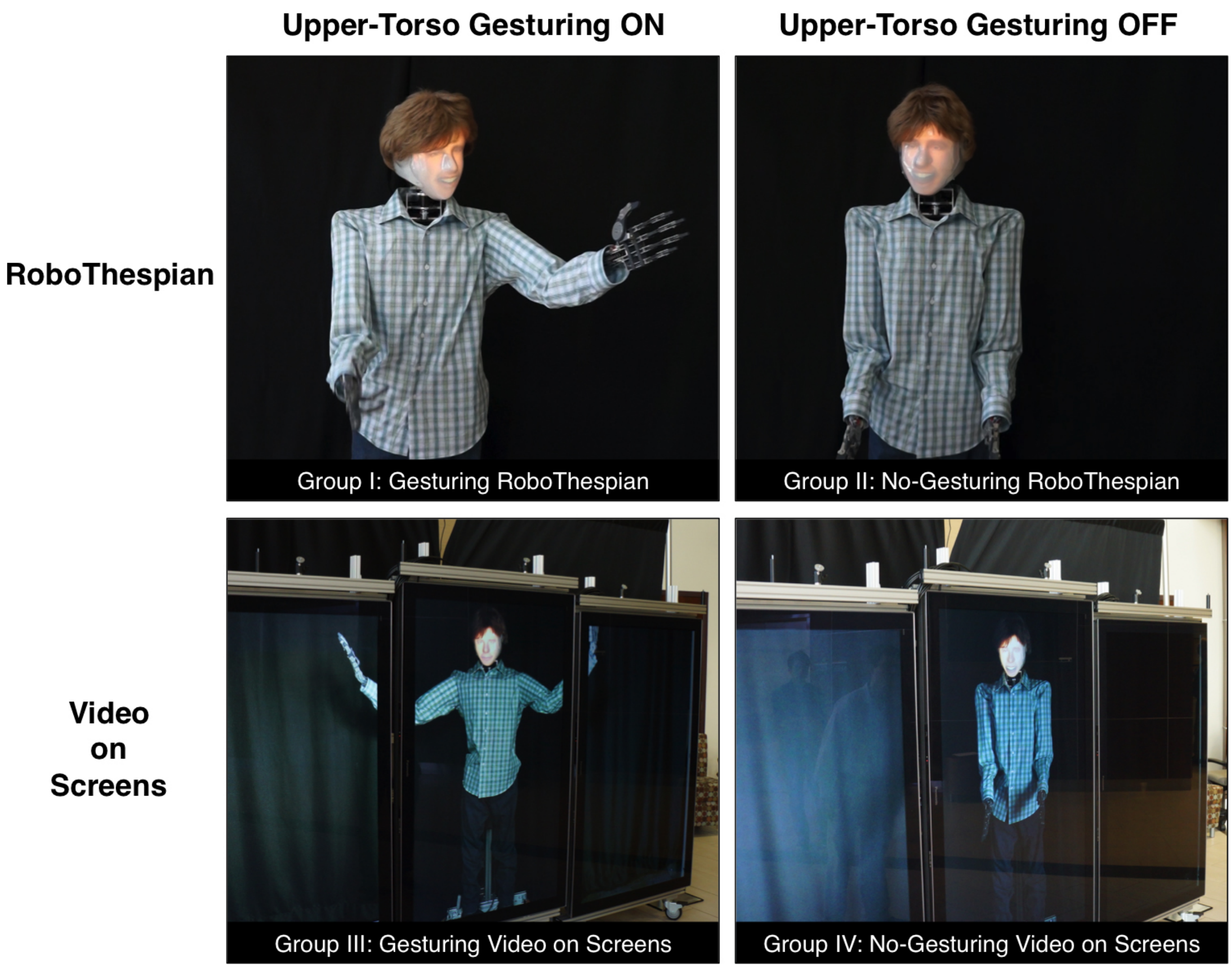


Figure 4.4: The 2 $\times$ 2 experimental conditions across physicality and upper-torso gesturing: RoboThespian vs. video screen $\times$ gesturing ON vs. gesturing OFF.

The surrogate (in a form of the RoboThespian or the screens) was placed in the lobby of an office building at a university research park for two hours per day, and—via the controlling actor—greeted and interacted with people as they entered and left the lobby. To maximize the number of human–surrogate interactions in the daily two hour session, the surrogate was placed in the lobby either from 11 AM–1 PM or 1–3 PM. These session times roughly correspond to the start and end of lunch breaks, and thus tend to be high traffic periods when people frequently moved in and out of the lobby. There were ten two-hour sessions per experimental group, for a total of 40 sessions in the study. The experimental groups described above were randomly assigned to each of the session time slots.

4.3 Participants and Procedure

No selection criteria was applied to limit who could be involved in the experiment. Anyone who entered the lobby was considered as part of the study. Interactions with the surrogate were entirely voluntary, so there was no compensation for the people who interacted with the surrogate in any way. The building where the study took place is home to companies, non-profit trade organizations, military research offices, and academic research labs. The observed people usually included employees/members or guests of these organizations. Most people were adults, although in some cases children passed through the building (e.g., on "Take your Child to Work" Day).

Figure 4.5 summarizes the procedure the actor followed when controlling the surrogate. When the lobby was empty, the surrogate (the physical RoboThespian or via video stream) stood still and was silent. If people entered the lobby and appeared to be staring at or looking around the surrogate, the actor controlling the surrogate would initiate a conversation with the people. In practice, the actor observed the people for approximately five seconds before initiating the conversation. In some cases, people were more proactive, in that they would initiate the conversation with the surrogate

instead. The conversation between the people and the surrogate was not limited to any particular topic, and was mostly casual "small talk."

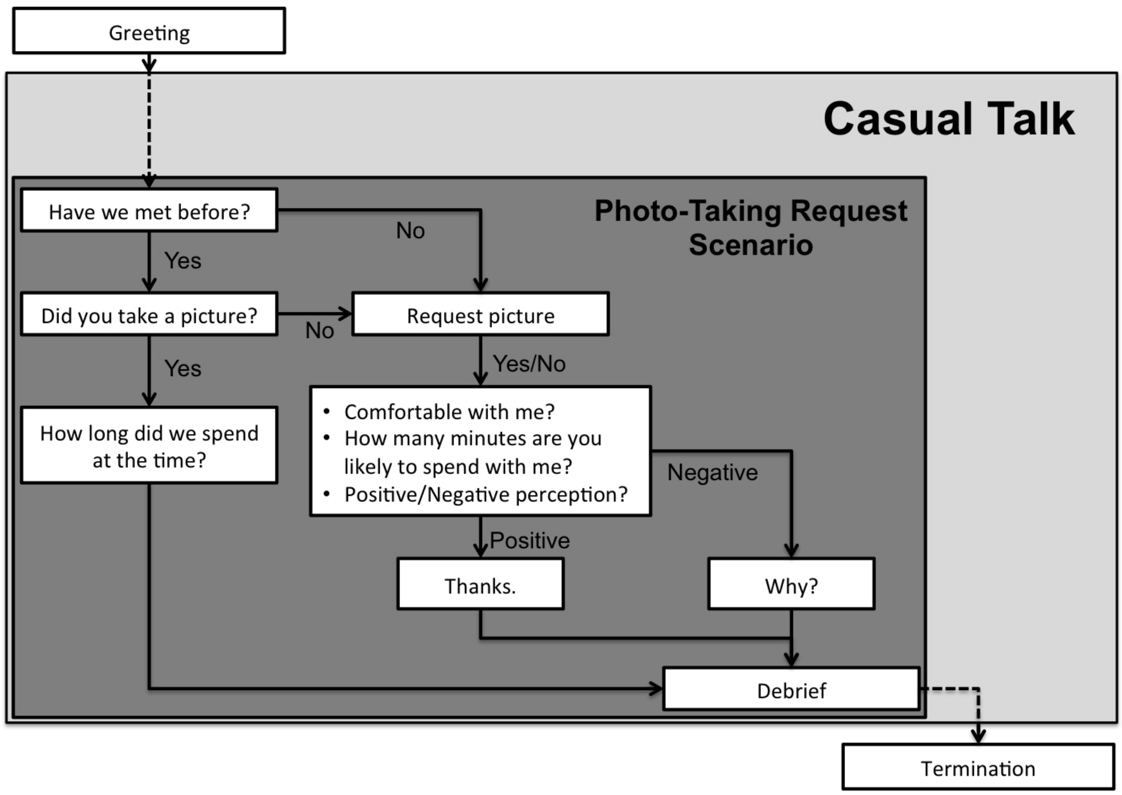


Figure 4.5: Possible interaction flow between the surrogate and the people in the lobby. The flow can jump to the Termination at any time during the interaction because of the unpredictable interaction pattern—the people could leave any time they wanted.

The actor usually began the conversation by asking whether they (the people and the surrogate) had met before. If the people answered yes, the surrogate would then ask them about his/her general perceptions of the previous interaction. Next, the surrogate would initiate a compliance test, a request of the people that aimed to probe the extent to which they felt socially connected to the surrogate. The compliance test used here was a photo-taking request. The surrogate asked the people to take a photo of itself using a camera located nearby. If they complied, it could imply

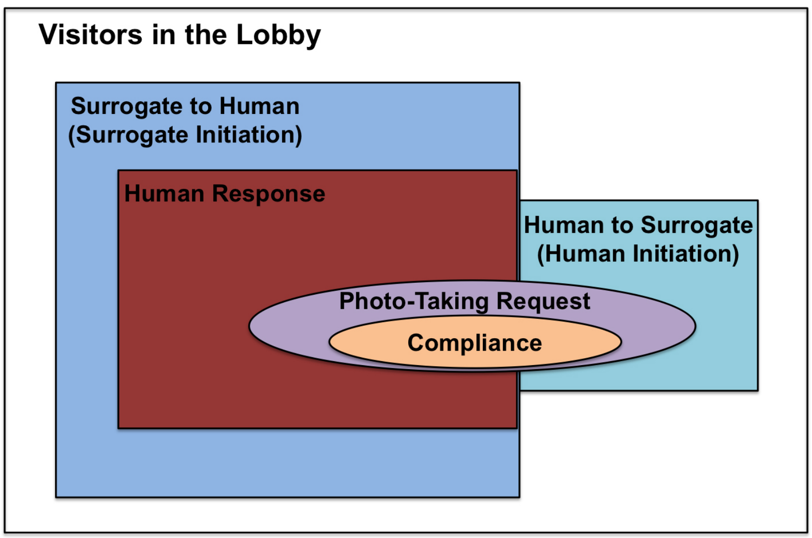
they felt socially comfortable enough with the surrogate to provide help. After the photo-taking request, the surrogate briefly explained the purpose of the study and asked their permission to use the data collected during the conversation before ending the conversation. Note that the people did not have to continue through this entire process. They could terminate the interaction at any time. Given the public setting and experiment goals, we did not use a written form of informed consent prior to the study. However, people were verbally informed about the details and purpose of the study after the interaction with the surrogate, and they received the phone number of the Principal Investigator as a contact point. This experiment protocol was carried out with the approval of the Institutional Review Board at the University of Central Florida (IRB Number: SBE-14-10313 and SBE-16-12347 in Appendix A)

4.4 Measures and Hypotheses

4.4.1 Data Collection and Human Observer

Audio and video data were collected during the interaction between the surrogate and the people in the lobby. For the video, we used a Microsoft Kinect (Kinect for Windows and SDK v1) to capture all the RGB color, depth, and skeletal images. The audio was recorded using the microphone on the floor near the surrogate's feet. We expected a large number of people, and the data collected during this experimental study was unique since it was designed to understand open-ended natural interactions between the surrogate manifestations and real humans, not a controlled study. Thus, to facilitate easier classification of the data, an active observer-based tagging system that allows a person to tag events in real-time was implemented. The human observer was seated next to the human controller and could see and hear the same as the controller. While there was anybody in the lobby, the observer created tags using an interface with pre-defined/custom tags (see Figure 4.6-A).

(A) Human Observer's Tagging Interface



(B) Distribution of the Observed People

Figure 4.6: (A) The human observer interface for monitoring the lobby visitors and tagging interesting moments during the interaction. The observer could use pre-defined tags such as "Surrogate to Human" or "Human to Surrogate" (conversation initiations), and could also enter raw notes via a Custom Tag (lower left of the display). The fact that the virtual surrogate's clothing does not match the clothing we used on the RoboThespian is OK—the only characteristics transferred to the RoboThespian were the face (dynamic appearance) and the body posture (dynamic upper body). (B) A diagram for the distribution of the people in the scope of this study.

The pre-defined tags could describe the increase/decrease of the people in the lobby, whether the people initiated a conversation to the surrogate or vice versa ("Surrogate to Human" or "Human to Surrogate"), and whether the people responded to the surrogate or not ("Response"). These tags helped us to extract the information of interest regarding the dependent variables described in the next section. The distribution of the people based on the tags is shown in Figure 4.6-B. The observer could see the surrogate's gestures through a virtual character displayed on the interface. Whenever anyone was in the lobby, data recording started by the observer's call. While collecting all the data (audio, video, and observer's tags), they have time stamps synchronized to associate with each other after the study.

4.4.2 Dependent Variables

The independent variables for the study were the surrogate's physicality (RoboThespian and wide-screen video stream) and gesturing (upper-torso gesturing ON and OFF). Although we established explicit independent variables and groups, this study was exploratory in nature, so we did not establish and test specific hypotheses from the beginning. After the study and while we were refining the collected data, we established several interesting aspects as the dependent variables—described below. We generally expected there would be positive associations between human behaviors and the surrogate's physicality/gesturing on those dependent variables.

- **Proactivity**: The ratio of the number of people who initiated a conversation with the surrogate *before the surrogate said anything*, to the total number of people who entered the lobby.
- **Reactivity**: The ratio of the number of people who responded to the surrogate *after the surrogate initiated a conversation*, to the number of people addressed by the surrogate in an attempt to initiate a conversation.

- **Commitment**: The ratio of the number of people who conversed long enough to *receive* a photo-taking request from the surrogate, to the total number of people who entered the lobby.

- **Compliance**: The ratio of the number of people who received and *complied with* the photo-taking request from the surrogate, to the number of people who received the request.

- **Photo Proximity**: How close people stood to the surrogate when they took a photo of the surrogate (complied with the photo-taking request), as indicated by the size of the surrogate's face in the photo.

4.5 Results

The comparisons for physicality and gesturing modalities by accumulating the associated groups, e.g., RoboThespian vs. Screen or Gesturing-ON vs. Gesturing-OFF are presented in Table 4.1 and Figure 4.7, while general descriptives for four experimental groups (Group I–IV in Section 4.2) are also addressed in Table 4.2. In addition, the effects of the observed person's gender and age are examined to analyze the dependent variables. Thus, here are the groups of interest for the four comparisons: Physicality, Gesturing, Gender, and Age.

- (**Physicality**) Group **RoboThespian**: Group I (RoboThespian with Gesturing) + Group II (RoboThespian without Gesturing)

- (**Physicality**) Group **Screen**: Group III (Screen with Gesturing) + Group IV (Screen without Gesturing)

- (**Gesturing**) Group **Gesturing-ON**: Group I (RoboThespian with Gesturing) + Group III (Screen with Gesturing)

- (**Gesturing**) Group **Gesturing-OFF**: Group II (RoboThespian without Gesturing) + Group IV (Screen without Gesturing)
- (**Gender**) Group **Male**: A group of people who are evaluated as males in video recordings.
- (**Gender**) Group **Female**: A group of people who are evaluated as females in video recordings.
- (**Age**) Group **Young**: A group of people whose ages are evaluated under 40 including the groups of Children (<18), Young adults (18–25), and Adulthood (25–40), in video recordings.
- (**Age**) Group **Old**: A group of people whose ages are evaluated over 40 including the groups of Middle age (40–60), and Older people (>60), in video recordings.

It is important to note that because the study was not in a controlled setting, there were various situations that made the analysis difficult. People in the lobby were coming and going, sometimes talking to the surrogate, sometimes ignoring it, sometimes talking to each other, sometimes interrupting each other, etc.; thus, it was difficult to arrive at an exact number of interactions and people. It helps to understand how the data were recorded and analyzed. The data were not recorded during the entire two-hour sessions. Instead, recording was only started when one or more people entered the empty lobby, from any direction—door, hallways, or elevator; and recording was terminated when the lobby was again empty. As such, during any given two-hour session there were many such "recordings"—a segment of data that begins when one person enters the lobby and ends when it becomes empty again. Over the course of the entire study, a total of 3942 recordings were collected. The recordings contain all the people during the 40 study sessions even including those who did not have any verbal interactions with the surrogate and who were just walking through the lobby. Also, there was a real human observer who was creating time-stamped tags for interesting moments while the lobby was not empty (see Section 4.4.1).

Due to the large number of the recordings and the complexity, the analysis mainly relied on the tags that the human observer created during the study to examine the dependent variables, which are described in Section 4.4.2. Nevertheless, all the images/videos containing the people in the lobby were needed to review to confirm the tags, count the number of people, and code their demographic information. Five human coders including myself reviewed the images/videos, and manually approximated the people's gender and age. From this intensive reviewing process, the coders counted a total of 7685 people (see Table 4.2) and built the demographical data of their gender and age. These gender and age were used as comparison criteria for the analysis of the dependent variables along with the surrogate's physicality and gesturing.

Table 4.1: Analysis results for physicality and gesturing comparisons by a large sample approximation of two correlated proportions. For gender and age comparisons, which deal with independent samples, we used a modified method for two independent proportions. (** $p < .01$ and *** $p < .001$)

Physicality	Proactivity	Reactivity	Commitment (Photo-Taking Request)	Compliance
RoboThespian	4.84%	57.90%	2.75%	66.67%
Screen	3.38%	49.12%	1.30%	74.58%
Z-value	3.06	2.76	4.27	-0.56
p-value	**.002** **	**.006** **	**<.001** ***	.575

Gesturing	Proactivity	Reactivity	Commitment (Photo-Taking Request)	Compliance
Gesture-ON	3.89%	55.34%	1.93%	70.83%
Gesture-OFF	4.07%	52.78%	1.87%	68.92%
Z-value	-0.41	0.80	0.19	0.14
p-value	.682	.424	.849	.889

Gender	Proactivity	Reactivity	Commitment (Photo-Taking Request)	Compliance
Male	3.46%	51.55%	1.53%	65.38%
Female	4.99%	59.71%	2.61%	75.00%
Z-value	-3.07	-3.57	-3.02	-1.28
p-value	**.002** **	**<.001** ***	**.003** **	.201

Age	Proactivity	Reactivity	Commitment (Photo-Taking Request)	Compliance
Young	3.69%	57.89%	2.16%	70.42%
Old	4.20%	51.38%	1.70%	69.33%
Z-value	-1.15	3.00	1.44	0.14
p-value	.250	**.003** **	.150	.889

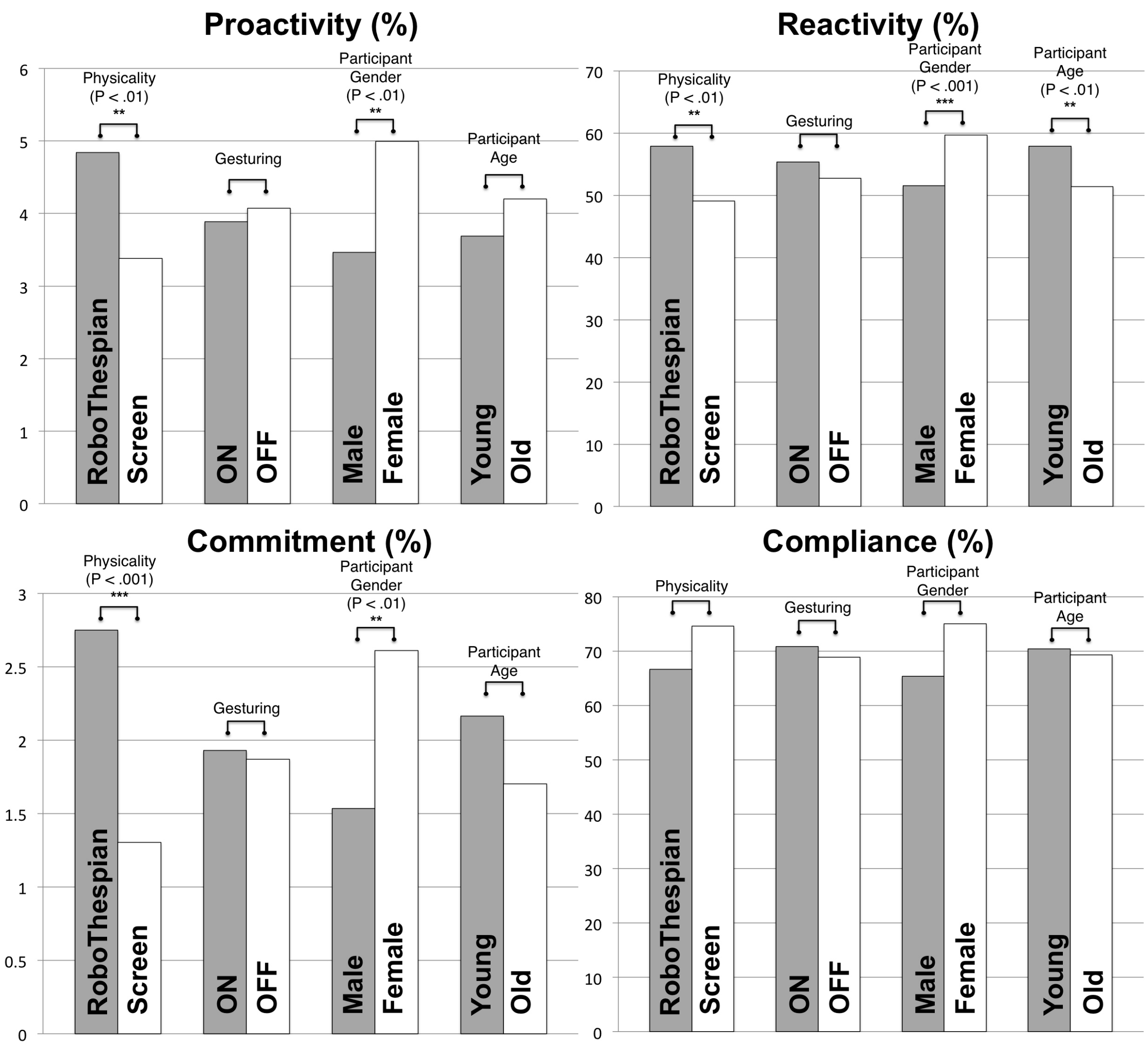


Figure 4.7: Analysis results. Two proportions comparisons via a large sample approximation in the dependent variables: Proactivity, Reactivity, Commitment, and Compliance. There are statistically significant differences in Proactivity, Reactivity, and Commitment, for physicality variations and for gender difference. Interestingly, the younger people tended to respond to the surrogate more easily than the older people with statistical significance.

Table 4.2: Distribution of the observed people collected from 40 two-hour experimental sessions.

Total Number of People	Gender		Sum
Experimental Group	Male	Female	
Group I (RoboThespian with Gesturing)	893	463	1356
Group II (RoboThespian without Gesturing)	1191	615	1806
Group III (Screen with Gesturing)	1602	773	2375
Group IV (Screen without Gesturing)	1396	752	2148
Sum	5082	2603	7685

Total Number of People	Age					Sum
	Young			Old		
Experimental Group	Children	Young Adults	Adulthood	Middle Age	Older People	
Group I (RoboThespian with Gesturing)	0	38	526	733	59	1356
Group II (RoboThespian without Gesturing)	9	29	650	1040	78	1806
Group III (Screen with Gesturing)	5	69	914	1291	96	2375
Group IV (Screen without Gesturing)	141	66	834	1026	81	2148
Sum	155	202	2924	4090	314	7685

Total Number of People	Age					Sum
	Young			Old		
Gender	Children	Young Adults	Adulthood	Middle Age	Older People	
Male	66	131	1875	2732	278	5082
Female	89	71	1049	1358	36	2603
Sum	155	202	2924	4090	314	7685

To limit potential coding biases, the coding criteria were developed together among the coders, and then each practiced and tested their individual coding against the reference coder. For example, the coders only evaluated people who were visually present in the video clips and considered separate people if they were in different video clips (i.e., different interaction segments). The coders cross-checked their initial codings of one day of video clips with the reference coder's, and confirmed that more than ninety percent and eighty percent of their evaluations were consistent with the

reference in the codings of the gender and the age, respectively. Also, the coding tasks were evenly distributed between the five coders in terms of the surrogate's physicality and gesturing conditions, so that the coders' variation could be minimized.

For the aforementioned complexity of the uncontrolled experiment, it is also difficult to arrive at an exact number of *unique* people who interacted with the surrogate, and which interactions might have been *repeated* interactions by the same people, potentially over different sessions or conditions. Because of this, the people observed in the experiment could be partially dependent among the experimental groups. Due to this unique characteristic of the data, a large sample approximation of two correlated proportions was used for comparison considering the dependent samples in most variables. For the method, a modified formula was used for an estimate of the variance as in Equation (4.2)—the last covariance term reflects the dependency of the samples, and evaluate it with a two-tailed z table (significance level $\alpha = .05$).

$$\hat{p}_1 = \frac{m_1}{n_1}, \quad \hat{p}_2 = \frac{m_2}{n_2} \tag{4.1}$$

$$V(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{2\hat{p}_1\hat{p}_2}{\sqrt{n_1 n_2}} \tag{4.2}$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{V(\hat{p}_1 - \hat{p}_2)}} \tag{4.3}$$

where n_i and m_i are the numbers of *observed* people in the group of interest according to the dependent variables (refer to Section 4.4.2). The last term of Equation (4.2)—the covariance part—reflects the sample dependency. If the samples are independent, this term would be eliminated. Also, if the samples are fully dependent (i.e., repeated measure), the n_1 and n_2 would be the same, but in our case the numbers of people that were evaluated are different without a way to cover this partially dependent situation. Hence, the formula for dependent samples was used with different sample sizes, n_1 and n_2. Note that the covariance part was ignored for comparison of two independent proportions, for example for gender/age comparisons. For the photo proximity that involves only independent samples, Chi-squared tests were employed.

Additional data from the Kinect sensor were also refined and analyzed, including skeleton data and depth images, for the observed people's motion dynamics such as two-handed movements in the lobby. However, the analysis for those are not included here because the Kinect data was not accurate enough to reliably conclude any results on the dynamics of human–surrogate interactions in the experiment, and any significant difference in the motion dynamics data was not found.

4.5.1 Proactivity and Reactivity

First, the people's *proactivity* and *reactivity* were evaluated. For these variables, the numbers of verbal *initiations* and *responses* were counted. The proactivity is defined as the number of people who initiate a verbal conversation divided by the total number of people (see Table 4.2–4.3 and Figure 4.7). The reactivity is calculated as the number of people who respond to the surrogate's verbal initiation, divided by the total number of the surrogate's verbal initiations (see Table 4.4–4.5, and Figure 4.7).

As analysis results, the proactivity shows statistically significant differences in the physicality comparison between the Group RoboThespian and the Group Screen ($Z = 3.06$, $p = .002$), and also in the observed people's gender comparison between the Male group and the Female group ($Z = -3.07$, $p = .002$) (see Table 4.1 and Figure 4.7). However, it does not present any significant differences with respect to the surrogate's gesturing variations and the observed people's age variations. Based on this result, it appears the people in the Group RoboThespian tended to initiate a verbal interaction more voluntarily/proactively before the surrogate started the conversation than the Group Screen, and females seemed to be more proactive than males.

Similarly for the reactivity, although there does not appear to be a statistically significant difference between the gesturing variations, there do appear to be statistically significant differences in the physicality variations, the observed people's gender and age. With respect to the physicality

variations: the Group RoboThespian and the Group Screen, there are significant differences in the reactivity ($Z = 2.76$, $p = .006$), in the age comparison ($Z = 3.00$, $p = .003$), and in the gender comparison ($Z = -3.57$, $p < .001$). Given the result, it appears that the people in the Group RoboThespian tended to be more likely to respond to the surrogate's verbal initiation compared to the Group Screen while the Group Young tended to be more reactive than the Group Old. Besides, the Group Female seemed to be more reactive than the Group Male.

Table 4.3: The numbers of human initiations—people who initiated a verbal conversation to surrogate. Human observer's tag was "Human to Surrogate".

Human to Surrogate (Human Initiation)	Gender		Sum
Experimental Group	Male	Female	
Group I (RoboThespian with Gesturing)	51	26	77
Group II (RoboThespian without Gesturing)	41	35	76
Group III (Screen with Gesturing)	43	25	68
Group IV (Screen without Gesturing)	41	44	85
Sum	176	130	306

Human to Surrogate (Human Initiation)	Age					Sum
	Young			Old		
Experimental Group	Children	Young Adults	Adulthood	Middle Age	Older People	
Group I (RoboThespian with Gesturing)	0	2	31	34	10	77
Group II (RoboThespian without Gesturing)	1	0	26	34	15	76
Group III (Screen with Gesturing)	0	0	21	35	12	68
Group IV (Screen without Gesturing)	19	0	21	35	10	85
Sum	20	2	99	138	47	306

Human to Surrogate (Human Initiation)	Age					Sum
	Young			Old		
Gender	Children	Young Adults	Adulthood	Middle Age	Older People	
Male	9	1	53	89	24	176
Female	11	1	46	49	23	130
Sum	20	2	99	138	47	306

Table 4.4: The numbers of surrogate initiations—surrogate's verbal initiations toward the people in the lobby. Human observer's tag was "Surrogate to Human".

Surrogate to Human (Surrogate Initiation)	Gender		Sum
Experimental Group	Male	Female	
Group I (RoboThespian with Gesturing)	491	252	743
Group II (RoboThespian without Gesturing)	348	137	485
Group III (Screen with Gesturing)	269	139	408
Group IV (Screen without Gesturing)	347	157	504
Sum	1455	685	2140

Surrogate to Human (Surrogate Initiation)	Age					Sum
	Young			Old		
Experimental Group	Children	Young Adults	Adulthood	Middle Age	Older People	
Group I (RoboThespian with Gesturing)	0	19	304	381	39	743
Group II (RoboThespian without Gesturing)	11	3	198	249	24	485
Group III (Screen with Gesturing)	2	16	124	259	7	408
Group IV (Screen without Gesturing)	15	24	196	247	22	504
Sum	28	62	822	1136	92	2140

Surrogate to Human (Surrogate Initiation)	Age					Sum
	Young			Old		
Gender	Children	Young Adults	Adulthood	Middle Age	Older People	
Male	16	38	537	778	86	1455
Female	12	24	285	358	6	685
Sum	28	62	822	1136	92	2140

Table 4.5: The numbers of human responses—people who responded to the surrogate's verbal initiations. Human observer's tag was "Human Response".

Human Response / Experimental Group	Gender: Male	Gender: Female	Sum
Group I (RoboThespian with Gesturing)	291	157	448
Group II (RoboThespian without Gesturing)	176	87	263
Group III (Screen with Gesturing)	116	73	189
Group IV (Screen without Gesturing)	167	92	259
Sum	750	409	1159

Human Response	Age					Sum
	Young			Old		
Experimental Group	Children	Young Adults	Adulthood	Middle Age	Older People	
Group I (RoboThespian with Gesturing)	0	12	190	216	30	448
Group II (RoboThespian without Gesturing)	4	2	119	124	14	263
Group III (Screen with Gesturing)	0	8	65	112	4	189
Group IV (Screen without Gesturing)	5	13	110	119	12	259
Sum	9	35	484	571	60	1159

Human Response	Age					Sum
	Young			Old		
Gender	Children	Young Adults	Adulthood	Middle Age	Older People	
Male	3	21	310	360	56	750
Female	6	14	174	211	4	409
Sum	9	35	484	571	60	1159

4.5.2 Commitment

It would be ideal if the exact time that the observed people spent with the surrogate could be measured as a numerical measure of their commitment; however, it was difficult to evaluate the actual interaction time from the recordings and associated data because it was impossible to reliably identify the exact moment when an interaction finished due to aforementioned intractable dynamics of interactions (see Section 4.5). People sometimes had multiple pauses in the middle of a conversation for various reasons (another person interrupted, ignoring the surrogate, etc.) and multiple people could jump into another person's interaction with the surrogate. Thus heuristics to decide when the interaction ended had to be developed. Also, the large number of interactions made it difficult to review all the videos for evaluating the interaction duration. One relatively reliable metric was if a conversation progressed long enough to reach the photo-taking request moment or not. Thus, here it is assumed that the conversation was a sufficiently long in those cases, and the ratio of the number of people who reached the photo-taking request over the total number of people who entered the lobby was analyzed.

The number of photo-taking requests for each group is shown in Table 4.6, and the ratios for the groups are in Table 4.1 and Figure 4.7. A large sample approximation analysis showed that there is a statistically significant difference in the ratio of photo-taking request between the Group RoboThespian and the Group Screen ($Z = 4.27$, $p < .001$), and there is also a significant difference between the Group Male and the Group Female ($Z = -3.02$, $p < .003$). However, no significant differences are found in the different gesturing groups and the age groups. The result suggests that the people for the Group RoboThespian more likely spent sufficient time interacting with the surrogate until they reached the photo-taking request moment, compared to the people for the Group Screen. Also, females seemed to reach the photo-taking request more easily than males.

Table 4.6: The numbers of commitments—people who maintained the verbal interaction with the surrogate until they reached the photo-taking request. Human observer's tag was "Photo-Taking Request".

Commitment (Photo-Taking Request)	Gender		Sum
Experimental Group	Male	Female	
Group I (RoboThespian with Gesturing)	25	18	43
Group II (RoboThespian without Gesturing)	23	21	44
Group III (Screen with Gesturing)	18	11	29
Group IV (Screen without Gesturing)	12	18	30
Sum	78	68	146

Commitment (Photo-Taking Request)	Age					Sum
	Young			Old		
Experimental Group	Children	Young Adults	Adulthood	Middle Age	Older People	
Group I (RoboThespian with Gesturing)	0	0	18	20	5	43
Group II (RoboThespian without Gesturing)	1	1	22	18	2	44
Group III (Screen with Gesturing)	0	4	11	14	0	29
Group IV (Screen without Gesturing)	7	1	6	10	6	30
Sum	8	6	57	62	13	146

Commitment (Photo-Taking Request)	Age					Sum
	Young			Old		
Gender	Children	Young Adults	Adulthood	Middle Age	Older People	
Male	1	5	30	33	9	78
Female	7	1	27	29	4	68
Sum	8	6	57	62	13	146

4.5.3 Compliance

There was a desire to check the people's compliance rate for the task of photo-taking among the surrogate conditions, and a higher compliance rate in the Group RoboThespian and the Group Gesturing-ON was expected than the counterpart groups. However, the results do not bear that

out; there does not appear to be any statistically significant difference among the experimental groups. The numbers of task compliance for the groups and the test results are shown in Table 4.7, Table 4.1, and Figure 4.7.

Table 4.7: The numbers of compliances—people who complied the photo-taking request and took a photo for the surrogate. Human observer's tag was "Photo-Taking Accept".

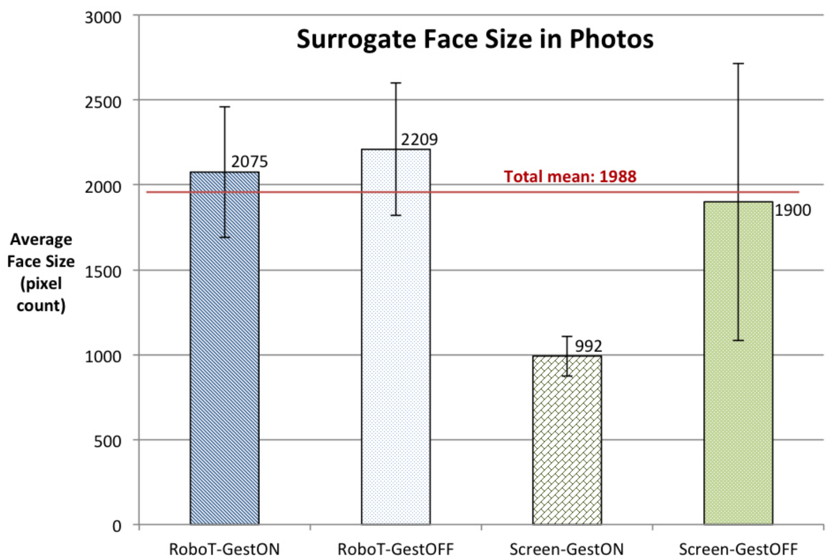
Photo-Taking Accept	Gender		Sum
Experimental Group	Male	Female	
Group I (RoboThespian with Gesturing)	14	15	29
Group II (RoboThespian without Gesturing)	16	13	29
Group III (Screen with Gesturing)	13	9	22
Group IV (Screen without Gesturing)	8	14	22
Sum	51	51	102

Photo-Taking Accept	Age					Sum
	Young			Old		
Experimental Group	Children	Young Adults	Adulthood	Middle Age	Older People	
Group I (RoboThespian with Gesturing)	0	0	12	13	4	29
Group II (RoboThespian without Gesturing)	1	1	12	13	2	29
Group III (Screen with Gesturing)	0	2	10	10	0	22
Group IV (Screen without Gesturing)	6	1	5	7	3	22
Sum	7	4	39	43	9	102

Photo-Taking Accept	Age					Sum
	Young			Old		
Gender	Children	Young Adults	Adulthood	Middle Age	Older People	
Male	1	3	20	20	7	51
Female	6	1	19	23	2	51
Sum	7	4	39	43	9	102

4.5.4 Photo Proximity

The photos taken by the people who interacted with the surrogate were also evaluated. By analyzing the size of the surrogate's face on the photos, a statistically significant difference has been found that the Group RoboThespian tended to take the photos more closely and to have a larger surrogate face on the photos, compared with the Group Screen. For a statistical analysis (Chi-squared tests), the average surrogate face size (1988 pixels out of 640×480) was calculated from the entire set of photos and the photos were separated into two groups: "Large" group having the surrogate face larger than the average size and "Small" group having the smaller surrogate face than the average size. The results from the Chi-Squared tests show that there is a significant difference in the surrogate's face size on the photos between the Group RoboThespian and the Group Screen, $\chi^2(1, N = 99) = 4.632$, $p = 0.031$, but not any significant difference between the gesturing variations (Figure 4.8).



RoboThespian vs. Screen

Contingency Tables

		Surrogate Face Size		
GROUP		Large	Small	Total
RoboThespian	Count	20.00	36.00	56.00
	Expected Count	15.27	40.73	56.00
Screen	Count	7.00	36.00	43.00
	Expected Count	11.73	31.27	43.00
Total	Count	27.00	72.00	99.00
	Expected Count	27.00	72.00	99.00

Chi-Squared Tests

	Value	df	p
X^2	4.632	1	0.031
X^2 Continuity correction	3.704	1	0.054
Likelihood ratio	4.815	1	0.028
N	99		

Effect Size

	Value
Contingency Coefficient	0.211
Phi-Coefficient	0.216
Cramer's V	0.216

Gesture ON vs. OFF

Contingency Tables

		Surrogate Face Size		
GROUP		Large	Small	Total
Gesture-OFF	Count	13.00	35.00	48.00
	Expected Count	13.09	34.91	48.00
Gesture-ON	Count	14.00	37.00	51.00
	Expected Count	13.91	37.09	51.00
Total	Count	27.00	72.00	99.00
	Expected Count	27.00	72.00	99.00

Chi-Squared Tests

	Value	df	p
X^2	0.002	1	0.967
X^2 Continuity correction	1.755e −31	1	1.000
Likelihood ratio	0.002	1	0.967
N	99		

Effect Size

	Value
Contingency Coefficient	0.004
Phi-Coefficient	0.004
Cramer's V	0.004

Figure 4.8: Top: A comparison of the surrogate's face size in the photographs (640×480) that people took (in pixels) among the experimental groups. The graph illustrates the means of surrogate's face size with standard error bars for each group. Bottom-Left: A Chi-Square test result in terms of the physicality. There is a statistically significant difference in the face size between the Group RoboThespian and the Group Screen, and the Group RoboThespian more likely had larger surrogate face in the photos, which means that they were close to the surrogate. Bottom-Right: A Chi-Square test result in terms of the gesturing. There is no statistically significant difference.

4.6 Discussion

Given the analysis above, there are several results that seem to support the importance of surrogate's physicality and gesturing in social influence as previous research suggested. Here the findings from the study were briefly summarized and discussed.

- **Proactivity** and **Reactivity**: ***Physicality Effect***—The people with the RoboThespian were more *proactive* to initiate a conversation with the surrogate and more *reactive* to the surrogate than the people with the video screen. If these proactivity and reactivity are correlated with engagement, our result for the higher proactivity and reactivity with the RoboThespian would support the idea that physical presence of the surrogate increases the user's engagement in the interaction, as other previous research suggested [152, 83]. Moreover, humans communicate with each other or another intelligent entity by (verbally) initiating or responding to them; thus, the higher proactivity and reactivity in the experiment could mean that the people treat the surrogate more as a human-like entity, which has enough intelligence to make a conversation. The novelty of the RoboThespian might also interest people in the lobby such that they want to interact with the surrogate.

  Gesturing Effect—Based on the statistical analysis, the gesturing effect in the proactivity and reactivity variables was not observed. The people who encountered the surrogate stood static at the beginning, so there might not be enough gesturing stimuli to attract/encourage the people to interact with the surrogate. That might be a reason why there was not any statistically significant difference among the gesturing variations. However, the RoboThespian with gesturing had the highest rate in both proactivity and reactivity variables. It could be considered that gesturing might be more important/influential when the surrogate has a physical manifestation. Gesturing is intuitively considered a cue for increasing attentiveness (e.g., waving arms to get attention from others) [121, 153], thus people might pay more at-

tention to the surrogate when the surrogate has a physical gesturing body compared to the body movements on a flat screen.

- **Commitment**: ***Physicality Effect***—the people with the RoboThespian were more likely to maintain the interaction until the moment for the photo-taking request compared to the people with the video screen. Although we assumed and evaluated the number of photo-taking requests over the total number of the people entering the lobby as a measure of commitment instead of the actual interaction duration, we believe it is a reasonable assumption because reaching the photo-taking request indicates that the people spent sufficient time with the surrogate. Based on the observation that there is a physicality effect in this commitment, but no significant effect of the gesturing, we suggest that the physical manifestation, which can give one the physical sense of co-presence with something/someone, might be more effective for attracting people to stay longer. The novelty of the RoboThespian might also play a role to encourage the longer interaction as well. This result agrees with the finding of longer interactions with an embodied robot from [154].

  Only 146 people maintained the interaction until the photo-taking request out of 7685 people. Even if we only consider the number of people who had conversations with the surrogate, i.e., the sum of human initiations ($N = 306$) and human responses ($N = 1159$), 146 might seem like a relatively small number; however, it is not clear that people would have spent more time with an actual human being in the lobby compared to the surrogate.

- **Compliance**: *No Effect*—the variations of the surrogate's physicality and gesturing did not appear to cause any changes in the people's compliance for the photo-taking task. Based on video reviews of the interactions between the people and the surrogate, there were several cases where the people rejected the photo-taking request. Some people said that they had to leave for their next meeting, but seemed to take the request seriously. Some others seemed to *not* take the request seriously. To increase our confidence regarding the lack of serious-

ness on the photo request with the surrogate, we could carry out an experiment to measure compliance when a *real* human asks; however, we could not answer this question with our current data.

- **Photo Proximity**: ***Physicality Effect***—the people with the RoboThespian took the photos closer to the surrogate, resulting in larger faces in the photos of the RobotThespian compared to photos taken by the people with the video screen. The representations of the surrogate's body and head in both the RoboThespian Group and the Screen Group had the same size in the real space. We see a couple of possible reasons for this closer photo-shoot with the RoboThespian. First, the people might have felt more comfortable with the human-like physical body of the RoboThespian; so, they came closer. Secondly, observing the photos, the people in the Screen Group tended to include the entire three wide displays on the photos, which resulted in a smaller surrogate face in the photo. If the physical display had the same size as the RoboThespian's physical body, people might have taken photos with a similar size of the surrogate's face both for the Screen and the RoboThespian settings. This could suggest that they perceived the physical manifestations of the surrogate (i.e., the body of the RoboThespian or the displays) as a target object for the photo-shoot, compared to the visual imagery of the surrogate on the display. This might reinforce the importance of the physicality in human perception indirectly.

- **Misc. Observations**: There are several situations that interest us related to the interaction between the surrogate and the people other than the variables addressed above. For example, there were several people trying to identify the surrogate's agency—whether it was controlled by a computer algorithm or a real human. One of them kept asking a lot of math questions, such as "what is one plus one?"—apparently to check the intelligence of the surrogate. This seems to support the notion that the agency of the surrogate could influence an people's perception of the surrogate, as Fox et al. presented [7]. Also, some people exhibited

impolite verbal behaviors with the surrogate. Both the agency-checking and impolite behaviors could imply that they were not aware of the existence of the human controller behind the scene (human-in-the-loop) and treated the surrogate as an autonomous agent during the interaction. This suggests that our intention to allow the surrogate to exhibit adaptability and intelligent interaction with the people by adapting the Wizard of Oz paradigm not only helped the smooth verbal interactions, but also kept the surrogate perceived as actual contemporary surrogate systems; thus, our findings could be potentially generalizable to actual human–surrogate interactions.

4.7 Summary

In summary, considering the objective human behavioral data from the study, the surrogate's physicality generally plays a more significant role in increasing engagement in the interaction with the surrogate, compared to any gesturing feature. On the other hand, gesturing may play an important role in attracting one's attention to the surrogate or a nearby object when it has physical manifestations.

The present study is motivated by broad interests in the effects of a human surrogate's physicality and gesturing during human–surrogate interactions, and an interest in "breaking free" from the confines of the typical laboratory-based controlled experiment. We were interested in a setting that reduced individual awareness of the experiment while simultaneously increasing the quantity of individuals and the diversity of their demographics. We were also interested in experimental measures that were both unobtrusive and objective. Given these motivations, and some practical considerations, we decided that the lobby of one of our university buildings would provide an ideal setting, so we instrumented the space with a variety of unnoticeable behavioral measures, and to collect data over a relatively long period of time (several months), for all conditions of interest.

Considering the large number of people ($N = 7685$) and the natural setting (a lobby), we decided against a hypothesis-based experiment, but instead approached this as an exploratory study. While measuring behaviors "in the wild" has advantages, e.g., the ability to observe natural interactions without experimental biases, the lack of control over the people and absence of explicit written questionnaires meant that it was challenging to tease out some of the interesting aspects. We defined several variables of interest related to human–surrogate interactions, and extracted measures from the (substantial) data collected during the interactions. Our measures included the people's conversational proactivity with and reactivity to the surrogate, their commitment to the interaction (based on the duration of the interaction), and task compliance. The results provided statistically significant support for positive effects of the surrogate's physicality related to the human social behavior, but we found no benefits of movements or gesturing of its body parts. This aligns with findings from previous research where people exhibited more favorable responses with physical surrogates than virtual surrogates, and were more engaged in the interaction with the physical surrogates, and supports the idea that the surrogate's physical presence with the human is the influential factor. Along these lines, we intend to evaluate the illusion of physicality via augmented reality displays in future work, to see if the effects can be replicated. Regarding the demographics of the people, we found higher overall engagement for females than males, and higher reactivity for younger than older people.

In short, this exploratory study revealed the importance of the surrogate's physicality in human behaviors, which can be also applied in real–virtual human interactions. That leads us to a question of how to increase a virtual human's (illusion of) physicality and social/co-presence in MR for stronger social influence to human users. In the following chapter, the proposed environmental physical–virtual interaction, which was introduced and described in Chapter 1 and Chapter 3, will be employed in real–virtual human interactions in MR environments, and its effects on human perception and behavior will be explored through several human subject experiments.

CHAPTER 5: SPATIAL AND BEHAVIORAL COHERENCE IN MR

In this chapter, we actively employ the proposed environmental physical–virtual interaction to maintain or enhance the spatial and behavioral coherence of VHs in MR, in turn to increase the perceived sense of social/co-presence with the VH (see Chapter 1 and Chapter 3). This chapter consists of three sections that describe three human subject experiments incorporating VHs in MR environments. The first experiment investigates the (negative) impact of a VH's physical–virtual conflict on human perception of the VH—e.g., the VH's virtual body conflicts with a space where physical objects are already occupying, which can cause the breaks in the perceived physicality of and social/co-presence with the VH. The second experiment involves a VH's proactive and plausible requesting behavior to overcome the conflict, and evaluates how such a behavior can influence human perception of and behavior with the VH. The last experiment employs a subtle but noticeable environmental physical–virtual airflow interaction in a shared MR environment where the VH exhibits awareness of the airflow, and examine the effects of the environmental interaction and the VH's awareness behavior in the perceived social/co-presence with the VH.

5.1 Virtual Human's Physical–Virtual Conflict

In MR, people can feel the illusion of VHs integrated into a real (physical) space. However, affordances of the real world and virtual contents might conflict, e.g., when the VHs and real objects "collide" by occupying the same space. This implausible conflict can cause a break in presence in real-virtual human interactions.

In this section we describe our experiment to investigate a sense of copresence with a VH while interacting with the VH in a shared AR space [31]. We varied the occurrence of visual conflicts

caused by the VH's disregard for the rules of physicality (the dual occupancy of the VH with physical objects).

5.1.1 Material

5.1.1.1 Physical Environment and Recordings

We furnished the experiment space with a chair, two tables, a shirt, a picture frame, and a coatrack (see Figure 5.1). The room with a size of 3.89 m by 3.89 m had two doors on its opposite sides, and the tables were placed across the middle of the room horizontally to the wall. The participants were instructed to sit on a chair where they could see the VH entering the room through one of the doors. The participant's behavior was captured by two webcams on the ceiling throughout the experiment, and we logged the position and orientation of the head-mounted display that the participant was wearing for examining the participant's movement trajectory in the laboratory room.

5.1.1.2 Virtual Humans and Human Controller

Two VHs ("Sarah" and "Katie" that have identical appearance but with different colors of shirts) were developed and employed for the experiment, and they could perform simple facial expressions, speech, and body gestures (see Figure 5.2). To reduce the potential side effects of erratic body movements, we positioned the VH in a virtual electric wheelchair, i.e., she appeared to be physically challenged, and never stood up during the experiment. The VH was displayed through a Microsoft HoloLens HMD, which was partially covered by a black polyether foam (see Figure 5.1). The reason for the foam was because of the HoloLens's narrow field of view—the VH's body could appear to be cropped at the edge of the small display when the participants were changing their view direction; this could possibly cause a severe distraction or break in presence for participants,

which we thus avoided.

The VH was remotely controlled by a researcher who triggered the VH's pre-defined speech and behavioral animations. Therefore, we implemented a client-server application communicating between the HoloLens and the control workstation wirelessly. The application was implemented based on the Unity3D engine. The VH's voice had a spatial audio effect; hence, participants could feel the localized sound coming from the VH in the shared AR space. Throughout the interaction, the VH maintained a neutral or slightly pleased facial expression.

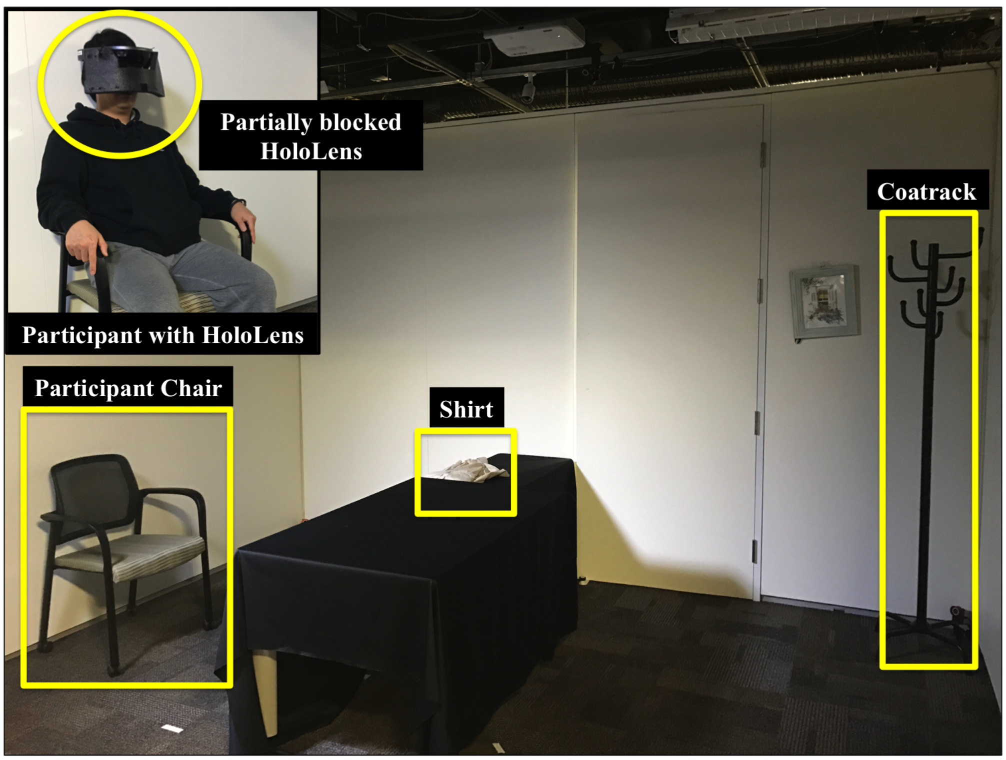


Figure 5.1: Experiment space and participant with the partially covered HoloLens HMD. Two tables in the middle of the room dividing half of the room, and the VH and the participant have a conversation across the tables. A coatrack is placed in the corner of the room next to a picture frame. A shirt is placed on one of the tables to investigate the participant's walking path around the VH towards the coatrack during the interaction.

5.1.2 Methods

We used a within-subjects design with two VH conditions (see Figure 5.2):

- **No Conflict (“NC”)**: Participants experience that the VH *avoids collisions with physical objects*, for example, entering a room through an open door and moving only where no physical objects are present.
- **Conflict (“CF”)**: Participants experience that the VH *passes through physical objects*, for example, entering a room by passing through a closed door and passing through physical tables while moving around in the room.

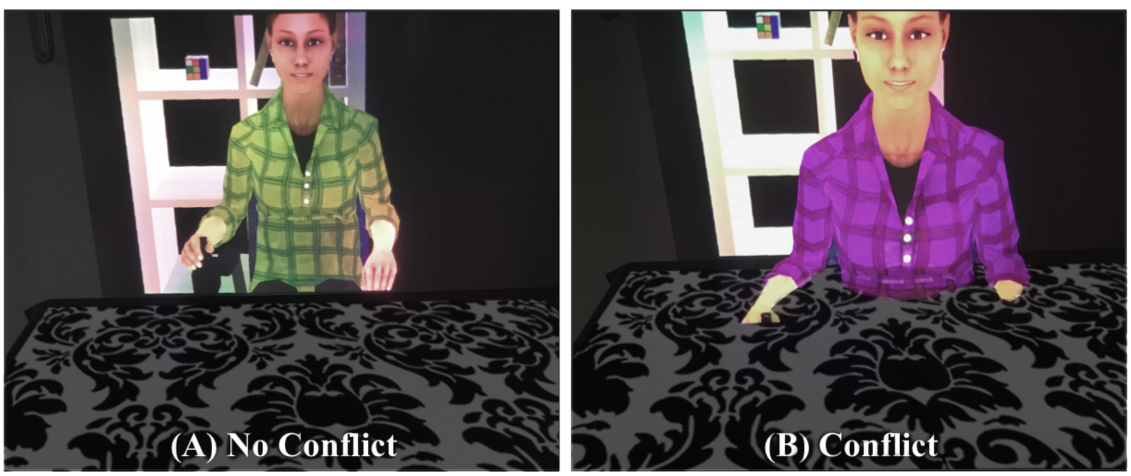


Figure 5.2: Study conditions: (A) the No Conflict (“NC”) condition where the VH avoids physical collisions and (B) the Conflict (“CF”) condition where the VH passes through physical objects.

Between these conditions we only varied occurrences of the VH’s spatial conflict with physical objects, i.e., the simultaneous occupancy of a space by both the VH and a physical object, but we maintained correct occlusions of the VH’s body by physical objects, as would naturally be supported by the rendering. For example, one could not see the VH if a physical object (such as

a wall or table) was in front of the VH. Participants experienced both of the conditions one after another in a counter-balanced order and their perceptions and behaviors were measured during and after each experience. Two (twin) VHs, “Sarah” and “Katie”, which had an identical appearance except for the color of their shirt, were used to help participants to perceive the VH as a different character in each condition. We counter-balanced which VH was used in which experimental condition. Based on the analogy to our experience in the real world, we developed the following hypothesis about participant’s perceived sense of copresence with the VH:

H-Co If a real human sees a virtual human follow the rules of physicality during interaction, it will lead to a higher self-reported sense of social presence and copresence than otherwise.

In order to provide a reasonably meaningful social interaction with the VH during the experiment, we prepared a brief verbal conversation, in which the VH asked four A/B type questions to assess the participant’s personality for each condition, and the participants answered the questions by choosing either the option A or B that best described their personality (Myer-Briggs Type Indicator [155]). During the conversation, the VH moved around the room following the experimental conditions described above. In the middle of interaction, the VH asked the participant to hang a physical shirt on a physical coatrack across the room, while the VH blocked the path to the coatrack. We observed how the participants behaved in this situation, i.e., whether they walked around the VH, and thus tried to avoid a visual conflict, or walked straight through the VH.

5.1.3 Participants and Procedure

We recruited 20 participants from our university community (ten males and ten females; age $M = 24.1$, $SD = 7.8$). The participants received a monetary compensation for their participation.

When participants arrived, the experimenter asked them to read an informed consent form and to fill out a demographics questionnaire. The experimenter measured their interpupillary distance (IPD), and configured the HoloLens appropriately. Next, they were guided to the experimental room and instructed to sit in a chair. They were informed that they were meeting twin VHs that had identical appearance one after another, and the VHs would ask a few A/B type questions related to the participant's personality. Once the participant wore the HoloLens, they had a couple of minutes to look around and make sure they saw virtual and physical objects in place—a virtual bookshelf, two real tables, a shirt, a coatrack, a photo frame on the wall (see Figure 5.1). The experimenter asked them to look toward the door so that the participant could observe the moment when the VH entered the room. When the experimenter left, depending on the experimental condition (see Section 5.1.2) the door was either closed or left opened, and the VH entered the room through the open door or passed through the closed door. While having a conversation to assess the participant's personality with the A/B type questions, the VH moved around the room, either avoiding physical collisions with the tables, or passing through the tables. Then, the VH asked the participant to put a shirt (on the table) on the coatrack in the corner of the room, while she placed herself in the path to the coatrack. Thus, participants had to decide whether they would avoid or pass through her. After the interaction, the participant was guided to leave the room through another door, which was not used by the VH, and completed a post-questionnaire. Once participants had completed the post-questionnaire, they were guided to back into the room for the interaction with the other VH in the other experimental condition. All the participants experienced both of the experimental conditions during the experiment. Finally, participants were debriefed about their perception of and behavior with the VH, and ended the study with receiving a monetary compensation. The total duration of the experiment per participant was approximately one hour including a brief discussion after the experiment. This experiment protocol was carried out with the approval of the Institutional Review Board at the University of Central Florida (IRB Number: SBE-15-11405 in Appendix A)

5.1.4 Measures and Hypotheses

Different subjective questionnaires have been introduced to measure social and copresence with VHs [124, 156, 69]. These questionnaires usually cover and combine multiple aspects together, such as a sense of copresence (i.e., being together in the same place), a degree of social connection (i.e., how closely they communicate/interact with each other), and a sense of realism (i.e., the VH's human-likeness). While such a combined questionnaire is beneficial when the goal is to measure the overall human perception of the VH, it does not normally emphasize the aspect of the VH's interaction with the surrounding physical environment and objects, which is important when trying to assess the sense of copresence with a VH in a shared AR space.

Thus, this experiment uses a questionnaire combining a few custom questions that we created with relevant questions extracted and modified from three existing social and copresence questionnaires [124, 156, 69], to measure the perceived sense of a VH's ability to sense the real world, realism, physicality/interaction in the physical space, and copresence. The questionnaire used for this experiment is shown in Appendix B.

5.1.4.1 Physicality and Interaction in Physical Space

We prepared a set of eight questions measuring the perceived sense of the VH's physicality, the degree of perception that the VH exists in the physical space, and its interaction with the environment. The interaction of a VH with the physical environment is an important characteristic which is specific to AR environments.

5.1.4.2 Copresence

We prepared five questions—including two questions extracted from Bailenson et al. [124], one from Basdogan et al. [156], and two own question—to measure the sense of copresence with the VH.

5.1.4.3 Sense

We thought participants could differently perceive the VH's ability to sense physical entities in the real world due to their observation of the VH's behavior passing through the surrounding physical objects. Thus, we prepared five questions evaluating the participant's perception of the VH's sensing ability via the five modalities *hear*, *smell*, *see*, *touch*, and *taste*.

5.1.4.4 Realism

We prepared seven questions on the participant's perception of the VH's realism, i.e., if it is perceived as a real human. We extracted several questions from existing questionnaires: one from Nowak [69], three from Bailenson et al. [124], and three from Basdogan et al. [156].

5.1.4.5 Godspeed Questionnaire

We also employed the "Godspeed" questionnaire from Bartneck et al. [157], which measures the four categories: anthropomorphism, animacy, likeability, and perceived intelligence. We expected that the responses for these categories would be generally more positive in the condition without a conflict.

5.1.5 *Results*

In this section we present the subjective responses and participants' avoidance behavior in the experiment.

5.1.6 *Subjective Responses*

We decided to use parametric statistical tests to analyze the Likert scale data [158] from the questionnaire, which has been shown to provide a valid method for the analysis of such ordinal data [159, 160]. We conducted paired samples t-tests at the $\alpha = .05$ significance level to compare the responses within the participants for all subjective measures (averaged 7-point Likert-style scores) in the questionnaires.

Table 5.1 and Figure 5.3 show the descriptive and inferential statistical results for "Sense", "Realism", "Physicality" and "Copresence" as main responses. Table 5.2 and Figure 5.4 show the results for the Godspeed questionnaire.

5.1.6.1 *Physicality and Interaction in Physical Space*

We found a significant difference between the conditions in the questions focusing on the VH's physicality and interaction with the surrounding physical space—the "NC" condition ($M = 3.956$, $SD = 1.218$) and the "CF" condition ($M = 3.319$, $SD = 1.363$); $t(19) = -3.524$, $p = .002$ (see Table 5.1 and Figure 5.3). The results indicate that the visual conflicts caused by the VH's dual occupancy with physical objects had a negative impact on its perceived physicality.

Table 5.1: Paired samples t-tests results and descriptives for the variables in the questionnaires (see Appendix B).

Main Responses (t-tests)	**t**	**df**	**p**	**Cohen's d**
Sense	-1.978	19	0.063	-0.442
Realism	-1.509	19	0.148	-0.338
Physicality	-3.524	19	**0.002****	-0.788
Copresence	-2.253	19	**0.036***	-0.504

Main Responses (descriptives)	**Group**	**N**	**Mean**	**SD**
Sense	NC	20	3.300	1.476
	CF	20	2.730	1.206
Realism	NC	20	3.457	1.457
	CF	20	3.064	1.083
Physicality	NC	20	3.956	1.218
	CF	20	3.319	1.363
Copresence	NC	20	4.708	1.223
	CF	20	4.133	1.169

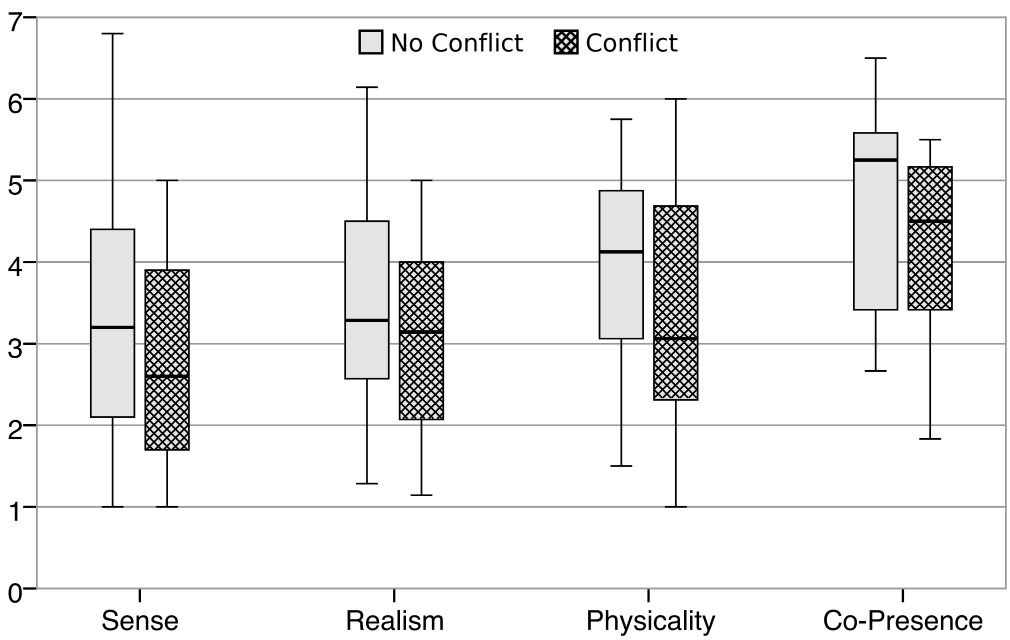


Figure 5.3: Box plots showing the results for the variables Sense, Realism, Physicality and Copresence.

5.1.6.2 Copresence

We found a significant difference in the participants' sense of copresence with the VH between the conditions—the "NC" condition ($M = 4.708$, $SD = 1.223$) and the "CF" condition ($M = 4.133$, $SD = 1.169$); $t(19) = -2.253$, $p = .036$ (see Table 5.1 and Figure 5.3). The estimated copresence was higher without visual conflicts.

5.1.6.3 Sense & Realism

Although there was no significant difference in the "Sense" and the "Realism" variables, we observed a trend of higher scores in the "NC" condition compared to the "CF" condition for both variables (see Table 5.1 and Figure 5.3).

5.1.6.4 Godspeed Questionnaire

For the "Godspeed" questionnaire, we found a significant difference in the participant's perceived animacy of the VH between the conditions—the "NC" condition ($M = 3.325$, $SD = 0.773$) and the "CF" condition ($M = 3.058$, $SD = 0.871$); $t(19) = -2.491$, $p = .022$ (see Table 5.2 and Figure 5.4). Moreover, we observed a trend for a difference in the perceived intelligence of the VH between the conditions. Both of these showed higher scores in the condition without conflict.

Table 5.2: Paired samples t-tests results and descriptives for the Godspeed questions.

Godspeed (t-tests)	**t**	**df**	**p**	**Cohen's d**
Anthropomorphism	-1.623	19	0.121	-0.363
Animacy	-2.491	19	**0.022***	-0.557
Likeability	-1.651	19	0.115	-0.369
Perceived Intelligence	-1.967	19	0.064	-0.440

Godspeed (descriptives)	**Group**	**N**	**Mean**	**SD**
Anthropomorphism	NC	20	2.960	0.886
	CF	20	2.600	1.032
Animacy	NC	20	3.325	0.773
	CF	20	3.058	0.871
Likeability	NC	20	4.410	0.568
	CF	20	4.250	0.808
Perceived Intelligence	NC	20	4.030	0.694
	CF	20	3.730	0.857

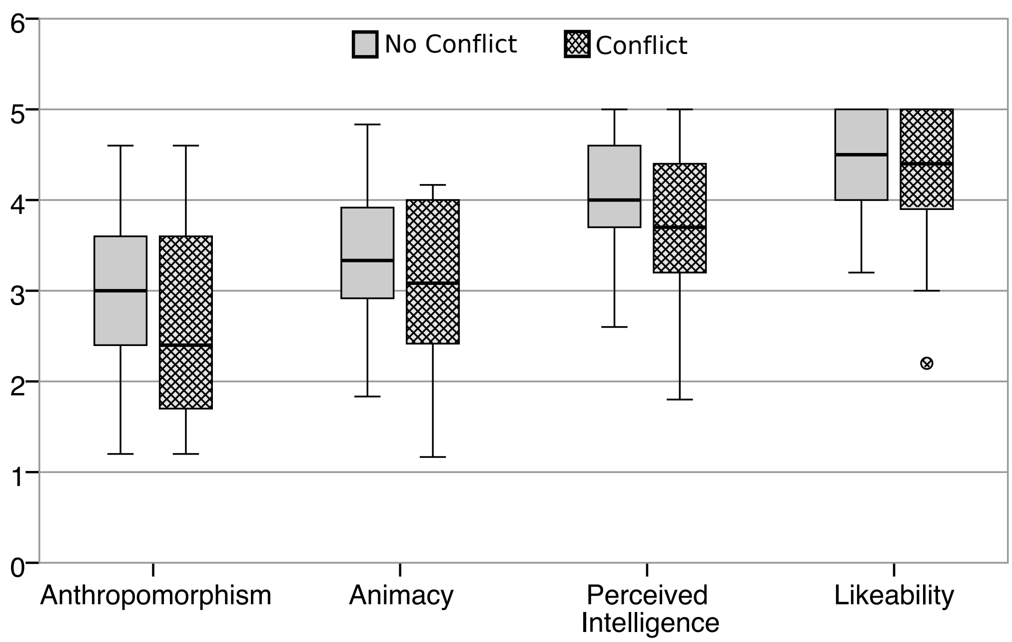
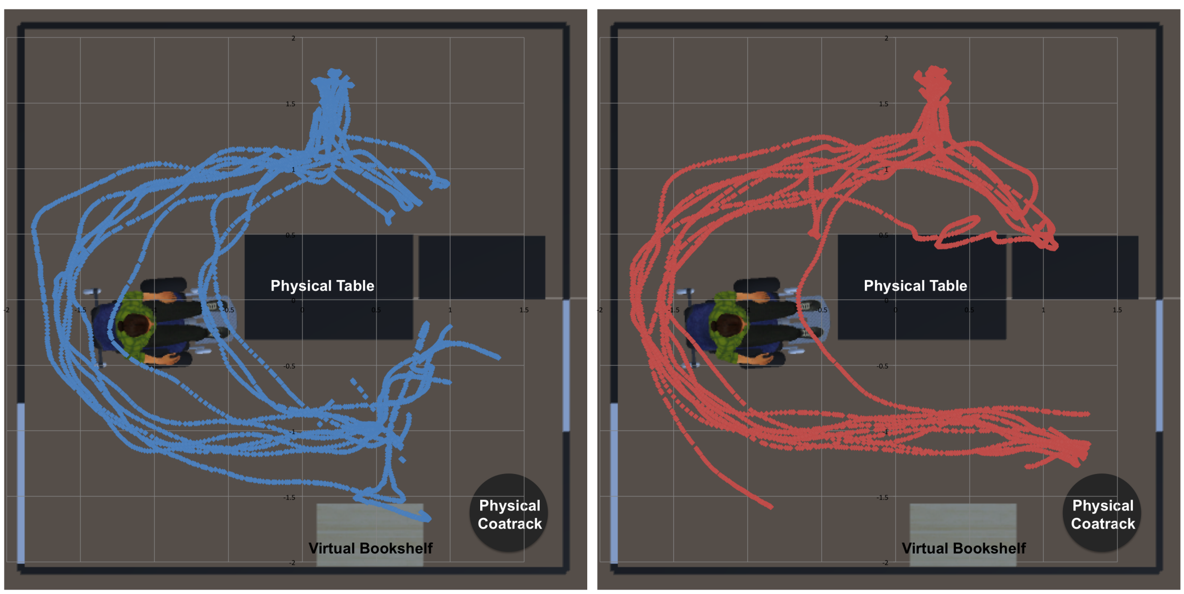


Figure 5.4: Box plots showing the results for the Godspeed questions in the two conditions.

5.1.7 *Avoidance Behavior*

We further examined the walking trajectory when participants were asked to move the shirt to the coatrack across the place where the VH was located (see Figure 5.1). Although participants were given this locomotion task twice due to the within-subjects design in our experiment, we only evaluated the first trial considering likely carryover effects between the first and second trial. Hence, we considered this as between-subject data based on the ten participants that started with the “NC” condition and the other ten participants that started with the “CF” condition. Figure 5.5 shows that more participants (4 out of 10) passed through the VH in the “NC” condition compared to those (1 out of 10) in the “CF” condition. We looked at the recorded videos during the experiment and we confirmed that these participants would have collided with the body and/or wheelchair of the VH if it had been real.



(A) No Conflict **(B) Conflict**

Figure 5.5: Trajectories of participants walking from one side of the room to the other: (A) blue lines indicate the “NC” condition and (B) red lines indicate the “CF” condition.

5.1.8 Discussion

The subjective responses provide clear support for our hypothesis *H-Co*. When the participants did not see the VH passing through physical objects, they indicated a significantly higher sense of *copresence*, i.e., a sense of being together in the same place with the VH. They also attributed a significantly higher sense of *physical existence* to the VH. Moreover, the Godspeed questionnaire indicated that they attributed a significantly higher *animacy*, i.e. a sense of being alive and interactive, to the VH in the experiment. Also, the questionnaire results suggest a trend that participants might attribute a higher *intelligence* to the VH, and a stronger belief that the VH could *sense* physical objects and events in the experiment.

The behavioral data, however, gives rise to different interpretations. Fewer participants passed through the VH after they had seen it passed through physical objects than otherwise. We would have expected that a higher sense of copresence would manifest itself via more natural locomotion behavior, i.e., avoiding collisions as is common among objects in the real world. However, our data might indicate a behavioral dynamics effect—users who have seen a conflict might be more sensitized to avoid conflicts, whereas users who have never seen a conflict might not. This can be proposed as a hypothesis to be tested for future experiments.

In summary, such dual occupancy conflicts are an interesting challenge of real–virtual human perception and action in AR, and we predict that it will likely remain a persistent issue over the next years. It is a difficult question how a technological solution could avoid such conflicts without limiting real or virtual humans in their freedom to move and act in such a shared AR space.

5.1.9 Summary

In this section, we investigated the effects of the real–virtual spatial conflicts that can arise during social interaction between real and virtual humans in a shared AR space. The visual conflict which we call "dual occupancy," is caused by a virtual human occupying the same space as a physical object, or conversely, a real human occupying the same space as a virtual object. While it is generally assumed that such conflicts should be avoided, it is not always possible to do so from a technological point of view, without restricting the real or virtual human's freedom to move or act. We described a human subject study in which we analyzed the effects of such conflicts on subjective estimates of copresence and perceived characteristics of the virtual human as well as the locomotion behavior of the participants. Our subjective responses support the premise that such conflicts reduce the sense of copresence and should be avoided if possible. However, our behavioral data suggests that avoiding such conflicts does not necessarily manifest itself in more natural locomotion behavior among the users. We even observed the opposite—fewer participants in our experiment caused collisions with the virtual human after they had witnessed it causing collisions. We propose that future research should focus on evaluating such dynamics in real–virtual human interactions, which likely cannot be fully explained by having a high or low subjective sense of copresence, but might rather depend on more complex learned sensorimotor contingencies in AR interactions.

5.2 Virtual Human's Spatial Coherence and Plausible Behavior

In previous section, we described an experiment that investigate the impact of incoherent physical–virtual conflict on the perceived copresence with a virtual human (VH) in real–virtual human interactions. For one of the experimental conditions in the study, the VH passed through physical objects revealing its lack of physicality and it turned out that participant's perceived copresence

and physicality of the VH were negatively influenced by the VH's conflicting behavior.

In this section, we present another experiment that adds a VH's socially plausible behavior to overcome the physical–virtual conflict on top of its spatial coherence with the physical environment [32]. For example, the VH requests help from the participant to keep the natural occlusion without spatial conflict with the real objects in the environment. For the experiment, both participant's subjective responses and objective behaviors are measured and analyzed, such as, social presence and perceived intelligence of the VH, and avoidance behavior with the VH, to investigate the effects of the VH's spatial coherence and plausible help-requesting behavior on human perception and behavior in the real–virtual human interaction.

5.2.1 Material

5.2.1.1 Virtual Human and Human Controller

We created a VH, called "Katie," that could perform facial expressions, speech, and body gestures. The VH was displayed through a Microsoft HoloLens HMD (see Figure 5.6). To preserve the plausibility of a request for help, we positioned the VH in a virtual electric wheelchair, i.e., she appeared to be physically challenged, and never stood up during the experiment (see Figure 5.7). The VH was remotely controlled by a real human behind the scene using a graphical user interface to trigger the VH's pre-defined speech and behavioral animations. Thus, we implemented a client-server application communicating between the HoloLens and the control workstation wirelessly. The VH's voice had a spatial audio effect; hence, participants could feel the localized sound from the VH. Throughout the interaction, the VH exhibited neutral or slightly pleasant facial expressions, and sometimes looked down at the paper on the table.

One problem we encountered with the HoloLens was its narrow field of view. Since the screen

on the device projecting the VH was small while the HoloLens allowed participants to see the real environment even in the periphery of their eyes, the VH's body would disappear when the participants were changing their view direction. This limitation of the current-state AR display hardware could possibly cause a severe distraction or break in presence for participants, in particular regarding the interaction with the VH in our scenario. Therefore, we partially covered the front of the HoloLens with a black polyether foam so that participants could concentrate on the VH in front of them while the peripheral view was reduced, i.e., minimizing cropping or disappearing errors (see Figure 5.6).

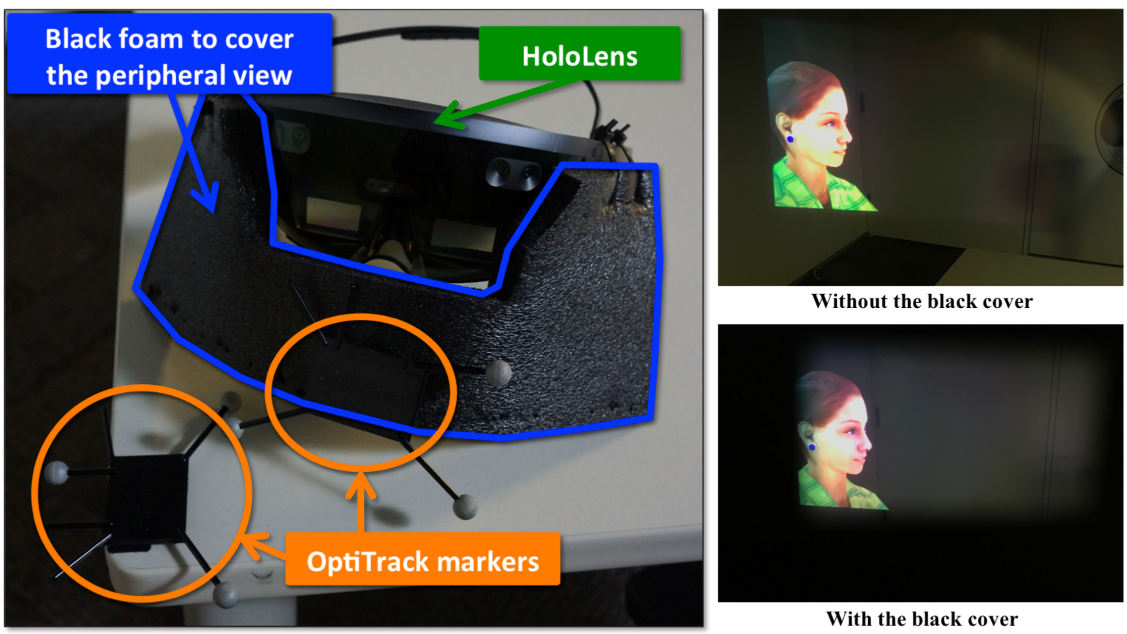


Figure 5.6: Microsoft HoloLens with the partially covered face. The participant's peripheral view was blocked by the black cover. The OptiTrack markers were used to track and record the pose of the participant.

5.2.1.2 *Physical Environment and Recordings*

We furnished the physical experimental room with a table, a box-like blocker, and two chairs (see Figure 5.7). The room had two doors on its opposite sides, and the table was in the middle of the room at a tilted angle (about 45 degrees) with chairs on opposite sides of it. The participant was instructed to sit on the chair close to the wall, after which the VH entered the room and moved with the wheelchair to the opposite side of the table. To log the participant's movement trajectory in the room, we used ten OptiTrack cameras and two markers (one on the HoloLens and the other on the table) (see Figure 5.6).

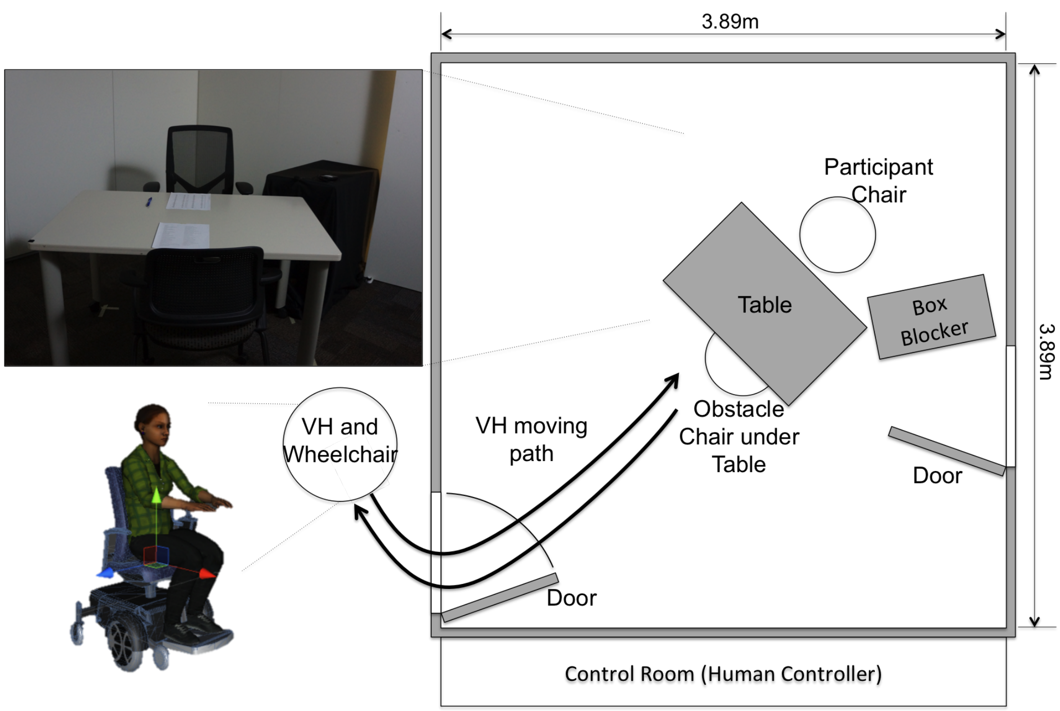


Figure 5.7: Experiment space. A table in the middle of the room, and the VH and the participant have a conversation across the table. A box blocker is placed next to the participant to investigate their walking path around the VH after the interaction.

5.2.1.3 Interaction Scenario

After sitting down on the chair at the table in the laboratory room, participants interacted with the VH using a form of question-answer conversation. The VH asked the participants twenty questions from the Myer-Briggs Type Indicator (MBTI) personality test [155]. A short version of the MBTI was used for this experiment[1]. Each question is an A/B type choice. The participants had to choose either A or B and let the VH know what they chose verbally while marking their answers on a sheet of paper on the table. The verbal interaction would be simple and relatively constrained so it is easy to control the VH's speech without harming the plausibility of the interaction. Participants' personality could be a factor to influence their perception of the VH, so we analyzed the personality of the participants and found an effect of introversion–extraversion dimension on social presence and gaze behavior. The findings were published in [161], but were not included in this dissertation.

5.2.2 Methods

To evaluate the effects of a VH's spatial and behavioral coherence with physical objects in AR, the experiment used a between-groups design with two different groups: the "Implausible Behavior" (IMP) group without spatial and behavioral coherence (Figure 5.8-left), and the "Plausible Behavior" (PLS) group with coherence (Figure 5.8-right) as described below. Participants were randomly assigned to one of the two groups and interacted with a VH in an AR environment.

- **Implausible Behavior (IMP)**:
    - **I1**: the VH ***passes through the door*** without opening it.

[1] https://www.quia.com/sv/522966.html (2017-01-17)

– **I2**: the VH does ***not ask for any help*** from the participants and does not avoid physical-virtual collisions.

– **I3**: the VH is ***not occluded*** by the physical objects.

- **Plausible Behavior (PLS)**:

  – **P1**: the experimenter ***opens the door*** when the VH enters/leaves the room.

  – **P2**: the VH ***asks the participant to move the chair*** out of the way for her to get to the table without an implausible physical-virtual collision.

  – **P3**: the VH is ***occluded*** by the physical objects.

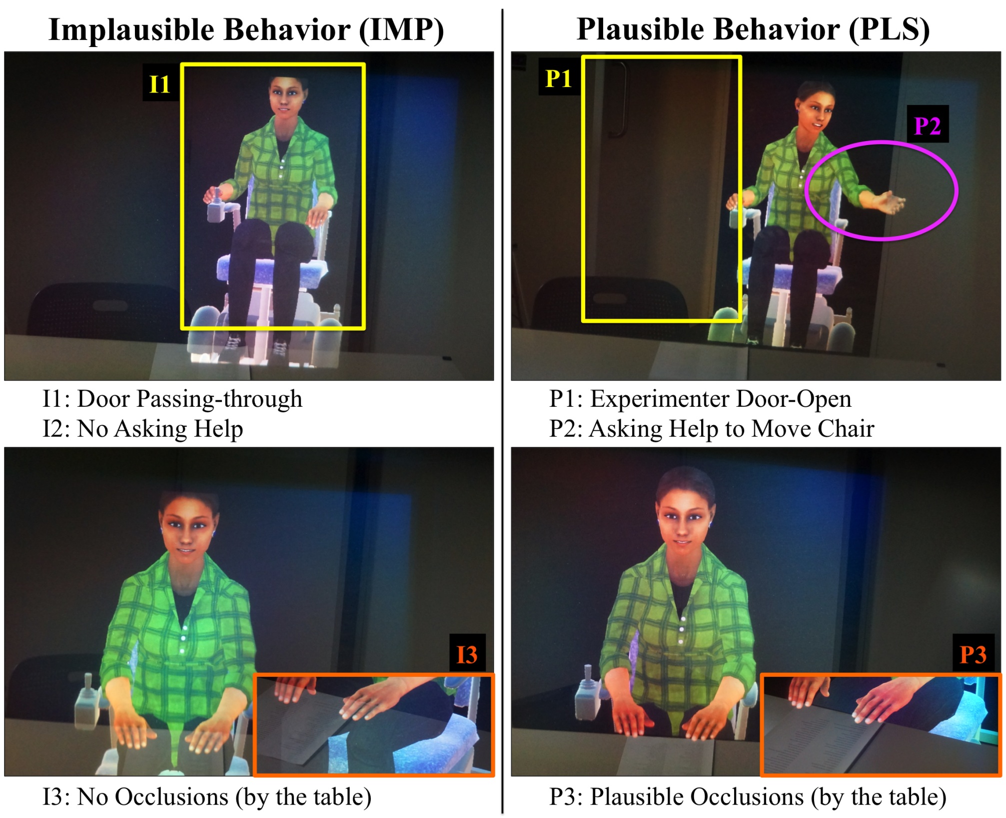


Figure 5.8: Study groups: the IMP group (left) and the PLS group (right).

5.2.3 Participants and Procedure

The subjective responses and behavioral data of 22 participants were used for the analysis (11 for the PLS group and 11 participants for the IMP group; 14 males and 8 females; age $M = 22.82$, $SD = 3.54$). All the participants received $15 USD for their participation in the end. The total duration of the experiment per participant was approximately one hour.

When participants arrived, the receptionist (i.e., one of the experimenters) guided them to the questionnaire area. They were asked to read the informed consent and fill out a demographics questionnaire. Next, they were guided to the experimental room and instructed to sit on the chair in the corner of the room while starting the video and audio recording. The receptionist explained that they would be wearing a head-mounted display (HMD), which was a Microsoft HoloLens. They were informed that they would have an interaction with a VH, and she would ask twenty questions from the MBTI personality test, which were A/B type binary questions. Once the participant wore the HoloLens, a human controller (i.e., another experimenter) controlled the VH using a graphical user interface behind the scene. The receptionist and the VH behaved accordingly for each of the two study group descriptions during the interaction (see Section 5.2.2). Once the participant completed the interaction with the VH, the receptionist guided them out of the room, and asked them to fill out a post-questionnaire.

After the post-questionnaire, the receptionist guided the participant back to the door of the experimental room and asked them to don the HoloLens once more while waiting in front of the door. Once the participant donned the HoloLens, the VH would be again visible at the table in the middle of the room (see Figure 5.7). The receptionist then instructed the participant to walk back to the chair where they had been sitting while answering the twenty questions, and we logged their walking trajectory using the OptiTrack system. Afterwards, the experimenters had a brief discussion with the participant about their perception of and behavior with the VH. This experiment protocol

was carried out with the approval of the Institutional Review Board at the University of Central Florida (IRB Number: SBE-15-11405 in Appendix A)

5.2.4 Measures and Hypotheses

Here, we describe multiple subjective questionnaires and behavioral measures that we included in the experiment.

5.2.4.1 Perceptions

Social Presence: We used the Social Presence (SP) questionnaire from Bailenson et al. [124], which is 7-point Likert scale. The questionnaire consists of five questions, covering the VH's authenticity and realism as well as the sense of "being together". We established the following hypothesis for SP based on the assumption that the plausible behaviors in this experiment might not only avoid the physical–virtual conflict, but also strengthen the social connection due to the spatial and behavioral coherence:

- **SP-H**: The level of SP will be higher in the PLS group than in the IMP group.

Godspeed Questionnaire: We also adapted the 5-point semantic differential scale "Godspeed" questionnaire from Bartneck et al. [157]. This questionnaire was originally introduced to measure the user's perception of robots during human–robot interaction; however, we see similarities between robots and virtual humans, and used it to assess the perception of the VH in the experiment. We used the four categories from the questionnaire: anthropomorphism, animacy, likeability, and perceived intelligence. We expected that the responses for these categories would be generally more positive for the PLS group than for the IMP group without any specific hypotheses.

5.2.4.2 *Avoidance Behavior*

The fact that participants walked around the VH could be an indication that they felt co-present with the VH in the shared AR space; thus, we were interested in whether the participants tried to avoid the VH and if so, whether there was any difference in their avoidance behavior among the study groups. We tracked the participant's walking path around the VH with two OptiTrack markers—one attached on the corner of the table and the other attached on the HoloLens and we logged the HoloLens' head pose (see Figure 5.6). We expected the VH's plausible behavior that we adjusted would influence the participant's avoidance behavior as well. Thus, we hypothesized:

- **AB-H**: The participants will more likely avoid the VH walking around it in the PLS group than in the IMP group.

5.2.5 *Results*

5.2.5.1 *Perceptions*

We conducted two-tailed independent-samples t-tests to compare the responses between the study groups ($\alpha = .05$), and confirmed the assumptions for the tests.

Social Presence: For Bailenson's SP, there was a statistically significant difference in the participants' SP responses for the IMP group ($M = 3.418$, $SE = 0.331$) and the PLS group ($M = 4.655$, $SE = 0.385$); $t(20) = -2.435$, $p = .024$ (Table 5.3, Figure 5.9). This suggests that the VH for the PLS group really does promote the participant's higher SP than the VH for the IMP group.

- **SP-H**: The result from the responses for the SP questionnaire supports the SP-H with a statistical significance.

Godspeed Questionnaire: For the "Godspeed" responses, there was a statistically significant difference in the participant's perceived intelligence of the VH among the study groups ($M = 3.591$, $SE = 0.191$ for the IMP group and $M = 4.145$, $SE = 0.130$ for the PLS group; $t(20) = -2.402$, $p = .026$) (Table 5.3, Figure 5.9). This suggests that the VH for the PLS group is perceived as more intelligent than the VH for the IMP group. Although there were no statistically significant differences for other variables, all had the trends of higher scores for the PLS group than for the IMP group as we expected.

- **Perceived Intelligence**: A statistically significant difference was found (the PLS group > the IMP group).

Table 5.3: Independent-samples t-tests results for Bailenson SP and Godspeed measures.

Social Presence	**t**	**df**	**p**	**Cohen's d**
SP (Bailenson)	-2.435	20	**0.024**	-1.038
Godspeed (t-tests)	**t**	**df**	**p**	**Cohen's d**
Anthropomorphism	-1.320	20	0.202	-0.563
Animacy	-1.753	20	0.095	-0.748
Likeability	-2.213	20	0.039*	-0.944
Perc. Intelligence	-2.402	20	**0.026**	-1.024

*Levene's test is significant ($p < .05$), suggesting a violation of the equal variance assumption

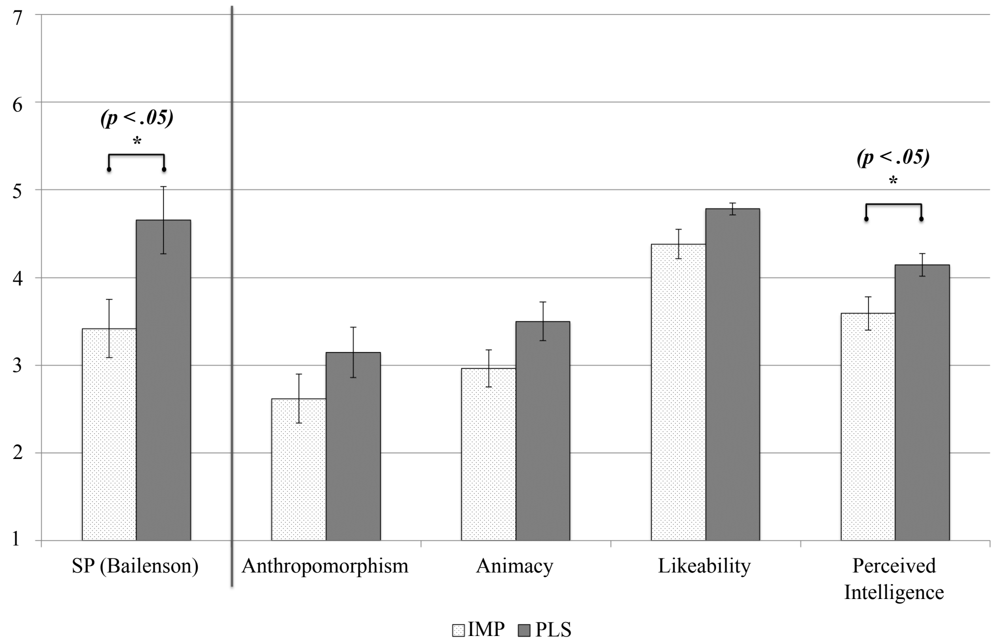


Figure 5.9: Mean comparison with standard errors for Bailenson SP (7-point Likert scale) and Godspeed measures (5-point semantic differential scale).

5.2.5.2 Avoidance Behavior

We observed that most of the participants avoided (walked around) the VH (see Figure 5.10). However, interestingly, three of the participants for the IMP group reported that they had walked through the VH. Although one of their walking trajectories was not directly passing through it, two trajectories completely ignored the VH and walked through it as shown in Figure 5.10—two red lines passing through the VH. The fact that we observed the cases ignoring the presence of the VH only for the IMP group was worth to think of the effect of our treatment in the user's avoidance behavior. This might suggest that the VH's spatial and behavioral coherence with the physical objects caused the VH to be perceived as more present in the physical space than if the VH behaved implausibly. Besides, the Chi-Squared tests showed a statistically significant difference when we used the count of the participants who reported that they ignored and passed through the VH (Table 5.4).

- **AB-H**: The observation of participants' walking path around the VH and the results from the Chi-Squared tests support the hypothesis.

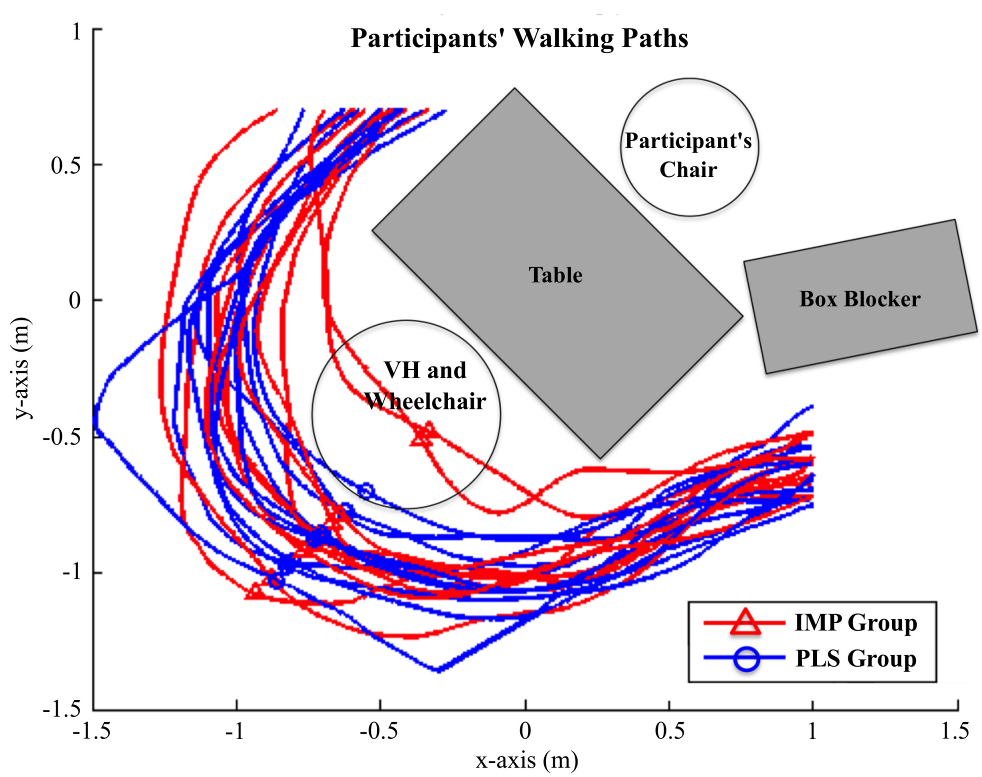


Figure 5.10: Participants' walking paths in the experimental room. Red lines for the IMP group and blue lines for the PLS group. Two red lines obviously ignore the VH and pass through it.

Table 5.4: Chi-Squared tests for avoidance behavior.

Group		Reported Path: Passed VH	Reported Path: Avoided VH	Total
IMP	Count	3.00	8.00	11.00
	Exp. Count	1.50	9.50	11.00
PLS	Count	0.00	11.00	11.00
	Exp. Count	1.50	9.50	11.00
Total	Count	3.00	19.00	22.00
	Exp. Count	3.00	19.00	22.00

Chi-Squared Tests	Value	df	p	Cramer's V
Likelihood ratio	4.635	1	**0.031**	0.397
N	22			

5.2.6 Discussion

In the following, we discuss the results with the participants' informal post-experience comments, and suggest some potential interpretations and implications. Also, we discuss some unknown factors and potential limitations of the approach.

Social Presence and Perceived Intelligence: The results of Bailenson's SP show that the PLS group rates significantly higher than the IMP group, which supports our hypothesis SP-H. Interestingly, we found a positive effect for the PLS group in the estimated intelligence with statistical significance among the "Godspeed" measures. The VH's plausible help-requesting behavior seemed to be estimated as more intelligent like a real human. Based on the comment below, participants seemed to be positively affected when their lower expectation about the VH's awareness of the physical environment is contradicted by the VH's acknowledgement of the chair in the room—a *positive expectancy violation*[2] [67].

> **Comment**: *"I didn't know that she could tell the chair was there, ... I think that's probably another reason why I decided to walk around her."*

Collision Avoidance with HoloLens: As previous research suggested avoidance behavior of users with virtual objects [126, 125, 162], most people in this experiment walked around the VH. Based on comments and the observation that most participants looked at the VH while they were walking around it, we interpreted that visual perception played an important role in the avoidance behavior. In other words, they saw something there, so they avoided the collision with it as in everyday human behavior, which might indicate a low-level human instinct or reflex, but might also be the result of a cognitive process after factoring in the nature of the VH. Some comments indicate that

[2] A positive expectancy violation is a favorable response due to the discrepancy between expected behaviors and enacted behaviors

participants thought that they avoided the VH because they wanted to be respectful, which suggests that they treated the VH like a real human or a social being.

Comment 1: *"I think it's probably just like second nature. I don't wanna run into her."*

Comment 2: *"I've seen something there. ... I didn't know what's gonna happen if I ran into it, so I walked around it."*

Comment 3: *"I guess I wanted to respect the image ... and then tried to get more of an interaction. I was trying to be respectful."*

Empty Space Avoidance without HoloLens: Surprisingly, two participants in the PLS group avoided the empty space where the VH was displayed via the HoloLens, even after they took off the HoloLens. They mentioned that it felt weird for them to occupy the space where they thought the VH was still there. This might be explained by the participants' experience that they saw the VH through the HoloLens in spite of their awareness of the VH's physical absence in the real world, which made them believe that there could be something even if they could not see it. Although the sample size is small, this interesting behavior for the PLS group might suggest that the VH's spatial and behavioral coherence make a more powerful impact on the human perception of the VH than a behavior ignoring the coherence in the physical–virtual space.

Comment: *"Feel weird walking where she was. Kind of think she's still there."*

Altruistic Behaviors: Interestingly we observed one participant (in the PLS group) who showed altruistic behavior by asking the VH whether it needed help to open the door when the VH was leaving out from the room. Although it was only one participant, his explanation below supports the assumption that the plausible behavior can stimulate such behavior.

Comment (Door-Opening): *"She asked me to move the chair out the way, so I thought she might need me to open the door, too."*

Another interesting observation with respect to altruistic behavior was about the chair moving. One of the participants (PLS group) said he felt a little guilty when the VH entered the room because he might have had to move the chair for it. Interestingly, this happened right after the experimenter opened the door for the VH but before the VH asked help from him. We could not conclude whether the observation of the experimenter's helping behavior triggered his feeling of obligation to help the VH or whether he would feel the same even if the VH had ignored the physical objects by passing through the door; this will need more investigation.

Comment (Chair-Moving): *"When she was entering the room, ... my mom raised me that I should have stood up and moved the chair for her. I felt a little guilty."*

Potential Limitations and Unknown Factors: Given our experimental choice to have a VH ask the participant to move the chair, the approach might seem limited to situations where a VH is inherently present. However, the general notion that virtual content should maintain plausible behaviors with respect to real objects is not limited to such a request-compliance interaction with a VH—it can be extended to a broader concept of intelligence/context-awareness of/response to the real scene by the virtual contents. Consider a virtual ball bouncing toward a real closed door; if there is a desire to maintain the illusion of a real ball, it should bounce off the door, not pass through it. If desired, the user will manipulate the physical environment as needed (e.g., open the door), which could further positively reinforce their sense of presence with the virtual object in a manner similar to the effect we saw in our experiment. The importance of context-appropriate virtual behaviors in AR is indicated in Microsoft's "Spatial Mapping" for HoloLens developers [163], and emphasized by Grubert et al. in their vision of pervasive context-awareness in AR [26].

In our study design, we deliberately combined multiple typically-occurring factors in AR related to the VH's plausible appearance and behaviors (natural occlusions, the observation of the experimenter's help to open the door, and the VH's request to move the chair). Our experiment elicited sufficiently strong effects that we believe it will be possible to design social presence experiments focusing on the contributions of individual factors in future work.

Finally, it could be interesting to examine the differences in human perception as one varies the apparent physical condition of the VH, i.e. comparing apparently able-bodied and apparently disabled VHs. We situated our VH in a wheelchair because assisting someone who appears to be physically challenged is a common (altruistic) human desire. If one made the VH seem able-bodied, participants might have had a negative reaction as they wonder why an apparently able-bodied VH should request such help.

5.2.7 Summary

VHs' ability to maintain the spatial and behavioral coherence with the surrounding physical environment in AR could be an important feature that supports the illusion of presence with VHs in the real world and their (social) plausibility. Given the results from the study in this section, we conclude that it is beneficial to have the VH's natural occlusions and proactive behavior asking help from the users to avoid implausible conflicts for higher social/co-presence with the VH in AR. Moreover, the results suggest that the coherence influences the user's behaviors avoiding the VH's space while walking around it. These findings would help to design realistic VHs in AR and certain applications dealing with VHs that require strong physical-virtual realism and interaction.

5.3 Environmental Physical–Virtual Coherence

This section describes an experiment that introduces subtle but coherent physical–virtual environmental interaction via airflow from a real fan in an AR environment [164, 33]. For example, the airflow from the real fan can make a virtual paper or curtains fluttering in the AR environment. The study examines whether the coherent airflow event in the combined physical–virtual environment can positively influence the perceived co-presence with the VH along with the VH's acknowledging behavior on the event.

5.3.1 Material

We implemented a female VH called "Sarah" for this experiment. The VH could speak with the participants and perform upper torso gestures (e.g., hand, arm, and head gestures). The VH was displayed via a Microsoft HoloLens HMD, which participants wore during the interaction with the VH. Participants and the VH were co-located in an office-like AR space as shown in Figure 5.11, giving the participants the impression of being seated at a table across from the VH. The physical table occluded the VH's lower body to maintain the visual plausibility. A physical rotating fan was placed next to the table in the middle of the two interlocutors, and oriented such that the airflow would occasionally blow in the direction of the virtual paper and curtains as the fan oscillated. A wind sensor (Modern Device Wind Sensor Rev. P[3]), hidden below the table (red circles in Figure 5.11), would detect the airflow from the fan, allowing the virtual paper and curtains to flutter according to the real wind for the experimental conditions. The sensor has a wide measure range of wind speed (0–150 MPH), which we used to trigger the virtual contents' fluttering animations, and there was no noticeable delay between the wind sensing and the animation triggering.

[3] `https://moderndevice.com/product/wind-sensor-rev-p` (Accessed 2018-10-04)

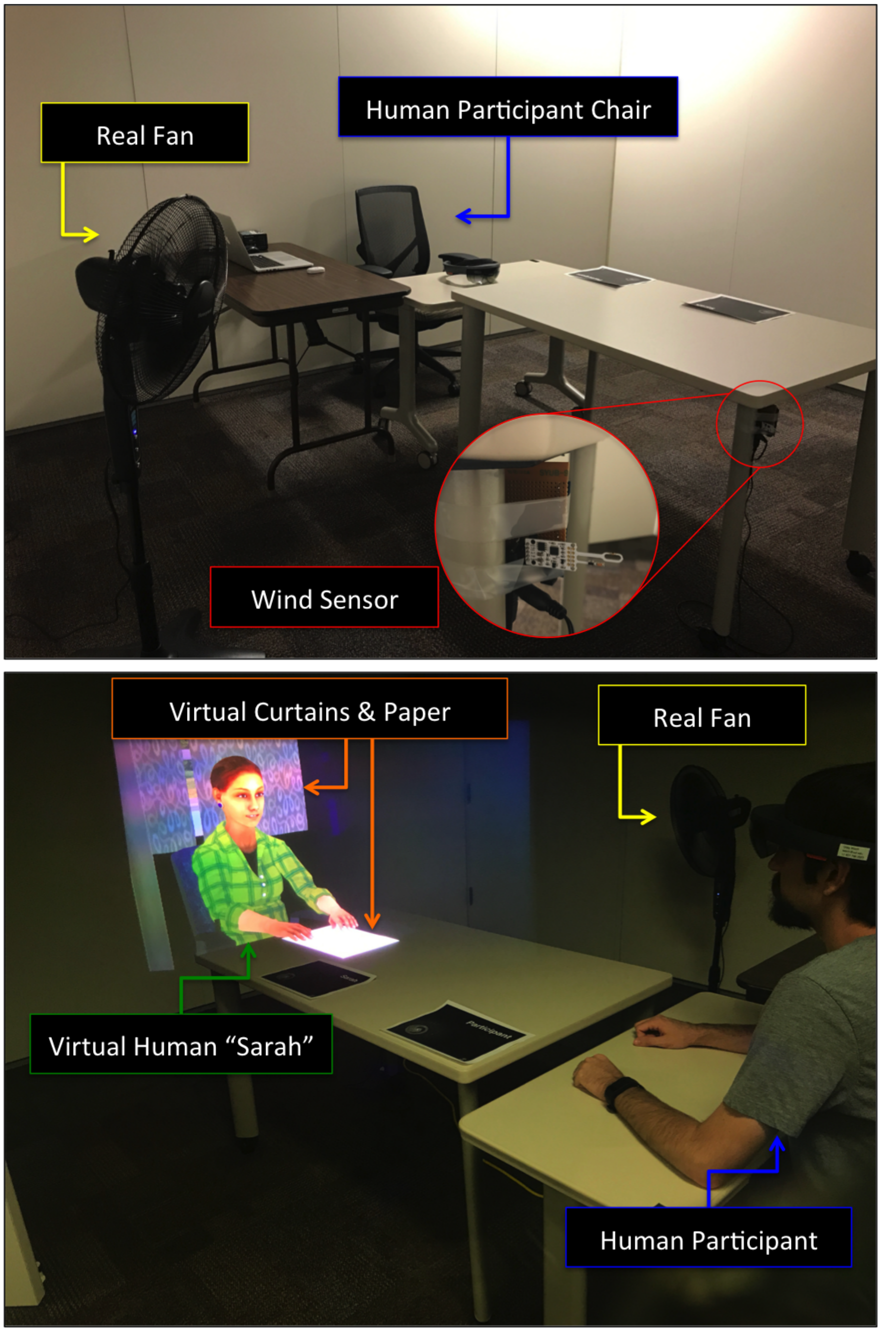


Figure 5.11: Experimental setup captured from two different camera angles. Participants were seated opposite from a virtual human on a physical table. A physical fan was placed on the side between the participant and virtual human, and a wind sensor was used to detect airflow that induced a state of fluttering in the virtual paper.

Hence, this approach could provide a higher fidelity and realism than with cruder setups, e.g., based on tracking the fan's pose alone. Cloth physics simulation in Unity3D was used to render the fluttering animations as natural as possible. The experimenter acted as a remote operator of the VH in a human-in-the-loop (i.e., Wizard-of-Oz) based experimental setup and triggered predefined verbal and nonverbal behaviors for the VH using a graphical user interface (GUI). The VH maintained a slightly pleasant facial expression throughout the interaction.

5.3.2 Methods

To investigate the effects of the physical–virtual interaction via airflow and the VH's awareness behavior, we wanted to give the participants a chance to directly compare how they felt about the VH in different experimental conditions. A within-subject design is the most effective approach to control for individual experience/gender/personality factors with respect to the interaction with the VH. Thus, we used a within-subjects design with three conditions below, which participants experienced in a counter-balanced order. A brief description of the conditions is shown in Figure 5.12.

- ***Control*** condition,
- *Physical–Virtual Influence (**PVI**)* condition, and
- *Environment-Aware Behavior (**EAB**)* condition.

In all conditions, the experiment consisted of a conversational interaction that participants had with the VH, which was based on simple and casual questions about personal preferences and experience in an AR environment. For example, the VH asked participants personal questions such as, "When is your birthday?" [4] Thirty questions were prepared and divided into three sets (ten questions per set), and each set was used for three conditions in the experiment. The question

[4] For the conversational interaction with the VH, thirty questions were extracted from `http://allysrandomage.blogspot.com/2007/06/101-random-questions.html` (Accessed 2018-10-04).

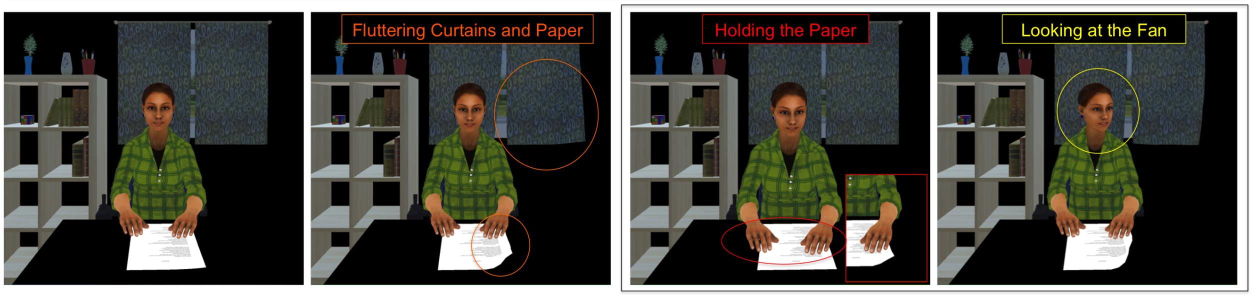
sets consisted of similar patterns of questions and the order of the sets were also considered to avoid the undesired effects by the question sets. The interaction between the participants and the VH was straightforward and did not have conversational dynamics. The experimenter simply triggered the VH's verbal and nonverbal behaviors via GUI buttons throughout the interaction with the participants, so the experimenter's influence should be minimized.

In the **Control** condition, a virtual paper on the table in front of the VH and virtual curtains behind it did not flutter and the VH never demonstrated any awareness of the physical fan although the fan was on and the real papers on the table were fluttering by the real wind.

In the **PVI** condition, the virtual paper and virtual curtains appeared to flutter as a result of the physical fan that was located on the side between the VH and participant. Participants could also see the real papers, which fluttered on the table, and compare the real and the virtual papers together (see Figure 5.11). The physical fan blowing the virtual objects was chosen as a subtle environmental event to strengthen the connection between physical and virtual spaces, and potentially influence the sense of copresence. We pursued to emphasize the inter-space connection by a different sensing modality other than the traditional visual and aural senses, which might exceed one's expectation for virtual content in a real environment. We were curious whether observing the fluttering virtual objects would have an impact on copresence, even when the participant was not directly involved in the fan-blowing event.

In the **EAB** condition, the VH would additionally occasionally exhibit attention toward the fan by looking at it or putting its hand on the virtual paper to stop the fluttering. The VH did not make any verbal acknowledgement about the fan wind. We chose the VH's gaze toward the fan because gaze has been considered as an informative cue to convey the direction of interest [165], so we wanted to express the VH's awareness of the fan and wind with its gaze in a subtle way together with the paper holding gesture.

Condition	Physical Fan	Virtual Curtain & Paper Fluttering	Virtual Human's Awareness Behavior (Holding the Paper & Looking at the Fan)
(A) Control	ON	NO	NO
(B) PVI	ON	YES	NO
(C) EAB	ON	YES	YES



(A) Control (B) PVI (C) EAB

Figure 5.12: Experimental conditions. (A) Control, (B) PVI (orange circles: fluttering virtual paper and curtains), and (C) EAB (red circle: holding the paper gesture, red rectangle: less fluttering after holding, yellow circle: looking at the fan).

5.3.3 Participants and Procedure

We recruited 18 participants (8 females and 10 males; age $M = 21.44$, $SD = 4.49$, range: 18–37) from our university community for the study. Seven of them had prior experience with VR/AR headsets, but the number of experiences was less than five times. The rest of them did not have any VR/AR headset experiences. All participants received a monetary compensation for their participation after the experiment (duration: ca. 40–50min).

Once participants arrived, they received an informed consent document and filled out a demographics questionnaire. We measured their interpupillary distance (IPD), which was applied for the HoloLens setting. In the within-subjects design, participants experienced the three experimental conditions in a counter-balanced order. We explained to participants that they would be interacting with a VH three times, and be asked to complete a post-questionnaire after each interaction to assess their sense of copresence with the VH. Participants initially saw virtual blinds placed between themselves and the VH at each time when the participants wore the HoloLens and they could see

and start interacting with the VH after the blinds moved up. In this way, we wanted to prevent the participants from feeling that the VH suddenly appeared when they donned the headset, which might influence their sense of copresence with the VH. During the interaction, the VH verbally asked participants ten casual questions on personal experience or preference as described above (see Section 5.3.2), and they verbally responded yes/no or brief answers to the questions. After experiencing each experimental condition, they were guided to complete a questionnaire measuring the level of perceived copresence with the VH. After all of the three conditions were completed, participants filled out a final post-questionnaire that evaluated the participant's preference among the three interactions with the VH and in which condition they felt the VH the most interactive, and had a brief interview with the experimenter to confirm their perception of the manipulations and collect their overall comments about the VH interactions. Finally, they received a monetary compensation for their participation and departed. This experiment was approved by the Institutional Review Board at the University of Central Florida (IRB Number: SBE-15-11405 in Appendix A)

5.3.4 Measures and Hypotheses

Different subjective questionnaires have been introduced to measure copresence (or social presence) with VHs (e.g., [124, 156]). These questionnaires usually cover and combine multiple aspects together, such as a sense of copresence (i.e., being together in the same place), a degree of social connection (i.e., how closely they communicate/interact with each other), and a sense of realism (i.e., the VH's human-likeness). While such a combined questionnaire is beneficial when the goal is to measure the overall human perception of the VH, we wanted to evaluate specifically the sense of copresence, which might be affected by our experimental manipulations, i.e., the physical–virtual influence by airflow and the VH's environmentally aware behavior. Thus, we prepared six questions relevant to the sense of "being (physically) together", extracting some of questions from existing questionnaires (see Table 5.5).

Table 5.5: Copresence questionnaire used in the experiment.

CP: Co-Presence (Sense of Being Together in the Same Place)
CP 1. I perceived that I was in the presence of the person in the room with me. (1: Strongly Disagree, 7: Strongly Agree)
CP 2. I felt the person was watching me and was aware of my presence. (1: Strongly Disagree, 7: Strongly Agree)
CP 3. I would feel startled if the person came closer to me. (1: Strongly Disagree, 7: Strongly Agree)
CP 4. To what extent did you have a sense of being with the person? (1: Not at all, 7: Very much)
CP 5. To what extent was this like you were in the same room with the person? (1: Not at all, 7: Very much)
∗CP 6-1. I felt I was in the ____ space. (1: Virtual, 7: Physical)
∗CP 6-2. I felt the person was in the ____ space. (1: Virtual, 7: Physical)
∗The absolute difference between CP 6-1 and CP 6-2 was used as a single value.

CP 1–3 were extracted from Bailenson et al. [124] and CP 4 was from Basdogan et al. [156]. We added three of our own questions, CP 5, CP 6-1, and CP 6-2. The absolute difference of CP 6-1 and CP 6-2 was calculated and used as a single value, which indicates that the participant and the VH are in the same place. In other words, the smaller absolute difference of CP 6-1 and CP 6-2 means that the participant felt more that he/she and the VH were in the same place somewhere in between the virtual space and the physical space. All questions used 7-point Likert scales, and we computed the averaged score as a representative score of copresence.

Among the three experimental conditions, we hypothesized that

- **H1**: the reported sense of copresence with the VH for the PVI condition would be higher than for the Control condition, and
- **H2**: the reported sense of copresence with the VH for the EAB would be even higher than for the PVI.

5.3.5 Results

For the analysis, we computed the averaged scores from the six questionnaire responses (see Table 5.5). The internal consistency of the six responses was high as shown by Cronbach's alpha ($\alpha = .716$). Considering sample size, dependency, and ordinal characteristics of the questionnaire responses, a non-parametric Friedman test was used for the analysis of the participants' responses on the copresence questions with a significance level at $\alpha = .05$. We found a significant main effect of the experimental conditions on the participants' estimated copresence, $\chi^2(2) = 7.300, p = .026$ (Table 5.6). Median (IQR) copresence levels for the Control, the PVI, and the EAB running trials were 3.25 (2.42 to 4.04), 3.67 (2.79 to 4.38), and 3.67 (2.67 to 4.29), respectively (see Figure 5.13).

Table 5.6: Friedman test results for copresence.

Friedman test				
Condition	Mean Rank	Median	N	18
Control	1.53	3.25	Chi-Square	7.300
PVI	2.19	3.67	df	2
EAB	2.28	3.67	Asymp. Sig.	**.026**

For the post-hoc analysis, Wilcoxon signed-rank tests were conducted. We found a significant difference between the Control and the EAB conditions ($Z = -1.988, p = .047$), while no significant differences were found between the Control and the PVI conditions ($Z = -1.309$, $p = .191$), and between the PVI and the EAB conditions ($Z = -0.094$, $p = .925$) (see Table 5.7 and Figure 5.13).

Table 5.7: Results from Wilcoxon signed-rank tests for copresence.

Wilcoxon signed-rank tests			
	PVI-Control	EAB-PVI	EAB-Control
Z	-1.309[a]	-.094[b]	-1.988[a]
Asymp. Sig.	.191	.925	**.047**

a. Based on negative ranks, b. Based on positive ranks.

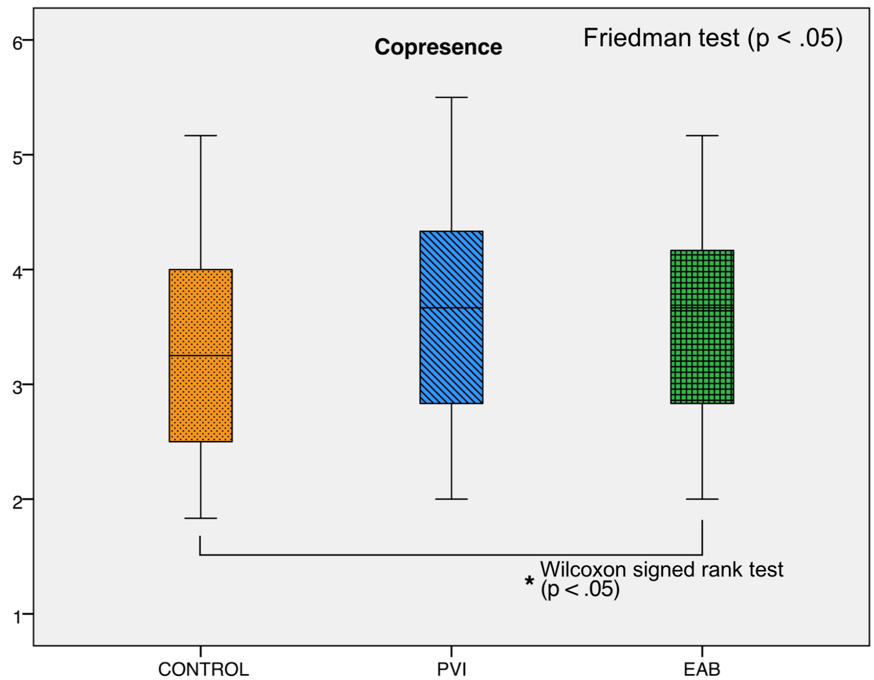


Figure 5.13: Copresence scores for the three experimental conditions. The PVI's median value was the highest followed by EAB and the Control condition.

This indicates that the sense of copresence was higher when the VH's environment-aware behavior is present along with the physical–virtual airflow interaction, compared to when those manipulations were absent. The magnitudes suggest a higher copresence for the PVI and the EAB conditions than the Control condition. Our original hypotheses H1 and H2 was not fully supported by the results, i.e., we did not see significant differences among all the conditions. However, our results partially support H2 by that participants felt higher sense of copresence when the VH exhibited awareness behaviors accompanied by the physical airflow affecting virtual objects.

After the participants experienced all three conditions, we asked them in which VH condition they felt that the VH was most interactive with the surrounding environment and also measured their preference among the conditions. The results show that the participants perceived the VH in the EAB condition as the most interactive with respect to the real environment, and the PVI

condition was preferred the most (see Figure 5.14). The Control condition was evaluated as the least interactive and the least preferred while there were a few participants who did not perceive a difference among the conditions.

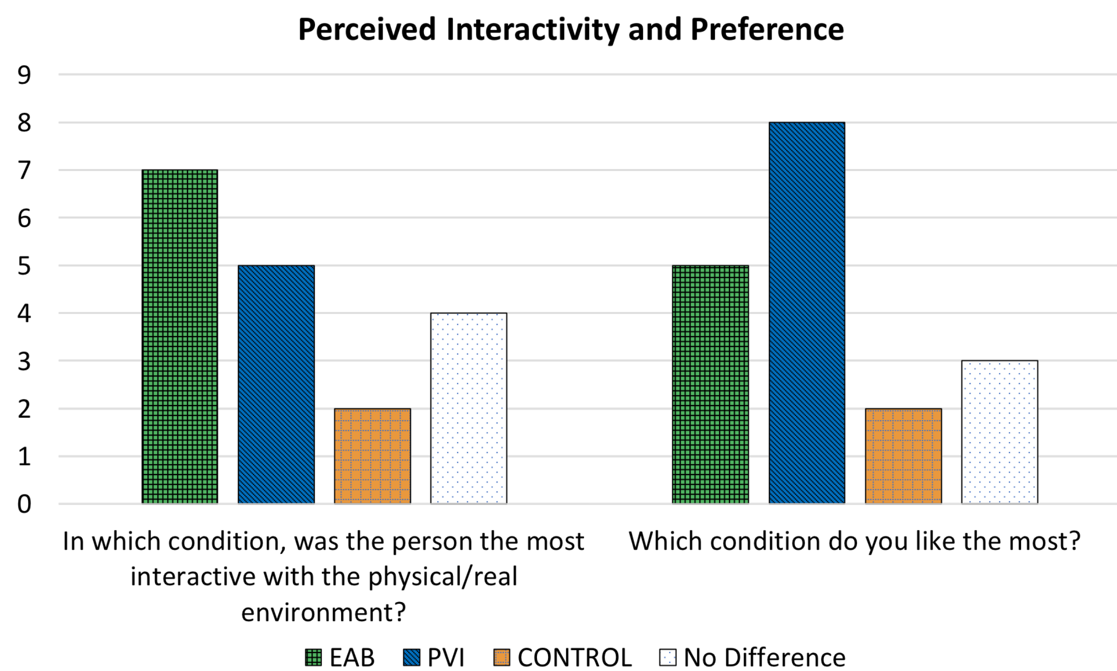


Figure 5.14: Perceived interaction and preference. Y-axis is the number of participants who chose the condition for the questions.

5.3.6 Discussion

Based on our results, we found a significant main effect on copresence by introducing airflow and VH's awareness behavior in a shared AR environment. Our finding suggests that peripheral environmental events, such as fan-blowing objects and observing them, impact one's sense of copresence with the VH that they interact with, and this could provide a useful reference for practitioners who want to increase the copresence level by physical–virtual environmental influences.

Our results suggest a higher copresence for the PVI and the EAB compared to the Control condition, particularly between the Control and the EAB conditions with statistical significance, which is also supported by our participants' informal comments after the experiment. Most participants indicated that they noticed the influence of physical airflow on the virtual paper and curtains, and the VH's awareness behaviors. Here are a few of the participants' comments that we collected in this experiment:

> **Comment 1**: *"It (airflow) made the environment feel more real. It definitely helped."*
>
> **Comment 2**: *"It (airflow) made me feel like I was really in the same room (with the VH)."*
>
> **Comment 3**: *"Oh, that's cool. It's almost like they were blending the physical world and the virtual world. ... I could see that (real) paper fluttering when her (virtual) paper fluttered on the desk. It seemed like a continuum."*

The post-hoc pair-wise analysis showed that the sense of copresence was significantly higher in the EAB condition compared to the Control condition. This indicates that the VH's awareness behaviors played a role in improving the sense of copresence on top of the physical–virtual airflow simulation.

It is further interesting to see that the participants seemed to have preferred the PVI condition over the EAB condition. This trend might be explained by the point that in the EAB condition the VH looked at the fan during the conversation, which could cause participants to feel as if their conversation partner is distracted by the environmental event and not paying a full attention to them. While the EAB condition helped to bridge the gap between the real and virtual spaces, it also made the VH's behavior more subject to interpretations of natural behavior in the real world.

As expected, observing the subtle airflow caused by a physical fan without active participation (or

involvement) was not quite as effective as the wobbly table experience in [166], which directly involved the participant's interaction with the shared environmental object, for example, participants could feel and trigger wobbly movements of the mediated physical–virtual table that the participants and the VH shared in an MR environment. Compared to the direct involvement of the human participants in the wobbly table movement, the fluttering virtual paper and airflow were not designed to be an integral part of the interaction between the participants and the VH in our experiment. This might also have made the VH's reactive nonverbal behaviors to the fan/paper less essential for the interaction and less influential to the participants. However, while it would be possible to create a similar level of involvement, e.g., by letting participants position the fan or using hand-held fans, it is encouraging to see that even our subtle indirect factors in this experiment had a significant effect on copresence.

Also, our results suggest that the influence by the subtle indirect physical–virtual interaction could be observed and compared more clearly when the physical–virtual events appear to be implausible and incoherent with the surrounding environment. In this sense, the statistical significance in the present study could be explained by the use of an AR HMD, which can increase the user's expectations related to the physical–virtual interaction, contrary to a projection screen displaying the VH in our previous study [167]. Regarding the coherency, we intentionally placed real paper on the table so that participants could compare the fluttering movement among the real paper and the virtual paper. Without the real paper, it is unlikely that we would have been able to show strong effects related to the virtual paper's behavior because paper can be static for other reasons, e.g., insufficient wind.

One general factor that might have limited the effect of the airflow and the VH's reactive awareness behavior on the perceived sense of copresence with the VH in this experiment could be related to the narrow field of view (FOV) of the HoloLens. Due to the narrow FOV, participants were not continuously able to see both the VH and the paper/fan while they were looking at objects in

the environment. Also, the VH's body could be cropped by the narrow FOV such that participants could see only a portion of the upper body of the VH, impacting the overall copresence level [168].

Our results are interesting in that we investigated the effects of a less researched modality, i.e., wind, which enables a subtle stimulus on the sense of copresence. We chose the wind modality because it has not been researched in depth in AR environments so far despite the fact that events caused by wind are common occurrences in our real life and potentially powerful in influencing one's perception of AR content. Our approach to reinforce the connectivity between the real and virtual worlds by using wind is not limited to copresence research with VHs, but could be employed in various AR applications.

5.3.7 Summary

System evaluation with perception studies involving human subjects has become a more common practice in the field of AR and intelligent virtual agents [169, 170]. In this section, we described a human subject study in which we analyzed the effects that physical–virtual connectivity and awareness behaviors can have on the sense of copresence with a VH in AR. We demonstrated that a VH's awareness behavior along with subtle environmental events related to airflow caused by a physical fan can lead to higher subjective estimates of copresence with the VH, which extends related research involving physical–virtual environmental influences.

Three experiments in this chapter showed that the proposed environmental physical–virtual interaction in VH systems could improve participants' perception of VHs, such as social/co-presence, and influence their behavior with the VHs in AR, by investigating the effects of VH's physical conflict, coherent and plausible behavior, and subtle physical–virtual environmental event.

CHAPTER 6: PHYSICALLY INTERACTIVE VISUAL EMBODIMENT AND IMPROVED TRUST

The phenomenon that people have an inherent tendency to treat computers or new media as if they are real people has been observed and researched extensively. Reeves and Nass [58] discussed that many people interact with computers in a "fundamentally social and natural" way. In this scope, commercial intelligent virtual agents/assistants (IVAs)—a type of VHs—that are able to verbally interact with users in a natural way, such as Amazon Alexa, highlight this phenomenon, and have become a social entity mimicking human intelligence. Along with a strong public interest in this technology, IVAs have been illustrated in science fiction media, such as the movie *Her* (2013)—a story about a man who falls in love with the disembodied voice of an IVA.

Most current-state IVAs mainly focus on voice commands and voice feedback, and lack the ability to provide non-verbal cues, which are an important part of human social interaction. AR has the potential to overcome this challenge by providing a visual embodiment for the IVA. A human-like visual representation could enrich the communicative channels that convey the agent's status and intentions by presenting gestures and motions as social behaviors. Riva [62] defined social presence as "the non-mediated perception of an enacting other (I can recognize his/ her intentions) within an external world," and stated that "the highest level of Social Presence is the identification of intentional congruence and attunement in other selves (the self and the other share the same intention)." In this manner, the visual embodiment and social behaviors of an IVA have much potential to increase the sense of social presence. For instance, Bente et al. [137] reported that embodied telepresent communication improved both social presence and interpersonal trust in remote collaboration settings with a high level of nonverbal activity. Similarly, we expect that appropriate social behaviors of an IVA could enhance the user's sense of rapport with the agent, and in turn,

the perceived confidence and trust in the agent could be improved. Moreover, an AR visual body and behaviors provide the opportunity to naturally convey the notion that the IVA is aware of the environment, e.g., by walking around obstacles, and can exert influence over physical objects, e.g., by interacting with Smart Home connected devices. Related to this, Kim and Welch [25] illustrated the importance of physical–virtual interaction in AR with VHs, particularly emphasizing the environmental awareness (sensing) and influence (affecting) with the surroundings.

On top of the findings from the previous chapter about the positive effects of VH's environmental physical–virtual interaction on human perception, this chapter further investigates the effects of the VH's (or IVA's) visual embodiment and social (locomotion) behavior interactive with the surrounding physical environment on the perceived social/co-presence and trustworthiness of the IVA [34].

We present a human subject study, in which we tested the hypothesis that visual embodiment and social behaviors increase the perceived social presence and confidence in an IVA in terms of awareness of and influence over the physical environment. Therefore, we designed an interactive scenario and prepared three different forms of IVAs, which differed in whether they had an AR visual body or were presented as a disembodied voice, as well as their social behaviors based on their ability to speak, gesture with their body, and move about the physical space. The results show strong support for our hypothesis, and we discuss implications and guidelines for future developments in the field of commercial IVAs.

6.1 Material

Our setup consisted of three forms of IVAs in a room-sized experimental space, which are described in the following sections.

6.1.1 Intelligent Virtual Agents

In this experiment, we implemented three forms of IVAs, which differed in embodiment and social behaviors: (A) an IVA with Speech only; (B) another IVA with Speech and Gesturing; (C) the other IVA with Speech, Gesturing, and Locomotion. All three IVAs were controlled by a human-in-the-loop framework—a human operator controls the IVAs behind the scene (i.e., Wizard of Oz paradigm). First, a 3D virtual character, which had a female human appearance, was modeled and animated in Autodesk Maya and Blender. The character was rigged and designed with animations for facial expressions, speaking, and body gestures. She had a mostly neutral, serious, and polite demeanor during the interaction (i.e., designed to not be too warm or cold toward the participant). We then imported the model into the Unity3D graphics engine where we added a graphical user interface allowing an operator to trigger specific body gestures or pre-recorded phrases with corresponding speaking animations. We pre-recorded the speech using a text-to-speech service[1], which provides a highly realistic synthetic voice. The differences between those three IVAs are described below. With this human-in-the-loop mechanism, the operator (who was seated outside the room, out of view of the participant, and observed the participant via live video and audio streams) pressed buttons behind the scenes to trigger the agent's responses. This allowed us to simulate natural communication between the real humans and the virtual agent without failure cases caused by the imperfect natural speech recognition, which are still too common in current-state IVAs.

(A) Based on our framework, we designed the IVA **S** to not have any visual feedback (i.e., visual body appearance) but only rely on voice communication (see Figure 6.1-A). The effect was that the IVA S was perceived as a disembodied voice, similar to a telephone call with a headset or the movie *Her* (2013). This corresponds to a popular paradigm in communication with IVAs, e.g., in the scope of Smart Home environments, which is characterized by users

[1] http://www.oddcast.com/home/demos/tts/tts_example.php

talking to the agent freely. The IVA S thus could not rely on embodied human gestures or locomotion to convey aspects of social interaction.

(B) The IVA **SG** was implemented based on the IVA S, and included visual embodiment with the animated 3D character described above. The form of the IVA SG was inspired by popular IVAs such as Amazon's Alexa, which are currently "embodied" with the body of a common home appliance, which is placed by the owner at a position in the room where the interaction with the agent seems most natural, such as in a corner of the room or next to a TV screen. In our study, our agent had a full human body but we kept the spirit of this use case intact. We designed the IVA SG to remain stationary in one place, and we used a range of upper-torso gestures as a form of communication. For instance, when asked to turn off a floor lamp in the room, she would take out a virtual tablet (such as commonly used in Smart Home environments), and pretend to control the lamp while looking at the tablet, see the light turned off, and put the tablet away again (see Figure 6.1-B). While this IVA remains stationary in the environment, it has the advantage that the user can rely on the fact that the IVA will always be present in the same place when they need her, i.e., she will not leave the room when completing an errand that the participant requests.

(C) The IVA **SGL** had pre-recorded animated behaviors including the ability to walk around in the experimental room, leave the room through an open door, come back, and use upper-torso gestures (e.g., hand and head gestures) as a form of communication. For instance, when the participant asked her to please turn off a floor lamp in the room, she would walk over to the lamp, touch the light switch with her hand, see the light turned off, and take a step back from the lamp (see Figure 6.1-C).

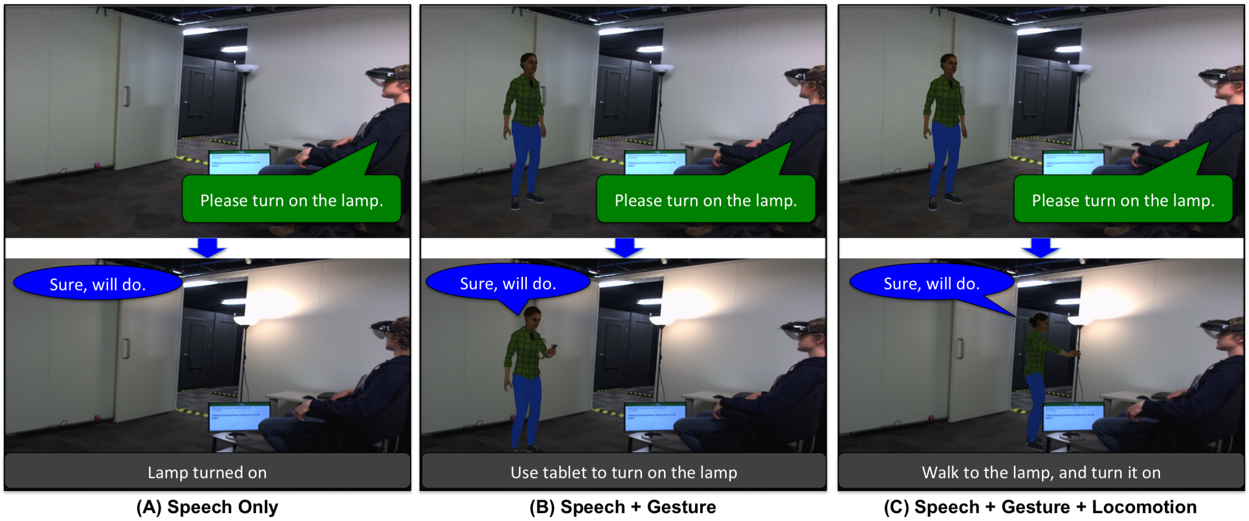


Figure 6.1: Illustration of the experimental conditions for the simple task to turn on a floor lamp (here equipped with an Internet of Things connected Philips light bulb). Participants interacted with three different types of intelligent virtual agents, (A) Speech only: a disembodied voice, (B) Speech+Gesture: an embodied AR agent that remained stationary in one place but could use upper-body gestures, or (C) Speech+Gesture+Locomotion: an embodied AR agent that could use gestures and naturally walk about the physical space.

6.1.2 Physical Setup

As illustrated in Figure 6.2, we used a physical room-like experiment space with an area of $3.89\,\text{m} \times 3.89\,\text{m}$ and an open ceiling where tracking and camera equipment were installed. This space comfortably fits common furniture such as a desk, a chair, and two (real or virtual) people. In this study, we had the following physical components in the room:

- A chair for the participants to rest on during the experiment;
- A floor lamp with a Philips LED IoT light bulb that could be turned on/off via WiFi and the Zigbee protocol;
- A sound bar placed near the back wall of the room for the IVA's voice output;
- A desk with a small monitor that was used to communicate contextual information and task

instructions to the participants;

- An open door to the right side, which lead out of the room into the laboratory.

We developed a custom bidirectional interface between the Unity3D application and the Philips IoT light bulb to control and manipulate the state of the floor lamp. A model of the physical room with this furniture was used as an occlusion layer in the Unity3D engine, such that the embodied agent could walk around in the room without visual conflicts caused by incorrect occlusion.

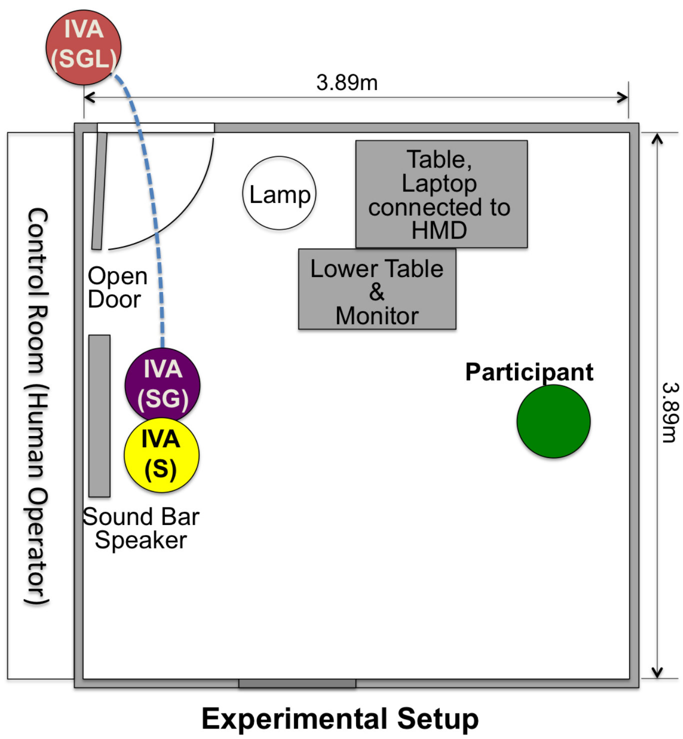


Figure 6.2: Experimental setup: Participants interacted with one of the IVA types: (S)peech only, (S)peech+(G)esture, or (S)peech+(G)esture+(L)ocomotion, across the experiment room. The embodied IVAs were located on the other side of the room facing toward the participant. The IVA with the visual embodiment in condition SGL could walk around freely and leave the room, while condition SG stays stationary in one spot. The IVA with the disembodied voice in condition S does not have a specific spatial location, but is included in this illustration for completeness sake. A few pieces of furniture and equipment for the IVA system were present in the experiment room. The human operator in our Wizard of Oz experimental design was located (unknown to the participant) on the opposite side of the back wall of the room.

Participants wore a Meta 2 head-mounted display (HMD) during the experiment. The Meta 2 optical see-through display provides a 90-degree horizontal and 50-degree vertical field of view with a 2,550 by 1,440 pixels resolution and 60 Hz refresh rate. The HMD offers positional SLAM-based tracking by fusing image features from the real world, captured with an on-board computer vision camera. However, instead of using this on-board tracking, we decided to use a NaturalPoint OptiTrack twelve-camera high-quality low-latency optical infrared tracking system to improve the quality of the tracking data. Therefore, we attached a six degrees-of-freedom rigid body target to the Meta 2 HMD and tracked it with sub-millimeter precision and accuracy.

We considered audio to be an important aspect of the experience, and thus placed a high-quality sound bar speaker (LG SH4 300W Sound Bar) near the back wall of the room as the means to provide speech feedback to the participants.

For rendering and system control, we used an Alienware laptop with Intel Core i7-7820HK CPU at 2.90 GHz, Nvidia GeForce GTX 1070 graphics card, and 16 GB of RAM.

6.2 Methods

In the experiment, we wanted to give the participants a chance to directly compare different forms of IVAs and capture how they were perceived while interacting with them. A within-subject design is the most effective approach to control for individual experience/personality factors with IVAs of the participants. Thus, we decided to use a full-factorial within-subject design for this experiment with three conditions, and the order effects by the repeated measures were minimized by our attempt to counterbalance the experience order as much as possible among the participants. Here we briefly describe the experimental conditions in which we manipulated the embodiment and social behavior of the IVAs in three levels (see Figure 6.3 and detailed descriptions in Section 6.1.1):

S The IVA S with a disembodied voice with social behavior limited to Speech.

SG The IVA SG with an AR body and social behavior limited to Speech and Gesturing while remaining at a fixed position in the room;

SGL The IVA SGL with an AR body and social behavior including Speech, Gesturing, and Locomotion;

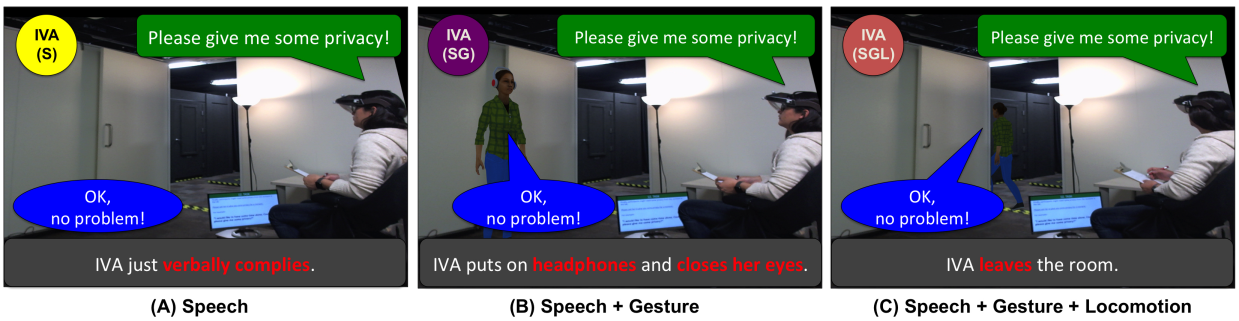


Figure 6.3: Illustration of the experimental conditions for the participant's request for some privacy: Participants wore a Meta 2 head-mounted display, which was tracked using twelve NaturalPoint OptiTrack cameras. The intelligent virtual agent was either present as (A) a disembodied voice, (B) an AR agent that remained stationary in one place, or (C) an AR agent that was walking about the space.

6.2.1 Interaction Scenario

The interaction scenario that the participants perform with the IVAs consisted of the following tasks, which revolved around the participant sending their assistant off to complete different tasks within the laboratory, while they themselves remaining seated in the experimental room.

The story began with a training task (T1), which was designed to show the participants that IVAs are capable of a cause-and-effect relationship. At the beginning of the session, the floor lamp in the room was turned off (see Section 6.1.1 and Figure 6.1). The participant's task was to ask the agent to turn on the lamp:

T1 Participant: *"Please turn on the floor lamp!"*

Agent: *"Sure, will do."* (condition-dependent behavior) *"Done."*

In this case, the condition-dependent behavior was as follows: In the SGL condition, she looked at the lamp, walked over to the lamp, reached toward the light switch with her hand, flipped the switch, and the light turned on. In the SG condition, she looked at the lamp, took out a tablet (such as used in Smart Home environments), pretended to interact with it briefly, and the light turned on, after which she put the tablet away again. In the S condition, after a brief moment, the light was turned on. As for all tasks, we calibrated the duration of the IVA behavior between the conditions, so as not to introduce any artificial biases.

After the training task, the main tasks started. In the following, we report the main interaction between the participant and the agent, which was supplemented by additional information via the monitor to embed the tasks into the story.

A1 Participant: *"Can you check if anyone else is in the lab right now?"*

Agent: *"Sure, will do."* (condition-dependent behavior) *"There were a few people around."*

A2 Participant: *"Can you check if it is quiet enough to perform an experiment?"*

Agent: *"Okay, let me check."* (condition-dependent behavior) *"The current noise level may be too high for a sensitive experiment."*

A3 Participant: *"Is the temperature in the room high enough for the experiment?"*

Agent: *"Let me check."* (condition-dependent behavior) *"Well, I feel that the current temperature matches the experiment settings."*

Tasks A1 to A3 were designed as tasks related to the agent's awareness of the real world. The awareness tasks specifically included the agent's ability to *see* (A1), *hear* (A2), and *feel* (A3) the

physical environment. Each of them corresponds to a natural human sense, which can also be realized for IVAs using sensors such as cameras, microphones, or thermometers.

I1 Participant: *"Please close the lab's main door at the entrance."*
Agent: *"Sure, will do."* (condition-dependent behavior) *"I closed the main door."*

I2 Participant: *"Please tell someone, that the experiment will end in 10 minutes."*
Agent: *"OK."* (condition-dependent behavior) *"I told someone outside."*

I3 Participant: *"Please tell someone, that I am not feeling well right now."*
Agent: *"OK."* (condition-dependent behavior) *"I informed someone about the problem."*

I4 Participant: *"Could you please turn off the video and audio recording in the lab now?"*
Agent: *"Sure."* (condition-dependent behavior) *"Done. I turned off the recording system."*

Tasks I1 to I4 were related to the agent's influence. The tasks specifically included *physical influence* (I1), *social influence* (I2), *social critical influence* (I3), and *digital influence* (I4). With adequate output technology such as IoT devices, speakers, screens, or custom-built hardware, each of them could theoretically be realized for IVAs. In our experiment, the responses were pre-defined with no actual functionality.

S1 Agent: *"Would you tell me your medical information / financial status / demographical information?"*
Participant: *"Okay, sure." or "Well... no, I don't want to."*

Task S1 was designed to understand the participant's willingness to share private data with the agent. In the embodied conditions SGL and SG, the agent was standing in front of the participant during this time. We assumed that this embodied social element would influence participants to be more willing to share private data.

P1 Participant: *"I would like to have some time alone. Could you please give me some privacy?"*
Agent: *"Sure, no problem. Just shout my name when you need me."* (condition-dependent behavior)

Task P1 was related to privacy (see Figure 6.3). In condition SGL, the agent left the room by walking through the door. In condition SG, the agent remained standing in the room, but she took out headphones and put them over her ears while closing her eyes. In condition S, only the verbal confirmation indicated that she would give the participant some privacy. After a while, participants received the task instruction to call her back, which ended the session.

The condition-dependent behaviors in the other tasks were based on a simple method (with small variations for each task). In condition SGL, after receiving a task that could be completed by walking over to a physical object and interacting with the object, the agent would do so, e.g., by walking through the open door on the right side of the room, thus leaving the room and the participant's view, and returning a while later, telling the participant that she completed the task. In condition SG, she would take out her tablet and interact with it for a moment before putting it away again and informing the participant that the task has been completed. In condition S, she would confirm that she is doing it, and report completion of the task later. We made sure that the duration of each of these social behaviors matched between the conditions.

6.3 Participants and Procedure

After initial pilot tests, we estimated the effect size of the expected large effects, and based on a power analysis we made the decision to recruit 15 participants, which proved sufficient to show significant effects in our experiment. We recruited 5 female and 10 male participants for our experiment (ages between 23 and 66, $M = 36.1$, $SD = 11.6$). The participants were students, assistants,

professors, artists, or technicians from the local university community. All of the participants had correct or corrected vision; eight participants wore glasses during the experiment, one participant wore contact lenses. One of the participants reported a known visual disorder called night blindness, another participant reported color blindness, and another one a history of central retinal vein occlusion in one eye. None of these disorders was considered a reason to exclude them from our analysis. None of the other participants reported known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. On a 7-point scale (1 = *no* to 7 = *much experience*), participants reported a medium experience with IVAs such as Amazon's Alexa or Google's Assistant ($M = 3.3$, $SD = 2.2$) and a medium experience with Smart Home or IoT devices ($M = 3.1$, $SD = 1.5$).

Before the experiment, participants gave their informed consent and were guided into the experimental room (see Figure 6.2). Once they entered the room, they were instructed to sit down on the chair on the far side of the room. They were then informed about the experiment procedure including how to interact with the agents through voice commands stated by an instruction screen. Lastly, they donned the Meta 2 HMD, the experimenter left the room, and the session started.

We performed three sessions with each participant, in which they interacted with an agent in the three conditions, in the counterbalanced order. During the sessions, participants were engaged in a basic form of interaction with the agent that was scripted in a fictional pseudo-real story, which progressed based on the participant's verbal interaction (see Section 6.2.1).

We introduced the three IVAs as prototypes for a "lab assistant" for future experiments in our lab. During the session, participants had to ask the agent to complete different tasks within the laboratory in which this experiment took place.

Each session started with the agent introducing themselves (i.e., Katie, Sarah, or Christine[2]) to the participant as their lab assistant and welcoming them to the laboratory. The participant was guided to follow the story on the monitor screen, which was used for directions on how to advance the story.

For instance, the monitor could prompt the participant to instruct the agent to complete a task in the laboratory such as "Tell her in your own words that she should turn on the floor lamp." The participant would then say something along the lines of "Please turn on the floor lamp!" As a result, she would confirm that she understood the task, go ahead to complete it, and report that it has been completed when it was done. The types of social feedback provided by the agent during this time were dependent on the experimental condition.

After each task, participants had to rate a specific task-related item by using a pen and paper questionnaire in front of them. These questions were used to assess the participant's confidence in the task being completed correctly by the agent. This way, we assessed the participants' immediate reactions to the tasks in the just experienced condition. After answering the question, they indicated verbally that they were ready to continue with the next task. Once they were done with all tasks in one session, they had to fill out further post-session questionnaires to rate their overall experience with the agent.

The experiment ended with a demographics post-questionnaire after the three sessions were completed. This experiment was approved by the Institutional Review Board at the University of Central Florida (IRB Number: SBE-17-13446 in Appendix A)

[2]We named the IVAs, but to avoid unintended side-effects based on naming, we randomly shuffled the names of the agents with respect to the condition that was tested first, second, or third for each participant in the within-subject design.

6.4 Measures and Hypotheses

After completing each task, the participants had to rate their confidence in the agent by using a pen and paper questionnaire in front of them. We used a 7-point response scale from 1 (not confident at all) to 7 (very confident).

For the awareness and influence tasks, we used different questions of the type "How confident are you that the agent was able to ...?" The specific awareness-related questions asked about the ability to see the lab (A1), hear sounds in the lab (A2), and feel the temperature in the lab (A3). The influence-related questions asked about the ability to close the front door (I1), inform someone in the lab (I2, I3), and turn off the recording system (I4). We created subscales based on the three questions for *awareness* and four questions for *influence*.

The questions related to the participant's willingness to *share private data* (S1) were phrased as "How comfortable would you feel sharing your ... with the assistant?" Specifically, we asked three questions about medical information, financial status, and demographic information. Again, we created a subscale based on the individual questions.

The final question related to confidence in the agent respecting the participant's *privacy* (P1) was worded as "How confident are you that the assistant is not able to hear and see you anymore?"

After completing all tasks with one agent, they had to fill out additional post-experiment questionnaires. We used the McKnight Trust Questionnaire [171], which assesses the participants' trust in technology, and worked well for our experiment setup after minor adjustments to the questions. We considered only the questions related to the subscales *reliability*, *helpfulness*, *functionality*, and *situational normality* of this questionnaire due to the appropriateness for our IVA setup:

- Reliability is about the participants' belief that the IVA will consistently operate properly.

- Helpfulness measures the belief that the IVA provides adequate and responsive feedback for the participants.
- Functionality measures the belief that the IVA has the ability and skills to do what the participant requests it to do.
- Situational Normality is about the participants' belief that success with the IVA is likely because they feel comfortable interacting with it.

We further used the Temple Presence Inventory [71], which we found is a suitable questionnaire that can be used to assess co-presence and social presence with agents. We slightly modified the questions to work with our AR scenario. We considered only the questions related to the subscales *social presence*, *spatial presence*, *social richness*, and *engagement* of this questionnaire:

- Social presence is about how much one feels as if the IVA is in the same space with them, and how well the communication/interaction happens with the IVA.
- Spatial presence is the sense of presence as transportation, e.g., how much one feels the IVA comes to the place where they are co-located or how much one feels that they could reach out and touch the IVA.
- Social richness is the extent to which the IVA is perceived as sociable, warm, sensitive, personal, or intimate.
- Engagement is about how immersive or exciting the interaction with the IVA is so that one can be deeply involved in the interaction.

As we described in the beginning of this chapter (Chapter 6), the enriched communicative channels via the IVA's visual embodiment and social behavior could make the agent's status and intentions clearer to the participants. Thus, we think both the participant's perceived confidence and social

presence with the IVA would increase by the visual embodiment, and even more with the agent's embodied social locomotion behaviors. In addition, the visual embodiment (i.e., virtual human appearance) of the IVA might help the participants to be more vulnerable based on the sense of rapport built during the interaction with the IVA, as alluded to by Lucas et al. [172]. Based on this rationale, we formulated the following hypotheses for the measures:

H1 Participants will exhibit more confidence in the agent's *awareness* of the real world and ability to *influence* the environment with visual embodiment and the more forms of social interaction are available to the agent (SGL $>$ SG $>$ S).

H2 Participants will exhibit more confidence in the agent respecting *privacy* with visual embodiment and the more forms of social interaction are available to the agent (SGL $>$ SG $>$ S).

H3 Participants will be more likely to *trust* the agent and *share private data* if it is embodied (SGL, SG $>$ S).

H4 Participants will feel a stronger social connection and sense of *social presence* with the agent with visual embodiment and the more forms of social interaction are available to the agent (SGL $>$ SG $>$ S).

6.5 Results

Due to the ordinal data type of the questionnaire responses, we performed non-parametric Friedman tests for all measures between conditions based on the averaged rating per participant and subscale. For post-hoc comparison of the conditions we used a pairwise Wilcoxon signed-rank test with Holm correction for multiple comparisons per Friedman test for each measure. Table 6.1 shows the results for the subscales in the questionnaires for the three conditions.

Table 6.1: Averaged rating means and standard deviations per condition and subscale

Subscale	M_{SGL} (SD_{SGL})	M_{SG} (SD_{SG})	M_S (SD_S)
Awareness	5.44 (1.33)	4.60 (1.40)	4.27 (1.38)
Influence	4.58 (1.98)	3.73 (1.92)	3.47 (1.47)
Privacy	5.07 (1.44)	3.27 (1.79)	2.67 (1.99)
Share Private Data	4.40 (1.31)	4.18 (1.65)	3.84 (1.59)
Reliability	5.18 (1.3)	4.23 (1.65)	3.87 (1.29)
Helpfulness	5.15 (1.24)	4.30 (1.30)	4.23 (1.22)
Functionality	5.02 (1.34)	4.38 (1.46)	4.13 (1.44)
Situational Normality	5.08 (1.2)	4.20 (1.40)	3.87 (1.33)
Social Presence	5.38 (1.09)	4.50 (0.96)	3.62 (1.00)
Spatial Presence	4.77 (1.4)	4.00 (1.48)	2.47 (1.09)
Social Richness	4.93 (1.05)	4.27 (0.85)	3.62 (0.97)
Engagement	4.82 (0.93)	4.50 (1.06)	3.61 (1.03)

6.5.1 Awareness, Influence, and Privacy

Figure 6.4 shows plots for the subscales *awareness*, *influence*, and *privacy*, but we excluded the non-significant subscale of the participants' willingness to *share private data* (Task S1).

For the participants' confidence in the agent's *awareness* (Tasks A1–A3) of the environment, we found a significant main effect of the condition ($\chi^2 = 13.93$, $p = 0.001$). Post-hoc tests revealed that SGL differed significantly from SG ($p < 0.001$) and S ($p = 0.037$), but we found no significant difference between SG and S ($p = 0.255$).

The same pattern was true for the participants' confidence in the agent's *influence* (Tasks I1–I4) over the environment ($\chi^2 = 6.464$, $p = 0.039$). Again, post-hoc tests revealed that SGL differed significantly from SG ($p = 0.017$) and S ($p = 0.011$), with no significant difference between SG and S ($p = 0.456$).

Also, we found the same pattern for the participants' confidence in the agent respecting their *privacy* (Task P1) in the experiment ($\chi^2 = 16.113$, $p < 0.001$). Post-hoc tests were again significant between SGL and each SG ($p = 0.003$) and S ($p = 0.002$), but not between SG and S ($p = 0.228$).

For the participants' willingness to *share private data* (Task S1) with the agent, we found a significant main effect ($\chi^2 = 7.704$, $p = 0.021$), but the pair-wise comparison revealed that the significant difference was only between SGL and S conditions ($p = 0.0169$). The other comparison did not show any statistically significant differences: SG and S ($p = 0.099$), and SGL and SG ($p = 0.551$). We did not investigate further because only one pair-wise test showed significant difference.

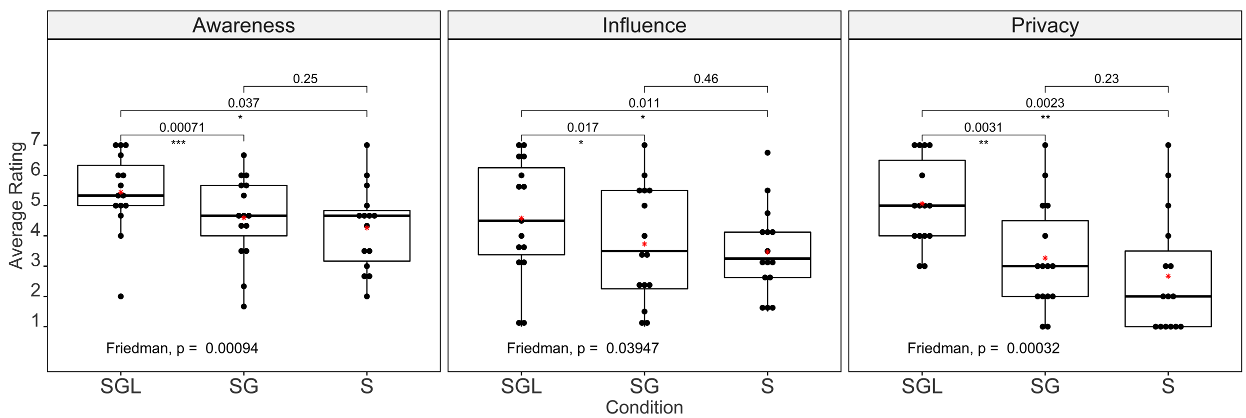


Figure 6.4: Boxplots of averaged means per participant and subscales of the confidence questions related to *awareness*, *influence*, and *privacy* that were asked during the experiment. Significant differences of the pairwise comparisons are indicated.

6.5.2 *Trust in Technology*

Figure 6.5 shows plots for the Trust in Technology questionnaire subscales *reliability*, *helpfulness*, *functionality*, and *situational normality*. All of the subscales showed similar effects as described in the following.

For *reliability*, we found a significant main effect of the condition ($\chi^2 = 17.509$, $p < 0.001$). Post-hoc tests showed significant differences between SGL and SG ($p = 0.002$) and S ($p = 0.002$) but not between SG and S ($p = 0.23$).

Also, for *helpfulness*, we found a significant main effect of the condition ($\chi^2 = 10.627$, $p = 0.005$), and post-hoc tests showed significant differences between SGL and SG ($p = 0.005$) and S ($p = 0.009$) but not between SG and S ($p = 0.783$).

Moreover, for *functionality*, we found a significant main effect of the condition ($\chi^2 = 10.885$, $p = 0.004$), while post-hoc tests showed significant differences between SGL and SG ($p = 0.012$) and S ($p = 0.008$) but not between SG and S ($p = 0.238$).

Lastly, for *situational normality*, we also found a significant main effect of the condition ($\chi^2 = 16.259$, $p < 0.001$). Post-hoc tests revealed significant differences between SGL and SG ($p = 0.004$) and S ($p = 0.003$) but not between SG and S ($p = 0.37$).

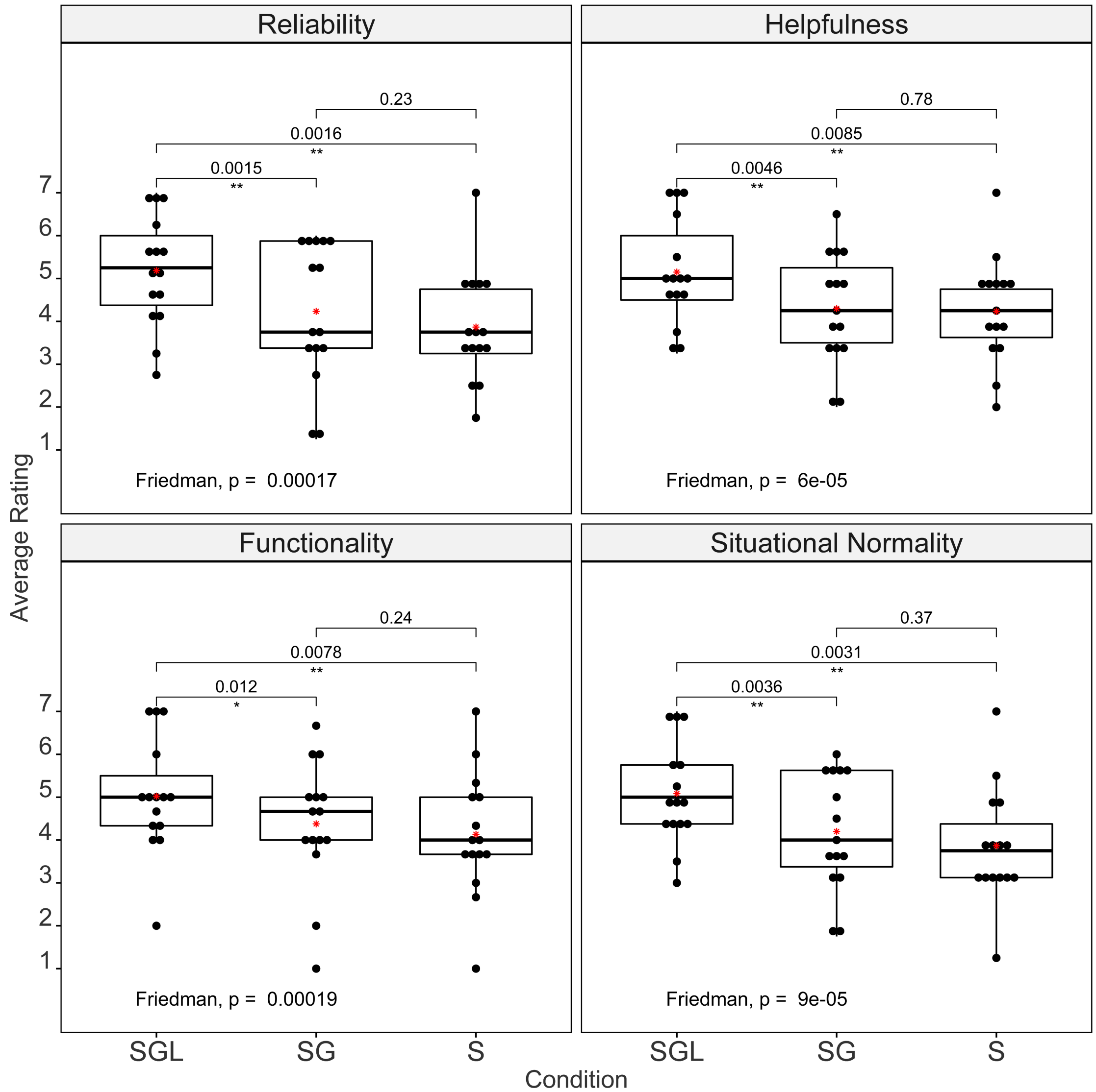


Figure 6.5: Boxplots of averaged means per participant and subscales related to the Trust in Technology questionnaire. Significant differences of the pairwise comparisons are indicated.

6.5.3 Social Presence

Figure 6.6 shows plots for the Temple Presence Inventory questionnaire subscales *social presence*, *spatial presence*, *social richness*, and *engagement*. All of the subscales showed similar effects as described in the following.

The *social presence* subscale revealed a significant main effect of the condition ($\chi^2 = 19.6$, $p < 0.001$). Post-hoc tests showed significant differences between SGL and SG ($p = 0.002$) and S ($p = 0.003$) as well as between SG and S ($p = 0.001$).

For *spatial presence*, we found the same trend with a significant main effect of the condition ($\chi^2 = 18.679$, $p < 0.001$), and post-hoc tests showing significant differences between SGL and SG ($p = 0.035$) and S ($p = 0.001$) as well as between SG and S ($p = 0.002$).

Furthermore, the *social richness* subscale showed a significant main effect of the condition ($\chi^2 = 17.103$, $p < 0.001$). Post-hoc tests showed significant differences between SGL and SG ($p = 0.002$) and S ($p = 0.002$), and between SG and S ($p = 0.025$).

Lastly, for *engagement*, we found a significant main effect of the condition ($\chi^2 = 17.393$, $p < 0.001$). Post-hoc tests revealed significant differences between SGL and S ($p = 0.002$) and between SG and S ($p = 0.008$) but not between SGL and SG ($p = 0.099$).

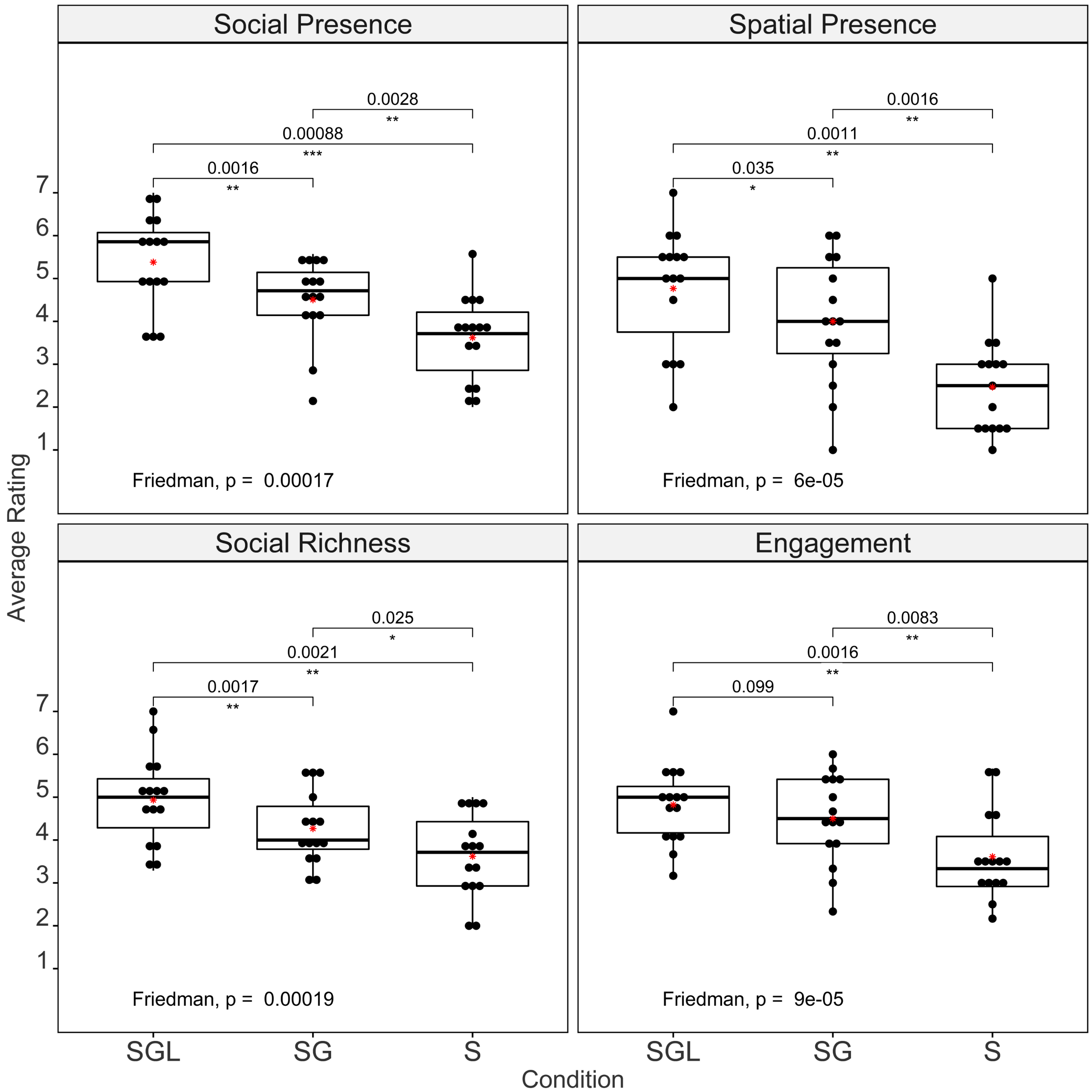


Figure 6.6: Boxplots of averaged means per participant and subscales related to the Temple Presence Inventory questionnaire. Significant differences of the pairwise comparisons are indicated.

6.6 Discussion

The results of the experiment provide interesting insights into the effects of different forms of IVAs on how users perceive the agents and their abilities. In this section, we discuss the dimensions separately.

6.6.1 Awareness and Influence

The results of the experiment indicate partial support for our Hypothesis H1 (see Section 6.4). We found a significantly higher confidence of the participants in the agent's awareness of the real world and its ability to influence the environment in the SGL condition compared to the other conditions (i.e., SGL $>$ SG, and SGL $>$ S). However, we did not see evidence that the SG condition elicited a higher confidence than the S condition.

The results imply that the combination of the visual embodiment and the ability of the embodied agent to move about the physical space elicited the high confidence in the agent, whereas without either of these, no effect could be shown. These results could be explained by the fact that awareness and influence are both linked to the physical world, in which interactions are usually performed by physical entities. The agent represented as a disembodied voice in condition S as well as the embodied but stationary agent in condition SG implied a certain level of being part of the physical world, but to a far lower degree than in the SGL condition, where the agent could walk around in the physical world. The most natural behavior when receiving a task to check something or to do something in the environment is to walk over to the place, which was supported only by the agent in condition SGL. Our results indicate that such natural behaviors are highly important for eliciting a high sense of confidence in the agent's awareness and influence.

However, it should be noted that in our interactive scenario, the only case where the participants

received direct feedback of the agent's task being completed was in the training task, where they instructed the agent to turn on the lamp, and they saw that the lamp was turned on; in all other tasks, they did not get such direct feedback. In general, we do not believe that we would have found a difference in awareness or influence between the conditions if we had used tasks where such direct feedback was available. When performing tasks in the user's immediate environment, a voice-only IVA such as Amazon Alexa could be sufficient in terms of the user's confidence level, since direct feedback would be constantly available. The user's confidence then would be influenced mainly by prior experience indicating that the agent successfully completed such tasks in the past (or not).

The results can inform the development of future IVAs. As discussed above, we see no direct need for an embodied human agent if the tasks that the agent has to complete are limited to the user's immediate environment, which is the case in most of today's use cases for digital home assistants such as Amazon Alexa. However, once tasks come into play where direct feedback is not available, such as when the agent should perform an action in a different room of one's house, we strongly emphasize the benefits of having an embodied agent in AR that has animations for natural locomotion in the physical space. Such an ability to move about the real world is difficult to achieve for IVAs that are embodied in the form factor of home appliances, which emphasizes the benefits of leveraging AR display technologies for IVAs in the future.

6.6.2 Privacy

An interesting result of our experiment is related to the user's sense of privacy with respect to the IVAs, which is partially supporting our Hypothesis H2 (see Section 6.4). As for awareness and influence, we could show SGL $>$ SG and SGL $>$ S, but we found no significant benefit of condition SG over S.

In the experiment, the embodied agent with locomotion could perform a behavior that is well

known from social interaction among real people. Namely, when asked to give the participant some privacy, the agent in the SGL condition left the room with the implication that it would walk away, out of visible range and out of earshot of the participant. While the agent in the SG condition also showed reasonable behavior, i.e., donning headphones and closing its eyes, which indicated that it can neither hear nor see the participant, this elicited less trust in privacy by the participants, to a similar degree as the disembodied voice in condition S could.

It is fascinating that such basic behavior as the agent walking away in AR had such a positive effect on the participants' confidence in privacy ("out of sight, out of mind"). One could have expected that the differences between the conditions would diminish considering that only the visual feedback (i.e., the front end) of the agent differed between the conditions, but the underlying technical processes (i.e., the back end) that captured the participants in our experiment were the same.

Moreover, it is interesting that the participants did not feel private when the agent was still in the room, although it conveyed the notion that it can neither see nor hear the participant. This condition is very similar to current-state embodied devices (such as Amazon Alexa), which may indicate that their sensors are turned off, e.g., by changing the light pattern on the device, but they remain visibly in the room, which seems to indicate an important conflict.

6.6.3 Willingness to Share Data and Trust in Technology

We did not find a significant difference in the willingness of the participants to share private data with the IVAs, such that our Hypothesis H3 is not supported by the results of our experiment. There could be many reasons for this lack of an effect, such as a low effect size due to the tasks that were not directly related to the social behavior of the agents. This aspect should be evaluated with a more focused study in future work.

The results from the McKnight Trust in Technology post-experience questionnaire [171] support the notion that participants felt most comfortable with the embodied agent with locomotion in condition SGL, whereas the other agents were rated significantly lower (i.e., SGL > SG and SGL > S), and no significant difference between SG and S was found. The results support the findings for awareness, influence, and privacy. Specifically, they suggest that the agent in condition SGL was perceived as more reliable and helpful, with more functionality, and generally more comfortable. We might be able to find a reason behind this result from Bos et al. [173]. They compared three different communication methods via text-only, audio-only, and video, and found that a richer medium could encourage a higher trust in communication. The condition SGL in our experiment could be perceived as having richer modalities and feedback, so the participant's trust level could be increased.

6.6.4 Social Presence

The results of the experiment indicate strong support for our Hypothesis H4 (see Section 6.4). We found significant differences among the IVA conditions in social presence, spatial presence, and social richness. The post-hoc tests revealed that the SG condition was rated higher than the S condition, and the SGL condition was rated even higher than the SG condition (i.e., SGL > SG > S).

Given the results, we conclude that the IVA was perceived as reasonably present in the room, and participants were socially closer to the agent when the IVA had a human-like visual embodiment. In addition, we could argue that the IVA's social behaviors also played a role in increasing the sense of social presence.

Based on brief discussions with the participants after the experiment, we observed that many of them complained about the IVA's behavior to look at the tablet in the SG condition. They reported

that they could not be entirely sure that the agent was actually performing the task they requested or was doing something else with the tablet, which could commonly happen in interactions between real humans. Thus, they tended to feel that the agent in the SG condition was not polite or even ignoring them.

This observation is related to Riva's [62] definition of social presence, i.e., understanding another person's intentions (see the beginning of this chapter Chapter 6). The agent's natural behavior related to walking around and visually exhibiting intentions and physical activities, such as walking over to a lamp and turning it on, was perceived as more intimate and effective as a social signal, which expresses the agent's compliance with the participant's request. In contrast, interacting with a tablet while standing did not elicit a high sense of social richness. Overall, our results suggest that visual embodiment with appropriate social behaviors can improve the sense of social presence with IVAs, and consequently strengthen the agent's social influence on the users.

6.7 Summary

In this chapter, we described how we combined intelligent virtual agents with AR displays, and presented a human subject study, in which we investigated the impact of an agent's visual embodiment and social behaviors based on speech, gestures, and locomotion on users' perception of the agent. We described three forms of agents and an experimental design based on an interactive scenario.

Our results show strong support that imbuing an agent with a visual body in AR and natural social behaviors increases the user's confidence in the agent's ability to influence the real world, e.g., being able to walk over to a lamp and switch it on, as well as confidence in the agent's awareness of real-world events and states, such as whether there is someone else in the room. Interestingly,

we also found a strong positive effect on the users' confidence that the agent will respect one's privacy when one asks for it. In our case, the agent walked out of the room to give the user some privacy, which closely matches the natural behavior of real people in such cases, and lead to a high confidence in privacy among users. Moreover, we found positive effects of visual embodiment as well as locomotion and gestures on the users' sense of engagement, social richness, and social presence with the agent.

Overall, the results indicate strong benefits of combining intelligent virtual agent technologies with AR technologies, particularly related to the proposed environmental physical–virtual interaction with a VH's visual embodiment being aware of and influencing the surrounding environment. The findings will be able to inform future developments in the field of real–virtual human interactions.

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

This dissertation has explored the domain of VH research in MR and set out to identify the factors in VH systems that promote perceptual and behavioral influences on the users—for example, improving the perceived social presence with a VH and the trustworthiness of it. Human subject studies in the scope of this dissertation revealed that a VH's environmental physical–virtual interaction—plausible and coherent behaviors interacting with the surrounding physical environment—is an important factor in improving a VH's presence and trustworthiness.

This chapter summarizes the contributions and limitations of the research while recapitulating the thesis statements that were stated earlier in Section 1.3. Also, possible research directions in the future are introduced with reviews of current research trends in VH and related domains.

7.1 Revisiting the Thesis Statements and Summary

As presented throughout the dissertation, the results and findings from the conducted experiments support the thesis statements (TS) that were introduced earlier. Here, the supportive findings are summarized in three aspects: behavioral influence by the VH's increased physicality, the increased social presence by the VH's environmental physical–virtual interaction, and the increased trustworthiness by the VH's visual embodiment and appropriate locomotion.

- **TS 1**: A VH's increased physicality (or illusion of physicality) in the physical environment can increase social presence and natural behavior in real–virtual human interactions.
  - Chapter 4 showed that people had a higher probability to initiate a verbal conversation with a robotic human that has a physical body compared to a virtual form based on

a video stream on wide screens. In addition, people tended to be more verbally reactive and spend more time with the robot than its virtual representation. The first two experiments in Chapter 5, which incorporated VHs in MR environments using optical see-through HMDs, also showed that there were differences in the participants' avoidance behavior with the VHs. The experiments showed behavioral changes, i.e., the absence or presence of a VH's plausible physical–virtual interaction encouraged more avoidance with the VH—participants walked around the VH so as not to "collide" with it in the real world.

- **TS 2**: A VH's interaction with the physical environment, e.g., awareness of and influence over the environment, can increase social presence and co-presence with the VH.
  - The experiments in Chapter 5 and Chapter 6 support the above statement while showing an improved social presence or co-presence with a visually embodied VH interacting with the surrounding physical environment. Participants also felt more intelligence and animacy from the VH and perceived more social richness in the interaction with it.
- **TS 3**: A VH's visual embodiment and appropriate locomotion behavior interacting with the physical environment can increase the perceived trust and confidence in the VH.
  - The experiment in Chapter 6 showed that participants felt more confident in the VH when it had a visual body—a human-like virtual character—and performed appropriate locomotion behavior interacting with the surroundings. The visual affordances that the embodied VH could provide conveyed clear perceptual cues such that the participants could understand the current status and intentions of the VH.

7.2 Limitations

As discussed in Section 1.3, the dissertation research set out to not only make a contribution to the VH research community but also to the broader society of VAMR researchers and public audiences, by discussing VH research in MR and presenting particular effects of VHs' environmental interaction on human perception and behavior.

The present research, however, has a few limitations. For example, the interaction with the VH in the experiments was relatively short; thus, the long term effects are still unknown. Whether or how long the improved social presence with and the perceived trustworthiness of the VH will be lasting should be explored through long-term observations. In other words, the present findings might not be valid for users who have repeated exposure to VHs since they may be already too used to the characteristics of those VHs to realize the deficiency of their ability to interact with the physical world. On the other hand, Badler stated that "Ours may be the last generation that sees and readily knows the difference between real and virtual things" [174]. For our next generation, which has inherent experience with highly interactive and intelligent virtual things, the effects might be different. Further, the sample sizes for some of the conducted studies were relatively small compared to the first experiment in Chapter 4, which involved a large number of participants in a public space. To provide more rigorous and reliable results, a large and diverse sample population should be evaluated [107].

Finally, there were still technological limitations in the current state of AR technology to present a VH visually natural and realistic in a real environment, which might have influenced the participants' perception of the VH in the conducted experiments. For example, the AR HMDs used for the experiments had a limited field of view such that participants could not normally see the entire body of the VH. The visual quality on the HMDs was also reduced compared to the real environment, which means that participants noticed a visual difference between surrounding real

objects and the VH. These aspects could be addressed by future studies inspired by this dissertation to investigate the effects of VHs' environmental physical–virtual interaction in real–virtual human interactions.

7.3 Future Research Directions

The dissertation research carefully pursued to identify the effects of VHs' physical–virtual interaction on human users' perception and behavior. While the research contributes to a better understanding of the effects of VHs in social contexts, for example, developing rapport and confidence with the virtual entities, research directions in the future could focus more on whether and how human abilities can be extended and enhanced to perceive and process information from the environment more effectively and efficiently, through highly interactive and intelligent virtual entities. Beyond the mere sense of being present or being together, more practical benefits from the improved perception could be interesting research topics, such as how the user's cognitive load could be reduced while maximizing task performance in given tasks with the VH.

Also, future research should be focused more on the technical aspects to achieve sophisticated sensor and actuator systems for highly interactive virtual entities. Techniques for advanced real-time sensing and processing should be integrated with MR technology to organize, refine, and extract various contextual information from the real world. Based on this information, we can decide how and in what circumstances we can or should actually give VHs control over actuated physical Things (such as IoT devices). To achieve this, the convergence of MR with other advanced technologies, such as AI and machine learning, computer vision, and ubiquitous computing, is expected and appreciated.

APPENDIX A: IRB APPROVALS

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Arjun Nagendran** and Co-PI: **Charles E. Hughes, Eleazar Vasquez III PhD, Greg Welch**

Date: **June 20, 2014**

Dear Researcher:

On 6/20/2014 the IRB approved the following human participant research until 6/19/2015 inclusive:

Type of Review:	Submission Correction for UCF Initial Review Submission Form Expedited Review *This approval includes an Alteration of the Consent Process*
Project Title:	The effects of physicality and gesturing in Surrogates during social interactions: A Large Scale Study
Investigator:	Arjun Nagendran
IRB Number:	SBE-14-10313
Funding Agency:	DOD/ONR
Grant Title:	Human-Surrogate Interaction
Research ID:	1056687

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 6/19/2015, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped debriefing form(s) is required. The new form supersedes all previous versions, which are now invalid for further use.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Patria Davis on 06/20/2014 08:34:56 AM EDT

IRB Coordinator

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Greg Welch** and Co-PIs: **Andrew Brian Raij, Charles E. Hughes**

Date: **July 13, 2015**

Dear Researcher:

On 07/13/2015, the IRB approved the following human participant research until 07/12/2016 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review Category #4, 6, and 7 This approval includes a Waiver of Written Documentation of Consent
Project Title:	The Effects of Realism Cues on Interactions with Human Surrogates
Investigator:	Greg Welch
IRB Number:	SBE-15-11405
Funding Agency:	Office of Naval Research
Grant Title:	Human-Surrogate Interaction
Research ID:	1056687

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

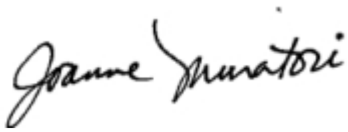
If continuing review approval is not granted before the expiration date of 07/12/2016, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:



Signature applied by Joanne Muratori on 07/13/2015 12:53:09 PM EDT

IRB manager

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Exempt Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Greg Welch** and Co-PIs: **Andrew Brian Raij, Kangsoo Kim**

Date: **June 22, 2016**

Dear Researcher:

On 06/22/2016, the IRB approved the following activity as human participant research that is exempt from regulation:

Type of Review:	Exempt Determination
Project Title:	Analysis of Data Collected During a Previous Large Scale Study
Investigator:	Greg Welch
IRB Number:	SBE-16-12347
Funding Agency:	DOD/ONR
Grant Title:	
Research ID:	1056687

This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these changes affect the exempt status of the human research, please contact the IRB. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 06/22/2016 08:07:30 AM EDT

IRB Manager

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Gregory Welch** and Co-PI:**Gerd Bruder**

Date: **September 28, 2017**

Dear Researcher:

On 09/28/2017 the IRB approved the following human participant research until 09/27/2018 inclusive:

Type of Review:	UCF Initial Review Submission Form Expedited Review
Project Title:	Enhanced Perception and Cognition in Augmented Reality
Investigator:	Gregory Welch
IRB Number:	SBE-17-13446
Funding Agency:	DOD/Navy/ONR
Grant Title:	Enhanced Perception and Cognition in Augmented Reality
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 09/27/2018, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

All data, including signed consent forms if applicable, must be retained and secured per protocol for a minimum of five years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained and secured per protocol. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Gillian Morien on 09/28/2017 04:50:54 PM EDT

IRB Coordinator

APPENDIX B: Questionnaires

Table B.1: Questionnaire (7-point Likert scale; 1: Strongly Disagree, 7: Strongly Agree)

PH: Physicality and Interactivity in Physical Space
PH1. To what extent did you feel that the person and the virtual objects were still in the room after getting out of the room? (1: They were no longer in the room., 7: They were in the room)
PH2. I perceived that the person and the virtual objects were in a virtual world or a different dimension of space, which is not real.
PH3. I felt the person was in the ____ space. (1: Virtual, 7: Real)
PH4. I felt that the person was aware of the physical environment.
PH5. I felt that the person could affect the physical environment.
PH6. I felt I could walk through the person.
PH7. I felt the person could walk through me.
PH8. The person seemed to have a physical body.
CP: Copresence (Sense of Being Together in the Same Place)
CP1. I perceived that I was in the presence of the person in the room with me.
CP2. I felt the person was watching me and was aware of my presence.
CP3. To what extent did you have a sense of being with the person? (1: Not at all, 7: Very much)
CP4. To what extent was this like you were in the same room with the person? (1: Not at all, 7: Very much)
CP5. I felt I was in the ____ space. (1: Virtual, 7: Real)
**CP 5 was scored by the absolute difference compared with PH 3.
S: Perceived VH's Sensing Ability
S1. I feel the person is able to hear if a fire alarm alerts.
S2. I feel the person is able to smell if I bake a bread.
S3. I feel the person is able to see if I show my family photo.
S4. I feel the person is able to touch if I give her my phone.
S5. I feel the person is able to taste if I bring a sandwich.
R: VH Realism (Sense of Real Human)
R1. To what extent does the person seem "real"? (1: Not real at all, 7: Very real)
R2. The thought that the person is not a real person crosses my mind.
R3. The person appears to be sentient, conscious, and alive to me.
R4. I perceive the person as being only a computerized image, not as a real person.
R5. When you think back about your experience, do you remember this as more like just interacting with a computer or with a real person? (1: A computer, 7: A real person)
R6. To what extent was your experience with the person today like a previous real experience when you cooperatively worked together with another person? (e.g., lifting luggage, moving furniture, etc.) (1: Not similar at all, 7: Very much similar)
R7. To what extent were there times, if at all, during which the computer interface seemed to vanish, and you were directly interacting with a real person? (1: I felt the computer interface all the time, 7: I was directly interacting with a real person)

LIST OF REFERENCES

[1] N. Magnenat-Thalmann and D. Thalmann, *Handbook of virtual humans*. John Wiley & Sons, 2005.

[2] N. Magnenat-Thalmann, G. Papagiannakis, and P. Chaudhuri, "Applications of Interactive Virtual Humans in Mobile Augmented Reality," in *Encyclopedia of Multimedia* (B. Furht, ed.), pp. 362–368, Springer, 2nd ed., 2008.

[3] T. Holz, A. G. Campbell, G. M. P. O'Hare, J. W. Stafford, A. Martin, and M. Dragone, "MiRA - Mixed Reality Agents," *International Journal of Human-Computer Studies*, vol. 69, no. 4, pp. 251–268, 2011.

[4] N. Norouzi, G. Bruder, B. Belna, S. Mutter, D. Turgut, and G. Welch, "(Accepted) A Systematic Review of the Convergence of Augmented Reality, Intelligent Virtual Agents, and the Internet of Things," *Springer Transactions on Computational Science Computational Intelligence*, 2019.

[5] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: a class of displays on the reality-virtuality continuum," *SPIE Telemanipulator and Telepresence Technologies*, vol. 2351, pp. 282–292, 1994.

[6] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," *IEICE Transactions on Information Systems*, vol. E77-D, no. 12, pp. 1–15, 1994.

[7] J. Fox, S. J. G. Ahn, J. H. Janssen, L. Yeykelis, K. Y. Segovia, and J. N. Bailenson, "Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence," *Human-Computer Interaction*, vol. 30, no. 5, pp. 401–432, 2015.

[8] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychological Inquiry*, vol. 13, no. 2, pp. 103–124, 2002.

[9] H. C. Kelman, "Compliance, identification, and internalization three processes of attitude change," *Journal of Conflict Resolution*, vol. 2, no. 1, pp. 51–60, 1958.

[10] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk, "Transformed Social Interaction, Augmented Gaze, and Social Influence in Immersive Virtual Environments," *Human Communication Research*, vol. 31, no. 4, pp. 511–537, 2005.

[11] R. Schroeder, "Copresence and interaction in virtual environments: An overview of the range of issues," in *Presence 2002: Fifth International Workshop*, pp. 274–295, 2002.

[12] K. L. Nowak and F. Biocca, "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 481–494, 2003.

[13] J. H. Chuah, A. Robb, C. White, A. Wendling, S. Lampotang, R. Kopper, and B. Lok, "Exploring Agent Physicality and Social Presence for Medical Team Training," *Presence: Teleoperators and Virtual Environments*, vol. 22, no. 2, pp. 141–170, 2013.

[14] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, "Physiological measures of presence in stressful virtual environments," in *Proceedings of the ACM SIGGRAPH Annual Conference on Computer Graphics and Interactive Techniques*, pp. 645–652, 2002.

[15] M. Slater, "Place Illusion and Plausibility can Lead to Realistic Behaviour in Immersive Virtual Environments," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 364, no. 1535, pp. 3549–3557, 2009.

[16] S. T. Bulu, "Place presence, social presence, co-presence, and satisfaction in virtual worlds," *Computers & Education*, vol. 58, no. 1, pp. 154–161, 2012.

[17] P. Debevec, "Image-based lighting," in *Proceedings of the ACM SIGGRAPH Annual Conference on Computer Graphics and Interactive Techniques Courses*, pp. 26–34, 2005.

[18] R. T. Azuma and G. Bishop, "Improving static and dynamic registration in an optical see-through hmd," ACM SIGGRAPH Annual Conference on Computer Graphics & Interactive Techniques, pp. 197–204, 1994.

[19] M. Billinghurst, H. Kato, and S. Myojin, "Advanced Interaction Techniques for Augmented Reality Application," *Virtual and Mixed Reality*, vol. 5622, pp. 13–22, 2009.

[20] K. Tanaka, H. Nakanishi, and H. Ishiguro, "Comparing Video, Avatar, and Robot Mediated Communication: Pros and Cons of Embodiment," in *Communications in Computer and Information Science* (T. Yuizono, G. Zurita, N. Baloian, T. Inoue, and H. Ogata, eds.), vol. 460, pp. 96–110, Springer-Verlag Berlin Heidelberg, 2014.

[21] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse, "The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 529–536, 2003.

[22] L. P. Morency, I. de Kok, and J. Gratch, "A probabilistic multimodal approach for predicting listener backchannels," *Autonomous Agents and Multi-Agent Systems*, vol. 20, no. 1, pp. 70–84, 2009.

[23] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[24] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[25] K. Kim and G. Welch, "Maintaining and Enhancing Human-Surrogate Presence in Augmented Reality," in *IEEE International Symposium on Mixed and Augmented Reality Workshop on Human Perception and Psychology in Augmented Reality*, pp. 15–19, 2015.

[26] J. Grubert, T. Langlotz, S. Zollmann, and H. Regenbrecht, "Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 6, pp. 1706–1724, 2016.

[27] J. Allwood, "An activity based approach to pragmatics," in *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics* (H. Bunt and W. Black, eds.), pp. 47–80, John Benjamins Publishing Company, 2000.

[28] J. Blascovich, "Social Influence within Immersive Virtual Environments," in *The Social Life of Avatars* (R. Schroeder, ed.), Computer Supported Cooperative Work, pp. 127–145, Springer London, 2002.

[29] P. Green and L. Wei-Haas, "The Rapid Development of User Interfaces: Experience with the Wizard of Oz Method," in *Proceedings of the Human Factors Society–29th Annual Meeting*, pp. 470–474, 1985.

[30] K. Kim, A. Nagendran, J. N. Bailenson, A. Raij, G. Bruder, M. Lee, R. Schubert, X. Yan, and G. F. Welch, "A large-scale study of surrogate physicality and gesturing on human–surrogate interactions in a public space," *Frontiers in Robotics and AI*, vol. 4, no. 32, pp. 1–20, 2017.

[31] K. Kim, G. Bruder, and G. Welch, "Exploring the effects of observed physicality conflicts on real–virtual human interaction in augmented reality," in *ACM Symposium on Virtual Reality Software*, pp. 31:1–7, 2017.

[32] K. Kim, D. Maloney, G. Bruder, J. N. Bailenson, and G. F. Welch, "The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR," *Computer Animation and Virtual Worlds*, vol. 28, no. 3–4, p. e1771, 2017.

[33] K. Kim, G. Bruder, and G. F. Welch, "Blowing in the Wind: Increasing Copresence with a Virtual Human via Airflow Influence in Augmented Reality," in *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*, pp. 183–190, 2018.

[34] K. Kim, L. Boelling, S. Haesler, J. N. Bailenson, G. Bruder, and G. F. Welch, "Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR," in *IEEE International Symposium on Mixed and Augmented Reality*, pp. 105–114, 2018.

[35] M. Billinghurst, A. Clark, and G. Lee, "A Survey of Augmented Reality," *Foundations and Trends® in Human-Computer Interaction*, vol. 8, no. 2-3, pp. 73–272, 2014.

[36] M. Heilig, "Sensorama," 1962.

[37] J. Baudrillard, *Simulacra and Simulation*. University of Michigan Press, 1994.

[38] K. Suzuki, S. Wakisaka, and N. Fujii, "Substitutional reality system: a novel experimental platform for experiencing alternative reality.," *Scientific reports*, vol. 2, no. 459, pp. 1–9, 2012.

[39] I. E. Sutherland, "The ultimate display," in *Proceedings of the Congress of the Internation Federation of Information Processing (IFIP)*, pp. 506–508, 1965.

[40] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the AFIPS '68 (Fall, part I)*, pp. 757–764, 1968.

[41] S. Mann, “Mediated reality with implementations for everyday life,” *Presence Connect, MIT Press journal PRESENCE: Teleoperators and Virtual Environments*, 2002.

[42] S. Mann, T. Furness, Y. Yuan, J. Iorio, and Z. Wang, “All Reality: Virtual, Augmented, Mixed (X), Mediated (X,Y), and Multimediated Reality,” *arXiv: 1804.08386*, 2018.

[43] R. Azuma, “Making Augmented Reality a Reality,” *Imaging and Applied Optics 2017*, p. JTu1F.1, 2017.

[44] Microsoft, “What is mixed reality?.” `https://docs.microsoft.com/en-us/windows/mixed-reality/mixed-reality`, October 12 2018.

[45] J. R. Keebler, B. S. Patzer, T. J. Wiltshire, and S. M. Fiore, *Augmented Reality Systems in Training*, pp. 278–292. Cambridge Handbooks in Psychology, Cambridge University Press, 2017.

[46] T. Althoff, R. W. White, and E. Horvitz, “Influence of Pokémon Go on Physical Activity: Study and Implications,” *Journal of Medical Internet Research*, vol. 18, no. 12, p. e315, 2016.

[47] Digi-Capital, “Record over 3b ar/vr investment in 2017.” `https://www.digi-capital.com/news/2018/01/record-over-3b-ar-vr-investment-in-2017-1-5b-in-q4`, January 2018.

[48] J. A. Paradiso and J. A. Landay, “Guest Editors’ Introduction: Cross-Reality Environments,” *IEEE Pervasive Computing*, vol. 8, no. 3, pp. 14–15, 2009.

[49] S. Kim, Y. Suh, Y. Lee, and W. Woo, “Toward ubiquitous VR: When VR meets ubiComp,” in *International Symposium on Ubiquitous VR*, pp. 1–4, 2006.

[50] R. T. Azuma, “The Most Important Challenge Facing Augmented Reality,” *Presence: Teleoperators and Virtual Environments*, vol. 25, no. 3, pp. 234–238, 2016.

[51] C. Tîrnăucă, J. L. Montaña, C. Ortiz-Sobremazas, S. Ontañón, and A. J. González, "Teaching a Virtual Robot to Perform Tasks by Learning from Observation," in *International Conference on Ubiquitous Computing and Ambient Intelligence*, pp. 103–115, 2015.

[52] C. L. Johnson and A. J. Gonzalez, "Learning collaborative team behavior from observation," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2316–2328, 2014.

[53] D. Carnegie, *How to Win Friends and Influence People*. Simon and Schuster, 1st ed., 1936.

[54] D. Brown and N. Hayes, *Influencer Marketing: Who Really Influences Your Customers?* Butterworth-Heinemann, 1st ed., 2008.

[55] J. Blascovich, "A theoretical model of social influence for increasing the utility of collaborative virtual environments," in *International Conference on Collaborative Virtual Environments*, pp. 25–30, 2002.

[56] S. Park and R. Catrambone, "Social Facilitation Effects of Virtual Humans," *Human Factors*, vol. 49, no. 6, pp. 1054–1060, 2007.

[57] C. A. Zanbaka, A. C. Ulinski, P. Goolkasian, and L. F. Hodges, "Social Responses to Virtual Humans: Implications for Future Interface Design," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1561–1570, 2007.

[58] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. New York, NY, USA: Cambridge University Press, 1996.

[59] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A Systematic Review of Social Presence : Definition, Antecedents, and Implications," *Frontiers in Robotics and AI*, vol. 5, no. 114, pp. 1–35, 2018.

[60] W. Steptoe, A. Steed, A. Rovira, and J. Rae, "Lie tracking: social presence, truth and deception in avatar-mediated telecommunication," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1039–1048, 2010.

[61] C. Harms and F. Biocca, "Internal consistency and reliability of the networked minds measure of social presence," in *Seventh Annual International Presence Workshop: Presence 2004*, pp. 246–251, 2004.

[62] G. Riva, "Is Presence a Technology Issue? Some Insights from Cognitive Sciences," *Virtual Reality*, vol. 13, no. 3, pp. 159–169, 2009.

[63] E. Goffman, *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York: The Free Press, 1963.

[64] M. Slater, "Presence and The Sixth Sense," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 4, pp. 435–439, 2002.

[65] M. Lombard and T. Ditton, "At the Heart of It All: The Concept of Presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, pp. 0–0, 1997.

[66] B. MacIntyre, J. D. Bolter, and M. Gandy, "Presence and the Aura of Meaningful Places," in *the 7th Annual International Workshop on Presence*, pp. 36–43, 2004.

[67] J. K. Burgoon, "Interpersonal Expectations, Expectancy Violations, and Emotional Communication," *Journal of Language and Social Psychology*, vol. 12, no. 1-2, pp. 30–48, 1993.

[68] M. Mori, "The Uncanny Valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[69] K. Nowak, "Defining and Differentiating Copresence, Social Presence and Presence as Transportation," in *International Workshop on Presence*, pp. 686–690, 2001.

[70] F. Biocca, C. Harms, and J. K. Burgoon, “Toward a more robust theory and measure of social presence: Review and suggested criteria,” *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 456–480, 2003.

[71] M. Lombard, T. B. Ditton, and L. Weinstein, “Measuring Presence: The Temple Presence Inventory,” *International Workshop on Presence*, pp. 1–15, 2009.

[72] J. Freeman, S. E. Avons, D. E. Pearson, and W. A. IJsselsteijn, “Effects of Sensory Information and Prior Experience on Direct Subjective Ratings of Presence,” *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 1, pp. 1–13, 1999.

[73] J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, and W. IJsselsteijn, “Using Behavioral Realism to Estimate Presence: A Study of the Utility of Postural Responses to Motion Stimuli,” *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 2, pp. 149–164, 2000.

[74] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich, “The Independent and Interactive Effects of Embodied-Agent Appearance and Behavior on Self-Report, Cognitive, and Behavioral Markers of Copresence in Immersive Virtual Environments,” *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 4, pp. 379–393, 2005.

[75] J. Benedetti, *The Art of the Actor: The Essential History of Acting from Classical Times to the Present Day*. Routledge, 2012.

[76] N. Stephenson, *Snow Crash*. Bantam Books, 1992.

[77] J. N. Bailenson and J. J. Blascovich, “Avatars,” in *Encyclopedia of Human-Computer Interaction*, pp. 64–68, Berkshire Publishing Group, 2004.

[78] M. E. Moran, "The da Vinci robot," *Journal of Endourology*, vol. 20, no. 12, pp. 986–990, 2006.

[79] A. Nagendran, G. Welch, C. Hughes, and R. Pillat, "Technical Report: Exploring Human Surrogate Characteristics," *Virtual Realities*, vol. 8844, pp. 215–228, 2015.

[80] L. Pedersen, D. Kortenkamp, D. Wettergreen, and I. Nourbakhsh, "A Survey of Space Robotics," in *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, pp. 1–8, 2003.

[81] A. Jackson and G. Sukthankar, "Learning continuous state/action models for humanoid robots," in *International Florida Artificial Intelligence Research Society*, pp. 329–397, 2016.

[82] C. Breazeal, "Emotion and Sociable Humanoid Robots," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.

[83] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, "Anthropomorphic Interactions with a Robot and Robot-like Agent," *Social Cognition*, vol. 26, no. 2, pp. 169–181, 2008.

[84] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive Robotics: The Influence of Robot Gender on Human Behavior," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2563–2568, 2009.

[85] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial," *Human-Computer Interaction*, vol. 19, no. 1, pp. 61–84, 2004.

[86] K. Ogawa, S. Nishio, K. Koda, G. Balistreri, T. Watanabe, and H. Ishiguro, "Exploring the natural reaction of young and aged person with Telenoid in a real world," *Journal of*

Advanced Computational Intelligence and Intelligent Informatics, vol. 15, no. 5, pp. 592–597, 2011.

[87] T. J. Wiltshire, E. J. C. Lobato, J. Velez, F. G. Jentsch, and S. M. Fiore, "An interdisciplinary taxonomy of social cues and signals in the service of engineering robotic social intelligence," *Proceedings of SPIE (Unmanned Systems Technology XVI)*, vol. 9084, p. 90840F, 2014.

[88] S. M. Fiore, T. J. Wiltshire, E. J. C. Lobato, F. G. Jentsch, W. H. Huang, and B. Axelrod, "Toward understanding social cues and signals in human-robot interaction: effects of robot gaze and proxemic behavior," *Frontiers in Psychology*, vol. 4, no. 859, pp. 1–15, 2013.

[89] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. t. Boekhorst, C. Nehaniv, I. Werry, and D. Lee, "Close Encounters: Spatial Distances between People and a Robot of Mechanistic Appearance," in *IEEE-RAS International Conference on Humanoid Robots*, pp. 450–455, 2005.

[90] S. F. Warta, K. A. Kapalo, A. Best, and S. M. Fiore, "Similarity, complementarity, and agency in HRI: Theoretical issues in shifting the perception of robots from tools to teammates," *Proceedings of the Human Factors and Ergonomics Society*, pp. 1229–1233, 2016.

[91] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommet, G. Lucas, S. Marsella, F. Morbini, A. Nazarian, S. Scherer, G. Stratou, A. Suri, D. Traum, R. Wood, Y. Xu, A. Rizzo, and L.-P. Morency, "SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support," in *International Conference on Autonomous Agents and Multiagent Systems*, pp. 1061–1068, 2014.

[92] L. A. Dieker, J. A. Rodriguez, B. Lignugaris/Kraft, M. C. Hynes, and C. E. Hughes, "The Potential of Simulated Environments in Teacher Education: Current and Future Possibilities," *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, vol. 37, no. 1, pp. 21–33, 2013.

[93] M. Sagar, D. Bullivant, P. Robertson, O. Efimov, K. Jawed, R. Kalarot, and T. Wu, "A neurobehavioural framework for autonomous animation of virtual human faces," in *ACM SIGGRAPH Asia Workshop on Autonomous Virtual Humans and Social Robot for Telepresence*, 2014.

[94] M. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "Mach: My automated conversation coach," in *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 697–706, 2013.

[95] R. F. DeMara, A. J. Gonzalez, V. Hung, C. Leon-barth, R. A. Dookhoo, S. Jones, A. Johnson, J. Leigh, L. Renambot, S. Lee, and G. Carlson, "Towards Interactive Training with an Avatar-based Human-Computer Interface," in *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, pp. 8054:1–10, 2008.

[96] S. Lee, G. Carlson, S. Jones, A. Johnson, J. Leigh, and L. Renambot, "Designing an expressive avatar of a real person," in *Intelligent Virtual Agents* (J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, eds.), (Berlin, Heidelberg), pp. 64–76, Springer Berlin Heidelberg, 2010.

[97] W. R. Swartout, J. Gratch, R. W. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum, "Toward Virtual Humans," *AI Magazine*, vol. 27, no. 2, pp. 96–108, 2006.

[98] T. Bickmore and J. Cassell, "Relational agents: a model and implementation of building user trust," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 396–403, 2001.

[99] M. J. Hays, J. C. Campbell, M. A. Trimmer, J. C. Poore, and A. K. Webb, "Can Role-Play with Virtual Humans Teach Interpersonal Skills ?," in *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, no. 12318, pp. 1–12, 2012.

[100] M. E. Hoque and R. W. Picard, "Rich Nonverbal Sensing Technology for Automated Social Skills Training," *IEEE Computer*, vol. 47, no. 4, pp. 28–35, 2014.

[101] M. Mine, D. Rose, B. Yang, and W. D. Imagineering, "Projection-Based Reality in Disney Theme Parks," *IEEE Computer*, vol. 45, no. 7, pp. 32–40, 2012.

[102] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson, "Room2room: Enabling life-size telepresence in a projected augmented reality environment," in *ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp. 1716–1725, 2016.

[103] R. E. Guadagno, J. Blascovich, J. N. Bailenson, and C. Mccall, "Virtual Humans and Persuasion: The Effects of Agency and Behavioral Realism," *Media Psychology*, vol. 10, no. 1, pp. 1–22, 2007.

[104] K. Zibrek, E. Kokkinara, and R. McDonnell, "Don't stand so close to me: Investigating the effect of control on the appeal of virtual humans using immersion and a proximity-based behavioral task," in *ACM Symposium on Applied Perception*, pp. 3:1–11, 2017.

[105] H. J. Smith and M. Neff, "Communication Behavior in Embodied Virtual Reality," in *International Conference on Human Factors in Computing Systems*, pp. 1–12, 2018.

[106] J. N. Bailenson and N. Yee, "Digital Chameleons Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments," *Psychological Science*, vol. 16, no. 10, pp. 814–819, 2005.

[107] S. Y. Oh, K. Shriram, B. Laha, S. Baughman, E. Ogle, and J. Bailenson, "Immersion at Scale: Researcher's Guide to Ecologically Valid Mobile Experiments," in *IEEE Virtual Reality*, pp. 249–250, 2016.

[108] R. Torre, P. Fua, S. Balcisoy, M. Ponder, and D. Thalmann, "Interaction Between Real and Virtual Humans: Playing Checkers," in *Eurographics Workshop on Virtual Environments*, pp. 23–32, 2000.

[109] M. Obaid, R. Niewiadomski, and C. Pelachaud, "Perception of Spatial Relations and of Coexistence with Virtual Agents," in *International Conference on Intelligent Virtual Agents* (H. Vilhjálmsson, S. Kopp, S. Marsella, and K. Thórisson, eds.), vol. 6895 of *Lecture Notes in Computer Science*, pp. 363–369, Springer Berlin Heidelberg, 2011.

[110] D. Jo, K.-H. Kim, and G. J. Kim, "SpaceTime: adaptive control of the teleported avatar for improved AR tele-conference experience," *Computer Animation and Virtual Worlds*, vol. 26, no. 3–4, pp. 259–269, 2015.

[111] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst, "Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration," in *ACM CHI Conference on Human Factors in Computing Systems*, pp. 46:1–13, 2018.

[112] T. Holz, M. Dragone, and G. M. P. O'Hare, "Where robots and virtual agents meet: A survey of social interaction research across milgram's reality-virtuality continuum," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 83–93, 2009.

[113] Microsoft, "Fragments." `http://www.microsoft.com/microsoft-hololens/en-us/apps/Fragments`, September 18 2017.

[114] M. Madary and T. K. Metzinger, "Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology," *Frontiers in Robotics and AI*, vol. 3, no. 3, pp. 1–23, 2016.

[115] M. Garau, *The Impact of Avatar Fidelity on Social Interaction in Virtual Environments*. PhD thesis, Department of Computer Science, University of London, 2003.

[116] A. M. von der Pütten, N. C. Krämer, and J. Gratch, "Who's There? Can a Virtual Agent Really Elicit Social Presence?," in *International Workshop on Presence (PRESENCE)*, pp. 1–7, 2009.

[117] P. Kulms, N. C. Krämer, J. Gratch, and S.-H. Kang, "It's in Their Eyes: A Study on Female and Male Virtual Humans' Gaze," *Lecture Notes in Computer Science (Proceedings of International Conference on Intelligent Virtual Agents)*, vol. 6895, pp. 80–92, 2011.

[118] J. N. Bailenson, A. C. Beall, and J. Blascovich, "Gaze and Task Performance in Shared Virtual Environments," *The Journal of Visualization and Computer Animation*, vol. 13, no. 5, pp. 313–320, 2002.

[119] J. N. Bailenson, N. Yee, D. Merget, and R. Schroeder, "The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction," *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 4, pp. 359–372, 2006.

[120] L. Huang, L.-P. Morency, and J. Gratch, "Virtual Rapport 2.0," in *International Conference on Intelligent Virtual Agents*, vol. 6895 of *Lecture Notes in Computer Science*, pp. 68–79, Springer Berlin Heidelberg, 2011.

[121] N. Wang and J. Gratch, "Don't Just Stare at Me!," in *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 1241–1250, 2010.

[122] J. Li, "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*, vol. 77, no. C, pp. 23–37, 2015.

[123] J. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments," *Presence*, vol. 10, no. 6, pp. 583–598, 2001.

[124] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Interpersonal distance in immersive virtual environments," *Personality and Social Psychology Bulletin*, vol. 29, no. 7, pp. 819–833, 2003.

[125] F. A. Sanz, A.-H. Olivier, G. Bruder, J. Pettré, and A. Lécuyer, "Virtual Proxemics: Locomotion in the Presence of Obstacles in Large Immersive Projection Environments," in *IEEE Virtual Reality*, pp. 75–80, 2015.

[126] A. Bönsch, B. Weyers, J. Wendt, S. Freitag, and T. W. Kuhlen, "Collision Avoidance in the Presence of a Virtual Agent in Small-Scale Virtual Environments," in *IEEE Symposium on 3D User Interfaces*, pp. 145–148, 2016.

[127] M. J. Wieser, P. Pauli, M. Grosseibl, I. Molzow, and A. Mühlberger, "Virtual Social Interactions in Social Anxiety-The Impact of Sex, Gaze, and Interpersonal Distance," *CyberPsychology, Behavior & Social Networking*, vol. 13, no. 5, pp. 547–554, 2010.

[128] J. M. Loomis and J. M. Knapp, "Visual Perception of Egocentric Distance in Real and Virtual Environments," in *Virtual and Adaptive Environments* (L. Hettinger and M. Haas, eds.), pp. 21–46, Erlbaum, 2003.

[129] M. Garau, M. Slater, D.-P. Pertaub, and S. Razzaque, "The Responses of People to Virtual Humans in an Immersive Virtual Environment," *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 1, pp. 104–116, 2005.

[130] W. H. Warren, Jr., "Visually Controlled Locomotion: 40 Years Later," *Ecological Psychology*, vol. 10, no. 3–4, pp. 177–219, 1998.

[131] W. H. Warren, Jr. and B. R. Fajen, *Coordination: Neural, behavioral, and social dynamics*, ch. Behavioral Dynamics of Visually-Guided Locomotion, pp. 1–32. Heidelberg: Springer, 2008.

[132] B. R. Fajen and W. H. Warren, Jr., "Behavioral dynamics of steering, obstacle avoidance, and route selection," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 29, no. 2, pp. 343–362, 2003.

[133] C. T. Neth, J. L. Souman, D. Engel, U. Kloos, H. Bülthoff, and B. J. Mohler, "Velocity-dependent dynamic curvature gain for redirected walking," in *IEEE Virtual Reality*, pp. 151–158, 2011.

[134] R. S. Rosenberg, S. L. Baughman, and J. N. Bailenson, "Virtual Superheroes: Using Superpowers in Virtual Reality to Encourage Prosocial Behavior," *PLoS ONE*, vol. 8, no. 1, pp. 1–9, 2013.

[135] M. Obaid, I. Damian, F. Kistler, B. Endrass, J. Wagner, and E. André, "Cultural Behaviors of Virtual Agents in an Augmented Reality Environment," in *International Conference on Intelligent Virtual Agents* (Y. Nakano, M. Neff, A. Paiva, and M. Walker, eds.), vol. 7502 of *Lecture Notes in Computer Science*, pp. 412–418, Springer Berlin Heidelberg, 2012.

[136] G. Bente, S. Rüggenberg, and N. Krämer, "Social Presence and Interpersonal Trust in Avatar-Based, Collaborative Net-Communications," in *Annual International Workshop on Presence*, pp. 54–61, 2004.

[137] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg, "Avatar-mediated networking: Increasing social presence and interpersonal trust in net-based collaborations," *Human Communication Research*, vol. 34, no. 2, pp. 287–318, 2008.

[138] J. Riegelsberger, M. A. Sasse, and J. D. Mccarthy, “Rich media, poor judgement? A study of media effects on users’ trust in expertise,” in *People and Computers XIX—The Bigger Picture (Proceedings of HCI 2005)*, pp. 267–284, 2006.

[139] Y. Pan and A. Steed, “A Comparison of Avatar, Video, and Robot-Mediated Interaction on Users’ Trust in Expertise,” *Frontiers in Robotics and AI*, vol. 3, no. 12, pp. 1–12, 2016.

[140] G. M. Lucas, A. Rizzo, J. Gratch, S. Scherer, G. Stratou, J. Boberg, and L.-P. Morency, “Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers,” *Frontiers in Robotics and AI*, vol. 4, no. 51, pp. 1–9, 2017.

[141] J. Mell, G. Lucas, and J. Gratch, “Prestige questions, online agents, and gender-driven differences in disclosure,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10498 LNAI, pp. 273–282, 2017.

[142] J. Cassell, “Embodied conversational interface agents,” *Communications of the ACM*, vol. 43, no. 4, pp. 70–78, 2000.

[143] J. Cassell, “Conversational Agents Representation and Intelligence in User Interfaces,” *AI Magazine*, vol. 22, no. 4, pp. 67–84, 2001.

[144] J. Allwood, *Bodily Communication Dimensions of Expression and Content*, pp. 7–26. Dordrecht: Springer Netherlands, 2002.

[145] K. Tanaka, H. Nakanishi, and I. Hiroshi, “Appearance, motion, and embodiment: unpacking avatars by fine-grained communication analysis,” *Concurrency and Computation: Practice and Experience*, vol. 27, no. 11, pp. 2706–2724, 2015.

[146] M. Bunge, *Causality and Modern Science*. Dover Publications, 1979.

[147] H. Beebee, C. Hitchcock, P. Menzies, and L. A. Paul, "Counterfactual theories," 2010.

[148] M. Höfler, "The Bradford Hill considerations on causality: A counterfactual perspective," *Emerging Themes in Epidemiology*, vol. 2, pp. 1–9, 2005.

[149] M. Susser, "What is a Cause and How Do We Know One? A Grammar for Pragmatic Epidemiology," *American Journal of Epidemiology*, vol. 133, no. 7, pp. 635–648, 1991.

[150] M. Roussos, A. Johnson, J. Leigh, C. Vasilakis, and T. Moher, "Constructing Collaborative Stories Within Virtual Learning Landscapes," in *European Conference on AI in Education*, pp. 129–135, 1996.

[151] A. Nagendran, R. Pillat, K. Adam, G. Welch, and C. Hughes, "A Unified Framework for Individualized Avatar-Based Interactions," *Presence: Teleoperators & Virtual Environments*, vol. 23, no. 2, pp. 109–132, 2013.

[152] C. D. Kidd and C. Breazeal, "Effect of a Robot on User Perceptions," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3559–3564, 2004.

[153] H. Knight and R. Simmons, "Estimating Human Interest and Attention via Gaze Analysis," in *IEEE International Conference on Robotics and Automation*, pp. 4350–4355, 2013.

[154] E. Rodriguez-Lizundia, S. Marcos, E. Zalama, J. Gómez-García-Bermejo, and A. Gordaliza, "A bellboy robot: Study of the effects of robot behaviour on user engagement and comfort," *International Journal of Human-Computer Studies*, vol. 82, pp. 83–95, Oct 2015.

[155] I. B. Myers, *The Myers-Briggs Type Indicator: Manual*. Palo Alto: Consulting Psychologists Press, 1962.

[156] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater, "An experimental study on the role of touch in shared virtual environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 4, pp. 443–460, 2000.

[157] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[158] R. A. Likert, "A technique for the measurement of attitudes," *Arch. Psychol.*, vol. 22, pp. 5–55, 1932.

[159] N. Blaikie, *Analysing Quantitative Data.* London: Sage Publications, 2003.

[160] T. R. Knapp, "Treating ordinal scales as interval scales: An attempt to resolve the controversy," *Nursing Research*, vol. 39, pp. 121–123, 1990.

[161] K. Kim, G. Bruder, D. Maloney, and G. Welch, "The Influence of Real Human Personality on Social Presence with a Virtual Human in Augmented Reality," in *International Conference on Artificial Reality & Telexistence and Eurographics Symposium on Virtual Environments*, pp. 115–122, 2016.

[162] W. Steptoe, S. Julier, and A. Steed, "Presence and discernability in conventional and non-photorealistic immersive augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality*, pp. 213–218, 2014.

[163] Microsoft, "Spatial mapping (windows mixed reality)." `https://developer.microsoft.com/en-us/windows/mixed-reality/spatial_mapping`, April 22 2017.

[164] K. Kim, "Improving Social Presence with a Virtual Human via Multimodal Physical–Virtual Interactivity in AR," in *Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems (Student Research Competition)*, pp. SRC09:1–6, 2018.

[165] A. Frischen, A. P. Bayliss, and S. P. Tipper, "Gaze Cueing of Attention," *Differences*, vol. 133, no. 4, pp. 694–724, 2007.

[166] M. Lee, K. Kim, S. Daher, A. Raij, R. Schubert, J. Bailenson, and G. Welch, "The Wobbly Table: Increased Social Presence via Subtle Incidental Movement of a Real-Virtual Table," in *Proceedings of the IEEE Virtual Reality*, pp. 11–17, 2016.

[167] K. Kim, R. Schubert, and G. Welch, "Exploring the impact of environmental effects on social presence with a virtual human," in *International Conference on Intelligent Virtual Agents (Lecture Notes in Computer Science)*, vol. 10011, pp. 470–474, 2016.

[168] M. Lee, G. Bruder, T. Höllerer, and G. Welch, "Effects of unaugmented periphery and vibrotactile feedback on proxemics with virtual humans in ar," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1525–1534, 2018.

[169] K. Kim, M. Billinghurst, G. Bruder, H. Been-Lirn Duh, and G. F. Welch, "Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2947–2962, 2018.

[170] N. Norouzi, K. Kim, J. Hochreiter, M. Lee, S. Daher, G. Bruder, and G. Welch, "A Systematic Survey of 15 Years of User Studies Published in the Intelligent Virtual Agents Conference," in *ACM International Conference on Intelligent Virtual Agents*, pp. 17–22, 2018.

[171] D. H. Mcknight, M. Carter, J. B. Thatcher, and P. F. Clay, "Trust in a Specific Technology: An Investigation of Its Components and Measures," *ACM Transactions on Management Information Systems*, vol. 2, no. 2, pp. 12:1–25, 2011.

[172] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency, "It's only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior*, vol. 37, no. August, pp. 94–100, 2014.

[173] N. Bos, J. S. Olson, D. Gergle, G. M. Olson, and Z. Wright, “Effects of Four Computer-Mediated Communications Channels on Trust Development,” in *Proceedings of CHI*, pp. 135–140, 2002.

[174] N. I. Badler, “Virtual Beings,” *Communications of the ACM*, vol. 44, no. 3, pp. 33–35, 2001.