Does a Digital Assistant Need a Body? The Influence of Visual Embodiment and Social Behavior on the Perception of Intelligent Virtual Agents in AR

Kangsoo Kim*
University of Central Florida

Luke Boelling†
University of Münster

Steffen Haesler‡
University of Würzburg

Jeremy N. Bailenson§
Stanford University

Gerd Bruder¶
University of Central Florida

Gregory F. Welch‖
University of Central Florida

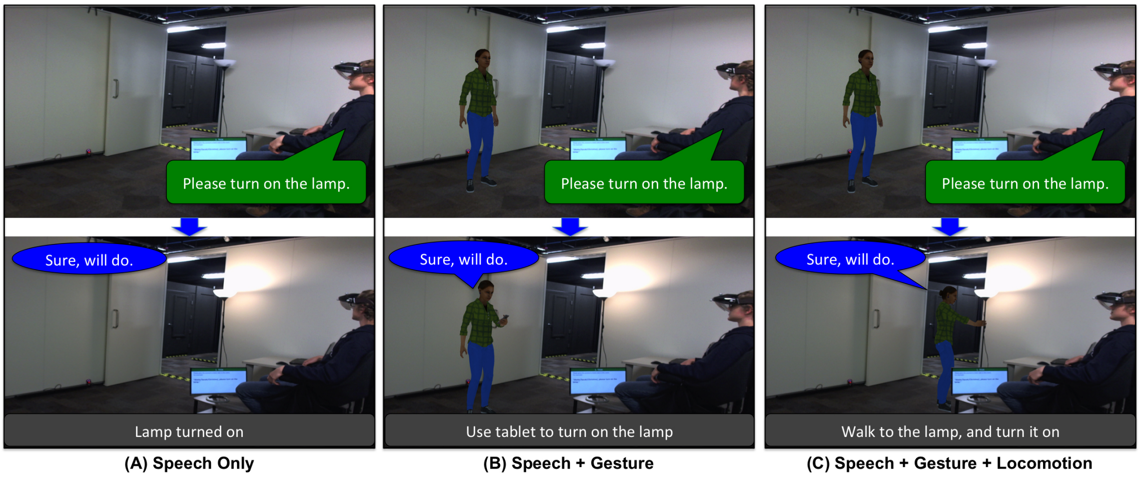


Figure 1: Illustration of the experimental conditions for the simple task to turn on a floor lamp (here equipped with an Internet of Things connected Philips light bulb). Participants interacted with three different types of intelligent virtual agents: (A) Speech only: a disembodied voice, (B) Speech+Gesture: an embodied AR agent that remained stationary in one place but used upper-body gestures, or (C) Speech+Gesture+Locomotion: an embodied AR agent that used gestures and naturally walked about the physical space.

Abstract

Intelligent Virtual Agents (IVAs) are becoming part of our everyday life, thanks to artificial intelligence technology and Internet of Things devices. For example, users can control their connected home appliances through natural voice commands to the IVA. However, most current-state commercial IVAs, such as Amazon Alexa, mainly focus on voice commands and voice feedback, and lack the ability to provide non-verbal cues which are an important part of social interaction. Augmented Reality (AR) has the potential to overcome this challenge by providing a visual embodiment of the IVA.

In this paper we investigate how visual embodiment and social behaviors influence the perception of the IVA. We hypothesize that a user's confidence in an IVA's ability to perform tasks is improved when imbuing the agent with a human body and social behaviors compared to the agent solely depending on voice feedback. In other words, an agent's embodied gesture and locomotion behavior exhibiting awareness of the surrounding real world or exerting influence over the environment can improve the perceived social presence with and confidence in the agent. We present a human-subject study, in which we evaluated the hypothesis and compared different forms of IVAs with speech, gesturing, and locomotion behaviors in an interactive AR scenario. The results show support for the hypothesis with measures of confidence, trust, and social presence. We discuss implications for future developments in the field of IVAs.

Keywords: Intelligent virtual agents, digital assistants, social interaction, presence, confidence, trust in technology, augmented reality.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities; J.4 [Computer Applications]: Social and Behavioral Sciences—Psychology

*e-mail: kskim@knights.ucf.edu (corresponding author)
†e-mail: l.boelling@uni-muenster.de
‡e-mail: steffen.haesler@stud-mail.uni-wuerzburg.de
§e-mail: bailenso@stanford.edu
¶e-mail: bruder@ucf.edu
‖e-mail: welch@ucf.edu

1 Introduction

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 [33] discussed that many people interact with computers in a "fundamentally social and natural" way. In this scope, Intelligent Virtual Agents (IVAs) 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.

978-1-5386-7459-8/18/$31.00 ©2018 IEEE
DOI 10.1109/ISMAR.2018.00039

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. Augmented Reality (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 [35] 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. [6] 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 IVA, and in turn, the perceived confidence and trust in the IVA could be improved. Moreover, an AR visual body and behaviors provide the opportunity to naturally convey the notion that the agent 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. Kim and Welch [21] illustrated the importance of physical–virtual interactivity in AR with IVAs, particularly emphasizing the environmental awareness (sensing) and influence (affecting) with the surroundings.

In this paper, we investigate how an IVA's visual embodiment in AR and social behaviors in the physical environment influence the perception of the IVA. 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 three 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 support for our hypothesis, and we discuss implications and guidelines for future developments in the field of IVAs. A large body of literature focused on human perception of virtual content, including virtual agents, but most research was conducted in Virtual Reality (VR) environments and lack the important aspect of physical–virtual interactivity and its influence in AR. Here we focus on human perception, particularly trust or confidence in and social presence with IVAs, which are influenced by the agent's embodiment and social behavior in AR.

2 Related Work

In this section, we resume previous work on agent appearance and embodiment, awareness and influence, as well as presence, trust, and confidence in the field of IVAs.

2.1 Appearance and Embodiment

The psychological implications of the level of embodiment of IVAs have been studied for decades, with an early qualitative review [13] showing mixed benefits for interfaces that featured embodied agents—sometimes they enhanced an application while otherwise not. A quantitative meta-analysis by Yee et al. [39] demonstrated that the effect size for adding a face (as opposed to just voice or text) was larger than the effect of realism (i.e., there was more gain in influence from having a face than for making that face more photographically or behaviorally realistic). Dozens of studies have examined the psychological gain of embodied agents in VR, with recent work demonstrating the gains in rapport and nonverbal realism [19, 38]. Demeur et al. [14] evaluated the impact of an agent's embodiment and emotion over the perceived social believability in the agent, and found that appropriate emotions conveyed through the agent's embodiment, particularly related to the sense of competence and warmth, could lead to higher believability. Latoschik et al. [22] investigated the effects of the appearance of avatars, and higher body ownership when using a visually realistic avatar compared to an abstract wooden mannequin avatar.

Although there have been many studies with virtual avatars and agents in VR, few studies examined the implications of agent appearance and embodiment in AR. In VR, the levels of realism of an agent are typically roughly matched to the rest of the scene, in terms of polygons, textures, and general scene complexity, since they are all being rendered by the same system. In AR, the embodied agent has the disadvantage of being contrasted with the real world, and there are huge implications for how an agent is utilized and perceived in this context.

2.2 Awareness and Influence in Context

Dey and Abowd [15] described *context* as "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." Blascovich et al. [8] stated that "context" is one of the parameters that determine how a user evaluates an agent in their theoretical model, and determines how influential an embodied agent should be. Maedche et al. [26] particularly highlighted the importance of context-awareness in intelligent assistant systems, e.g., virtual assistants like Apple Siri, to support individuals in daily life. Especially, in AR where the users interact with IVAs in physical environments, the physical context of the agent and its contextually-appropriate behaviors (context-awareness and influence) are more important for the perception of the agent than in VR, given the large discrepancy between digital humans superimposed on real objects while sharing the physical environment. Despite its importance, "context" is a variable that has received little empirical attention compared to other parameters such as agency (i.e., if a human or computer is controlling the virtual character) or realism (i.e., how detailed the textures and movements are) [9].

In VR and robotics, research has shown that agents that are aware of the user elicit more social presence and better performance than those that are not. For example, early work by Bailenson et al. [2, 3] showed that embodied agents who maintained mutual gaze with a user caused the user to respect the agent's personal space and also elicited higher self-report measures of influence. Mutlu et al. [28] found that participants performed better in recalling a robotic agent's story when the agent looked at them more in an engaging storytelling scenario. While there were some studies examining situationally aware and influential agents in VR and robotics, examining what awareness and influence of agents means in AR still remains largely unexplored beyond the mere use of virtual characters in AR [37].

Only comparatively few studies focused on an agent's interaction with the surrounding environment in AR. Chuah et al. [12] discussed (1) awareness of changes to the environment and (2) ability to influence the environment with an embodied conversational agent consisting of a physical mannequin's legs under a monitor showing a computer graphics representation of the agent's upper-body. For awareness of the environment, Kim et al. [17, 18, 20] demonstrated that an embodied agent in AR exhibiting awareness of objects in a physical room (e.g., awareness of wind from a fan or accurate collision avoidance) elicited higher social presence ratings than an agent who did not demonstrate awareness. Abilities to influence the environment encompass *physical influence*, when an agent is affecting a physical object, *digital influence*, when a digital process such as a video recording is started or stopped, and *social influence*, when a person's emotions, opinions, or behaviors are affected by the agent [8]. For instance, Lee et al. [23] showed that special-purpose devices built to enable agents in AR to move physical objects such as a wobbly table can have positive effects on co-presence.

O'Hare et al. [31] presented an AR agent framework that virtually sensed the world and could blur the traditional boundaries between the virtual and physical domain. Amores et al. [1] built a smart lamp that was controlled by an AR agent. Nowadays, we can more easily give IVAs basic abilities to control things in the physical world without the need to develop special-purpose hardware by repurposing existing Internet of Things (IoT) devices, e.g., switching an IoT light bulb on/off, which is just a state change in the connected device that can be triggered via the network.

We predict that future generations being used to IVAs will be more trusting and confident that agents can influence the physical world, but assume that most users today will benefit from such physically-grounded behaviors. Thus, in this paper, we set a scenario where IVAs are interacting or at least pretend to interact with the physical environment, e.g., IVAs having awareness of environmental status and ability to change the environment, such as controlling a real light bulb via IoT technology. Given the interactive setup, we explore the question how the perceived trust and confidence in these abilities of IVAs can change based on their embodiment and social behavior with respect to the user and environment.

2.3 Presence, Trust, and Confidence

Different concepts have been introduced to understand and measure the effectiveness of embodied agents in conveying an illusion of being real. In particular, Slater [36] introduced the concepts of *place illusion* and *plausibility illusion* which together define the sense of being *co-present* together with the agent. According to Slater [36], the former illusion is largely related to factors such as agent appearance and embodiment as well as its awareness of the real world, whereas the latter illusion 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." Different measures of social or co-presence with embodied agents or objects have been proposed over the years [4, 7, 16, 29, 30]. In this paper, we focus on the user's self-reported sense of social presence with the agents using the well-established Temple Presence Inventory post-experience questionnaire [24].

The perceived trust and confidence in IVAs are often researched together with the level of social presence because of their potential correlation. Bente et al. [5, 6] observed an increased interpersonal trust along with a strong social presence in network-based communications using embodied virtual representations, while Riegelsberger et al. [34] showed that an embodied virtual representation still elicits a lower level of trust than a video conference setting with real humans. Pan and Steed [32] also compared three different forms of communication including embodied virtual 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. These results emphasize the importance of the IVA's social behaviors with respect to the surrounding physical environment, which we focus on in this paper.

3 Experiment

In this section we present the experiment which we conducted to investigate the effects of embodiment and social behaviors on the perception of IVAs during social interaction.

3.1 Participants

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 from 1=no to 7=much experience, participants reported a medium experience with IVAs such as Amazon Alexa or Google Assistant ($M = 3.3$, $SD = 2.2$) and a medium experience with Smart Home or IoT devices ($M = 3.1$, $SD = 1.5$).

3.2 Material

Here we describe the details of our IVA implementation and physical setup in a room-sized experimental space that we prepared for the experiment.

3.2.1 Intelligent Virtual Agents

In this experiment, 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 not to 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—who was seated outside the room and observed the participant via live video and audio streams—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. This human-in-the-loop mechanism (i.e., Wizard of Oz paradigm) 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. We prepared three forms of IVAs under this human-in-the-loop framework, but they differed in embodiment and social behaviors:

(A) IVA **S** (**S**peech): was designed not to have any visual feedback (i.e., no visual body appearance) but only rely on voice communication (see Figure 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). The IVA S thus could not rely on embodied human gestures or locomotion to convey aspects of social interaction.

(B) IVA **SG** (**S**peech and **G**esturing): was implemented based on the IVA S, but included visual embodiment with the animated 3D character described above. The IVA SG was inspired by popular IVAs like Amazon Alexa, which are "embodied" with the body of a common home appliance, which is placed by the owner at a position in the room, such as in a corner of the room or next to a TV screen. In our study, our agent had a full virtual 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 1-B). While this IVA remains stationary in the environment, it has

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

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.

(C) The IVA **SGL** (**S**peech, **G**esturing, and **L**ocomotion): 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 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 1-C).

3.2.2 Physical Setup

As illustrated in Figure 2, we used a physical room-like experiment space with an area of 3.89 m × 3.89 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.

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.

3.3 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

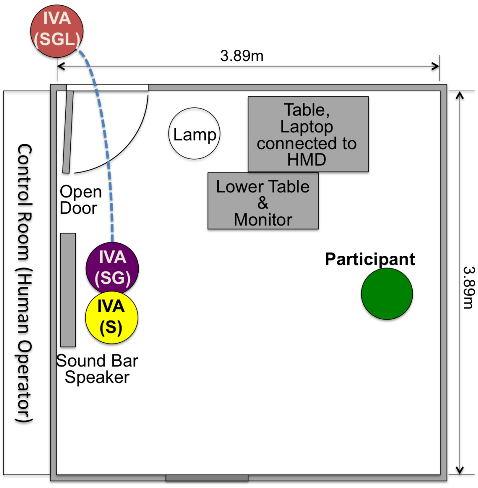


Figure 2: Experimental setup: Participants interacted with one of the IVA types: (S)peech only, (S)peech+(G)esture, or (S)peech+(G)esture+(L)ocomotion, across the experiment room. The embodied IVAs were located on the other side of the room facing toward the participant. The IVA with the visual embodiment in condition SGL could walk around freely and leave the room, while the IVA in condition SG remained stationary in one spot. The IVA with the disembodied voice in condition S did 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.

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 3 and detailed descriptions in Section 3.2.1):

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;

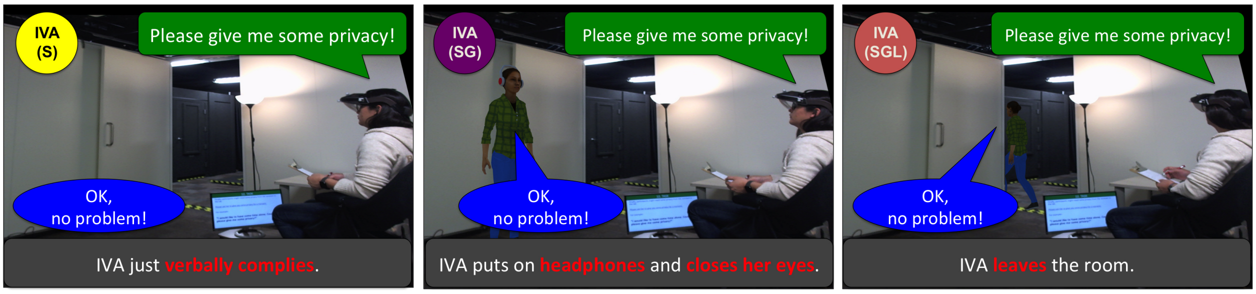
SGL The IVA SGL with an AR body and social behavior including Speech, Gesturing, and Locomotion.

3.3.1 General Procedure

Before the experiment, participants gave their informed consent and were guided into the experimental room (see Figure 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 randomized 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.

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.



(A) Speech **(B) Speech + Gesture** **(C) Speech + Gesture + Locomotion**

Figure 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.

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.

3.3.2 Interaction Scenario

The main story 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, 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. The participant's task was to ask the agent to turn on the lamp:

- 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 durations of these behaviors 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

[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.

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. 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 durations of each of these social behaviors matched between the conditions.

3.3.3 Measures

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* (T1) 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 [27], 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 [24], 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.

3.3.4 Hypotheses

As we described in Section 1, 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 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. [25]. 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).

4 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 1 shows the results for the subscales in the questionnaires for the three conditions.

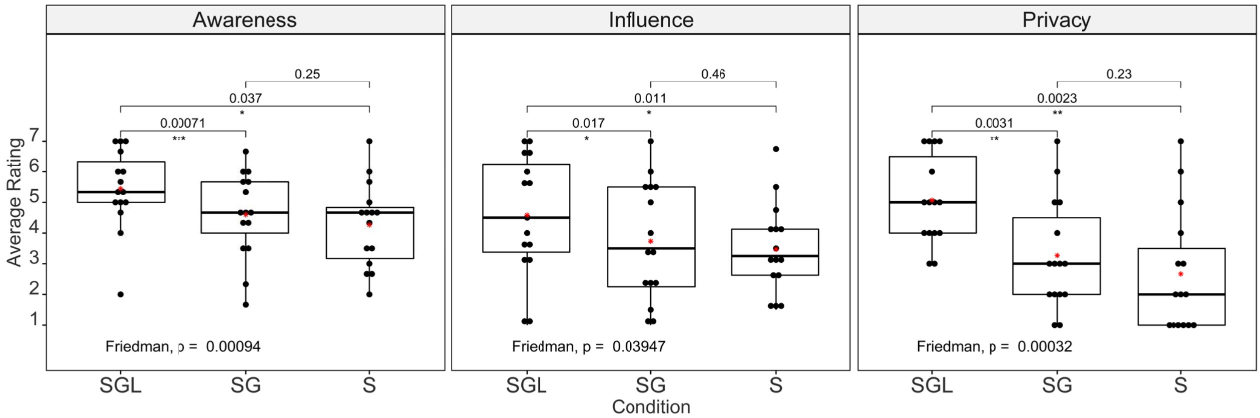


Figure 4: Boxplots of averaged means per participant and subscales of the confidence questions related to *awareness*, *influence*, and *privacy* that were asked during the experiment. Significant differences of the pairwise comparisons are indicated in the plots.

Table 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)

4.1 Awareness, Influence, and Privacy

Figure 4 shows plots for the subscales *awareness*, *influence*, and *privacy*. In this figure, we excluded the non-significant subscale of the participants' willingness to *share private data* (see below).

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 the SGL and S conditions ($p = 0.0169$). The other pair-wise comparisons did not show any statistically significant differences, SG and S ($p = 0.099$), and SGL and SG ($p = 0.551$), such that we did not investigate further.

4.2 Trust in Technology

Figure 5(a) 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 ($p = 0.37$).

4.3 Social Presence

Figure 5(b) 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 ($p = 0.099$).

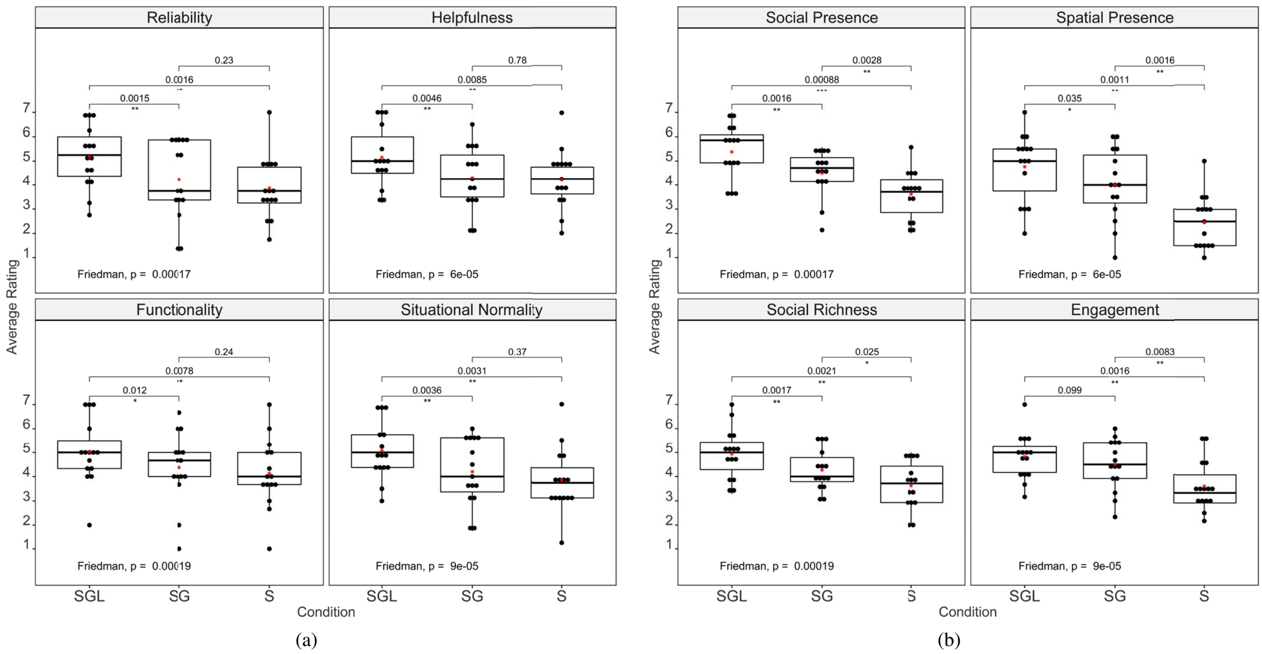


Figure 5: Boxplots of averaged means per participant and subscales related to the (a) Trust in Technology questionnaire and (b) Temple Presence Inventory questionnaire. Significant differences of the pairwise comparisons are indicated in the plots.

5 Discussion

The results of our 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.

5.1 Awareness and Influence

The results of the experiment indicate partial support for our Hypothesis H1. 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.

One possible interpretation of the results is that it was the *combination* of the (1) visual embodiment and (2) ability of the embodied agent to move about the physical space that caused the higher confidence, whereas without either of these, the confidence was significantly decreased. Support for this possible interpretation comes from 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. Locomotion and, in particular, *walking* is generally considered the most natural and intuitive form of behavior when performing an action between different spatial locations in one's immediate environment. This type of interaction was supported only by the agent in condition SGL. Our results suggest that such natural behaviors could be highly important for eliciting a high sense of confidence in the agent's awareness and influence.

In condition SGL, the agent's representation and behavior provided the most visual social cues to the participants, suggesting that the task was understood, performed, and completed successfully, in line with *seeing is believing* (e.g., [11]). 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 the participants' confidence in the agent's 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 in the action being completed successfully, 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 might not need an embodied virtual 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 might benefit from 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.

5.2 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. 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. Future studies should look into this aspect, which might be correlated with situational factors, prior experience, and the user's demographics, e.g., age, education, or technological understanding.

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.

5.3 Willingness to Share Private 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 observed 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 [27] 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. These results might be related to a study conducted by Bos et al. [10]. 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 have been increased.

5.4 Social Presence

The results of the experiment support our Hypothesis H4. 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 had the latent feeling that the agent in the SG condition was not polite or even ignoring them.

This observation is related to Riva's [35] definition of social presence, i.e., understanding another person's intentions (see Section 1). 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.

5.5 Limitations

The results of our study suggest interesting effects related to human social interaction in the presence of current-state or future implementations of IVAs in AR. However, it is important to note that there are limitations inherent to our and related studies.

For one, studies involving technologies such as IVAs, IoT, and AR are subject to preconceived notions, which might stem from personal experience with commercially available devices and their various limitations, the way the current state of technology is perceived due to its presentation in the media, or even futuristic visions of these technologies as presented in science fiction formats such as the movie *Her* (2013) or the movie *Blade Runner 2049* (2017). We tried our best to avoid the confounds that we identified as most important, e.g., by using a human-in-the-loop design instead of relying on a commercial conversational agent—thus avoiding negative connotations such as cloud-related data gathering and sharing—but it is not possible to avoid all of these potential confounds related to preconceived notions or the fact that some of these realizations will be perceived as more novel than others. Our laboratory environment might have influenced the results by setting the participant's expectation on the IVA as an experimental system. Another limitation of our results is that they were collected in a controlled laboratory study, which is not able to assess whether our observed effects would persist over multiple days or weeks or would adjust based on positive or negative experiences. Impressions such as helpfulness and trust are likely built and shaped over time.

6 Conclusion

In this paper, 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 indicate that imbuing an agent with a visual body in AR and natural social behaviors could increase 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 positive effect on the users' confidence that the agent will respect one's privacy when one requests it. In our case, the agent walked out of the room to give the user some privacy, which closely matches the natural behavior exhibited by real people in such cases, and led to a higher confidence in privacy among the 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 guide us to interesting research directions about the effects of a virtual agent's embodiment and behavior on human perception in AR. In future work, we plan to further investigate the interplay between an agent's awareness of the real world, influence over physical objects and events, and resulting benefits for different application domains. Further interesting directions for future work include longitudinal studies, comparisons between abstract and real entities in AR, and investigations of the question whether a social human agent might increase the user's forgiveness in errors made by the computer.

Acknowledgments

The work presented in this publication is supported by the Office of Naval Research (ONR) Code 30 under Dr. Peter Squire, Program Officer (ONR awards N00014-17-1-2927). We also acknowledge Florida Hospital for their support of Prof. Welch via their Endowed Chair in Healthcare Simulation.

References

[1] J. Amores, X. Benavides, M. Comin, A. Fuste, P. Pla, and D. Miralles. Smart Avatars: Using Avatars to Interact With Objects. In *Proceedings of ACM CHI*, pages 1–3, 2012.

[2] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments. *Presence: Teleop Virt*, 10(6):583–598, 2001.

[3] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833, 2003.

[4] 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: Teleop Virt*, 14(4):379–393, 2005.

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

[6] 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*, 34(2):287–318, 2008.

[7] F. Biocca, C. Harms, and J. K. Burgoon. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleop Virt*, 12(5):456–480, 2003.

[8] J. Blascovich. Social influence within immersive virtual environments. In R. Schroeder, editor, *The Social Life of Avatars*, Computer Supported Cooperative Work, pages 127–145. Springer London, 2002.

[9] J. Blascovich and C. McCall. *Attitudes in Virtual Reality*, pages 283–297. New York, NY: Psychology Press, 2017.

[10] N. Bos, J. S. Olson, G. M. Olson, Z. Wright, and D. Gergle. Rich Media Helps Trust Development. In *Proceedings of ACM CHI*, pages 135–140, 2002.

[11] S. Brownlow. Seeing is believing: Facial appearance, credibility, and attitude change. *J Nonverbal Behav*, 16(2):101–115, 1992.

[12] 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: Teleop Virt*, 22(2):141–170, 2013.

[13] D. Dehn and S. van Mulken. the impact of animated interface agents: A review of empirical research. *Int J Hum Comput Stud*, pages 1–22.

[14] V. Demeure, R. Niewiadomski, and C. Pelachaud. How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? *Presence: Teleop Virt*, 20(5):431–448, 2011.

[15] A. K. Dey and G. D. Abowd. Towards a Better Understanding of Context and Context-Awareness. *Comp Sys*, 40(3):304–307, 1999.

[16] C. Harms and F. Biocca. Internal Consistency and Reliability of the Networked Minds Measure of Social Presence. In *Annual International Presence Workshop*, pages 246–251, 2004.

[17] 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 and Technology*, pages 1–7, 2017.

[18] 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. *Comput Animat Virt W*, 28(e1771), 2017.

[19] 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. *Front. Robot. AI*, 4(32):1–20, 2017.

[20] 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*, pages 470–474, 2016.

[21] K. Kim and G. Welch. Maintaining and Enhancing Human-Surrogate Presence in Augmented Reality. In *IEEE ISMAR Workshop on Human Perception and Psychology in Augmented Reality*, pages 15–19, 2015.

[22] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The Effect of Avatar Realism in Immersive Social Virtual Realities. In *ACM Symposium on Virtual Reality Software and Technology*, pages 39:1–10, 2017.

[23] M. Lee, K. Kim, S. Daher, A. Raij, R. Schubert, J. N. Bailenson, and G. Welch. The wobbly table: Increased social presence via subtle incidental movement of a real-virtual table. In *Proceedings of IEEE Virtual Reality (VR)*, pages 11–17, 2016.

[24] M. Lombard, T. B. Ditton, and L. Weinstein. Measuring presence: The temple presence inventory. In *International Workshop on Presence*, pages 1–15, 2009.

[25] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency. It's only a computer: Virtual humans increase willingness to disclose. *Comput Human Behav*, 37:94–100, 2014.

[26] A. Maedche, S. Morana, S. Schacht, D. Werth, and J. Krumeich. Advanced user assistance systems. *Bus Inf Syst Eng*, 58(5):367–370, 2016.

[27] 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 Trans. Manage. Inf. Syst.*, 2(2):12:1–25, 2011.

[28] B. Mutlu, J. Forlizzi, and J. Hodgins. A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior. In *IEEE-RAS International Conference on Humanoid Robots*, pages 518–523, 2006.

[29] K. Nowak. Defining and Differentiating Copresence, Social Presence and Presence as Transportation. In *International Workshop on Presence*, pages 1–23, 2001.

[30] K. Nowak and F. Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleop Virt*, 12(5):481–494, 2003.

[31] G. M. P. O'Hare, A. Campbell, J. Stafford, and R. Aiken. NeXuS: Behavioural realism in mixed reality scenarios through virtual sensing. *Annual Conference on Computer Animation and Social Agents*, 2005.

[32] Y. Pan and A. Steed. A Comparison of Avatar, Video, and Robot-Mediated Interaction on Users' Trust in Expertise. *Front. Robot. AI*, 3(12), 2016.

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

[34] 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)*, pages 267–284, 2006.

[35] G. Riva. Is Presence a Technology Issue? Some Insights from Cognitive Sciences. *Virtual Reality*, 13(3):159–169, 2009.

[36] M. Slater. Place Illusion and Plausibility can Lead to Realistic Behaviour in Immersive Virtual Environments. *Philos Trans R Soc Lond B Biol Sci*, 364(1535):3549–3557, 2009.

[37] D. Wagner, M. Billinghurst, and D. Schmalstieg. How Real Should Virtual Characters Be? *Conference on Advances in Computer Entertainment Technology*, 2006.

[38] Y. Wang, G. Lucas, P. Khooshabeh, C. M. de Melo, and J. Gratch. Effects of emotional expressions on persuasion. *Social Influence*, 10(4):236–249, 2015.

[39] N. Yee, J. N. Bailenson, and K. Rickertsen. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In *Proc. of ACM CHI*, pages 1–10, 2007.