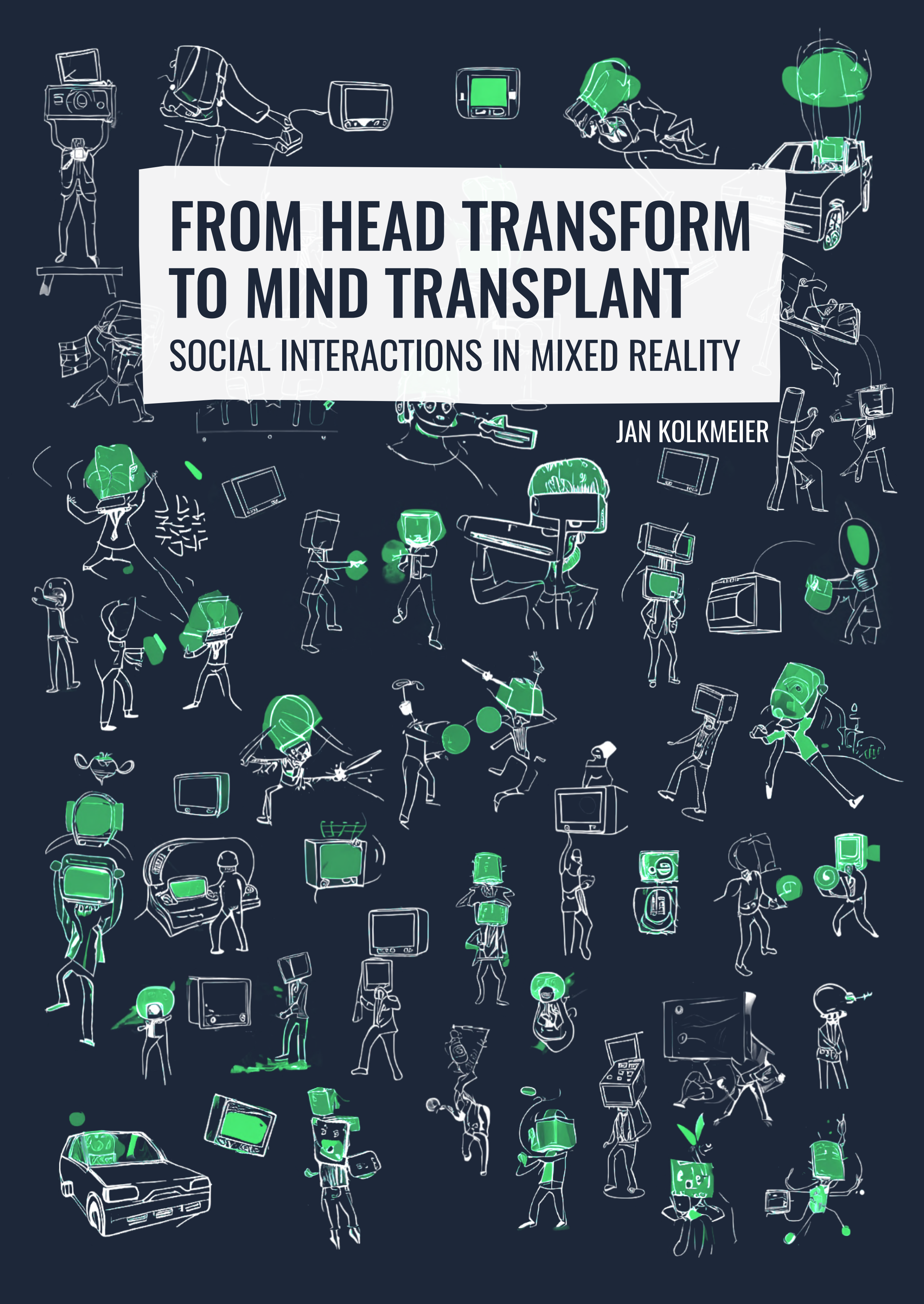


FROM HEAD TRANSFORM TO MIND TRANSPLANT

SOCIAL INTERACTIONS IN MIXED REALITY

JAN KOLKMEIER

Graduation committee

Chairman and Secretary:
prof.dr. J.N. Kok

Supervisor:

prof.dr. D.K.J. Heylen	Universiteit Twente

Members:

prof. dr. V. Evers	Universiteit Twente
dr.ir. D. Reidsma	Universiteit Twente
prof. dr. W.A. IJsselsteijn	Technische Universiteit Eindhoven
prof. dr. T. Bosse	Radboud Universiteit
prof. dr. M.P. Neff	University of California Davis
prof. dr. C. Pelachaud PhD	Université Pierre et Marie Curie

Paranymphs:
Thomas Beelen
Emiel Harmsen
David Goedicke

FROM HEAD TRANSFORM TO MIND TRANSPLANT: SOCIAL INTERACTIONS IN MIXED REALITY

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board,
to be publicly defended
on Thursday 29 September 2022 at 16:45 hours

by **Jan Kolkmeier**

born on the 19th of February 1988
in Osnabrück, Germany

29.09.2022

This dissertation has been approved by:
Supervisor:
prof.dr. D.K.J. Heylen

The research reported in this dissertation was carried out at the Human Media Interaction group of the University of Twente.

This work was supported by the 4TU Research Center Humans & Technology.

SIKS Dissertation Series No. 2022-29. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

DSI Ph.D. Thesis Series No. 22-006. ISSN 2589-7721. Digital Society Institute. P.O. Box 217, 7500 AE Enschede, NL

Illustrations & Cover design:	Sarah Stegmann, Jan Kolkmeier & Stable Diffusion
Typeset:	With XeLaTeX
Printed by:	Ipskamp Printing

ISBN: 978-90-365-5451-0
DOI: 10.3990/1.9789036554510

ABSTRACT

Voice and video calls are part of most people's everyday life. These are examples of technology being used to mediate social interaction between humans. Virtual and augmented reality technologies (VR and AR) play an increasing role in new applications for mediated social interaction. VR and AR can be used to improve upon conventional communication technologies by creating an increased sense of being together in a shared (virtual) space. For interactions where user's physical surroundings are important, mixed reality (MR) can offer further improvements by integrating the physical world into the experience. Another addition can be artificial virtual humans that are controlled by the application, but that can be interacted with using language and gestures. Examples are VR training systems with intelligent virtual humans that allow for practicing skills such as effective negotiation or public speaking. MR telepresence systems allow for people to solve complex problems remotely that previously required physical presence.

In this thesis, we have explored different and novel forms of mediated social interactions that are using these technologies. We have looked at how they could play out in the future, in a number of different contexts. These included (simulated) face-to-face interactions with (virtual) humans, social conflict situations, as well as remote collaboration tasks with other humans. Besides interactions with artificial virtual humans, and interactions with virtual avatars that embody real humans, we have also explored the concept of interacting with humans that are embodying the mind of artificial actors.

Most research was conducted in the form of lab studies, using working prototypes. We have employed qualitative and quantitative methods for evaluation, and a variety of measures across these studies. These measures included physiological responses, how participants perceived the interaction, how they perceived other interlocutors, their feeling of immersion and their perception of social presence with other actors.

The varied findings across these studies are reflected upon through the lens of a conceptual, science-fiction-inspired telepresence system: the SoIC-TV. By developing this concept throughout the thesis, we illustrate how in the future, mediated interaction experience using MR technology might transform beyond what is familiar both from co-located and present technology-mediated interactions.

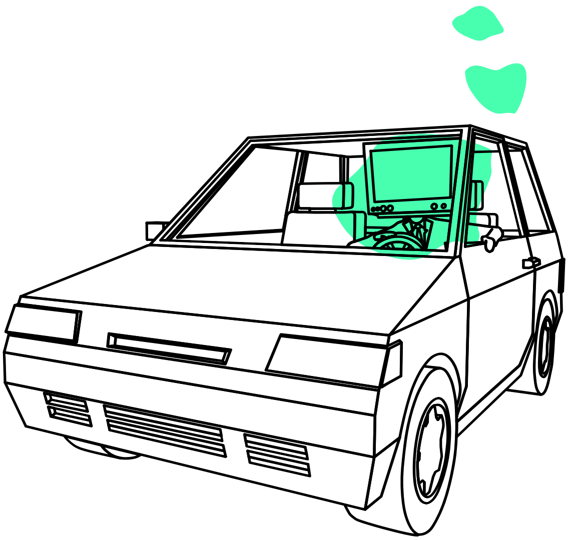
CONTENTS

PART *I*

INTRODUCTION

1

TECHNOLOGY, SCIENCE FICTION & DREAMS



A new camera sensor is released that tracks human pose and motion. A Virtual Reality (VR) headset comes out with a plugin to be used with the game engine I am already familiar with. Software for easy tracking of Augmented Reality (AR) markers, and rendering of AR content on mobile devices is published. When I started studying and working in the field of HCI (Human Computer Interaction), several hallmark products such as the Kinect sensor, the Oculus Rift headset, or the Vuforia AR engine followed each other. Not specialized hardware, but customer-ready products which were also friendly to work with for any amateur developer or student with some limited programming knowledge. These releases still spark my imagination, getting me to think about what I can now build or experience that I could not before.

Figure 1.1. The Star Trek Holodeck (left) and Star Wars Holograms.

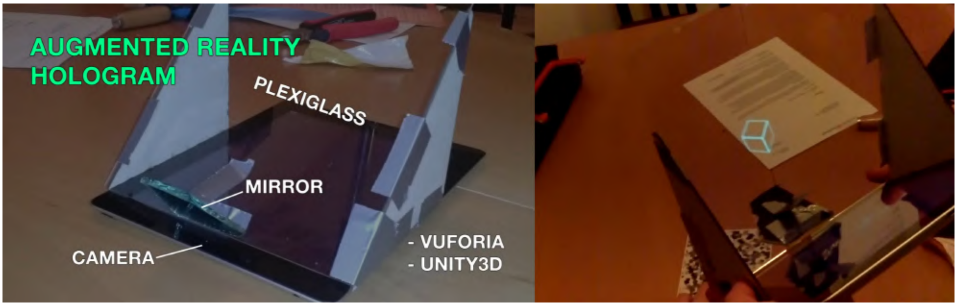


Figure 1.2. Combining Pepper's Ghost illusion with marker-based AR to build my first *holographic* display.

Thinking back, the first applications I wanted to try and realize with these new technologies were often modeled on those that were already planted in our heads from the world of science fiction media. For example, the desire to explore holographic displays, such as my first DIY attempt shown in Figure 1.2, was certainly informed by R2D2's famous hologram in Star Wars (Figure 1.1).

These explorations were most satisfactory when they would achieve something beyond their intended, out-of-the-box use, perhaps by misusing them or recombining them with something else. This happened on a regular basis during our weekly *WorkSPACE* workshops in the *SmartXP-Lab*, where me and a couple of other enthusiasts would explore these VR and AR technologies.

For example, the first generation of VR headsets only allowed for a seated or standing interaction, as the headset's translation through space could not yet be tracked. But everybody knows, for a real *Holo-Deck* (Figure 1.1), you need to be able to walk around in it! During one of the first *WorkSPACE* workshops, we combined the rotation-only headset to with the lab's Motion Capture stage, capable of tracking an object's position through an entire room using small reflective markers, and feeding that information back to the game engine. This allowed us to walk around whatever world

Figure 1.3. First Virtual Human I placed in AR, using a pattern printed on a pillar as a 3D marker for tracking the tablet's camera extrinsics at 'room scale' (Oct. 2013)

we could build out of colored cubes and cylinders within 10 minutes. That generation of headsets also came without any type of sensor to track movements of the hands. So we added more markers to a cardboard tube, which could be picked up and used as a Lightsaber in VR, and be used to push over and bat around all those colored cubes.

Now, in hindsight, it seems as if during this time, we almost gaplessly indulged in recreating all the most iconic concepts from our childhood science fiction media consumption. Often, it was not about replicating these concepts in complex or complete ways, but just building out enough to obtain a grasp of the essential experience, before the next NullPointerException[1] would ruin it again. What does it feel like to stand in front of a full sized virtual human (Figure 1.3)? What does it feel to have it respond to my own motion (Figure 1.4)? What does it feel like to have an out-of-body experience (Figure 1.5)?

Certainly, I am far from the first to realize that my ideas are inspired by media and science fiction. This is also not unusual for the field of Human Computer Interaction. Jordan et al. surveyed publications in a popular venue for HCI research for mentions

[1]A common type of runtime error caused by a mistake in a program, usually resulting in an abrupt end of the entire application.

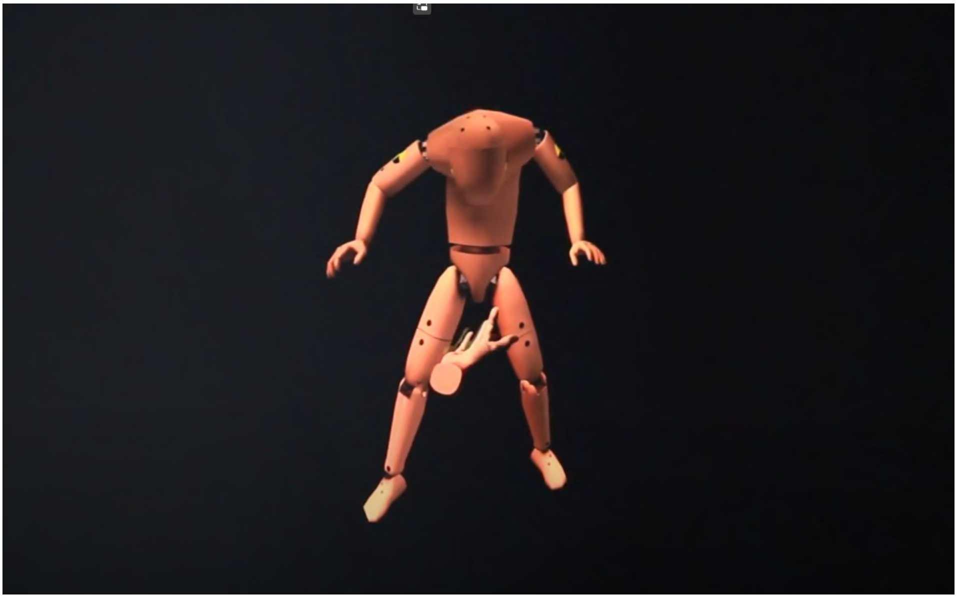


Figure 1.4. Virtual Ragdoll that would respond to hand movements based on an inverse kinematics rig, installation made for BYOB Twente 2016

Figure 1.5. Using a combination of point cloud sensors that allow for looped recordings, the wearer of the HMD could virtually stand behind themselves. Here, we tried to use this out-of-body experience to create a full-body Rubber Hand Illusion, during the 4TU Hectic Haptics Hackathon, March 2018

of Sci-Fi concepts, finding increasing numbers since the 2010s, and expecting implicit inspiration to be even higher (Jordan et al., 2018).

As perhaps illustrated by my personal examples above, Mixed Reality (MR, comprising Augmented, AR, and Virtual Reality, VR) technology is a field in which this inspiration may be most vivid. It is a highly visual medium. One could argue that MR technology has roots in those two, perhaps most salient science fiction inventions already mentioned above; The Holodeck from Star Trek, and the Holograms in Star Wars (see Figure 1.1). The HoloDeck in Star Trek is primarily used for training simulations, but also for leisure and relaxation. The holograms in Star Wars are 3D projections of recorded and live data alike, used for communication with others, as well as for examining complex data such as maps or starship blueprints.

The work in this thesis was driven by the desire to explore the boundaries of MR technology for use in mediated social interactions, similar to those that would also be possible on a HoloDeck or with holograms. In hindsight, many of the interesting questions that we ended up facing in the works discussed in this thesis are not only rooted in these popular examples of science fiction, but also in a piece of science fiction that my promotor, Dirk, and I invented at the beginning of my time as a PhD student. This invention was the *SoIC-TV*, which I will present in the following chapter. Throughout this thesis, I will reexamine questions raised by this invention, and also flesh out and extend the concept with the new things learnined along the way.

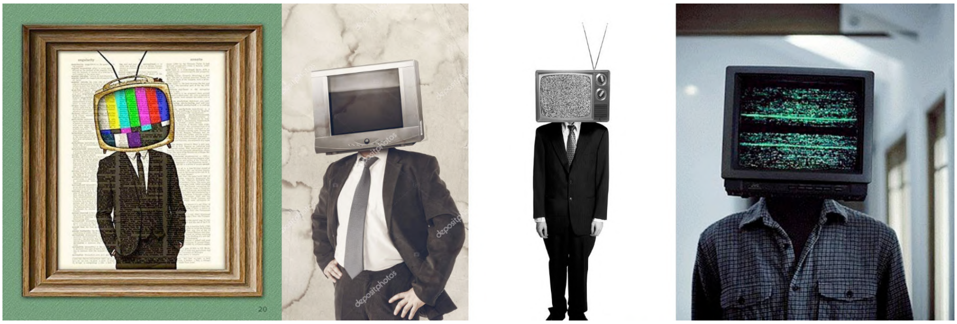
1.1 A DREAM: SOIC-TV

Perhaps you've seen an illustration of a person wearing a TV on their shoulders in place of a head. I'm not sure of the original source of this idea. It seems to represent some trope of mindless consumerism, or consumption of media. It is often shown wearing a black suit and tie. So perhaps it represents corporations or politicians, broadcasting messages to the masses (see Figure 1.6)?

Maybe one of these images got stuck in Dirk's head before he revealed to me that, in a recent dream, he had an encounter with a TV wearing person, and was talking to them, or at least talking to a person's face that was shown on the TV. Dirk was so kind as to provide a retelling of this dream for us:

> *At five o'clock in the afternoon the workshop on embodied dialogue agents at the DesignLab of the University of Twente would end. As usual, I did not attend the content part because of other duties but just went there for the drinks to meet up with people. I entered the room and talked to the organizers about how things went and looked at the attendants who were now mingling with each*

Figure 1.6. Men in suits wearing TVs on their heads

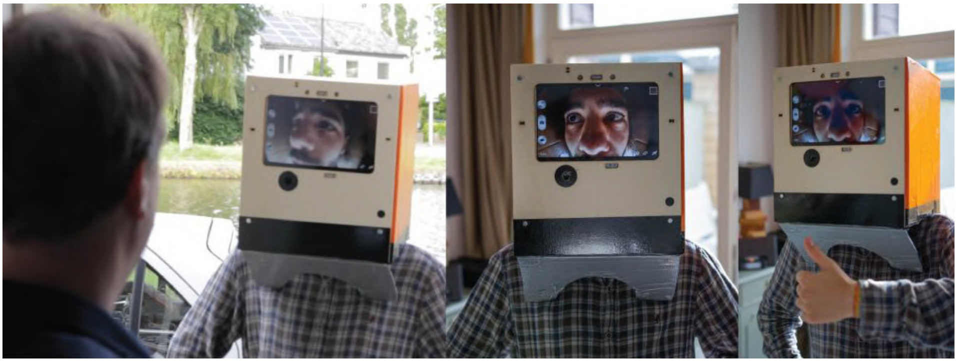


> *other at the drinks. I waved to some and indicated that I would talk to them later. Then I saw a familiar face talking with other familiar faces in a small group and thought "Is this really Gerard?" It was a long time since I had seen him and there was something peculiar about his appearance.*
>
> *After my chat with the organizers, I moved in with the crowd and went up to talk to whom I thought must be Gerard. When I approached him, it became clear what it was that had appeared a bit strange. Instead of Gerard's head there was an old-fashioned television on his shoulders on which his face was talking to me. I asked him how he was doing and what had happened. He told me that he had been involved in an accident as a motor cyclist and his head was cut off. Fortunately, he was lucky that they were able to download his mind and replace his head by this television. I was not surprised at all as this reminded me of Steve Austin, the six-million-dollar astronaut about whom I had seen many documentaries when I was young*[2]*.*
>
> *He guessed that I was not able to recognize him before because they had built in face aging software and as it was now already more than five years since we met last time he must be looking much older now. I asked him whether it was comfortable wearing the television on his shoulders. He told me it was no problem at all, except when he wanted to sleep. Turning his body and head while in bed was a bit of a nuisance. Fortunately, he told me, he could take his head off in the evening and put it on again when he woke up.*

Taking both my supervisor and his dreams very seriously, I was open to engage in a brainstorm on how to make this experience a reality. With just a couple of days to go until the next Intelligent Virtual Agents (IVA) conference would start, we quickly

[2]https://youtu.be/oCPJ-AbCsT8 The Six Million Dollar Man Opening and Closing Theme

Figure 1.7. The SoIC-TV, configured to pass through the inside-facing camera onto the main screen to show the wearer's face.



came up with a name, registered the domain[3], and built a prototype, so we could show it off at the IVA demo session.

1.1.1 THE PROTOTYPE

The prototype consists of a tablet and a smartphone taped back to back inside the front of a cardboard box, with the tablet screen facing outwards and the phone screen facing towards the "wearer". There are cutouts for both the cameras of the phone and tablet to face the outside, and the cameras facing the inside had wide-angle lenses attached to them. This allows for various experiences both for bystanders and for the wearer.

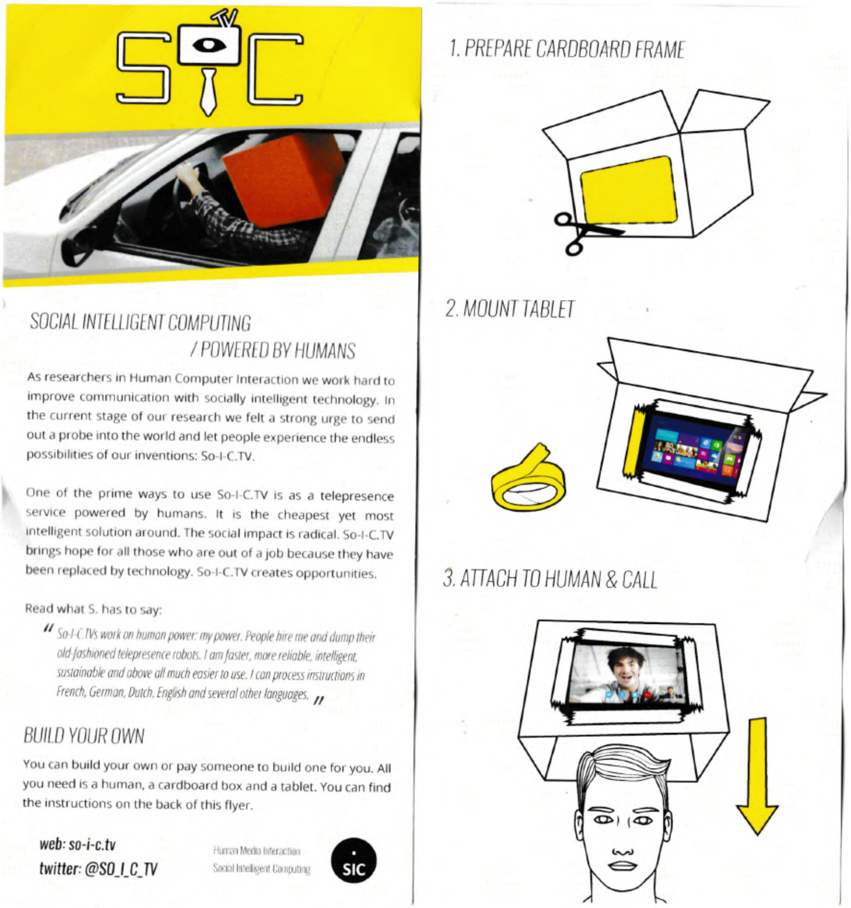
The inward-facing smartphone screen is typically used as a window to the outside world for the wearer, showing the outward facing camera image. This can be done by just turning on the camera app of the smartphone. In a more advanced version, we could also overlay AR content on this screen, or messages and instructions from a remote user.

The outward facing tablet screen can be used to show various content to bystanders. Most trivially, it could show slideshows or movie clips for the entertainment of bystanders. Typically, it would be used to run a videoconferencing software, showing the face of the remote caller on the screen, and showing the outward facing tablet camera image to the remote caller.

If needed, the SoIC-TV wearer can also be shown on the front facing screen by displaying the camera image from the inward-facing tablet camera, preferably with a

[3]http://so-i-c.tv/

Figure 1.8. A flyer that explains how to build your own (simplified version of a) SoIC-TV



wide angle lens attachment, due to the required field of view in the tight TV enclosure. At the IVA demo workshop, we have also used it to show the face of an Embodied Conversational Agent, which bystanders could interact with.

Our vision for this prototype went beyond what could truly be achieved with it, which is where the science-fiction-dream-turned-reality now turns back into science-fiction: A website was built praising this technological revolution, describing its various use cases and validating it with some fictional testimonials.

For example, one user ("Q") solved an argument with their partner using the SoIC-TV, writing:

> *"My wife and I always argued about what to watch on TV. She likes nature movies and I like fantasy. Now we put on nature movies on my screen and fantasy on hers. As I watch my movie, I also see how she is really enjoying hers, and she becomes part of my fantasy."*

Over the years, we have often referenced back to the idea of the SoIC-TV, and developed it further. Next we will discuss two envisioned applications of the system.

1.1.2 SOIC-TV AS A TELEPRESENCE SERVICE

Telepresence technology makes it possible to visit a remote location without physically traveling there. Besides creating the feeling of really *being there*, telepresence technology aims to provide an experience that affords better communication and even collaboration between visitors and those who are being visited. A common implementation of this uses telepresence robots, allowing visitors to even drive around in the remote space instead of being constrained to a single location, such as with a stationary teleconferencing setup.

One envisioned application of the SoIC-TV was to act as a more human alternative to telepresence robots. Users of the service could rent a SoIC-TV near the location of a remote social meeting that they want to attend. We draw an analogy to *mobility as a service* applications, such as *Uber* or *Lyft*. Using the same terminology, the SoIC-TV owner and service provider is the *driver*, and the customer, or remote visitor is the *passenger*.

With this service, the *passenger* is represented at the remote location with a more human-like appearance than with a telepresence robot – at least from the neck down. What is more, several of the shortcomings of telepresence robots are avoided. A SoIC-TV *driver* can easily take stairs or other small obstacles, and can navigate through crowds in a safe, socially acceptable way. *Drivers* even have fully articulated arms for object manipulation – they can even shake hands on behalf of the *passenger*!

A human *driver* may also be more aware of how to behave in a given social setting, adapting their posture and non-verbal behavior differently at a cocktail party than at a funeral. Through the interface inside the SoIC-TV, visible only to the *driver*, further instructions on how to act or behave could be shown on behalf of the *passenger*. For example, the *passenger* could indicate to the *driver* where to walk next, either in the form of a general direction, or using language to encode more complex instructions, such as: *"hide behind that pillar!"*, which are orders that robots may not always be able to interpret and realize correctly.

Figure 1.9. A SoIC-TV *driver* trying to get to work.

1.1.3 SOIC-TV AS A SERVICE FOR MACHINES

As an extension of the *telepresence* idea, we can also conceive of machines employing SoIC-TV *drivers*' services to achieve their goals. While the development of robots that can fulfill all tasks that humans can accomplish with their bodies may be far off, AI software for specific, specialized tasks has long been a reality - even for social contexts. Personal assistant AIs, for example, can already make appointments on behalf of a user, calling a local restaurant with a computer generated voice, acting out a conversational protocol to book a table.

We can imagine that there are situations where a phone call alone is not enough to achieve a given goal, where requirements are physical presence, the ability to manipulate objects, and the ability to recognize and adapt to complex social interactions.

With the SoIC-TV, an advanced Auctioning AI could bid at auctions that requires buyers to be physically present, and use hand signs to bid. Using the interface inside the SoIC-TV described above, the AI would instruct the *driver* on when to bid and how much, following an optimal strategy based on auction theory. What's more, after the auction, the AI can instruct the *driver* to package and post the auctioned items to the AI's storage facility, from where it could be resold at a later point.

1.1.4 SCIENCE FICTION SCENARIOS RAISE QUESTIONS

The SoIC-TV, while functional for demonstration purposes, served as a device to generate a multitude of fictional scenarios and ideas, such as the ones just described. These scenarios also caused us to formulate some questions regarding the implications of a device like the SoIC-TV – perhaps more wearable, and less obtrusive than the head mounted TV – to those who wear it and interact with it.

What is the experience for the wearer when acting as a *driver* for the *passenger* in social interactions? To what extent does it feel that one is yielding control over their own body to the *passenger*, be it a human or a machine? To what extent do they feel involved in the interaction, how would they describe their role? Do *passenger* and *driver* collaborate, or is the *driver* a mere vessel?

What about the perspective of *passengers*? Are they concerned about how they are being represented by the *driver*? After all, it's different from renting an Uber and being concerned about the Uber drivers driving style – which isn't something that other users of the road would attribute to the Uber's passenger. In the case of the SoIC-TV, those around the *driver* might perceive the relationship between *driver* and *passenger* differently.

Going back to the example of a machine using the SoIC-TV service, depending on how the interface is designed for the *driver's* instructions, the *driver* might not even be aware that the instructions are coming from a machine. In the scenarios sketched above, the *passengers* would typically have to portray themselves somehow on the SoIC-TV screen. Perhaps an AI would use a character, or even an animated virtual head of a human. However, consider a less comical scenario, where it's not a literal TV on the *driver's* head. When a more covert system is used to relay an AI-*passengers'* instructions to the *driver*, those around the *driver* might also never become aware that they are talking to an AI. The *driver* is lending their credibility as a human to the machine.

1.1.5 CONCLUSIONS

Before we start with the main part of this thesis, I should mention that the SoIC-TV and the questions generated surrounding the experiences it may afford were not the primary subject of this thesis. However, much of the research we will discuss was motivated by the desire to explore novel experiences in mediated social interaction given what we could afford to realize with combinations of new technology and existing knowledge. Still, the SoIC-TV will return in relevance throughout this thesis, and maybe helps in finding the incidental answer to the multitude of questions that were raised in this section, or find new applications and features that make the SoIC-TV an even more compelling product in the future.

1.2 THESIS SUMMARY

People use technology to engage in social interactions in both private life and work settings. Talking on the phone or using video conferencing software are examples of technology mediated social interactions. What's more, technology can be used to mediate social interactions with machines, in the form of chatbots and virtual characters. In this thesis, we have explored different forms of social interactions mediated through mixed reality (MR) technology, such as virtual and augmented reality (VR & AR), with both human and virtual human (VH) interaction partners. Interaction contexts addressed are simulations of co-located (i.e. face-to-face) interactions and remote collaboration tasks using telepresence (i.e. using MR to mediate interaction over distance). We have employed qualitative and quantitative methods, and a variety of measures across these studies. Measures included physiological responses, how participants perceived the interaction, how they perceived other interlocutors, their perception of social presence, and immersion. The varied findings across these studies are reflected upon through the lens of a conceptual, science-fiction-inspired telepresence system: the SoIC-TV. By developing this concept throughout the thesis, we illustrate how in the future, mediated interaction experience using MR technology might transform beyond what is familiar both from co-located and present technology-mediated interactions. Following is a summary of the main parts of the thesis.

Our first focus was on the experience of being face-to-face with VHs in shared virtual spaces. We looked at appropriate social distance and eye-contact as social norms that interlocutors maintain during an interaction. The first study employed a VR setup where participants were able to walk freely in a room-scale environment. Two VHs engaged in a simulated social interaction with the participant. They maintained either normal amounts of mutual gaze and interpersonal distance, or verged into staring at the participant and standing too close to the participant. We measured both perception and attitude towards these VHs as well as behavioral responses during these manipulations in terms of head orientation and distance between participants and VH. We found that these violations elicit responses in participants' own behavior similar to what we would expect from collocated human-human interaction. We also investigated how this affects the overall perception of and attitude towards these VHs, and concluded that even in simulated, or mediated experiences social norms apply. In a second study, rather than violating social norms through the VH's non-verbal behavior, we looked at changes in the conversation topics initiated by the VH to something more personal, intimate and stressful – creating a context with a quality that may not directly violate social norms, but something that would be unexpected to talk about with a VH. For this study, we extended the apparatus used to include physiological measures as well as a better measure for gaze behavior. While there were some minor effects on the physiological responses when the discussed topic became more serious, gaze behavior was not significantly affected. A more in-depth qualitative analysis however gave some interesting insights into how participants felt about having this type of conversation with a machine or VH.

In the second part of this thesis, we change the type of interaction from mediating social interaction with VHs in VR to mediating interactions through avatars between remote humans in mixed reality. Here, we investigate user experience when engaging in collaborative social interaction over distance using different user interfaces. More specifically, our focus is on remote expert, local novice scenarios. We surveyed existing work on designing such systems and implemented our own system with variations on certain design aspects, such as the interfaces to engage with the remote location, the presentation of the remote location and the representation (avatar) of the interlocutors in the remote environments. Two user studies have been conducted with this system. In the first one, we evaluated the effect of the design aspects on the relevant user experience metrics and performance in a collaborative puzzle task. In the second study we looked at how such a system may perform in life critical circumstances: In a design fiction scenario, a first responder has to learn CPR in an emergency situation using an AR system. We found that an embodied presence in AR can not only give better instructions, but might also make the experience less stressful for the first responder than when being coached by a disembodied voice over the phone.

Where the first two parts focus on mediating social interaction with virtual humans or avatars of remote real humans, the final part is blurring these concepts together. We built an apparatus called the Multimodal Echoborg that allows VH to engage in social interactions through the human body of a confederate instead of through their virtual embodiment - unbeknownst to other interactants. In an exploratory study, we looked at how this machine-controlled confederate is perceived, compared to when the machine is embodied through a virtual human. While most participants did not suspect our deception, they did start to speculate whether our confederate had some form of personality disorder. We discuss the merits of such a system as a research tool for embodied conversational agent research and reflect on what we have learned about how our SoIC-TV science-fiction scenario may play out in reality.

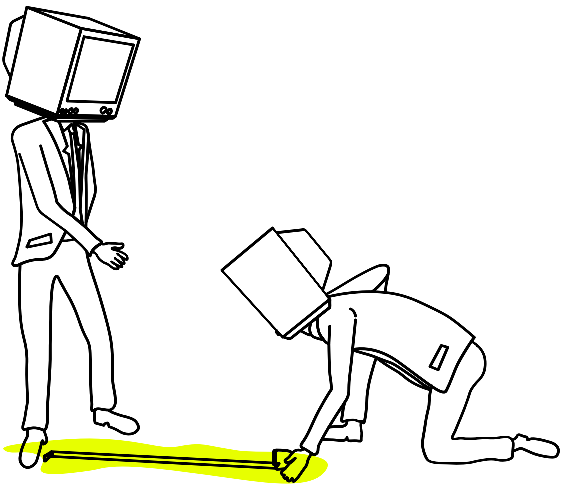
Finally, we take a broader view on our work and discuss our contributions to and future directions of the field of mediated social interactions in mixed reality.

PART *II*

FACE-TO-FACE IN VIRTUAL SPACE: INTIMATE ENCOUNTERS WITH VIRTUAL HUMANS

2

PERSONAL SPACE INVADERS



In this chapter[1] we investigate the effects of the changing dynamics in mutual gaze and interpersonal distance when interacting with virtual characters in shared environments. The distance humans keep between themselves and others in different social settings is moderated by a number of factors, such as how close they feel to each other, and the context of the interaction. While revisiting this chapter throughout the year 2020, it got a different twist from when we first investigated it: most public health institutions throughout the world issued recommendations for

[1]Based on our paper J. Kolkmeier, J. Vroon, and D. Heylen (2016). "Interacting with virtual agents in shared space: Single and joint effects of gaze and proxemics". In: *International Conference on Intelligent Virtual Agents*. Springer, pages 1–14.

social distancing. The goal was to reduce airborne transmission of the Coronavirus by maintaining increased physical distance in public and private encounters.

Especially for those in high risk groups, social distancing would often mean social isolation. And while effective in reducing transmissions, some concerns were raised that there may be unforeseen negative side effects to social distancing. Technology has quickly stepped in to compensate for the lack of face-to-face social experiences, and for many, both their professional and personal life got colored by the daily "Zoom"[2] calls. The technology that was accessible to everyone in 2020 was still limited to videoconferencing.

In this chapter however, we look at social interactions where all participants inhabit a shared virtual space through the use of virtual reality technology. In particular, we are interested in the social norms that hold in this virtual space when engaging in a social interaction with virtual characters. We investigated how characters are perceived that violate these norms in a user study. We study how humans respond behaviorally, and how this affects mutual gaze behavior, a known modulator of the perception of social distance.

2.1 A MODEL OF SOCIAL USE OF SPACE

A classical model for how humans perceive and manage interpersonal distance was introduced by Hall's studies on *Proxemics* (Hall, 1969). It describes four zones of interaction distance and relates them to different kinds of interaction when implicit cultural and social norms are adhered to.

The four zones are the *intimate*, *personal*, *social*, and *public* zones with sizes shown in Figure 2.1. The intimate zone is reserved for interactions like touching, hugging, while the personal zone is reserved for interactions with good friends and family. The social zone is where most interactions with acquaintances happen, and the public zone is reserved for public speaking.

Of course in social situations it may not always be possible to arrange oneself in these preferred distances for the given interaction: on a crowded street or in a packed elevator, it becomes acceptable to stand closer. Imagine, however, someone starting to stare at you while being packed in an elevator - this would make the ride much more uncomfortable than it may already be. There has long been speculation in the social psychology literature that the proxemic model is not a static model but is moderated by other modalities and by the context of an interaction. The Equilibrium theory (ET, Argyle and Dean, 1965) makes predictions on the relationship between

[2] zoom.com, a videoconferencing service that became increasingly popular during the COVID-19 pandemic in 2020

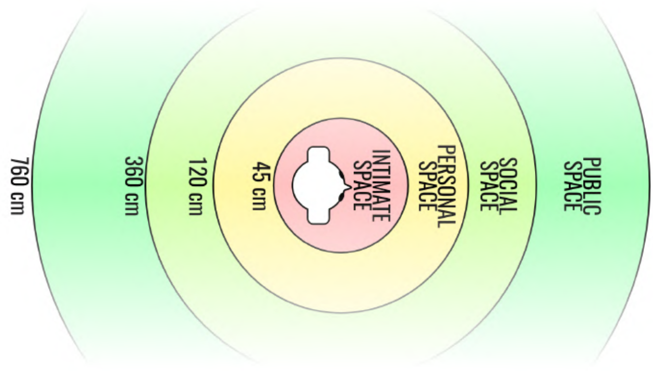


Figure 2.1. Interpersonal distance zones and their radii used in Hall's model

social distance, eye-contact behavior and the level of intimacy between those that share space. For example, high levels of perceived intimacy induced by reduced interpersonal distance can be compensated by regulative behaviors, such as averting gaze or by increasing interpersonal distance through a change in posture or position.

This theory has been tested and extended in various studies with varying methodologies and results supporting its general validity (Coutts and Schneider, 1976; Patterson, 1977; Rosenfeld et al., 1984). ET has been revisited in virtual reality in a prior study by Bailenson, Blascovich, et al., where first evidence was found that the theory also applies to interactions between humans and virtual agents (Bailenson, Blascovich, et al., 2001).

2.2 A PLAUSIBLE EXPERIENCE

But why is it worth investigating social distancing behavior in VR in the first place? On what basis do we speculate that virtual characters – or virtual avatars of other humans for that matter – can even violate what we think is the correct interpersonal distance and amount of eye contact? After all, they are made up out of polygons that are rasterized to pixels, shown on the display of a wearable headset, without any effect on the physical space that the user is actually inhabiting.

In his seminal work, Slater asks why users "tend to respond realistically to situations and events portrayed within [VR]" (Slater, 2009). He proposes that users respond realistically to events that transpire in immersive virtual reality systems based on two contributing components. First, the sense of being in the environment mediated by the VR technology, or the *place illusion* (PI). Second the sense of what is depicted in that environment is actually occurring, or the *plausibility illusion* (Psi). Slater further proposes that if both of these illusions occur, or are at least not being constrained by the system, users will respond realistically to what is transpiring in the virtual environment (Slater, 2009).

Slater defines PI as "[...] the strong illusion of being in a place in spite of the sure knowledge that you are not there" (Slater, 2009). Here the *place* is not limited to any specific experience, it could entail a virtual reality experience but also the experience of teleoperating a robot. The main constraint for PI is the level to which the system is responsive to actions of the user that should change the perception of the environment, or that should change the environment itself. For example, if the user rotates or translates their head, the system should update the rendering to render the virtual world accordingly. The stimuli the system generates should be contingent on the users' actions that are "meaningful in terms of perception within the virtual environment depicted" Slater, 2009. Slater calls these the *sensorimotor contingencies* (SC). The amount and fidelity of the SCs that a system supports is an objective property of that system, and can be used to classify the degree of *immersion* the system affords. Again, the most basic SC in VR systems is that the virtual world is rendered and displayed to the user consistently with how their own perspective would change in the physical world, i.e. with how they rotate and translate their head through space. Similarly, spatial audio and haptic feedback can be implemented consistently as to afford an *increasingly immersive* experience. Slater further qualifies the difference between immersion and PI: "immersion provides the boundaries within which PI can occur". While the immersiveness of a system is an inherent property that does not change within or between users of the system, the level of PI may very well depend on how the user interacts with the system. For example, examining objects up close may increase or decrease the PI depending on the rendering resolution of the system. Different users will encounter more or fewer of the boundaries of what the system can afford depending on how they use the system, and thus will experience different levels of PI.

The second component that Slater describes is the *plausibility illusion*, which he defines as the "illusion that what is apparently happening is really happening (even though you know for sure it is not)". Slater proposes that the experience become plausible, i.e. more real, if events in the virtual environment refer directly to the user. Virtual characters, for example, that respond to the behavior of the user, make the experience more real: "since you are as real as can be, and this external sensed world appears to be addressing you, the reality of that external world is itself enhanced" (Slater, 2009). Slater further qualifies that "Psi does not require physical realism", citing a number of experiments that show how interactions with virtual characters - however simple - elicit similar responses in participants to what we would expect from a *real* version of the experience.

Concluding, neither being transported into a remote space nor rendering an artificial embodied agent or avatar into a user's local space means that they are physically there. However, responding *"as if they were there, and real"* would indicate that a plausible VR environment with interactive social embodied agents would have real implications on the interaction and perception of these actors. We thus hypothesize that the social space related behaviors displayed by characters in a shared virtual space will elicit *realistic* responses in the user, and thus follow the predictions of models that have been created in the context of co-located interactions of humans, such as the equilibrium theory model.

2.3 INTERPERSONAL SPACE & GAZE SIMULATOR

To investigate interpersonal distance and eye contact further, we developed an apparatus that allows us to simulate a small social experience with three actors - the user together with two virtual avatars - in a shared space big enough to walk around in. To summarize, using the tracked position and rotation of the user's head, or the *head transform*[3], we can both drive the position of our virtual avatars to maintain a desired interpersonal position, having them gaze towards the user's head or away from it, and also estimate the movement and gaze direction of the user in response to the behaviors of the virtual avatars.

2.3.1 SETUP

The apparatus was installed in our lab, shown in Figure 2.2. An *Oculus Rift DK2* head mounted display (HMD) is tethered to the experiment PC which is situated in the truss. The tether is 2.6 meters in length from the top center of the room. Since this work was done before consumer VR equipment came with room-scale tracking technology, we used a NaturalPoint OptiTrack with six IR cameras to track the translation of the VR headset. This way, free movement and tracking is possible in most of the 4 by 5 meters truss area, the extreme corners being the exception.

The room displayed in the virtual environment was a generic apartment asset with a bigger empty space next to the living room area, which was mapped onto the experiment space. A transparent 3D model of the truss was placed in correspondence with its real-world position and dimensions to give users a reference in VR of where they were situated in the physical world (see Figure 2.2).

[3]Yes, this what the first part of the thesis title is referring to!

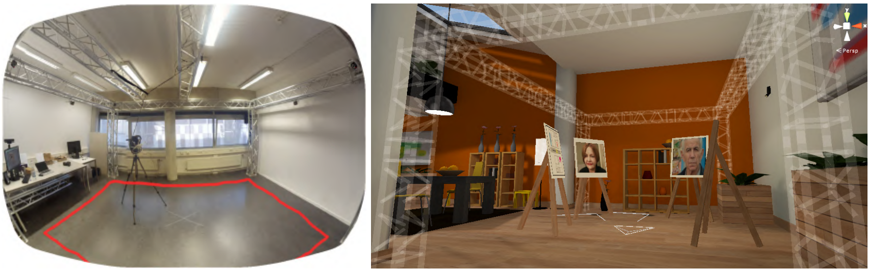


Figure 2.2. Left: the physical room. Right: the virtual room.

2.3.2 VIRTUAL AGENTS

The virtual agents used in this setup were generated using the Unity Multipurpose Avatar system (UMA, Ribeiro, 2015). The avatars were generated from the same base mesh to look very similar, but slight adjustments were made to face, hair and attire. To conform to the characters from the scenario described in the next section, both agents were chosen to be male.

To control for possible effects of the size of the agents, a scaling function was added that would scale the agents to have the same size as the current user (based on the height of the headset while participants are standing). The animation of the two agents was driven by the following three components.

Pose Component Walking and stepping animations were realized using Unity3Ds Mecanim animation blend-tree system[4]. A blend tree allows one to procedurally blend animations together. For example, it can be used to blend between forward and sidewards step animations to generate a diagonal step.

Gaze Component Using inverse kinematics, the head orientation can be controlled to have the agent look at various targets, such as at the user's or another agent's eyes. For this, the FinalIK[5] inverse kinematics plugin for Unity3D was used. FinalIK allows one to give different weights to rotation of chest, head and eyes towards a target.

Proxemics and Locomotion Component Using a blend tree, we can blend from the rest animation into a walking animation. This component is driven by a model for the desired group formation, which will be described in more detail in Section 2.4.1.

[4] http://docs.unity3d.com/Manual/AnimationOverview.html
[5] root-motion.com/final-ik.html

Figure 2.3. Agent gaze and positioning behavior.

2.4 USER STUDY

Using the components described in the previous section, we carried out a user study regarding user's responses various gaze and proxemics related behaviors of the agents.

We introduced *Equilibrium Theory* (ET) in Section 2.1 as a model for the relationship between social distance, eye-contact behavior and the level of intimacy between those that share space (Argyle and Dean, 1965). In summary, interlocutors regulate the amount of eye contact and the position between each other (by moving or leaning) to establish a desired level of intimacy.

In this study we draw from ET to make predictions on the outcome of a simulated social interaction in a room-scale VR setup. Participants engaged, as listeners, in a social interaction with two virtual embodied agents that vary in their use of mutual gaze and social distancing behaviors. These manipulations of agent behavior occurred several times throughout the interaction, and we measured how users responded to these manipulations in terms of their own behavior, i.e., if they regulate their interpersonal distance or amount of eye-contact. We call this regulation the *user response*.

Based on the ET, we hypothesized that after a change in the behavior of an agent that impacts the intimacy level of the situation - for example coming closer to the user - the human user would perform compensation behaviors - for example stepping back or averting gaze - to maintain the same level of intimacy.

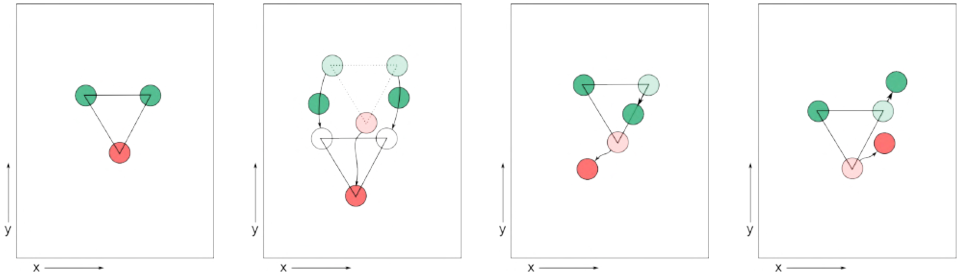
2.4.1 AGENT BEHAVIORS

We defined the behavior of each agent as a combination of gaze and proxemic behaviors. For gaze, we defined behaviors with neutral (G°), high (G^{+}), and low (G^{-}) intimacy. Similarly, we defined proxemic behaviors with neutral (P°), high (P^{+}), and low (P^{-}) intimacy.

The precise manifestations of these behaviors were based on a pilot study with colleagues (n=5) that were aware of the study goals. Participants were placed in a prototype of the experiment apparatus with a virtual agent. The experimenter let the agent alternate between different prepared versions of each behavior, interviewing the participant on how they perceived the behavior of the agent in terms of intimacy compared to the other realizations.

For proxemic behaviors, we let the agent move across the zones in Hall's model. We found that keeping a distance of 75 cm between users and agents was perceived as neutral (P°). Decreasing the distance to 40 cm was perceived as noticeably more intimate (chosen for P^{+}, see Figure 2.3c). This coincides with Hall's *intimate space* and the distance used in Kastanis and Slater (2012). At a distance of 110 cm the agent was found to be noticeably less intimate and was used for P^{-} (see Figure 2.3d).

For gaze, we found that having the agent switch between gazing at the user and averting its gaze in random intervals between 2 and 5 seconds was perceived as neutral (G°). Participants found it more intimate when the agent would always respond with mutual gaze if directed gaze at the agent by the user was detected (chosen for G^{+}, see Figure 2.3a). In this version of gaze behavior, the agent would also prolong that gaze for 1.5 seconds even after directed mutual gaze was interrupted by the user. Note that this version was chosen over a version where the agent would continuously direct its gaze at the user, as this was perceived as 'creepy'. Conversely, for G^{-}, we selected a behavior where the agent would always avert its gaze if directed



(a) **(b)** **(c)** **(d)**

Figure 2.4. The dynamic locomotion and social distance model. Agents can be configured to track a target position on a predefined group configuration - here shown is an equilateral triangle (a). The group formation remains relative to the user position (red circle), even if the user moves (b). To manipulate interpersonal distance, the agents' target positions can additionally be interpolated along the triangle's legs (c and d). During such manipulations the formation will not track the users movements.

gaze by the user was detected (see Figure 2.3b), which was found to be less intimate than the neutral version by the participants.

For the final manipulations in the experiment, we chose six combinations of the gaze and proxemic behaviors described above where one or both modalities would deviate from the neutral behavior ($G^{\circ}P^{\circ}$) in terms of increasing and decreasing intimacy: $G^{-}P^{-}$, $G^{\circ}P^{-}$, $G^{-}P^{\circ}$, and $G^{+}P^{\circ}$, $G^{\circ}P^{+}$, $G^{+}P^{+}$.

2.4.2 BEHAVIORAL MEASURES

As our behavioral measure, we compute the change in amount of eye-contact and the magnitude of interpersonal distance between user and agent before and after an agent performs a manipulation. The *Gaze Response* R_G of a user is the change in their head angle towards the agent. This may be looking more towards the agent (smaller angle) or looking more away from it (larger angle). We call compensating displacement of the user's whole or upper body the *Proxemic Response* R_P of the user. This may be moving away from the agent (positive response) or towards an agent (negative response) [6].

[6] The use of *negative* and *positive* purely refers to the reduction or increase of distance, and is not meant to carry any other valuation, i.e. whether these responses are *good* or *bad*

2.4.3 HYPOTHESES

We formulated our hypotheses as predictions of user's behavioral responses to different gaze and proxemic behaviors exhibited by a virtual agent. Our main hypothesis is:

H1 Users regulate their gaze and interpersonal distance during interaction differently towards agents that exhibit either high or low intimacy behaviors.

We made predictions of the single and joint effects of the behaviors that the high and low intimacy agents exhibit based on the ET:

$H1_a$ Increasing proximity of the agent ($G^{\circ}P^{+}$) will be compensated for by the user with a more positive R_P (moving away) – compared to a smaller, possibly negative R_P (moving closer) when agents perform $G^{\circ}P^{-}$.

$H1_b$ Increasing gaze of the agent towards the user ($G^{+}P^{\circ}$) will be compensated for by the user with higher R_G towards the agent (looking away) – compared to smaller R_G when agents perform $G^{-}P^{\circ}$.

$H1_c$ Besides R_P, also different levels of R_G will be observed in response to $G^{\circ}P^{+}$ and $G^{\circ}P^{-}$ manipulations.

$H1_d$ Besides R_G, also different levels of R_P will be observed in response to $G^{+}P^{\circ}$ and $G^{-}P^{\circ}$ manipulations.

$H1_e$ When the two non-contradicting behaviors are combined, user's responses will 'add up', i.e. R_P to $G^{+}P^{+}$ is higher than to $G^{\circ}P^{+}$ and R_G to $G^{+}P^{+}$ is lower than to $G^{+}P^{\circ}$, etc.

The assumptions of the ET is that a perceived *intimacy level* is mediating changes in behavior. We therefore checked if our manipulations would also be reflected in the user's of the agent's personality and perceived interpersonal attitudes. Our hypothesis was:

H2 Users rate agents that exhibit high intimacy behaviors higher on items related to intimacy.

These hypotheses were tested using the setup described in Section 2.3. We included intimacy of the agent as a within subject variable. One agent had the *high* intimacy manipulations assigned, the other had *low* intimacy manipulations. They did not change their assigned role during the experiment. This choice was made to be able to compare how the different more and less intimate behaviors would affect the user's perception of the agent (H2).

2.4.4 TASK & SCENARIO

Participants were not told that the experiment was about examining their movement and gaze behavior. Instead, they were given a listening task to focus on, based on a scenario that the two agents would act out. The scenario was taken from the 1957 movie *12 Angry Men*. In this movie, 12 male members of a jury have a discussion about whether they were presented sufficient evidence during the court case to sentence the defendant to death. Audio clips of speech segments were extracted from parts of the movie. To prevent dominance mediated by voice to be a factor in the perception of the agents, segments were selected where the argument was less heated. This resulted in 30 clips arguing for 'not guilty' (avg. length=11.49s) and 29 clips arguing for 'guilty' (avg. length=11.51s) side of the argument. The clips were spoken by the agents chronologically, alternating between the sides of the arguments to make up a consistent conversation between the agents (total duration=12m). It was suggested to the participant that the two agents would each attempt various 'strategies' in order to convince the participant of their side of an argument. The task given to the participant was to listen carefully, as they would be asked for their decision afterwards.

2.4.5 AGENT BEHAVIOR MANIPULATIONS

During the experiment, the agents formed a group with the user by positioning themselves on the base corners of an equilateral triangle. The length of the triangle's legs was 75 cm, corresponding to the *neutral* distance found in the pilot study. The triangle did not rotate with the user. It always faced the long side of the room. The angle of the user's corner was 60 degrees, which was to ensure that when the user centered their view between the agents, both were in view.

Agents would change their gaze and proxemic behaviors at moments that coincided with the dialog turns from the scenario. The behavior changed for the entirety of the turn. This resulted in *episodes* of different agent behavior. At the beginning of every second dialog turn, both agents would employ the *neutral* behavior to 'reset' the group formation (neutral episode, see Figure 2.4b). On every other turn, exactly one of the two agents manipulated its behavior (manipulation episodes), by performing one of the three behavior combinations that corresponded with its assigned *agent intimacy*. For example, the 'high' agent chose from G^+P°, $G^\circ P^+$ and G^+P^+ - as described in Section 2.4.1. Each of the three assigned behavior combinations were shown exactly four times throughout the experiment, in randomized order.

Which of the two agents would manipulate its behavior during manipulation episodes was alternated in turns. Since these depended on the dialog turns, the between subject variable 'talking agent' was unintentionally introduced: Within subjects, one of the agents changed its level of intimacy only when it was also the currently talking agent, whereas the other changed its level of intimacy only when it was not currently talking. Whether it was the 'high' or the 'low' agent that manipulated only during talking was randomized between subjects.

2.4.6 PARTICIPANTS

Thirty five participants were convenience-sampled from students and staff, all of whom were completely naive to the study's goals. They were between 19 and 30 years old (Mean = 21.4). Five were female. Of the 35, two were discarded from the data. One decided to stop the experiment early because of motion sickness, and another misunderstood the instructions, continuously moving around and exploring the room also during the main experiment.

2.4.7 BEHAVIORAL MEASURES

During the experiment we recorded the participants' and agents' head positions and orientations in the virtual world using the tracking system of the VR setup. We continuously calculated the distance between the user's head and the individual agents' heads as well as the angle of the user's gaze away from the individual agents (see Section 2.4.2).

We observed significant outliers in the proxemic responses of the remaining 33 participants. By reviewing video material and experiment notes, some of these outliers were found to be strong responses at the beginning of the experiment. Towards the end of experiment runs, outliers were also found to be caused by participants stepping around agents when 'cornered' by them at the bounds of the tracking area. Although these changes in position seem motivated by the intimate situation, they diverged significantly from the typical proxemic response in other episodes, where participants would either lean or take one or two small steps. From the analysis were excluded all episodes where R_P was bigger than 50 cm (n=6).

2.4.8 QUESTIONNAIRE

In addition to the behavioral measures, we were interested in participants' perception of the agents' personalities. A 13 item agent-personality questionnaire which has been successfully used before to measure perception of personality and interpersonal attitudes in both human (Guadagno and Cialdini, 2002) and virtual human (Guadagno, Blascovich, et al., 2007; Huisman, Kolkmeier, and Heylen, 2014a) communication partners was used. One extra item on politeness was added (Maat,

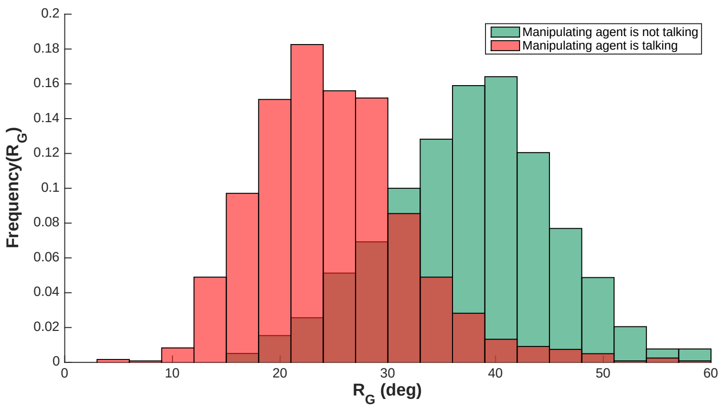


Figure 2.5. Histograms of R_G for episodes where the manipulating agent *is* the talking agent and *is not* the talking agent.

Truong, et al., 2010), and one for 'intimacy'. For each agent, identified by a picture, participants would indicate their agreement with the item on a 7-point Likert-scale. Scores of the intimacy related construct measured by this questionnaire were used to answer H2.

2.4.9 DATA ANALYSIS

The experiment was designed so that we could compare the effects of the six agent manipulations as a six level within-subject factor 'Agent Intimacy' on the two user measures R_G and R_P. However, due to the introduction of the talking agent as a between subject variable (see Section 2.4.5), this approach would not be sound, as one might expect the talking agent to gain more attention than the non talking agent, biasing the gaze response. Indeed, we found that users would typically look towards the currently talking agent (Figure 2.5). Consequently, we chose to focus on comparing participants' gaze responses only inside the group of agents that manipulated their behavior during their own turn of speech, and only allow for comparison of R_P regardless of the talking agent variable.

2.5 RESULTS

In Table 2.1 we present the mean R_P and R_G of all participants and episodes where the manipulating agent was also the talking agent. The measurements violate the assumption of sphericity and normality for both measures at many levels, therefore, under the assumption that the between subject variable *talking agent* does not

Manip.	Mean R_G (SD) in $^\circ$	Mean R_P (SD) in cm	n	outliers
G^+P^+	30.17 (6.41)	8.56 (11.70)	55	1
G^oP^+	28.57 (7.38)	8.43 (13.89)	53	2
G^+P^o	27.01 (7.73)	0.36 (9.50)	56	0
G^-P^o	25.23 (6.16)	-0.37 (5.79)	75	1
G^oP^-	25.16 (7.56)	-2.97 (8.89)	76	0
G^-P^-	23.52 (5.72)	-3.48 (6.51)	74	2

Table 2.1. Mean gaze response R_G and proxemic response R_P per agent manipulation from all episodes where the manipulating agent was also the talking agent. SD = Standard Deviation.

represent a bias (for the R_P measures), we used the non-parametric Friedman and Wilcoxon signed-rank tests to test our hypotheses.

BEHAVIORAL MEASURES

To test for significance, we performed tests for each of the two measures R_G and R_P. As explained in Section 2.4.9, we only allow comparison between all six manipulations to examine the difference between R_P regardless of the *talking agent*. Then we performed tests comparing only pairs of high and low manipulations respectively, and only when the manipulating agent was also the *talking agent* (to reveal differences without the bias it introduces). Since here, we are not comparing high and low intimacy agent, these do only allow to test aspects of $H1_e$.

As described above, the agent would act out each manipulation four times during the experiment. The non-parametric test compares pairs of responses to these manipulations. To not artificially inflate our sample size, we compared only one of the four participant responses. These samples however can only be compared when they are not paired with removed outliers. In our data, as described in Section 2.4.7, outliers were observed in the first and second instance of each manipulation, possibly because of a novelty effect, as well as in the fourth, because of the effect of the borders of the experiment area (participants being 'cornered'). Responses during the third instance of each manipulation contain no outliers. Therefore, the responses to the third instance of each manipulation were used for comparison in the non-parametric tests.

Differences between all six manipulations A Friedman test revealed that there was a statistically significant difference in displacement magnitude ($H1_{a,d,e}$) as a response to different levels of agent behavior intimacy, $\chi^2(5)$=32.84, p <.001. A Wilcoxon signed-rank test showed that in the 33 participants, the displacement magnitude in response to $G^{\circ}P^{+}$ behaviors was significantly more positive (i.e.: moving away) than that to $G^{\circ}P^{-}$ (Z=-3.368, p=.001).

Differences between high manipulations ($H1_e$) A Friedman test revealed that there was a statistically significant difference in response displacement magnitude between the high-intimacy behaviors, $\chi^2(2)$=7.00, p=.030. A Wilcoxon signed-rank test showed that in the 14 participants where the *high* agent manipulated its behaviors while also being the talking agent, the displacement magnitude in response to $G^{+}P^{+}$ episodes was significantly greater than the displacement magnitude in response to $G^{+}P^{\circ}$ (Z=-2.542, p=.011). In the same population, between the pair of $G^{+}P^{\circ}$ and $G^{\circ}P^{+}$ manipulations, we found that the former would elicit significantly less positive displacement magnitude (Z=-2.229, p=.026) than the latter, meaning that those in $G^{+}P^{\circ}$ would move away significantly less. The difference between the pair of $G^{+}P^{+}$ and $G^{\circ}P^{+}$ behavior was not found to be significant (Z=-.910, p=.363). No significant difference in gaze angle was revealed ($\chi^2(2)$=2.29, p=0.319).

Differences between low manipulations ($H1_e$) Between the low intimacy behaviors, a Friedman test did not reveal a significant difference in displacement magnitude response of the 19 participants where the *low* agent manipulated its behaviors while also being the talking agent ($\chi^2(2)$=2.95, p=0.229). No further tests comparing the individual pairs were performed.

The Friedman test, however, did reveal that there was a marginally significant difference in the participant gaze response between the *low* behaviors ($\chi^2(2)$=6.42, p=.040). Upon inspection, it appears the difference is due to asymmetry of the difference of the pairs, excluding it from further examination with the Wilcoxon signed-rank test. A sign test revealed no significant difference.

AGENT PERSONALITY QUESTIONNAIRE

To identify intimacy related constructs in the agent personality questionnaire (H2), we performed a principal component analysis with Varimax rotation and Kaiser normalization on the 15 items. Three factors were identified that explain 69.15% of the variance (Table 2.2). The factors 'Warmth' and 'Trustworthiness' are similar to those found in a previous study using a similar questionnaire (Huisman, Kolkmeier, and Heylen, 2014a), a new third factor emerged with the items 'intimate', 'interesting' and 'confident'. We name this new factor 'Intimacy'.

For each respondent, we calculated factor scores given to the two agents by averaging out those items that were associated with the respective factors. We performed repeated measures ANOVA with the intimacy of the agent (*high* or *low*) as the within

Factor	Item	Factor loading
Warmth	Friendly	.88
(α = .92)	Approachable	.83
	Warm	.83
	Likeable	.82
	Polite	.79
	Modest	.79
Trustworthiness	Informed	.82
(α = .87)	Credible	.82
	Competent	.76
	Honest	.71
	Trustworthy	.58
	Sincere	.56
Intimacy	Intimate	.78
(α = .57)	Interesting	.68
	Confident	.66

Table 2.2. Three factors identified in PCA and their corresponding item factor loadings. For each factor, Cronbach's alpha is reported.

subjects variable and agent side, the talking agent, and agent appearance as between subject variables, and the three computed factor scores as measures.

We found a main effect for the intimacy behavior of the agents on 'Warmth' ($F(1,24) = 21.45$, $p < .01$) and 'Intimacy' ($F(1,24) = 6.61$, $p < .05$). No interaction effects of agent appearance and agent side were found on either of the scores. There was however an interaction effect for the talking agent on 'Intimacy' scores ($F(1,24)=4.31$, $p < .05$). Pairwise comparison revealed that participants scored the agent with low intimacy higher on 'Warmth' related items than the high intimate agent (m_L^W=4.97 vs m_H^W=3.57). 'Intimacy' scores align with the intimacy behavior of the agents. Participants scored the agent with low intimacy lower (m_I^L=4.14) than the agent with high intimacy (m_I^H=4.90). For the interaction effect of the talking agent, pairwise comparison revealed that the high and low agents score similarly on intimacy scores when they are not the talking agent. However, when talking during the manipulation, the high agent scores higher on intimacy ($m_I^{H \times T}$=5.25) than the low agent ($m_I^{L \times T}$=3.86).

2.6 DISCUSSION

We found a number of differences in the behavioral response to the different agent behaviors (H1). While the overall means are in line with the predictions made based on the ET ($H1_{a\text{-}e}$), there is a high variance in the responses and only some could be supported with statistical significance.

We found that agents exhibiting higher proximity did cause participants to step away significantly more than agents exhibiting low proximity, where participants tended to step more towards the retreating agent ($H1_a$). As for the predicted effects of G^+P°, G^-P° on R_G ($H1_b$), we could not find significant differences.

In contrast to Bailenson, Blascovich, et al. (2001), our study did not find a notable effect of different agent gaze behaviors on the proxemic response ($H1_d$). This may be explained by their use of a more sensitive measure (minimum distance rather than the mean), and the different interaction between agent and participant (walking around rather than listening). Another possible explanation could be a ceiling effect of how comfortable individuals were with moving in the iVR setup - possibly also depending on whether they were already at the edge of the tracking area. This interpretation is also in line with the personality scores of the high agent. Scores were low on 'Warmth', which had loadings of the 'politeness' and 'friendliness' items.

If a smaller displacement was not sufficient to compensate intimacy, we would expect the remainder to be compensated with gaze. Given the approximate measure of gaze, such compensation may not have been sufficiently captured with the current apparatus.

Some joint effects were found. Participants stepped away more when both gaze and proxemic behaviors were manipulated in a high-intimate fashion, compared to the responses to only high gaze manipulation, supporting some aspects of $H1_e$. Lastly, we found that indeed, participants rated the high agent higher on intimacy related items, supporting H2.

LIMITATIONS & RECOMMENDATIONS

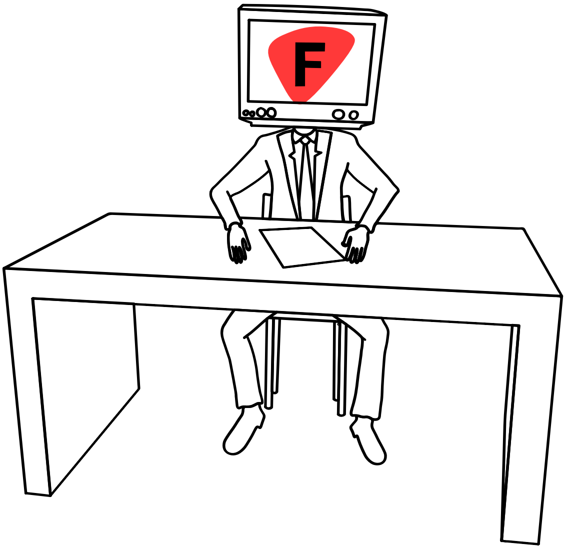
We recommend some changes to the experimental protocol to those that aim to replicate the experiment or adapt aspects of this study design. The extent to which head direction can serve as a proxy for eye-gaze is questionable. Slight gaze aversions away from the agent's face may only be captured with true eye-gaze tracking inside the headset. The two agent design may mitigate this shortcoming as more head movement is required when gaze is averted from one agent to another. For single agent designs however, actual eye-gaze tracking is recommended. For group interactions, we recommend being aware of the effect of the talking agent on the gaze of dialog partners that we observed, as participants might not notice stimuli by non-talking agents in the group. In this study, the introduction of the talking agent variable was a limitation as it complicated the analysis. Advantages of single agent designs are better generalizability of the findings as compared to the group setting in the present work. It is further suggested to examine the effects of the high and low agent by implementing an agent with mixed intimacy behaviors.

CONCLUSIONS

Proxemic and gaze behaviors deserve attention when designing virtual humans in immersive VR settings where users and agents share the same space. On the one hand, these behaviors can affect how users position themselves in the space, and given the spatial restrictions that most virtual (and physical) spaces have, the desire to change space may not always be possible to satisfy, which may lead to extreme responses, such as the outliers we observed. What is more, proxemic behaviors also affect how the agent's personality is perceived, which may change the outcome of the interaction in other undesired ways. This ET-inspired approach is a useful tool for human-agent interaction design and analysis in shared spaces. It may benefit from advances in VR technology such as in-headset gaze estimation and physiological sensors, which may be used to reveal more on the interaction between proxemics, gaze and intimacy.

3

A VIRTUAL SOCIAL CONFLICT



In the previous chapter, we established how experiences with virtual characters in VR can be perceived as plausible to the point that users' behavior is, in some facets at least, consistent with what we would expect from co-located interactions. We used the Equilibrium Theory (ET) model to predict this response – an increase of the intimacy of the situation, by virtue of the agent stepping closer, was compensated by re-establishing a more comfortable interpersonal distance. There, we focussed on manipulations of the intimacy levels through non-verbal behavior cues by the VH, while keeping the subject of the interaction consistent in terms of its level of intimacy.

Argyle and Dean however also considered that compensation behavior would occur in social interactions when the intimacy level of the subject changed, that is to say, when a sensitive topic is discussed or a conflict arises. This has informed the design of the study discussed in this Chapter[1], where we will look at whether we can mediate an interaction that is perceived as plausible and stressful through the content of the interaction alone (i.e., *what* the agent is saying) - as opposed to the pure non-verbal stimuli used by the VH in the study discussed in the previous chapter. We are assuming that if the content of the interaction is increasing in intimacy (or stressfullness), we will observe behavioral responses that compensate for it.

There are few examples of research on VHs intentionally inducing stress in humans. Blankendaal et al. compare how people react differently to aggressive (virtual) humans (Blankendaal et al., 2015). Bosse et al. have employed aggressive virtual *bad guys* to investigate whether threats received by the VH were taken more seriously if participants believed that the VH could cause harm to them in the real world (Bosse et al., 2018).

For the study we are discussing in this chapter however, we are not looking to induce an increased level of stress by the non-verbal behavior of the agent, or by a potential threat of physical consequences. Instead, we are looking to induce stress through the content of the conversation with the ECA, by raising a *social conflict* in the dialogue. A social conflict is *a process in which there is a perceived mismatch between one's and another person's (or people's) beliefs or interests* (De Dreu, Van Dierendonck, and Dijkstra, 2004). The study presented in this chapter was framed in an applied context; investigating a possible VR "social conflict training application" for teachers in academia, featuring an embodied conversational agent.

Participants of our study were chosen to be PhD students and teachers with some experience supervising students. This allowed us to design an experiment apparatus where the dialogue would converge towards invoking a relatable social conflict for the population of participants. There is a variety of social conflicts that teachers in academia face in their day-to-day work life (Findlen, 2000), and the type of conflicts that we chose for our experiment is *grade disputes*, where student and teacher disagree on what is a fair grade for delivered work - a particularly common type of social conflict encountered by teaching professionals (Carless, 2006).

In this study, to improve on the apparatus in the previous study, we included an in-headset eye-tracker to measure gaze behavior, rather than using head orientation as a proxy for gaze. As an additional control for elicited stress levels, we also included physiological measures, which we expected to reveal an increase in arousal if the agent could plausibly induce an increased stress level. A suite of subjective

[1]Based on our papers J. Kolkmeier, M. Lee, and D. Heylen (2017). "Moral conflicts in VR: Addressing grade disputes with a virtual trainer". In: *International Conference on Intelligent Virtual Agents*. Springer, pages 231–234; and M. Lee, J. Kolkmeier, et al. (2021). "Who Makes Your Heart Beat? What Makes You Sweat? Social Conflict in Virtual Reality for Educators". In: *Frontiers in Psychology* 12

questionnaires on agent perception as well as an interview on the experience and the concept of virtual trainers in academia was administered post-experiment.

Summarizing the results, we were not able to find differences in gaze patterns between conflict and non-conflict parts of the experiment. In terms of physical responses, the analysis only allowed for limited conclusions on the effect of the conflict aspect of the dialogue. An exploratory follow-up analysis on the additional subjective agent perception revealed some interesting correlations, suggesting agency and autonomy beliefs about the virtual trainer may be reflected in physiological responses. Finally, we also discuss some insights gained from a qualitative analysis on the post-experiment interviews on how the conversation with the agent was perceived by participants.

3.1 USER STUDY

We invited participants to have a conversation with a virtual human (VH) about their experience as teachers in academia. At one point in this conversation, a fictive grading dispute by an anonymous former student of the participant was raised by the VH. Our primary aim of this user study was to see if a VH could plausibly induce increased levels of stress through dialogue alone, by evoking a social conflict situation. Since conventionally, VHs or other (conversational) agents in HCI are seen and designed as helpful and polite, we were also interested in qualitatively investigating how participants perceived the conversation when the VH started evoking the social conflict, which is an accusatory act, containing some (indirect) blame towards the participant. To address these questions, we collected quantitative data on the behavioral responses in terms of participants' gaze behavior, based on the insights gained in the study discussed in the previous chapter.

Social conflicts involve emotional and physiological responses that manifest in cardiac and electrodermal activity (Hardy and Smith, 1988; Kreibig, 2010). There are examples of research measuring these physiological activities in response to environmental stressors in VR (e.g., Garau et al., 2005; Meehan et al., 2002), and also examples of research where the stressors are virtual humans with aggressive behavior (the aforementioned works of Blankendaal et al., 2015 and Bosse et al., 2018). In our study, we have measured Heart Rate (HR), Heart Rate Variability (HRV), Skin Conductance Level (SCL) and Skin Conductance Response (SCR) as indicators for how stressful the social conflict was perceived.

In addition, quantitative subjective measures in the form of post-experiment questionnaires on the agent perception as well as qualitative data from transcribed post-experiment interviews were used. This constitutes a mixed-methods approach with concurrent triangulation (Creswell et al., 2008).

3.1.1 METHODOLOGY

To create a conversation that features a social conflict as described in the research question, we used a virtually embodied conversational agent controlled through using the Wizard-of-Oz (WOz, Dahlbäck, Jönsson, and Ahrenberg, 1993) method. There were a set of agent utterances prepared to be triggered - unbeknownst to the participant - by the researcher to maintain a conversation. The utterances are designed to allow for a dialogue with three different stages. First a small talk conversation about the participant's day, then a conversation about the participant's professional experience as a teacher or supervisor, and finally the *conflict* stage, where the VH will bring up the grade dispute as a social conflict. The conflict stage is described in more detail in Section 3.1.2.

Thus, the experiment features one within-subject variable, the *Stage* of the conversation. Stage had three levels, *small-talk*, *teaching* and *conflict*. Stages were kept in the order as they would naturally occur in a conversation - that is, starting with the *conflict* stage and then switching to *small-talk* would seem unnatural and might otherwise influence the perception of the VH. For purposes of the physiological signals, a breathing exercise was included, which is treated as a fourth level in some analyses.

The main dependent variables we are looking at in our analysis are the physiological and behavioral responses per stage, specifically heart rate, heart rate variability and electrodermal activity as well as the gaze behavior, measured by the amount of eye contact made with the agent per stage (see Section 3.2 for more details).

Additionally, we administered a set of questionnaires regarding the participant's experience as a teaching professional, regarding their perception of the agent and of the overall interaction. More details about the procedure of the experiment, how the dialogue was wizarded and the post-experiment questionnaire are discussed in Section 3.1.5.

3.1.2 THE CONFLICT STAGE

The objective of the dialogue in the conflict stage was to evoke a social conflict through the dialogue, a grading conflict with a student in particular. We decided to have the VH evoke, in a plausible way for the participant, a grade conflict that supposedly occurred involving the participant and an anonymous former student. To initiate this conflict, we first had the agent change the subject from *teaching* in general to *grading*: *So, let's talk more about grading*. Followed by two more utterances introducing the conflict: *We have some records from the exam committee about a previous student of yours*, and *There is actually one student who was unsatisfied with the grade they were given*. A number of follow-up questions were included, intended for different responses the participant may give. For example, if participant would inquire about more information, or would just remain silent, the agent would follow up with *According to the student, you did not give timely feedback*. If a participant expressed outright refusal of the possibility that there was a complaint, we had prepared utterances to continue discussing the hypothetical, such as, *Would you change a grade of this student if your supervision was an issue?*.

While this dialogue may not represent a first-hand social conflict, but rather talking about it by proxy, it features many potential components involved. A former student that disputes the grade they were given is introduced, and accusations about the participant's poor professional behavior are relayed towards them. As such this conflict dialogue should represent a stimulus suitable for investigating the role of behavioral and physiological signals in social conflicts discussed with VH.

3.1.3 APPARATUS

The experiment apparatus consists of a virtual environment built in Unity, featuring a VH. The VH[2] was driven by ASAP-Realizer state-of-the-art behavior realizer (Kolkmeier, Bruijnes, Reidsma, and Heylen, 2017; Van Welbergen, Yaghoubzadeh, and Kopp, 2014), through our ASAP Realizer to Unity bridge (Kolkmeier, Bruijnes, Reidsma, and Heylen, 2017). Participants wore the Fove VR-headset to access the virtual environment, where they were seated opposite the VH (see Figure 3.1).

The VH used no hand gestures and kept a neutral facial expression throughout the interaction, meaning not overly positive or negative in behavior to focus on the content of the conversation as the driver of social conflict. Speech was realized using MaryTTS[3]. The utterances used by the VH were pre-defined and could be triggered by the experimenters (see procedure in Section 3.1.5). A Logitech webcam recorded video and audio of the participants from a frontal perspective. Participants eye-gaze in the virtual world was captured with the integrated eye-tracker of the *FOVE* HMD used in the VR setup. For physiological measurements of heart rate and electrodermal activity, the BioSemi ActiveTwo system was used.

3.1.4 PARTICIPANTS

After receiving ethical clearance, we recruited 42 participants from the University of Twente (n = 20) and the Eindhoven University of Technology (n = 22) through snowball sampling by mainly targeting PhD students with teaching duties. Their

[2]Created with the Unity Multipurpose Avatar system: https://github.com/umasteeringgroup/UMA

[3]http://mary.dfki.de/

Table 3.1. Some of the Agent's utterances in the WOz setup per part of the conversation.

Part	Utterances
Small talk How are you?;	What are you currently working on?; Could you describe a typical work day? What are you currently working on? Could you describe a typical work day?
Teaching	Have you given lectures?; Can you tell me more about a recent student project?; How do you feel about being in a teaching position? Can you tell me more about a recent student project? How do you feel about being in a teaching position?
Conflict	According to the student you did not give timely feedback.; Have you had such complaints from students before? Would you change a grade of this student if your supervision was an issue? Have you had such complaints from students before?

average age was 31 and 12 were women. Participants were naive to the experiment subject of grading conflicts, and we did not recruit based on prior experiences with those conflicts. Six participants revealed in the exit-interviews that they had such experiences in real-life, prior to the experiment.

3.1.5 PROCEDURE AND DATA COLLECTION

Before the study, participants read and signed the informed consent form that stated that they would talk about students with a VH and that their participation was voluntary. When participants were directed to the study set-up, an experimenter attached the electrodes for physiological measures, helped with putting on the headset, and performed the required calibration for the in-headset eye-tracking system. Next, participants were given instructions that the VH would begin the discussion after a breathing exercise. After the first breathing exercise of 120s, the conversation with the VH would begin.

From here on, VH utterances were *wizarded* by the experimenters from a separate area. For each *Stage* of the conversation there was a different set of predefined utterances (see Table 3.1). First *small-talk*, then *teaching*, and finally the *conflict* stage. Here, the VH asked for permission to talk about an anonymous, former student who voiced a complaint. The VH claimed that the student was unhappy about the received grade and blamed the participant. The wizard was not allowed to use utterances from an earlier stage. Utterances within a stage were selected in a semi-structured way. Each stage was sustained for 2-3 minutes each. On average, the first stage ended up shorter than the other two (Small Talk M = 112 s, SD = 30 s; Teaching M = 187 s, SD = 54 s; Conflict M = 199 s, SD = 54 s)[4].

[4] M = Mean, SD = Standard Deviation

Figure 3.1. The experiment setup (left). The participant is immersed in a virtual office that matches the real room, with the VH located across the desk (shown right).

In addition to the stage specific utterances, there were some generic utterances like "why is that?", "yes", or "no" to drive the conversational flow. All utterances were logged with a time stamp. In the analysis, stages were segmented based on the first use of an utterance specific to each stage.

The VH ended the discussion by thanking participants, communicating that the researcher will be back after a breathing exercise, and saying good-bye. Afterwards, all equipment for audio, video, VR and physiological measures were removed. We also assured participants that no actual students' complaints were involved in the study. Next, participants completed a questionnaire on demographics, their subjective stress levels for the past few days (Hashimoto, Mojaverian, and Kim, 2012; Suzuki, 1997), items on the perceived warmth and competence of the agent (e.g., *I thought the dialogue partner was approachable*; Guadagno, Blascovich, et al., 2007; Huisman, Kolkmeier, and Heylen, 2014b; Kolkmeier, Vroon, and Heylen, 2016). We also asked if they found the conversation with the VH stressful and if they would find similar conversations with an actual university exam board or a student stressful (modified scale (Mezo et al., 2005)). They then partook in semi-structured interviews on their teaching philosophy, how they felt about the VH and the conversation, the VH's negative assessment about them, and what they thought about the VH for training purposes.

3.2 ANALYSIS

We looked into potential outliers, failed recordings, and artifacts first. From the post-experiment questionnaires, no participant indicated heart related problems and the sample did not report to be highly stressed overall; the average subjective stress level was at 0.72 on a 4-point scale from 0 to 3 (SD = 0.46, MIN = 0.00, MAX = 1.56) according to the stress response scale (Suzuki, 1997).

From the total sample (n = 42), 8 were excluded from physiological analyses due to failed ECG recordings, likely due to conductivity failure of the electrodes, as manifested in either completely (n = 5) or partially (n = 3) missing data and prolonged sections showing artifacts[5]. From the remaining sample (n = 34), 2 more participants were excluded from EDA analyzes (SCL and SCR) for the same reason; failure of the skin conductance electrodes. We have audio recordings from 28 post-study interviews that were transcribed; 1 participant did not partake in the interview and recordings were not successful for the rest. When relevant, interview notes for all participants were referred to for analysis, alongside videos.

Gaze Behavior We obtained participants' eye-gaze through the in-headset eye-tracker of the HMD at 30Hz. We used the approach described by Diaz et al. to detect gaze shifts and fixations (see Figure 3.4). Due to the lower sample rate of our data, we selected smaller kernels for reducing noise in the gaze signal and for producing the exaggerated valleys before and after gaze shifts in the filtered gaze velocity signal. We further filled in gaps of gaze data during blinks with the nearest values before and after the blink. This way, when gaze is shifted while eyes are closed, the shift will still be detected. Using the independent data from the left and right eye trackers, we excluded gaze shifts that were not observed on both eyes at the same time.

Gaze fixations are determined as segments between identified gaze shifts. We compute the fixation durations as time spent gazing at the agents or elsewhere by thresholding the angle between vector of the participants gaze and the vector towards the center of the VH's head. A threshold value of $15°$ was chosen based on the histograms of the gaze angles, where two peaks could be identified, one around $0°$-$15°$ and another around $20°$-$30°$. We compute a *gaze proportion of time measure* in a given segment of the interaction by dividing the total duration of all gaze fixations in by the total duration of the gaze fixations that occurred under the $15°$ threshold (i.e. that are directed at the VH head) within that segment. A *gaze duration* measure for a segment of the interaction was computed by calculating the mean duration of all fixations on the VH's head within that segment.

Following the rational based on the Equilibrium Theory, as elaborated in Chapter 2, our hypothesis was that participants' gaze behavior would be increasingly avoidant between the non-conflict and conflict parts of the dialog, due the personal, accusative nature of the scenario. However, we expected gaze to also be significantly affected by whether the participant was listening or speaking themselves. Thus, the gaze data has been segmented and labeled based on the speech activity of the participant and agent, that is to say, based on who has the turn of speech. Agent speech activity

[5]Potential causes for these failures were experimenter error in applying the sensors or due to unexpected movements of the participants interfering with the setup.

has been extracted from the VH behavior realizer logs, while participant speech activity was extracted by thresholding the voice activity from the participant audio recordings.

Physiological Responses We computed heart rate variability per stage using the Kubios HRV analysis software (Tarvainen et al., 2014). The skin conductance data was processed using using Ledalab (Benedek and Kaernbach, 2010) using the Continuous Decomposition Analysis technique to decompose skin conductance into it's tonic (longer term skin conductance level, SCL) and phasic (short term peak skin conductance responses, SCR) components. For SCL, we average the level over the duration of the *stage*. To obtain an SCR measure, we normalize the number of detected peaks by the duration of the *stage*.

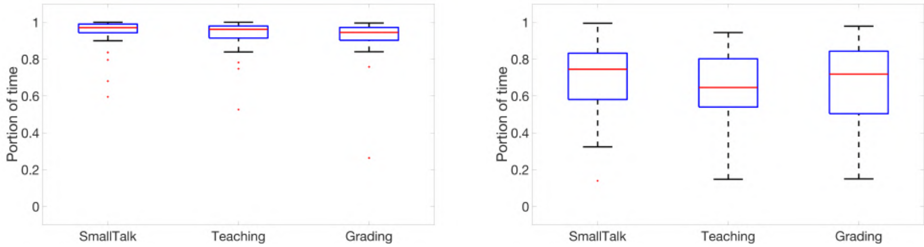
To formulate a hypothesis on how our manipulation affects physiological response, we looked at prior research on physiological responses to events in VR. It seems that skin conductance responds to the overall level of stressfullness of a virtual environment (i.e. Blankendaal et al., 2015; Bosse et al., 2018), while HR may be more sensitive to particular events (i.e. Blankendaal et al., 2015; Garau et al., 2005; Meehan et al., 2002), specifically the VH's social behaviors (Garau et al., 2005). The conflict stage of the VH dialogue consists of several small events (i.e. the confrontational, accusatory utterances) following each other. We hypothesized that there would be an increase in all four measures of physiological activity in the conflict stage compared to the previous two stages.

We analyzed the physiological data in three ways. First, by looking at average levels over the *stage* of the conversation (see Table 3.2). Second, by looking at individual trajectories of changes on a per-utterance basis (see Figure 3.5). Finally, we explored correlations between physiological data and participant responses to post-experiment survey data.

3.3 RESULTS

3.3.1 GAZE BEHAVIOR

When the VH was talking, participants mostly looked directly at the VH (*M=.93*, *SD=.10*), whereas during their own turn, participants also spent some proportion of the time averting their gaze, (*M=.67*, *SD=.21*). When the VH was talking, participants also had longer continuous fixations on the agent's face (*M=4.33s*, *SD=2.40s*). During their own turn, fixations on the VH were shorter (*M=1.86s*, *SD=1.14s*). Comparing the means between the three topics of conversation of the conversation however, we find



(a) During agent turns **(b)** During own turns

Figure 3.2. Proportions of time in turns spent by participants fixated on the VH's face (rather than averting gaze).

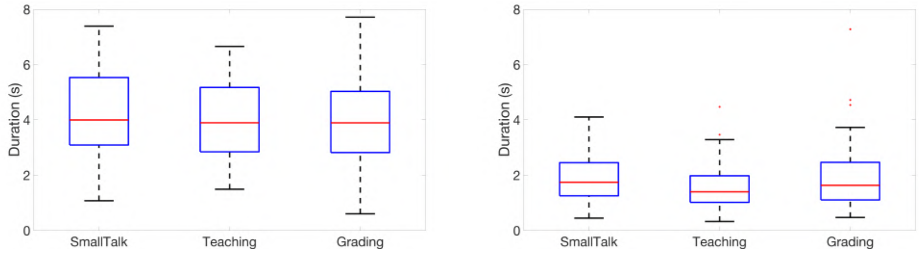
little difference. A repeated measures ANOVA with a Greenhouse-Geisser correction confirmed this for both the proportion and duration measures.

In Figure 3.2, the *proportion of time* measure is shown, representing the time participants held gaze focused at the VH's face, during the agent's turn and the participant's own turn. As expected, when the VH was talking, participants mostly seemed to have looked directly at the agent (over the entire conversation: *M=.93*, *SD=.10*), whereas during their own turn, participants also spent some proportion of the time averting their gaze, (over entire conversation: *M=.67*, *SD=.21*), notably there were greater differences between participants.

Figure 3.3 shows the distribution of the *average duration* of participants' fixations on the agent's face before averting gaze. When the VH was talking, participants appear to have had longer continuous fixations at the VH's face (over the entire conv.: *M=4.33s*, *SD=2.40s*). During their own turn, fixations on the VH appear shorter, before averting gaze again (over entire conv.: *M=1.86s*, *SD=1.14s*). Our distributions indicate that the differences between the three topics of conversation are very small. A repeated measures ANOVA with a Greenhouse-Geisser correction confirmed that there was no significant difference between three stages in Table 3.1. No pairwise comparisons were made.

3.3.2 PHYSIOLOGICAL RESPONSE

We observed a significant effect of stage for all physiological measures according to within-subjects repeated measures analyses of variance (ANOVA), HR: $F(3, 33) = 11.99$, $p < .001$; HRV: $F(3,33) = 3.71$, $p = .014$; SCL: $F(3, 33) = 26.45$, $p < .001$; SCR: $F(3, 33) = 15.13$, $p < .001$ (individual level differences are nested within-stage level analysis). Pairwise comparisons then showed that for most measures, the difference is between the baseline and each respective stage (Table 3.2). Only for HR and SCL were there



(a) During agent turns **(b)** During own turns

Figure 3.3. The average duration of participants' fixations on the VH's face before averting their gaze in three parts of the experiment dialog.

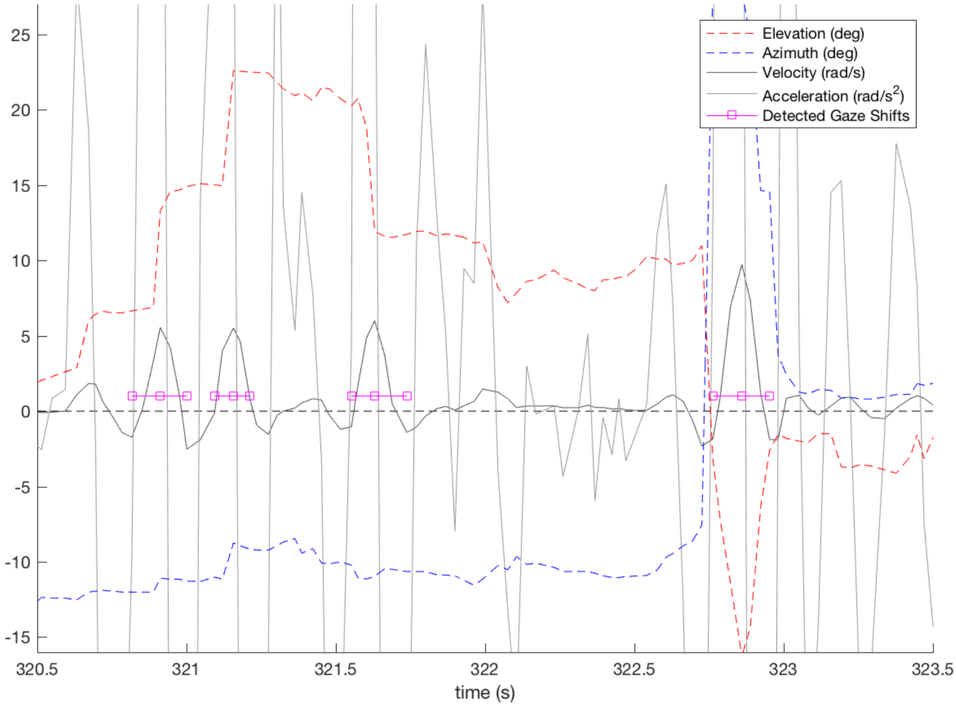


Figure 3.4. Identifying Gaze Shifts using the method from Diaz et al. (2013). The velocity and acceleration are derived from the change in eye-gaze direction over time. Zero crossings in the acceleration curve are considered as centers for gaze shifts if the velocity at the same point exceeds a threshold. After convoluting the velocity signal with a kernel function that produces negative valleys near gaze shifts, the borders of the start and end times of gaze shifts are determined at the closest nearby zero crossings in the velocity signal.

significant differences between the other stages. HR *decreased* significantly between Small Talk and Conflict by an estimated 2.45 beats per minute (SE = .93 bpm, p = .045), although also the difference manifested for the most part between Small Talk and Teaching near significance (est. decrease of 2.45 bpm; p = .085). SCL *increased* significantly by an estimated 0.487 µS (MicroSiemens) between small talk and Conflict (SE = 0.156 µS, p = .0003).

3.3.3 EXPLORATORY ANALYSIS

To investigate the variance in our physiological measures, we explored correlations with how the interaction and the agent were perceived, and with demographic data. To this end, we averaged the physiological measures across all stages and normalized them by the baseline measurements from the breathing exercise.

Highest correlations between the normalized physiological measures and subjective measures were found between HRV and the perceived level of the VH's autonomy. Although this negative correlation was not significant (r = -.35, p = .058), the tendencies suggest that those with higher normalized HRV reported they believed the VH was controlled by a human (instead of by a computer). Normalized SCL moderately negatively correlated with the extent to which participants reported that such a situation would be stressful to them in a real encounter, again though, this correlation was not significant (r = -.32, p = .085). The normalized HR was weakly linked to participants' self-report of finding the conversation with VH to be stressful, but was not statistically significant (r = .29, p = .12). Other self-report and demographics items such as gender or perceived interpersonal attitudes such as agent perceived warmth did not correlate with the normalized physiological measures.

DESCRIPTIVE ANALYSIS

The above analyses are based on averages of physiological signal over the time of the different types of conversation stage. We wondered if perhaps individual utterances in the conflict stage had induced stronger physiological reactions. In the following, we look at local changes on a per-utterance level by visualizing HR response trajectory density over the time around the utterance (see Figure 3.5). Darker areas indicate that responses of many participants followed a similar trajectory of change. In addition, two participants' trajectory lines are plotted to highlight individual differences.

No utterances shorter than 2 seconds were included (e.g., "yes"). Not all utterances were used for all participants to preserve conversational naturalness. For example, utterance U03 was not used with either one of the two highlighted participants.

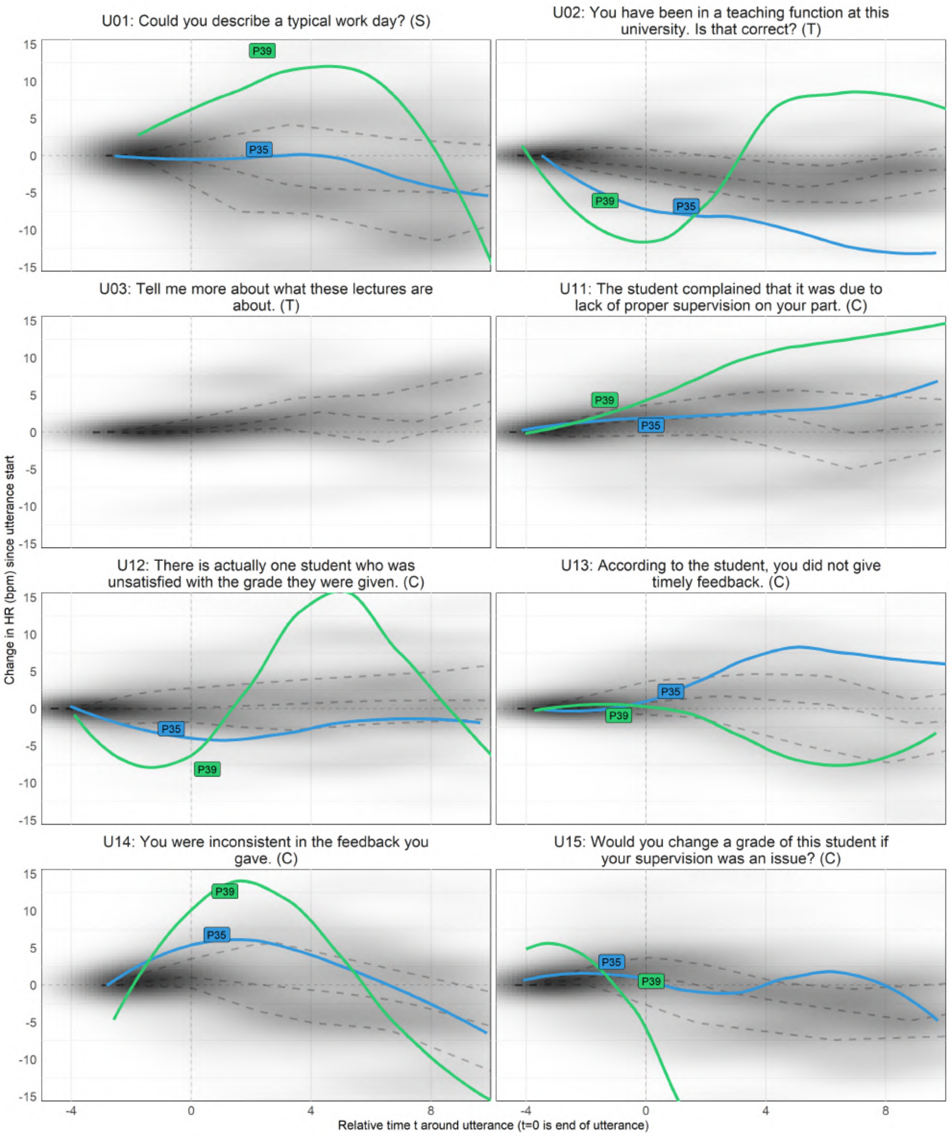
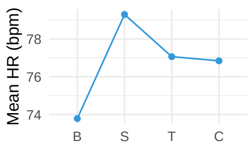
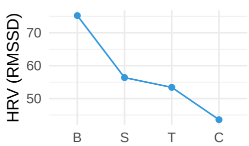
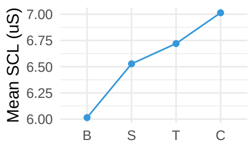
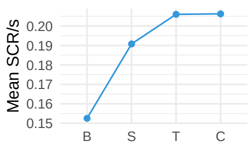


Figure 3.5. Density of changes in HR after hearing agent utterances over time, in seconds, per utterance. Utterance end-times are aligned at x=0. Quantile regression lines are shown as gray dashed lines, at 75%, 50%, and 25% from top to bottom (RQSS method, Koenker, 2019). For illustrative purposes, trajectories of two individually examined participants are shown as lines fit with local regression for their measured change in heart rate.

Table 3.2. Pairwise comparison of physio measures by stage (Tukey adjusted P-values). The four stages are B=Baseline, S=Small Talk, T=Teaching and C=Conflict. Estimated marginal means from the models are in the first column.

EMMeans	cntrst	est.	SE	df	t.ratio	p.value
	B - S	-5.51	.93	99	-5.96	< 0.0001
	B - T	-3.28	.93	99	-3.54	0.0033
	B - C	-3.06	.93	99	-3.31	0.0071
	S - T	2.24	.93	99	2.41	0.08
	S - C	2.45	.93	99	2.65	0.045
	T - C	.22	.93	99	.24	1.0
	B - S	18.86	9.73	99	1.94	0.22
	B - T	21.83	9.73	99	2.24	0.12
	B - C	31.66	9.73	99	3.26	0.0083
	S - T	2.96	9.73	99	.31	0.99
	S - C	12.80	9.73	99	1.32	0.56
	T - C	9.83	9.73	99	1.01	0.74
	B - S	-0.513	0.116	93	-4.44	0.0001
	B - T	-0.706	0.116	93	-6.11	< 0.0001
	B - C	-1.000	0.116	93	-8.65	< 0.0001
	S - T	-0.193	0.116	93	-1.67	0.35
	S - C	-0.487	0.116	93	-4.21	0.0003
	T - C	-0.296	0.116	93	-2.54	0.06
	B - S	-.038	.009	93	-4.16	0.0004
	B - T	-.053	.009	93	-5.81	< 0.0001
	B - C	-.054	.009	93	-5.84	< 0.0001
	S - T	-.015	.009	93	-1.65	0.35
	S - C	-.016	.009	93	-1.68	0.34
	T - C	-.000	.009	93	-.03	1.00

We identified three trends in the density of HR changes around individual utterances in Figure 3.5. First, HR peaks occurred at different points in time: Some peaks are located around the time where participants started answering, while other peaks formed already while the VH was still talking (e.g. "according to the student, you did not give timely feedback", U13). Second, patterns of HR increase and decrease follow different patterns per statement and between participants. While some conflict utterances indicate overall upwards trends for the majority of responses (e.g. U11), other conflict utterances feature overall downwards trends (e.g. U15). Lowering HR responses are not uncommon even in emotionally loaded responses, and may be associated with people redirecting their attention to the new stimuli or thinking process (Koruth et al., 2015; Lang et al., 1993) and some emotions decelerate the HR (e.g., suspense or fear Kreibig, 2010). Third, we also notice that statements outside of the *conflict* stage elicited increases in HR (e.g. U01 for the P39, and U03 with an upwards trend for the overall responses). So overall, it seems that there is large individual variance in participants, perhaps based on their own background and experiences on the topics covered per stage. This makes it difficult to identify, at least visually, clear patterns in local HR changes based on the type of utterance, or whether it addresses a conflict situation.

VIDEO ANALYSIS

As a final attempt to explore possible patterns in the data further, we decided to look back at the video data at specific moments with participants that exhibited particularly large changes in HR around utterances. Four participants in particular contributed a sizeable number of the highly ranked utterances (40/112 utterances = 35.7%): P24, 32, 39, and 40 (P39 is also shown in Figure 3.5). Three of the four participants dealt with the *conflict* utterances with denial (e.g. "that's impossible because I always give time", P40), and one accepted the possibility but was not particularly bothered ("that's too bad. I suppose life happens", P32). Two of these participants also indicated post-experiment that they felt discomfort with the system either due to problems with eye-sight in the HMD (P40), leading to often adjusting the headset, or due to difficulty with understanding the VH's speech synthesis (P32). Except for P39, none of these four participants indicated feeling particularly stressed about discussing the grade dispute with the VH.

Overall, we find again little clear patterns in this analysis. Some changes may be attributed to discomfort, but it is also notable that also self-described stress levels do consistently correlate with the observed data.

3.3.4 THEMATIC ANALYSIS

A thematic analysis was conducted on the participants' responses during the experiment and the post-experiment interview. In the following, we provide a summary of the two main themes relevant for this work, regarding perception of the VH and the perceived immersion during the interaction. A more extensive version of this analysis, also discussing the themes on cross-cultural differences in how the VH was perceived, as well as themes on the potential integrations of similar systems in real work environments can be found in our paper (M. Lee et al., 2021).

Our first identified theme was that *Expecting human-likeness leads to negative evaluation*. From the interviews, we found a number of reasons why participants evaluated the agent negatively. One reason was that participants did find that the VH was not perceived as empathetic, based on an observed lack of facial expressions. Rather than perceiving them as neutral, some noted that expressions looked cold or "mad" (P18). Another reason participants gave for negative evaluations was their uncertainty about the identity of the VH. P32 noted: "I have no information about whether my answers are simply being recorded or whether they're actually influencing the next question; whether it is in fact a conversation, or just filling in of a virtual form with my transcribed responses". P26 likened the VH to a "confrontational GIF", and noted that "sometimes I felt that it's stupid to talk to an avatar because I wasn't sure to who I was talking". In other words, participants were attempting to consolidate the human appearance with both what it may be (i.e. a capable conversational machine, or perhaps merely an embodied *Q&A* form) and who it would represent (i.e. if it was truly acting on behalf of the exam committee). Finally, negative evaluations were also attributed to the technical implementation of the VH and the VR environment. Some, as already noted above, found it difficult to listen to the VH due to the quality of the speech synthesis. One participant even noted that it was like being "interviewed by an ATM". P40 also noted that also the visual style was distracting, stating that the "environment is fictitious but [...] you notice this more because the graphics are not high end [...], which interrupts (the) conversation a lot".

Even though participants reported on the limitations of the VH in terms of realism, we also identified a second theme on how realistic and immersed participants felt about the overall conversation with the action. The theme is that *Immersion refers to situational or conversational realism, which can be decoupled from how "real" an environment or an agent looks*. Especially those participants who expected the VH to be purely machine-like were more positive on the overall experience during the experiment, allowing that human-like conversations could be had even though it was clearly a machine. Some noted that it was easier to talk about these kinds of conflict situations with a VH, as they "are virtual so I'm more open to talk about anything. But based on the assumption that they are not real" (P25). P28 went even further, stating that "It feels like that you can talk more about it. You feel more relaxed to talk more because it's fake. Not the conversation, I mean the whole situation can be like in a dream [...]". Others noted that they felt more engaged in the interaction precisely because it was about a conflict situation. They felt put on the spot as they were able to relate to past experiences regarding the content of the conversation, such as P19, who shared during their interaction with the VH that: "there was a student group that was unsatisfied with their grade we gave indeed. [...] um, we have a, it was a sort of misinterpretation or miscommunication [...] between the supervision team because I didn't supervise them all by myself". Others again noted that the precise moment when the conflict situation was brought up made them more engaged in the interaction, even though they were not able to remember an actual prior experience of a grade conflict: "even if you know it's a hypothetical situation, you don't want to be put on a spot like that. [...] I would feel pretty bad if the student did not find the grade fair, even in a hypothetical situation" (P03). Similarly, P01 brought up that the relatability of the context of the conversation caused them to get immersed: "At first, it was hard to take the VH seriously. You're really out of your context. At the same time when the VH raises questions that some people are not happy with you, then the problem becomes more realistic. So all of a sudden it became very immersive for me to talk with the VH because it doesn't really matter anymore (that it is virtual), and the problem becomes real and I have to pay attention and be serious".

3.4 DISCUSSION

We have looked in a number of ways at how our participants reacted to a grading dispute, a professional conflict, as evoked in an interview with our VH.

3.4.1 GAZE BEHAVIOR

Based on the work discussed in the previous chapter, we hypothesized that participants would show a change in gaze behavior once the subject of the conversation would switch to the conflict situation. Although we couldn't definitely demonstrate this effect in that previous experiment, we expected it was due to the lack of a fine-grained eye-tracker. Additionally, we expected such a response to be exaggerated in a seated interaction, where increases in perceived intimacy could not be compensated by increasing interpersonal distance. In our measures, we expected increased frequency of gaze aversion and overall less eye contact with the agent when addressing the conflict. The present analysis however did not reveal any significant differences between the non-conflict and conflict related parts of the conversation.

This may indicate that our manipulation may not have been successful for participants, meaning the VH was unable to believably evoke the conflict situation through the dialogue. To understand whether this was the case, we next discuss participants' physiological responses as well as how they perceived the interaction, as quantified by the questionnaire responses and by looking at their responses in the post-experiment interviews.

3.4.2 PHYSIOLOGICAL DATA

While we found significant main effects of conversational stage on some physiological measures, we must be careful when interpreting them due to the lack of counterbalancing their order in our study design. For example, the level of skin conductance, being related to the activity of sweat glands, is known to drift upwards over time, due to sensors covering the skin. This favors our hypothesis as the conflict stage is placed towards the end of our interaction. What is more, although the conflict stage measured higher than Small Talk on the HR and SCR measures, so does the teaching stage, which we did not expect to be stress inducing. Thus, similarly as with the gaze behavior, based on the physiological responses alone, we cannot conclude that participants found the parts of the conversation where the VH evoked the social conflict to be stressful.

To examine this more, we followed up with the exploratory analysis, looking at possible correlation of the physiological responses with subjective measures. First, we note the correlation between HRV and the perceived autonomy of the agent, which was unexpected. One explanation can be that people who believed that the agent was controlled by a human, such as the experimenter or perhaps someone from the exam committee, could trigger them to feel observed, influencing people's HRV. A more expected result was the positive correlation between normalized HR and how stressful the interaction was perceived. Interestingly, this was not in line with the decrease in overall heart rate towards the end of the interaction, further suggesting that order effects may be at play.

We further visualized local trends in changes of heart rate during the times after specific utterances and analyzed them descriptively, to examine whether possible effects might have been diluted by the averaging of per-stage measures. Although there were apparent differences in the trends between utterances, there was no pattern that would indicate that those agent utterances that were used to introduce the conflict, or used to relay the blame of the issue were responded to differently than to those utterances that were part of the less loaded parts of the interaction.

Finally, how participants perceive and respond to individual utterances may depend on the individual, as touched upon by in the exploratory correlation analysis. Another possible concern is that the manipulation was not successful in increasing the perceived intimacy. It may be that discussing grade conflicts simply wasn't inherently stressful to our participants, or that they were able to rationalize away a complaint from an *anonymous student* as being made up - especially when this is being brought up by a VH instead of a real human.

Unlike the purely non-verbal manipulations made by the VHs in this study discussed in Chapter 2, the (stressful) content of the conversation may have been processed more cognitively, causing the *plausibility illusion* to break. As discussed in the thematic analysis, some did find the VH lacking realism, others explicitly noted that exactly the real, serious and relatable topic of the grading was what got them engaged and put on the spot during the conversation with the agent.

3.4.3 LIMITATIONS AND FUTURE WORK

There are also some more limitations to the study design and analysis that need to be considered. A more rigorous study design with counterbalancing of conditions and more exact control of stimuli duration is required to be able to derive better quantitative insights from the physiological measures. In the future, other sources of social signals could be included in the measurements. For one, facial expressions are important for conveying people's opinions, comments, and conversational turn-taking (Poggi, D'Errico, and Vincze, 2013). While this was still a technical limitation at the time of conducting this study, recently facial expression recognition has even made its way into consumer hardware[6].

A lack of a human-human experimental condition or pre-study is another weakness of the present research, as we may first want to confirm whether participants would have plausibly reacted to the social conflict if evoked by a human interviewer, rather than by the VH, which would speak towards whether the manipulation, or rather the stress stimulus, was designed appropriately.

Concluding, we have explored whether evoking social conflicts through the content of a dialogue with a VH causes changes in humans' behavior and physiological responses. Concluding from the results, we can't quantitatively support that a social conflict involving the participant invoked merely through the content of the dialogue of a VH causes participants to show the behavioral and physiological responses that we expected based on the social models we used to formulate our hypothesis. Going back to the predictions of Equilibrium Theory, even if only one of the two expected outcomes of gaze aversion or physiological markers for increased levels of stress would have been observed, it would be compatible. Averting gaze would be a means of reducing the perceived increase in intimacy, and thus we should not have measured increased physiological responses. Similarly, if an increase in gaze aversion has not occurred, the increased stressfulness content of the conversation should have been observed in the physiological data. However, we did not find significant changes in line with these hypotheses in either physiological or behavioral measures, and we have discussed a number of possible explanations for this, also potentially including the limitations of the study design.

[6] See for example https://www.vive.com/eu/accessory/facial-tracker/

4

DON'T SHOOT THE MESSENGER

In the introductory chapter, we introduced the SoIC-TV, a system for social telepresence. On the basis of this system, we sketched out a number of areas where mediated social interaction technology may play perhaps unexpected roles in the future. In the remainder of this thesis, we will attempt to connect back to the SoIC-TV, and see if any previously raised questions can be answered with what was learned in the studies covered.

Although we designed the interaction in the previous chapter to feature a social conflict primarily for the sake of investigating how participants would respond, and the fact that the agent is virtually embodied and co-located is a necessity for being able to use gaze as a measure, the resulting apparatus and the experience that it affords also served as an interesting case study to capture how participants feel about such a system in general, that is, if it was employed in practice. Let us consider one of the potential envisioned use cases of the virtual human as a mediator: The possibility that the VH is employed by the exam committee to mediate an issue with a teacher on their behalf. In the following, we explore how this would relate to the SoIC-TV universe.

The SoIC-TV services for telepresence (see Section 7.1 and Section 1.1.3) are different from the virtual human mediator in several ways. In the case of the "telepresence as a service scenarios", a human is being employed by either a machine or another human to engage in a social interaction, or to achieve some task in the real world. In the studies reported on in the previous two Chapters however, the VH is an embodied machine, featuring a human-like appearance, and co-located in a virtual environment.

Now, would someone feel more comfortable as a student discussing a grade dispute with a supervisor, if they were using a SoIC-TV? Would it be better or worse to break up with a romantic partner using the SoIC-TV, or using an SMS? Could it be used in conflict situations where one party wants to remain anonymous, yet desires to engage in the argument in a real-time, interactive fashion?

The thematic analysis described in Section 3.3.4 captures some interesting findings on how participants responded to the experience during and after the experiment. While some participants questioned whether the machine could be empathetic (beyond surface appearance) and have true understanding of the conflict at hand, others remarked that this lack of capabilities is what would make it easier to talk about such conflicts in the first place, others remarked that raising the social conflict is what made the interaction feel “real” in the first place, and reportdd that their emotional state was affected immediately after the conflict was introduced.

A person engaging in a social conflict through the SoIC-TV service (i.e. as a *passenger*) may enjoy not having to directly confront the other, especially if it was possible to stay anonymous. The reaction of the other party in the conflict may be mixed, at least based on what we learned in our analysis. If the topic is serious and important, they might engage with it in a serious manner. They may even also prefer the level of anonymity that the *passenger* enjoys, and being aware that they are working out the conflict through a neutral mediator (the *driver*).

We have also sketched scenarios where there might be a machine in the loop, renting the SoIC-TV *driver*, and in this case, attempting to resolve the social conflict. Participants in the previous study raised that they wouldn't like being judged by a machine, so this scenario might be problematic, depending on how the *driver* is perceived. In general, participants raised concerns about recognizing the identity of the VH - not just in the context of the lab setting, that is to say, whether the agent was controlled by the experimenter or not, but also uncertainties on whether the VH might actually act on behalf of an authority. Confusion about the identity of the SoIC-TV *driver* may lead to negative consequences, when the proverbial messenger of bad news is confused for the source. The aforementioned mediated *break-up* situation might play out bad for the *driver*.

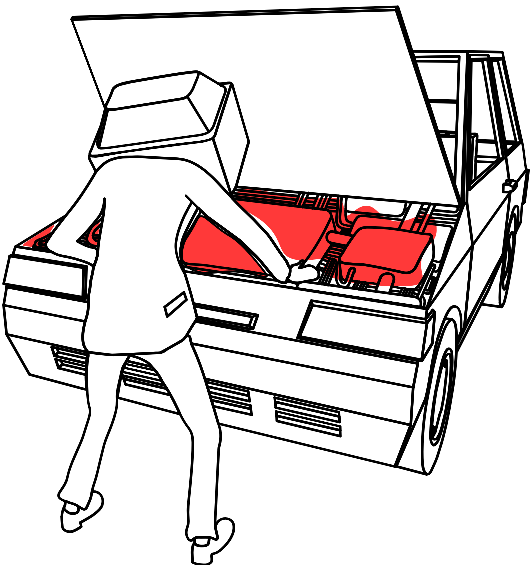
In the next part of the thesis, we focus more on technology mediated social interactions between humans.

PART *III*

PERSPECTIVE AND PRESENCE: ASYMMETRIC REMOTE COLLABORATION

5

MIXED REALITY REMOTE COLLABORATION SYSTEMS



Up to this point of the thesis, we used VR technology to create experiences of social interactions with machines, embodied by virtual humans. As sketched out in the SoIC-TV scenarios in the introduction, VR technology may also be employed to mediate social interactions between humans, for example to enable interactions between two humans in different physical locations. This is what we will focus on in this part of the thesis. In particular, we will look at interactions where humans use technology to interact over distance to *collaborate remotely* on a given task. What is more, we will employ *mixed reality* technology, allowing the mediation of social interaction to include a user's physical space as well.

In the remainder of this chapter, we will introduce these new aspects in a bit more detail. Then, in the following chapter, we will introduce our own novel mixed reality telepresence system, discussing its design and implementation. In Chapters 7 to 9, we will present two user studies conducted with this system and a followup analysis on the main measure of interest – *social presence* – combining the findings from both studies.

5.1 MIXED REALITY REMOTE COLLABORATION

Since humans are not capable of telepathic communication, we use technology to facilitate real-time communication over distance when solving problems. Although a telephone call may suffice for some tasks, more complex tasks may benefit from richer interfaces. For example, repairing a car engine may be difficult for a novice purely based on verbal instructions from an expert on a phone. Identifying one of the many parts of an engine may be easier if the remote expert could point towards the part, rather than attempting to verbally describe it, and trusting that the novice has identified the correct part for themselves. In the following Chapter, we will illustrate more benefits of using mixed reality technology in these contexts, such as allowing the mediation of pointing gestures, in a way that is similar to co-located interactions.

One core aspect that probably every remote collaboration system features is that it allows users to feel either present in a shared virtual environment, or, as in the case of physical tasks, feel present in the remote environment. Since our focus is on physical tasks, we will first introduce the latter case of *telepresence*.

Telepresence technology makes it possible to visit a remote location without physically traveling there. Besides creating the feeling of really *being there*, telepresence technology aims to provide an experience that affords better communication and even collaboration between visitors and visitees.

The ultimate telepresence system would empower a visitor to perceive, traverse and manipulate the remote environment as if they were really there. At the same time, it would allow the visitor to make their presence, attentions and intentions known to others in the remote environment - as if they were really there. Various parts of this complete experience have already been explored to enable certain applications that benefit from specific aspects of *being there*. For example, during a telepresence museum visit, the visitor wants to choose their own perspective of the exhibits and choose their own pace for traversing the large area of the museum, as for example in the TOURBOT and WebFAIR projects (Trahanias et al., 2005). On the other hand, for remote collaboration on tasks, navigation of such large unconfined spaces may be less important. Instead, users benefit from being able to perceive actions and intentions from other users through channels that cover multiple modalities, and from being able to manipulate virtual entities, such as in the work of Regenbrecht et al. (2004).

Current research towards telepresence systems can be divided into two categories; conferencing telepresence systems, where the focus lies on making remote people feel present with each other, and collaborative telepresence systems that, besides making people feel present together also focus on making them feel present in the remote environment.

In conferencing telepresence systems, the focus lies on supporting conversations between people by providing tools to better express themselves compared to a simple phone call, for example, by offering high-quality audio and high-definition life-size video (Feldmann et al., 2010) or stereoscopic displays (Maimone and Fuchs, 2011). A typical application of those systems is often found in conferencing over distance for companies but also in social contexts such as (video-)calling with friends and family. Conferencing telepresence is mostly implemented in a symmetrical fashion, which means that at both remote locations people are captured and visualized in the same way, as there is normally no reason to bias the system in favor of somebody by, for example, giving one user access to tools the other doesn't have; all participants are regarded equal.

Mixed Reality (MR) Remote Collaboration systems on the other hand allow physically dislocated users to be virtually co-located. The system must afford collaboration between people by virtually transporting a remote visitor into the visitee's environment. One example is ShowMe (Amores, Benavides, and Maes, 2015). Here a remote expert receives a live video stream of the first person view of a local user. The hands of the expert are captured using a depth camera. The local user sees the same view of their local environment that is streamed to the remote expert, however, with the hands of the expert rendered transparently into this view. Such mixed reality telepresence approaches are not limited to physically dislocated situations of visitors and visitees, value can also be created for interactants that are already physically co-located. For example, DollhouseVR allows for a playful interaction between a user in VR and non-VR user, by beaming the VR user (doll) into a virtual doll-house. On a table-like screen, the non-VR user can see a top-down view of the virtual dollhouse, with a representation of the doll. Meanwhile, when looking up through the open roof, the doll can see a camera feed of the non-VR user's face, using a camera connected to the table-like screen. This creates an impression for the VR user of being a small doll, while the non-VR user joins the interaction with a god-like representation, from the perspective of the doll (Ibayashi et al., 2015).

Collaborative telepresence systems are often asymmetrical (Steed et al., 2012). Just as in the example mentioned in the introduction about calling in an expert for reparing a car engine, or having a specialist remotely called in for the investigation of a crime scene (Poelman et al., 2012), the remote callers expertise is what's required at the location in question. The asymmetry is not just manifested in the users' experience, but usually goes along with who needs access to what information. The remote expert needs access to the environment of the novice but not the other way around. The visitee, for example, needs help with their car or with the crime scene they are at, while the environment of the visitor is of less interest to the visitee. The visitor needs to engage with and feel present in the visitee's environment, while the visitee needs to understand actions, instructions from the visitor to perform the task at hand.

5.2 CONCLUSION

Collaborative telepresence systems can employ mixed reality technology to allow remote users to work on tasks involving physical space. Scenarios where this is particularly useful are scenarios where there is an asymmetry in the relationship between the users, the access to the information about the physical space, and the modalities that the systems need to mediate for efficient communication between the users.

We specifically focus on this type of asymmetric scenarios in the research discussed in the following chapters. First, in the next chapter, we present the design and implementation of our own novel mixed reality collaboration system suited for these scenarios. Drawing from existing work, we present a toolkit that we developed and used for the user studies in the following chapters. The first user study focuses on evaluating how certain design aspects of a mixed reality telepresence system affects usability, performance, and social communication. For this user study, we conceived a puzzle task, where a remote assistant is virtually called in to support the physical user with a mixed-reality escape room. For the second user study on remote collaboration in mixed reality, we conceived a design-fiction scenario around an emergency situation requiring a novice to administer cardiopulmonary resuscitation (CPR) in a world where Augmented Reality displays are pervasive technology. Thus, the emergency-line dispatcher would not just give instructions by voice, but would also have virtual access to the emergency location and would be virtually visible to the novice in an AR overlay. Here, we again look into what the added values of such a system could be in terms of usability, performance and interaction quality when compared to a phone-bound CPR dispatcher.

6

OPENIMPRESS: IMMERSIVE PRESENCE SYSTEM

In this chapter[1] we describe the design and implementation of *OpenIMPRESS*, our open source[2] immersive presence system toolkit that we developed for building and researching mixed reality (MR) remote collaboration systems. It uses off-the-shelf MR hardware available to anyone with a small to medium budget, to afford interactions between physically dislocated users in collaboration and support scenarios.

The scenarios we will focus on are those asymmetric scenarios described in the previous chapter. In this context, we assign specific names for the different users and their roles in the system. Users who are physically present at the site that is being visited (the on-site location), we call On-Site Operator (OSO). Users not physically present in this environment, but only virtually from a remote location, we call Remote Operator (RO). Again, in these scenarios the visitor (RO) typically gains access to the physical environment of the visitee (OSO). This *on-site* location is of interest to both the RO and the OSO. The RO's physical environment (the remote location) is not so; however, the RO's actions and instructions directed towards the OSO are.

The basic OpenIMPRESS system is shown in Figure 6.1. The system spans two different locations: the on-site location and the remote location, where respectively the On-Site Operator (OSO) and the Remote Operator (RO) are located. At the on-site location, an AR headset (here the Hololens) worn by the OSO and a number of

[1]Based on our paper and the supervised theses S. Giesselink (2018). "Teach me CPR now! Augmented learning for one-shot teaching of cardiopulmonary resuscitation in out-of-hospital cardiac arrest". Master's thesis. University of Twente; E. Harmsen (2018). "OpenIMPRESS: an open immersive telepresence system". Master's thesis. University of Twente; J. Kolkmeier, E. Harmsen, et al. (2018). "With a little help from a holographic friend: The OpenIMPRESS mixed reality telepresence toolkit for remote collaboration systems". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11.

[2]OpenIMPRESS was made available under a permissive open source license: https://github.com/openimpress

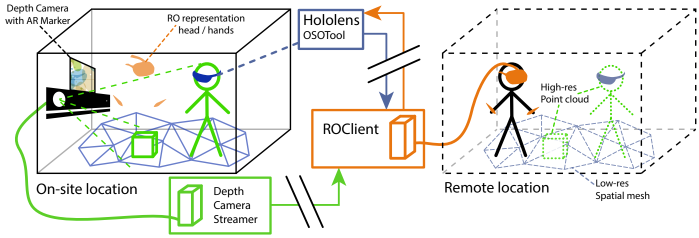


Figure 6.1. OpenIMPRESS basic system overview.

RGB+Depth (RGBD) cameras capture the on-site environment with as much detail as possible. The client of the remote user (ROClient) receives this data, and renders it for the RO through a head-worn VR device. The place and orientation of the RO's head and hands are in turn visualized for the OSO through their Hololens. Additionally, users can create annotations, such as virtual line drawings, in each other's environments. The system also allows the RO and OSO to talk with one another directly. One aim was to provide a modular toolkit, so that it can be easily configured for different conditions. We used a technology driven method to develop our toolkit. It works with emerging off-the-shelf mixed reality hardware that is available to anyone with a small to medium budget. The system should afford collaborative interactions between physically dislocated users, and support various scenarios. To this end, we informed the design of our system with findings from existing MR systems for remote collaboration.

The design of our toolkit is characterised by three main principles that determine the experience of and interaction between users. First, we allow the visitor to move freely in the on-site environment. Second, the visitor engages with the on-site environment in a natural way through an immersive, mixed reality interface that includes gesture tracking. Third, the visitor's presence and actions are visible to the visitee - who is physically present on-site - through a multimodal projection of the visitor and their gestures into the visitee's environment, using augmented reality technology. This allows communication between visitor and visitee, which is further enriched through the possibility to place persistent virtual annotation content in a shared digital layer over the environment. In the following we describe the system and the rationale behind it in more detail.

6.1 RELATED WORK

There is a rich body of work on the design of remote mixed reality collaboration systems. We will discuss this along the three design aspects that we deem crucial for the experiences that our toolkit should afford.

Mobility at the capture site Allowing the visitor to view the scene independently instead of being fixed to a static view (or to the viewpoint of a local user), is an important aspect in telepresence for remote collaboration. It allows users to simultaneously work in different parts of the scene and gives the remote visitor more situational awareness (Fussell, Setlock, and Kraut, 2003). One way to achieve this is to physically manipulate the camera at the capture site. Telepresence robots such as the *Double* have been used to this end (Kristoffersson, Coradeschi, and Loutfi, 2013; Rae, Mutlu, and Takayama, 2014). Although these provide a remote visitor with a physical body and complete freedom to navigate the environment, manipulating the robot can be awkward and can potentially harm others; therefore the controls of these robots typically err on the side of slow and safe. Other solutions synthesize a view based on imagery captured from the visitee's perspective. The visitee wears a camera that streams a first-person perspective to the visitor. This limits the navigational freedom of the visitor, although visitors can still be given some degree of independence from the local person. For example, *JackIn Head* (Kasahara and Rekimoto, 2015) uses a 360° degree camera mounted on a person's head to allow a remote visitor to look around in the visitee's environment. Gao et al. (2016) use a tracked head-mounted depth camera instead of a 360° camera. The images are oriented in such a way that when the visitee turns their head, the projected image for the visitor turns as well, requiring the visitor to turn their head to follow the image. This doesn't give the visitor any extended navigational freedom but helps them keep track of the visitee's current orientation. Tait and Billinghurst use a statically mounted depth camera above a table to create a live 3D scan of the workplace. The visitor is then free to navigate this using a desktop user interface. When tested in a collaborative setting this resulted in "faster task completion time, more confidence from users, and a decrease in the amount of time spent communicating verbally during the task" (Tait and Billinghurst, 2015). Similarly, the RemoteFusion system (Adcock, Anderson, and Thomas, 2013) creates a real-time reconstruction of a scene that can be accessed by a remote user using a touch screen for view manipulation and annotation.

Presentation of the capture site The way the capture site is presented to a remote visitor has an effect on the telepresence experience as well. A more immersive display makes the visitor feel more present (Baños et al., 2004). Telepresence systems can be distinguished by the degree of immersion they provide. Earlier systems used 2D monitors on which a video stream of the visitee is displayed, like the *HandsOnVideo* system (Huang and Alem, 2011). Newer systems try to immerse the visitor more, often through head mounted displays (Tecchia, Alem, and Huang, 2012). Other display systems are also used, like CAVE projection systems where multiple walls of a room are projected with live images from another location (Komiyama, Miyaki, and Rekimoto, 2017). Those systems are often used in combination with head tracking to make the movement of the images match the visitor's head movements.

Having an effect at the capture site Besides providing visitors with a view of the capture site, telepresence systems usually also allow visitors to *affect* it. This can be done by physically manipulating the environment (tele-operation), but more often it is done by adding virtual elements to it that are shown to the visitee in some way (e.g., the JackIn system, Kasahara and Rekimoto, 2014). RemoteFusion (Adcock, Anderson, and Thomas, 2013) uses projectors to project annotations made on the visitor's touch screen onto surfaces of the capture location. For complex assembly tasks, annotations can be used to support alignments of parts (Oda, Elvezio, et al., 2015).

To assist the visitee in tasks that require more fine control, an overlay of the visitor's hands is often used (Alem and Li, 2011; Amores, Benavides, and Maes, 2015; Chénéchal et al., 2016; Gao et al., 2016; Huang and Alem, 2011; G. Lee et al., 2017; Tecchia, Alem, and Huang, 2012). This allows the visitor to guide the visitee in a more natural way in object manipulation tasks, by using gestures.

We can identify a trend in these projects, from video based captures of gestures projected into the the visitee's environment (i.e. Alem and Li, 2011) to capturing and rendering more volumetric representations of the visitor's hands into the visitee's environment. Sodhi et al. (2013) control a 3D model of a hand through a touch interface, whereas the *3D Helping Hands* of Tecchia, Alem, and Huang (2012) capture the user's hands using cameras that also capture the depth of hand positions.

Komiyama, Miyaki, and Rekimoto (2017) show that capturing and streaming complete representations using RGBD cameras is also possible – although, in their case, this is limited to capturing visitees and displaying them to remote visitors. Approaches for end-to-end streaming of high-detail full body captures currently require elaborate capture and processing setups, as shown in, for example, Fairchild et al. (2017) and Orts-Escolano et al. (2016). The *Holoportation* system (Orts-Escolano et al., 2016) shows how full body representations could be streamed in both directions in symmetrical setups. Their system uses consumer grade AR and VR display technologies in combination with a new 3D capture system to capture full $360°$ 3D models of people and objects and visualize them at a remote location in real time. All of this is done with the goal of creating an experience that resembles physical presence as closely as possible. Users at both locations use a HoloLens augmented reality display to get an independent view of each other's environment. Although no physical manipulation of each other's environment is possible, the high resolution scans of the users allow for very precise gestural instructions that are properly aligned with their context.

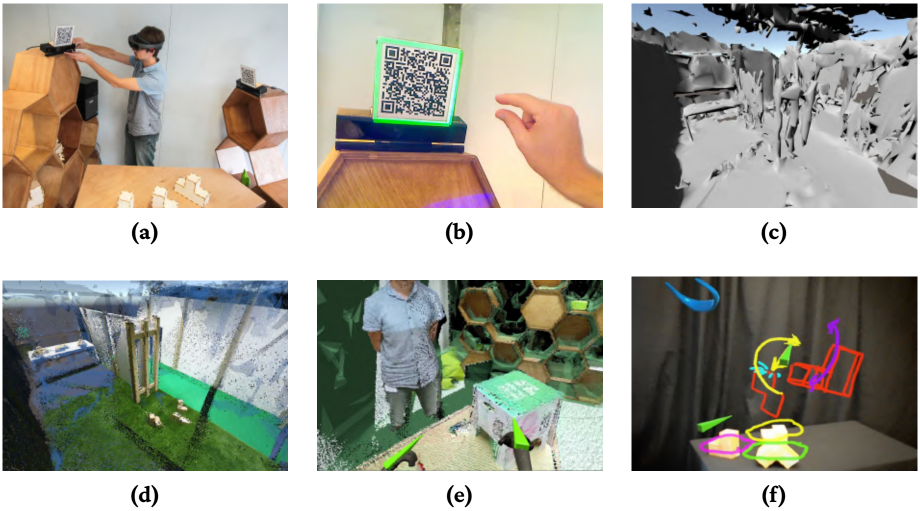


Figure 6.2. a) OSO installs depth camera. b) OSO calibrates depth camera. c) RO receives spatial mesh from HoloLens. d) RO receives point clouds, which are aligned with spatial mesh. e) RO points controllers at objects (green cones). f) OSO sees RO representation plus annotations through HoloLens.

Piumsomboon, Dey, et al.'s CoVAR system features hand tracking and sharing using the LEAP sensor and consistent spatial cues between remote and local environments – however does not provide any real-time streaming of a representation of a physical environment. In their exploratory study they investigate a number of enhancements to the communication and interaction between the dislocated parties, including sharing of gaze cues using eye-trackers in both headsets (Piumsomboon, Dey, et al., 2017).

6.2 IMPLEMENTATION

Our OpenIMPRESS toolkit features several components. Most of these were developed in C#, in the form of plug-ins for the Unity3D game engine. The components on the OSO side are all included – in the basic configuration of the toolkit – in a single Unity3D application running on the HoloLens. The components on the RO side similarly all run in a single Unity3D application connected to the VR HMD of the RO. Standalone components are the Depth Camera Streamer (written in C++) and the experimental Matchmaking Server of the network system (written in JavaScript/NodeJS).

6.2.1 ON-SITE ENVIRONMENT CAPTURE AND VISUALIZATION

Our module responsible for capturing the on-site location and re-visualizing it at the RO consists of several components (cf. Figure 6.2).

Depth Camera Streamer This is a standalone cross platform component written in C++ that captures depth and color image data from a depth sensing camera and transmits them over the network. Implementations are included for different brands of depth cameras and their respective libraries. At the time of writing, we support Microsoft Kinect V1 and V2 with the official software development kit (Kinect SDK[3]) and with the open source libraries from the OpenKinect Project[4], and the Intel RealSense cameras through *librealsense*[5]. For the Kinect V2 we also support streaming of data from the adaptive beam-forming microphone, skeleton detection and body indexing.

Depth Camera Tracker To properly align point clouds with each other and the rest of the scene requires information on sensor position and orientation (camera extrinsics) and on lens and sensor properties (camera intrinsics) of RGBD cameras in a shared coordinate system. Camera intrinsics for most RGBD sensors are known or can be obtained through the device's firmware. Camera extrinsics need to be (re-)estimated whenever the camera moves. Our solution was to estimate camera extrinsics with an on-the-fly procedure that can be performed at any time by an OSO at the on-site location. To this end, we attached a unique augmented reality marker to each RGBD camera, with a known offset to the camera's sensor. OSOs can *scan* these markers to estimate and share the RGBD camera's extrinsics with telepresent ROs (Figures Figure 6.2a and Figure 6.2b). These extrinsics are given in the internal coordinate system used by the OSO's HoloLens. To make a different area of the on-site location more visible to the RO, an OSO can simply move RGBD cameras there.

The typical error between point clouds and between a point cloud and the HoloLens coordinate system was around 3 cm at 2 meters from the camera. In the future, we could implement refinement procedures of camera extrinsics from multiple RGBD cameras (Kowalski, Naruniec, and Daniluk, 2015).

Spatial Mesh Streamer and Viewer The spatial mesh is the virtual geometrical representation that the HoloLens generates of its environment. It can be used as a rudimentary representation of the on-site location. To this end, the *Spatial Mesh*

[3]https://developer.microsoft.com/en-us/windows/kinect

[4]https://github.com/OpenKinect/

[5]https://github.com/IntelRealSense/librealsense

Streamer periodically serializes and streams it to the RO. The *Spatial Mesh Viewer* receives it and places a visualization in the 3D scene at the RO (Figure Figure 6.2c).

Point Cloud Viewer This plugin for the Unity3D game engine receives, processes and visualizes point-cloud data from multiple sources. An efficient implementation for drawing point clouds into virtual scenes was made to reduce CPU computations. Alignment of these point clouds to the basic spatial mesh (Figure Figure 6.2d) is done using the information from the Depth Camera Tracker.

6.2.2 REMOTE EMBODIMENT FOR RO HEAD AND HANDS

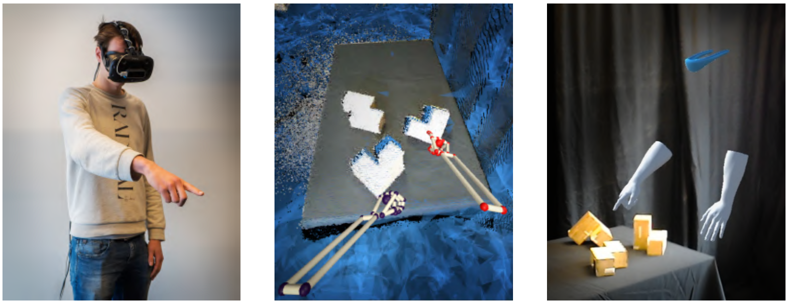
For communication, we want both OSO and RO to be able to see (a representation of) each other in a way that communicative acts such as gaze direction and gestures are possible. The OSO is captured as part of the remote environment and is visible to the RO through in the point cloud in real-time. However, we do not foresee a similar capture setup at the remote location in order to show the RO to the OSO.

Instead, we represent the RO through a virtual avatar embodiment consisting of a visualization of the RO's head and hand positions (see Figure 6.2 and Section 6.2.2). The head visualization always reflects the headset position, the hand positions reflect the position of the RO's controllers.

OpenIMPRESS can also be configured to allow ROs to use their bare hands with the *Hand Gesture System* (see Figure 6.3). An implementation of this for the *Leap Motion* sensor is available that can capture the position of the hand and individual fingers.

Leap Hand Sensor The Leap Motion sensor is a sensor that is used to track the hands of a user with high precision without the need of special trackers or gloves on the hands themselves. It makes use of infrared light and two infrared cameras to light up the hands and track them in 3D space. It is often placed facing upwards between the keyboard and the monitor in desktop set-up scenarios. The data the sensor provides are *frames* of poses, which can be sent to a *viewer*, so a visualization can be made.

Leap Hand Viewer The *Leap Hand Viewer* takes received hand frame data from the aforementioned component and applies it to 3D models for both local visualization for the RO (Section 6.2.2) and the remote visualization for the OSO (see Section 6.2.2).



(a) (b) (c)

Figure 6.3. (a) The RO using the gesture system to point. (b) The RO's hands as seen by the RO and (c) by the OSO.

6.2.3 RO VIEWPOINT TRANSFORM

By default, the RO can move around their virtual representation of the remote environment (i.e. the spatial mesh or point cloud) using their normal movements for as far as their own tracking area allows them. However, since the remote environment can potentially be larger than the tracking area of the RO's VR system, we have implemented another way of allowing the RO to manipulate their viewpoint. The *VRWorldDragger* component is activated when the RO presses and holds the trigger button on both the left- and right-hand controllers. Then, by changing the position of their hands, the RO can *drag* their position through space. That is, when the RO extends their arms, and then drags both hands towards them (keeping both trigger buttons pressed), they move forward through space. This works seamlessly along all axes of the world coordinate system. The *VRWorldDragger* also allows rotation around the *world-up* axis by dragging hands in opposite directions. Note that at all times, the other (head) motions of the RO in their tracking space still consistently affect their viewpoint. In other words, using the *VRWorldDragger* does not set a fixed viewpoint, but rather adds an offset to the *normal* tracking position and rotation. This offset is also taken into account for the RO's representation in the OSO's space, to keep the RO's own perspective and their remote representation consistent with each other.

Note that this system not only allows RO's to correct for their position once they are at the edge of their own tracking area. It also provides them with the ability to choose an optimal position in space given a task at hand – without needing to worry about physical constraints, and while maintaining whatever physical pose and position is comfortable to them in their local environment. For example, in a *car repair* scenario, the RO might find that they can help best with pointing out components to the OSO while virtually sitting in the car's engine room, which would be physically impossible in a co-located interaction.

6.2.4 ANNOTATION SYSTEM

Annotation functionality is added in the form of a line drawing tool. With this tool, the RO and OSO can draw lines in space, visible to each other (see Figure 6.2f). All "brush strokes" the RO makes are sent to the OSO in real time, where they are displayed as holograms in OSO's surroundings using the HoloLens. For the RO, this is done through the Vive controllers. For the OSO, lines are drawn by moving a bare hand in space while performing a *tap and hold* gesture, using the HoloLens' built-in hand tracking system. The current implementation creates lines at the tip of the controller or finger, respectively.

6.2.5 NETWORKING SYSTEM

Components communicate peer-to-peer using UDP messages. To this end, we provide a standardized interface, available both for native (C++) and Unity platforms (C#). Configuration of IP addresses and ports can be done manually. In addition, an experimental *matchmaking server* can be used to simplify configuration. Components using our networking interface can be configured to register themselves automatically at the matchmaking server. In a web-interface, components can be connected to each other; the matchmaking server then exchanges information between assigned components so they can establish a peer-to-peer connection.[6] Our current implementation allows for streaming data of multiple RGBD cameras at 30 fps and low-latency communication between the locations. However, when rendering more than two *depth camera streamers*, in order to maintain the HMD's frame rate of 90 fps and thus eliminate stutter, the refresh rate of the *Point Cloud Viewer* has to be limited.

6.2.6 VERBAL COMMUNICATION SYSTEM

We have also implemented Audio support to OpenIMPRESS, allowing the RO and OSO to talk to each other. A *VoIP Manager*[7] component was developed to this end. This component runs at both the OSO and the RO. It accesses the system's microphone using Unity's built-in Microphone class[8]. Because this class is supported on both the desktop version of Unity and the HoloLens, the same code can be used on both devices.

[6] Our matchmaking server further allows clients behind routers to connect to each other without configuring port-forwarding, using UDP hole punching (Ford, Srisuresh, and Kegel, 2005).

[7] VoIP: Voice over IP

[8] https://docs.unity3d.com/ScriptReference/Microphone.html

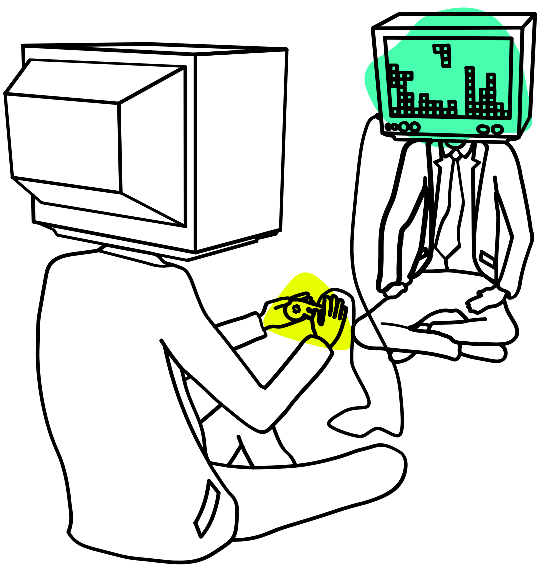
The latest audio samples are fetched from the microphone and are sent to the *VoIP Manager* of the other system. Both the sample frequency and number of channels that the samples are recorded with are sent as well. This meta information ensures that the receiving end can play the samples back again properly.

6.3 CONCLUSION

We have presented our OpenIMPRESS toolkit which can be used to build and research mixed reality telepresence systems for remote communication and collaboration. The modular approach allows us to configure the system in different ways. One way is configuring the ways in which users, in particular the Remote Operator, engage with representation of the environment, as well as how the ROis represented to the user - this is something we look at in the next chapter in particular. The second way in which OpenIMPRESS can be adapted to suit a particular task is by extending the annotation system with custom widgets meaningful for the task at hand, something that is made use of in the second user study, discussed in Chapter 8. The next step for us was to evaluate OpenIMPRESS. In particular we were interested in what aspects of the design contribute to how well it works. But how do we evaluate a remote collaboration system? It must allow disconnected users to solve a given problem correctly and within reasonable time, so metrics such as task completion time or number of errors are indicative of *how well* a system works. However, these may not be indicative of *why* the system works (or doesn't work). Can users communicate clearly and efficiently? Are they confident that the other understands them, or can follow their direction? In the following chapter we discuss a user study conducted to answer these questions for the system that we presented here.

7

EVALUATING DESIGN OPTIONS: A PUZZLE TASK



In this chapter[1] we will discuss our first study on a mixed reality remote collaboration system built with OpenIMPRESS. The main research question was whether the OpenIMPRESS system works as expected as a telepresence system that allows social interaction and collaboration over distance. In particular, we were interested in how much the three design aspects identified and implemented in the previous chapter contribute to performance and user experience. To this end, we built a mixed reality

[1]Based on our paper J. Kolkmeier, E. Harmsen, et al. (2018). "With a little help from a holographic friend: The OpenIMPRESS mixed reality telepresence toolkit for remote collaboration systems". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11, and the supervised thesis E. Harmsen (2018). "OpenIMPRESS: an open immersive telepresence system". Master's thesis. University of Twente

puzzle task that must be completed by two users that are physically dislocated, and where certain knowledge to complete the task is required by one user that only the other can provide. We tested this same task with different configurations of the telepresence system, each manipulating one of the design aspects implemented in OpenIMPRESS, in order to identify their contribution to how well the system performs. As we already raised in the conclusion of the previous chapter, evaluating how fast users will be able to complete the task or how easy they find it to use the system are two obvious facets of evaluation, but they only paint a partial picture. Our main research interest in this thesis is in novel, unique, and perhaps contrived interaction patterns and experiences that are only possible with mixed reality technology - and thus may have no equivalent in co-located situations. So besides the aforementioned functional qualities of the experience, we are also interested in more fundamental aspects. What is it like to give someone directions while being virtually in their head? What is it like to receive directions from someone whose body representation is not bound to any physical limitations, yet is situated (virtually) in the same physical space as you, having full (visual) access of your environment? These different kinds of sharing the same space and being with each other may affect not just the functional aspects of the interaction, but the relationship between users. How do they perceive each other - their social roles, attitudes or personalities - due to the different style of interaction that a novel system affords when it mediates the verbal and nonverbal social signals between users in ways that are unlike what users are used to from co-located interactions?

In the following section we discuss *social presence*, a metric that we employ in this study and in the remainder of this part of the thesis, and that should help us towards answering these questions.

7.1 SOCIAL PRESENCE

Social Presence theories attempt to describe and operationalize these phenomena in technology mediated social interactions. There is a body of work from Biocca and colleagues from the early 2000s that extensively reviews theories on social presence and develops validated measurement instruments. This definition they provide for social presence highlights its different facets well:

"*Social presence in a mutual interaction with a perceived entity refers to the degree of initial awareness, allocated attention, the capacity for both content and affective comprehension, and the capacity for both affective and behavioral interdependence with said entity.*" (Harms and Biocca, 2004)

First, social presence is defined to exist in a *mutual* interaction with a perceived entity – although the perceived social presence does not need to be mutual. So interactions with artificial entities are not excluded from this. Second, social presence is not thought of as a binary construct - one can feel a *degree* of social presence of an entity in technology mediated social interaction. Next, social presence is not just the degree to which one is aware of the entity's presence, but also includes the perception of the extent to which the other is able to comprehend content of what is being communicated, as well as the perceived capacity of the entity to be able to enter some affective and behavioral interdependency. This highlights a more psychological aspect to social presence theory - it requires some sense of the other having some intelligence and the ability to enter a relationship.

One of the most popular instruments to measure social presence is the *Networked Minds* social presence questionnaire by Harms and Biocca. It measures six sub-dimensions of social presence. In the following paragraphs, we provide brief definition as well as examples of the associated items for each of the six subscales, as specified in Harms and Biocca (2004).

Co-Presence *"Co-presence is the degree to which the observer believes he/she is not alone and secluded, their level of peripheral or focal awareness of the other, and their sense of the degree to which the other is peripherally or focally aware of them."* Some items measuring the Co-Presence construct are: "I noticed my partner", "My partner noticed me", "My partner caught my attention" and "My presence was obvious to my partner".

Attentional Allocation *"Attentional allocation addresses the amount of attention the user allocates to and receives from an interactant."* The instrument uses items such as "My partner was easily distracted from me when other things were going on" and "I remained focused on my partner throughout our interaction" to measure this construct.

Perceived Message Understanding *"Perceived message understanding is the ability of the user to understand the message being received from the interactant as well as their perception of the interactant's level of message understanding."* The instrument uses items such as "My partner found it easy to understand me" and "Understanding my partner was difficult" to measure this construct.

Perceived Affective Understanding *"Perceived affective understanding is the user's ability to understand an interactant's emotional and attitudinal states as well as their perception of the interactant's ability to understand the user's emotional and attitudinal states."* The instrument uses items such as "My partner could describe my feelings accurately" and "I could tell how my partner felt." to measure this construct.

Perceived Affective Interdependence *"Perceived affective interdependence is the extent to which the user's emotional and attitudinal state affects and is affected by the emotional and attitudinal states of the interactant."* The instrument uses items such as "My partner was sometimes influenced by my moods" and "My attitudes influenced how my partner felt" to measure this construct.

Perceived Behavioural Interdependence *"Perceived behavioral interdependence is the extent to which a user's behavior affects and is affected by the interactant's behavior."* The instrument uses items such as "The behavior of my partner was often in direct response to my behavior." and "I reciprocated my partner's actions" to measure this construct.

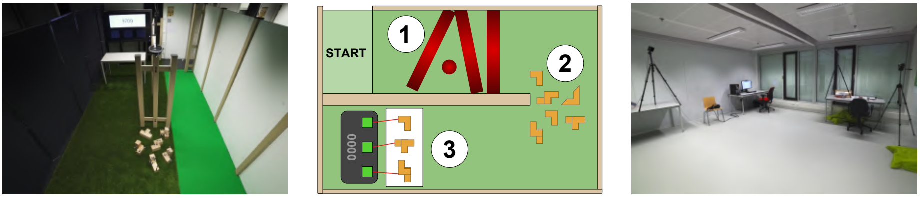
Third order measures For all social presence constructs in this instrument, questions are asked both in terms of an interactants perception of their own sense of social presence, and how they think their interaction partner perceived it. This *symmetry*, although not explicitly discussed in their validation study (Harms and Biocca, 2004), reflects the concept that there is some additional higher order construct of social presence about how symmetrical the observers perceive their relationship and the extent to which interactants correlate between each other on their perception of this symmetry (Biocca and Harms, 2002). We will discuss this further in Chapter 9.

7.2 STUDY DESIGN

In the experiment, pairs of participants completed an escape-room like task that required collaboration using our system. The details of the task are outlined in Section 7.2.1. The study was split into three sub-experiments: a *View independence*, an *Immersiveness of the Display* and a *Virtual Embodiment* experiment. There was a single within subject variable "configuration" at one of two levels: *baseline* and *experiment*. This means that each pair of participants completed the task twice: once in the *baseline* configuration, and once in one of the three experimental configurations, depending on the sub-study they were assigned to. The order of "configuration" was counterbalanced. In the experimental conditions, we changed the system configuration in one of the three related aspects (see Section 7.3 for details).

7.2.1 TASK

An escape room like environment was built in which the participants were tasked with retrieving a four-digit code by solving three puzzles. Escape rooms are designed to be a collaborative experience: "Escape rooms require teamwork, communication, and delegation" (Nicholson, 2016). As the tasks are presented as puzzles in a playful context, they can be designed to be more abstract in order to disconnect them from any real-world knowledge and remove any bias that this could give.



(a) (b) (c)

Figure 7.1. a) The escape room where the OSO is located, b) A map of the escape room, with 1) Radiation beam avoidance task; 2) Block collection task; 3) Block alignment task, and c) The room nearby where the RO is located.

One participant, the OSO, was located in the escape room (Figure Figure 7.1a) and the other participant, the RO, was located in a room nearby (Figure Figure 7.1c). From this nearby room, the RO was remotely present in the escape room to help the OSO complete the tasks.

The escape room consisted of the following three tasks that needed to be completed in order (see map in Figure Figure 7.1b).

1. **Navigation Task - Radiation beam avoidance** The RO was tasked to lead the OSO through a maze of virtual radiation beams that only the RO could see, while making sure the OSO's head never touched one of the beams. This puzzle can be thought of as a laser maze, but with the exception that only a remote partner can see the lasers. Also, to increase the visibility and difficulty, beams with a thickness of around 25 cm were used instead of thin laser rays.

2. **Recognition Task - Block shape collection** Six blocks with different shapes were on the ground and the OSO needed to collect three specific ones to use in the next step. Only the RO was shown a virtual display depicting which blocks needed to be collected; they had to communicate this to the OSO.

3. **Manipulation Task - Block alignment** Once placed on top of the table, the three collected blocks started to emit a virtual laser beam that only the RO could see. The blocks needed to be positioned in such a way that the lasers pointed to markers behind the table. To do this, the RO had to communicate to the OSO how to manipulate the blocks. Once each of the markers were lit by a laser beam, a part of the code was displayed on the screen.

In the instructions to the participants, no specific priority on speed or accuracy was given: participants were instructed to complete the experiment in the way that they felt the most comfortable with.

7.2.2 MEASURES

We had a single *performance* measure, which was the time it took to complete the entire task. There are three subjective measures related to *user experience*. First a *usability* measure to indicate the effectiveness, efficiency and satisfaction with which a system built with our toolkit can be used. We used the 10-item *System Usability Scale* (SUS, Brooke, 1996). Second we measured *spatial presence*, the feeling of being in a certain environment. Here, we used the 14-item *igroup presence questionnaire* (IPQ, Schubert, Friedmann, and Regenbrecht, 2001), which provides a subscale specifically on spatial presence. Finally, we administered the Networked Minds instrument (Harms and Biocca, 2004) discussed in Section 7.1 as a measure of social presence. These instruments were included in a questionnaire that was administered after each run with a given configuration. In addition, we asked participants' demographics, including experience with virtual reality systems and video gaming.

7.3 SYSTEM CONFIGURATIONS

Our baseline system was characterized by including all the core design aspects: an *independent view* for the RO, the RO using *immersive VR* technology to move in and engage with the on-site location, and a *virtual embodiment* of the RO rendered in augmented reality to the OSO. In this study, we investigated how performance, usability and spatial and social presence are affected when our system is configured with one of these aspects completely removed or implemented in a reduced way. That is, with a *dependent view*, with a *desktop set-up* for the RO, and with *leaving out the embodiment* of the RO that is normally shown to the OSO. In the following, we describe the baseline and experimental configurations in more detail.

Baseline: all features default In the *baseline* configuration, most features from the system described in Chapter 6 were present in their default implementation. Only the *Annotation System* was disabled, limiting the RO feedback to hand gestures and voice. The RO wore the VR headset, was free to walk around or use the controllers to shift their viewpoint (as described in Section 6.2.3), and the RO's head position and hands were visualized as holograms at the OSO.

Dependent view The *dependent view* configuration removed the ability for the RO to independently walk around, by fixing the RO's viewpoint to the viewpoint of the OSO, similar to the *JackIn Head* setup (Kasahara and Rekimoto, 2015) and to the configurations used in the studies of Gao et al. (2016) and Amores, Benavides, and Maes (2015). This introduced a mismatch between the RO's movements and the movements that were perceived in VR, as the latter were now copied from the movements of the OSO instead of the RO. This mismatch between perceived and actual movement is known to introduce nausea (Groen and Bos, 2008). We addressed this by asking participants to keep head movements to a minimum (as they would not affect the view of the RO anyway), and secondly by filtering movements of the OSO with a low-pass filter before copying them to the RO.

Desktop set-up The *desktop set-up* replaced the immersive VR set-up with a conventional 2D monitor, keyboard and mouse, similar to the set-up used by Tait and Billinghurst (2015). Instead of navigating by physically walking, the RO had to use the same mouse and keyboard controls that are often found in first-person shooter video-games; the arrow keys were used to translate the view to a different position and the mouse was used to rotate the view.

ROs could still use their hands for pointing or gesturing; the leap motion hand sensor that is otherwise attached to the front of the VR headset was now positioned at the keyboard, so the hands were detected when the RO held them up in front of the screen.

No embodiment: remove the RO's embodiment The *no embodiment* configuration disabled the holographic representation of the RO (visor and hands) for the OSO as well as for the RO, thus removing any visual representation of the RO from both views.

7.4 HYPOTHESES

We hypothesized that for dependent view, desktop set-up, and removed embodiment configurations the performance would be negatively affected compared to the baseline. ROs with an *independent view* were expected to spend less effort on adjusting their view which would decrease the total time spent on the task. With a *desktop set-up* configuration, the lack of certain spatial cues would make it more difficult for the RO to support the OSO, especially in tasks that required the ability to recognize three-dimensional features of objects, or required the RO to have an overview of a relatively large area. We expected the *virtual embodiment* to support the RO when giving explanations by increasing the number of channels that could be used for communication. This would decrease the amount of time required for communication, thus increasing the effectiveness.

For similar reasons, we expected usability to be negatively affected in the respective conditions. Specifically for the RO when using the dependent view and using a desktop set-up, and for the OSO when the embodiment of the RO was removed. For the spatial presence measure, we expected that both a dependent view and using a desktop set-up would reduce the RO's feeling of being at the on-site location.

For the co-presence measure, we mainly looked at the RO, expecting decreased scores with the dependent view and desktop set-up configurations as well as when the embodiment was disabled. For the OSO, we expected lower scores with the RO's embodiment disabled.

For the perceived message understanding, we looked both at the RO and OSO, hypothesizing that this measure would be negatively affected in all experimental conditions.

7.5 PROCEDURE

Before the experiment started, the system was set up to use the configuration following a counterbalancing table for the experimental conditions. Once both participants arrived, they were randomly assigned the RO and OSO roles. After a basic introduction to the procedure, they were asked to read and fill in the consent form. The experimenter then explained how the system worked to the participants, first RO then OSO, each in their respective rooms. The connection was tested and elements of the system were demonstrated – depending on the condition. After the participants finished the task, they were helped to take off their HMDs and asked to fill in the post-experiment questionnaire. Participants did not switch roles between the two runs. The procedure was then repeated (minus the introduction) with the other system configuration.

The system froze once on six different experiment runs due to technical issues. In these cases, the experiment was either restarted, or continued from the same point in the experiment at which the system froze, depending on how far the run had already progressed.

7.5.1 DATA PROCESSING

Task completion time was computed through an automatic parsing of the experiment log files. For the six cases where an experiment session was interrupted by a crash of the software, completion time was manually determined using the video recordings.

For the user-experience measures, we looked at the difference between the paired scores from the two runs by the same pair of participants. Score differences were considered outliers if they fell outside 1.5 times the interquartile range above the upper and lower quartiles and were removed from the analysis.

7.5.2 PARTICIPANTS

For each of the three sub-experiments, 10 pairs of participants were recruited using messages on social media and convenience sampling. Participants were asked to sign up, preferably in pairs, using an online form. Participants who signed up alone were paired with another participant who signed up alone. Of the sixty participants, 29 were female. Mean participant age was 25.9 years (SD=9.0, MAX=57). Eight pairs were male-male, seven female-female and fifteen were male-female.

7.6 RESULTS

In Table 7.1 we present the descriptives of performance, usability and social presence measures for the RO and OSO. For those differences that we formulated hypotheses for, the test statistics are reported. These results are discussed in Section 7.7.

Removing view independence by locking the RO's view to the OSO's did significantly increase the time it took participants to complete the task and decrease the usability, spatial presence, and the feeling of co-presence for the RO. The perceived message understanding for either the RO and OSO did not significantly decrease.

Changing the immersive VR set-up with a desktop screen, mouse and keyboard set-up (but keeping the hand gesture system) significantly increased the time it took participants to complete the task, and decreased the reported usability and spatial presence for the RO, and the perceived message understanding for both the OSO and RO as expected. It did not decrease how co-present the ROs reported they felt with the OSO.

Removing the virtual embodiment of the RO did not significantly increase the time it took participants to complete the task, the reported usability for the OSO or the perceived message understanding. For perceived co-presence, only ROs reported a lower sense of co-presence.

7.7 DISCUSSION

Three design features of our mixed reality telepresence system were tested to understand how they affected the performance and experience of the users in a collabo-

Table 7.1. Per sub-study on the three design aspects, we report the descriptives on the two conditions. Reported *n* smaller than 10 indicates outliers removed according to the procedure described in Section 7.5. One-tailed paired-sample t-test statistics are reported. In one case, the Shapiro-Wilk test revealed that the assumption of normality was violated. Here we report statistics of a paired-samples Wilcoxon test.

	View Independence		Immersiveness of the Display		Virtual Embodiment	
Measure	**Baseline (Independent View)**	**Dependent View**	**Baseline (immersive VR set-up)**	**Desktop set-up**	**Baseline (Embodiment)**	**No Embodiment**
Task Duration	M=301.040, SD=74.400, n=10	M=348.644, SD=66.133, n=10	M=298.029, SD=81.956, n=9	M=358.348, SD=119.836, n=9	M=331.437, SD=93.044, n=10	M=301.716, SD=103.732, n=10
	$t(9) = -2.614$, $\mathbf{p = 0.014}$		$t(8) = -2.105$, $\mathbf{p = 0.034}$		$t(9) = 1.229$, $p = 0.875$	
Usability (RO, $\alpha = 0.855$)	M=77.778, SD=10.266, n=9	M=66.944, SD=14.185, n=9	M=77.250, SD=15.831, n=10	M=59.000, SD=17.288, n=10	M=70.500, SD=11.714, n=10	M=73.250, SD=13.126, n=10
	$t(8) = 3.506$, $\mathbf{p = 0.004}$		$t(9) = 3.222$, $\mathbf{p = 0.005}$		–	
Usability (OSO, $\alpha = 0.775$)	M=77.000, SD=8.803, n=10	M=74.500, SD=10.055, n=10	M=77.250, SD=7.017, n=10	M=73.500, SD=15.193, n=10	M=79.444, SD=8.640, n=9	M=77.500, SD=7.289, n=9
	–		–		$t(8) = 1.306$, $p = 0.114$	
Spatial Presence (RO, $\alpha = 0.852$)	M=6.233, SD=0.610, n=10	M=5.150, SD=1.101, n=10	M=5.633, SD=1.165, n=10	M=4.017, SD=1.738, n=10	M=5.963, SD=0.558, n=9	M=5.444, SD=0.612, n=9
	$t(9) = 4.079$, $\mathbf{p = 0.001}$		$t(9) = 2.735$, $\mathbf{p = 0.012}$		–	
Co-Presence (RO, $\alpha = 0.819$)	M=4.407, SD=0.501, n=9	M=4.111, SD=0.577, n=9	M=4.367, SD=0.520, n=10	M=4.150, SD=0.506, n=10	M=4.350, SD=0.412, n=10	M=3.883, SD=0.798, n=10
	$t(8) = 3.411$, $\mathbf{p = 0.005}$		$t(9) = 1.361$, $p = 0.103$		$V = 28$, $\mathbf{p = 0.022}$	
Co-Presence (OSO, $\alpha = 0.902$)	M=4.483, SD=0.512, n=10	M=4.217, SD=0.369, n=10	M=4.283, SD=0.681, n=10	M=4.100, SD=0.851, n=10	M=4.167, SD=0.624, n=10	M=4.050, SD=0.478, n=10
	–		–		$t(9) = 0.606$, $p = 0.280$	
Perc. Msg. Understanding (RO, $\alpha = 0.86$)	M=4.100, SD=0.425, n=10	M=4.050, SD=0.324, n=10	M=3.833, SD=0.692, n=9	M=3.278, SD=0.607, n=9	M=4.067, SD=0.459, n=10	M=4.200, SD=0.414, n=10
	$t(9) = 0.340$, $p = 0.371$		$t(8) = 3.780$, $\mathbf{p = 0.003}$		$t(9) = -1.350$, $p = 0.895$	
Perc. Msg. Understanding (OSO, $\alpha = 0.785$)	M=4.000, SD=0.493, n=9	M=3.889, SD=0.464, n=9	M=4.250, SD=0.517, n=10	M=3.733, SD=0.903, n=10	M=4.017, SD=0.319, n=10	M=4.133, SD=0.205, n=10
	$t(8) = 1.512$, $p = 0.085$		$t(9) = 1.935$, $\mathbf{p = 0.042}$		$t(9) = -1.000$, $p = 0.828$	

rative context. Each feature was tested in a study in which the participants had to collaborate once in the system configured with all features implemented and once configured with the feature in question left out.

7.7.1 VIEW INDEPENDENCE

Our findings on how removing view independence affects the experience are in line with those of Tait and Billinghurst (2015), who showed the positive effects of view independence on the quality of the collaboration using a desktop interface. ROs could solve the task faster, and found the system more usable when their view was independent of the view of the OSO. ROs did report feeling less present when view independence was removed. Interestingly, this was not directly reflected in all social presence constructs. While ROs did report feeling more co-present with the OSO, they did not report significantly differently on how (well) they perceived messages from the other, and how they thought their own messages were perceived by the others.

7.7.2 IMMERSIVE VR SET-UP

Our results support most hypotheses on how using an immersive VR setup versus a desktop screen, keyboard and mouse setup for the RO would affect the performance, usability and (social) presence constructs of users. First and foremost, ROs indeed reported significantly lower spatial presence scores when not using the immersive VR setup.

From observations during the sessions we could see that many ROs had trouble navigating through the virtual representation of the on-site location in the condition using the 2D desktop interface. Those participants often tried to give instructions to the OSO from a perspective that made it difficult to see important parts of the task, for example by parts being obscured by other parts or just being far away. We observed for example some ROs trying to give instructions about how to navigate through the maze while already being positioned at the end of the maze, with the OSO still standing at the beginning. These observations are also supported by the perceived message understanding measure, which was reported to be significantly lower by both the RO and OSO when using the desktop set-up.

Correlating the performance of participants using the desktop interface with their responses in the video game experience item in the post-experiment questionnaire provides some explanation. Those with more game experience had faster task-completion times (M=283s, SD=54s, n=5) than those that played rarely to never (M=433s, SD=130s, n=5). This shows that, at least for the task-completion measure, the performance was affected not only by the diminishing of the immersiveness of the experience, but also by how comfortable users were with the other input modalities.

Observations of the ROs in this condition also seem to suggest that moving around a lot improves performance. ROs that were relatively quick to complete the tasks often seemed to move around a lot, especially in the maze task. We hypothesize this to be a technique that is used to compensate for the lack of stereoscopic depth perception when using a 2D monitor; by moving around a part of the scene, the motion parallax can create an impression of the relative depth when other cues are absent (Rogers and Graham, 1979). The lack of stereoscopic depth perception can also be a reason for the hand gesture system to remain unused relatively often. It looked as if people had trouble understanding exactly how far away their hands were from their position in the virtual representation.

7.7.3 VIRTUAL EMBODIMENT

Removing the virtual embodiment of the RO did not significantly increase the time it took participants to complete the task or decrease the reported usability for the OSO or the perceived message understanding. For the perceived co-presence measure, only ROs reported a lower sense of co-presence.

These results are surprising, since we anticipated the visible embodiment of to be an advantage. One observation made during the experiments was that the RO often navigated the OSO through the maze by only using their voice while the ability to use hand gestures was available. Some participants did not seem to realize they could use them until the last task, while others tried to use them for navigation but did not persist and fell back to verbal commands after it did not work for them as expected. Perhaps this resulted in a ceiling effect especially for social presence.

A technical reason could be the HoloLens' limited field of view, leading to the OSO not being able to see the RO's hands, especially when they were held close to the OSO's head. The RO would sometimes ask whether the OSO could see their hands and fall back to voice if not. This could be addressed by giving the RO more information on what the OSO can see – similar to how viewpoints are shared between hosts and guests in G. Lee et al. (2017), the shared virtual frustums in CoVAR (Piumsomboon, Dey, et al., 2017) or enhanced referencing support (Oda and Feiner, 2012). Another approach to solving this issue is the use of scaled down avatars, such as Mini-Me (Piumsomboon, Lee, et al., 2018).

In fact, Piumsomboon, Lee, et al.'s recent work addresses this issue specifically. In their approach, when the RO is not in view of the OSO, a smaller *Mini-Me* representation of the RO is rendered in the OSO's view instead, keeping gaze and pointing cues of the RO consistent with the local space of the OSO (Piumsomboon, Lee, et al., 2018).

Curiously, while the OSOs did not report a difference in their feeling of co-presence with the RO when the RO's embodiment was removed, the ROs did report feeling significantly less co-present with the OSO in that case. One explanation could be that, although the OSOs did not report any perceived difference, they still behaved differently towards the RO when a virtual embodiment was available, which made the RO feel more co-present with the OSO.

7.7.4 LIMITATIONS

One limitation of this work is the difficulty in building three implementations that isolated the different design aspects fairly. For example, the desktop condition not only removed the immersive display, but also came with different input modalities, as discussed above.

Some sub-tasks emphasized guidance through space, while another focusing on object identification. While total time still gives a valid measure of performance, it also averaged out potential interaction effects. It would have been interesting to see if certain design aspects interact with the type of sub-task. We were unable to perform this analysis as there was not a clear division between the second and third task – if the wrong puzzle piece was selected for the third part, users had to return to the second part to recover.

7.8 CONCLUSIONS

In the previous chapter, we presented our toolkit for mixed reality remote collaboration systems. OpenIMPRESS works with off-the-shelf hardware, and all software is publicly available under an open-source license. Our evaluation presented in this chapter served to evaluate some choices made when designing the user experience for a default configuration.

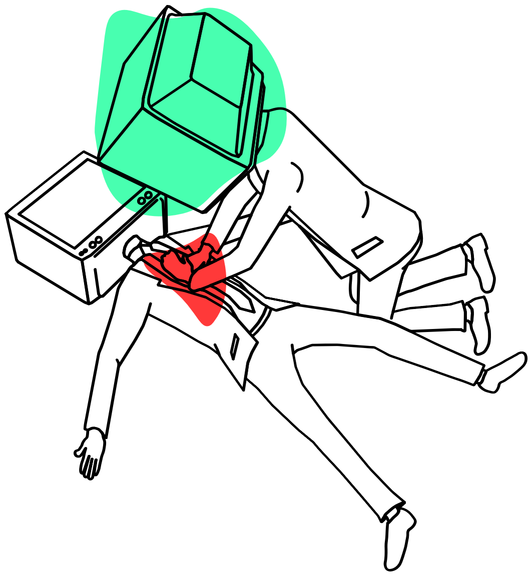
The results show how using an immersive VR display as opposed to a conventional desktop setup affects the experience for users in various ways, both in terms of performance, usability and (social) presence related metrics when used in a collaborative task with our system. Similarly, making use of free navigation in the virtual representation of a remote scene did increase presence, usability and performance significantly.

This confirms to some extent the design choices made for our toolkit, which were in great part made based on the existing literature on mixed reality remote collaboration systems, and indicate some comparative insights between them, as also our experimental conditions were informed by others' work.

However, we could not support all the hypotheses on the social presence measures. In fact, especially the results on the Virtual Embodiment sub-experiment are surprising, and leave some answers that we will try to address in more detail in Chapter 9.

8

REMOTE REANIMATION USING OPENIMPRESS



Imagine a world where wearable augmented reality glasses or contact lenses are part of most peoples daily lives, just like smartphones are today. These devices can render a virtual overlay over your surroundings. The first thing you might be imagining now is a dystopian world, dominated by cluttered, animated messaging and advertisements, competing for peoples attention. Perhaps you're thinking of the episode *Nosedive*[1] from the popular *Black Mirror* series, depicting a future society where eye implants are used to render an augmented reality social rating system. In this chapter, we take a more optimistic view of the applications that may be possible in a world where wearable augmented reality displays are a pervasive reality.

[1]"Black Mirror" Nosedive (TV Episode 2016) https://www.imdb.com/title/tt5497778/

The user study discussed in this chapter[2], evaluates a system that would allow wearers of future AR technology to call in a virtually embodied dispatcher for support in emergency situations, specifically when responding to a cardiac arrest. This is, again, an asymmetric scenario, where a local novice is supported by a remote expert in order to solve a problem in the novice's physical environment.

While the user study in the previous chapter focussed on different implementations of the design as the independent variable, with an artificially constructed scenario as a task, the study presented in this chapter sets out to investigate whether a real-world problem can benefit from a technology intervention. Accordingly, the baseline condition in this study reflects the state of the art for emergency dispatch systems: a telephone dispatcher. There are two main questions we want to address in this study. First, does the "performance" of the first responder improve? Do users learn to administer CPR quicker, and with fewer errors? Second, since emergency situations are stressful, can the presence of a virtual dispatcher improve the subjective experience of the task?

8.1 CARDIAC ARREST & FIRST RESPONSE

When imagining an emergency situation with an unconscious victim, one might immediately think of the potential need for cardiopulmonary resuscitation (CPR), meaning the massage of the heart muscle through chest compression and ventilation of the lungs. In cases of cardiac arrests, starting CPR immediately can double to quadruple survival chances (Holmberg, Holmberg, Herlitz, Gårdelöv, et al., 1998; Valenzuela et al., 1997; Waalewijn, Tijssen, and Koster, 2001). Besides factors such as age of the victim, location of the emergency, and the time of arrival of an ambulance, the quality and in particular the immediacy of CPR performed by the first responder can contribute to the survival chances (Abella et al., 2005; Holmberg, Holmberg, and Herlitz, 2001; Stiell et al., 2012). If determined necessary, emergency dispatchers will advise the first responder to start CPR immediately. In cases where the first responder is unsure of how to perform CPR, the dispatcher will instruct them as quickly as possible. However, the ability of the dispatcher to understand the situation and give appropriate support to this end is limited to what can be communicated over the phone.

[2] Based on our paper: J. Kolkmeier, E. Harmsen, et al. (2018). "With a little help from a holographic friend: The OpenIMPRESS mixed reality telepresence toolkit for remote collaboration systems". In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11; and on the supervised thesis: S. Giesselink (2018). "Teach me CPR now! Augmented learning for one-shot teaching of cardiopulmonary resuscitation in out-of-hospital cardiac arrest". Master's thesis. University of Twente

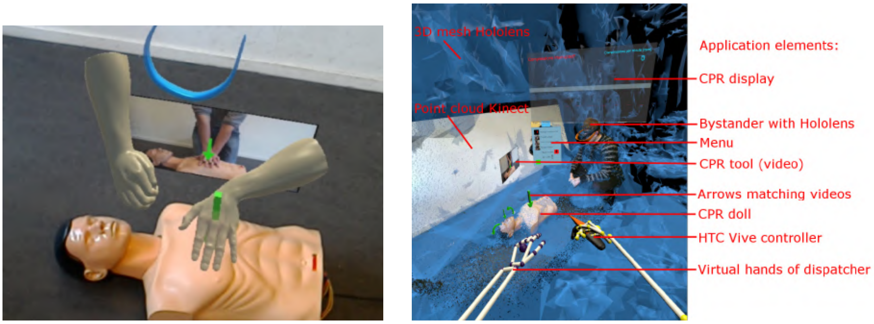
In the study described in this chapter, we investigate an envisioned scenario where wearable augmented reality glasses are pervasive technology. A first responder in an emergency situation could allow the dispatcher access to their AR glasses. The dispatcher gains access to the environment, as mapped by the AR glasses, and themselves becomes virtually present at the scene. We hypothesize that such a system, given it being technologically reliable, could be used to improve the quality of CPR significantly. One potential advantage is the ability for the dispatcher to give richer instructions on the process of CPR through the multimodal interface and having visual access to the on-site location. Another potential advantage is the richer social interaction between the dispatcher and first responder: being embodied at the remote location, dispatchers may be able to provide better emotional support, as the system mediates an increased feeling of the remote expert's social presence to the first responders in such a stressful situation.

8.2 THE TELEPRESENCE CPR SUPPORT SYSTEM

Using the OpenIMPRESS toolkit, we built the envisaged remote support system that allows dispatchers to be virtually present at an emergency site where a bystander needs to reanimate a cardiac arrest victim. We call it *ARC*, for Augmented Reality CPR. It was built with the goal in mind to evaluate it in a lab-study setting, where CPR is performed on a sensor equipped CPR doll (see Figure 8.1a).

ARC has a similar topology to the apparatus discussed in the study in Chapter 7. Here, the dispatcher is taking the role of the Remote Operator (RO), being fully immersed in a virtual environment that represents the environment of the bystander. The dispatcher has access to the annotation system discussed in Section 6.2.4, with additional annotation content added that can be used as visual aids specific for CPR. One aid is an *audiovisual metronome* that moves up and down and makes a sound at the correct compression rate. The other aids are *video animations* that show the different tasks required to perform CPR (see Figure 8.2). We further envisioned that in the future, through video analysis of the scene, or with the help of the accelerometer data from the AR display, extracting the compression frequency from the movements should be possible, and would be helpful information for the dispatcher. To simulate this in our prototype, we visualize the frequency and depth of the compression using the doll's force sensor and place it near the virtual representation of the bystander's head in the dispatcher's VR view. The dispatcher can use this information to gain quick insight into the performance of the bystander and determine suitable feedback.

The bystander's system is the same that was used by the On-Site Operator (OSO) in the previous study. Thus, a bystander wears a Microsoft HoloLens to render the telepresent dispatcher into the emergency scene. Any visual aids that the dispatcher



(a) (b)

Figure 8.1. Left: View of the bystander (OSO), showing the hands of the telepresent dispatcher (RO). Right: View of the RO showing the 3D virtual environment and application elements. The 3D spatial map of the HoloLens visible as the low resolution blue mesh. The higher resolution point cloud from the Kinect shows the bystander, CPR doll, floor and wall.

places into their virtual environment are also visible in augmented reality at the corresponding position at the emergency location. Voice communication is enabled using the verbal communication system (see Section 6.2.6).

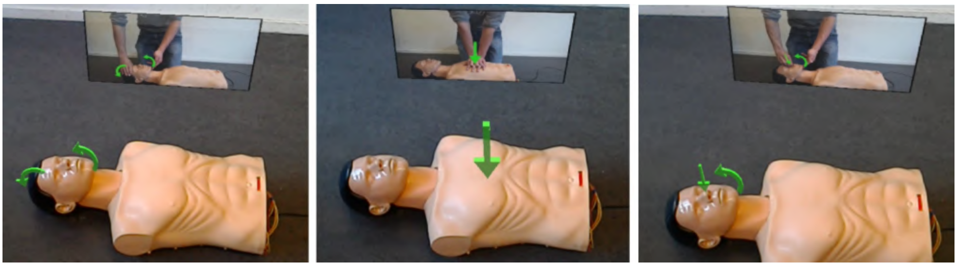


Figure 8.2. Augmented instruction videos with arrows: left; chin lift, middle; chest compressions and right; rescue breaths. As visible to the bystander using a HoloLens.

8.3 RESEARCH DESIGN

We conducted an experiment where participants were asked to perform CPR under the instruction of a remote emergency dispatcher. The experiment featured a between subject design with two conditions - the control condition *Phone-CPR*, where participants communicated with the dispatcher through an audio-only connection, and the *AR-CPR* experimental condition, where the dispatcher was virtually present using the ARC system described above to support the participant in the same way. The dispatcher role was always performed by the same experimenter, adhering to a strict protocol described in Section 8.3.3, based on a training the experimenter received from a professional dispatcher. The study did not take place in the wild, but in a lab setting with the aforementioned sensor-equipped CPR doll.

8.3.1 MEASURES

We collected data both on the quality of the CPR administered using the sensors in the doll, and subjective measures regarding the experience during the task and the perceived social relationship with the dispatcher. For the CPR quality, deviations from the optimal frequency and optimal compression depth were computed using the force sensor. The main performance-related measures were derived from the force sensor in the CPR doll, namely the frequency of compressions and compression depth. These measures were applied over multiple segments of the experiment, potentially revealing different rates of improvement.

Our primary focus in the context of this thesis however is in the subjective measures. Here, we used a similar set of measures to the one used in Chapter 7, including the Networked Minds instrument (Harms and Biocca, 2004). The usability instrument was deemed less meaningful for the *Phone-CPR* condition, as it does not include any interaction with technology besides setting the phone to use the loudspeaker. Instead, we elected to administer the perceived task load using the Raw-TLX instrument (Hart, 2006), which we expected to better capture potential effects of the increased physical and mental demands of the task due to the increase in equipment weight and added visual modalities in the experimental condition.

Not discussed in the present work are annotations that were done based on the video recordings regarding, for example, correct hand placements for a more in-depth analysis of the quality of the administered CPR. This data and accompanging analysis and can be found in the thesis of Giesselink (2018).

8.3.2 SETUP

The participant and experimenter were located in two separate rooms (see Figure 8.3). The overall technical setup was the same as the one introduced in Chapter 6 and used in Chapter 7. The *bystander room* corresponds to the On-Site Environment. This is where the participant was located, together with the CPR doll and two Kinect cameras that captured the scene. Additionally, a regular RGB camera was placed for recording purposes. Additional material for the participants were the HoloLens AR headset worn in the *AR* condition, or a smartphone for the *phone* condition.

The *dispatcher room* corresponds to the Remote Operator Environment, where the experimenter was located, acting as the emergency dispatcher. The rooms were separated by a wall, not allowing direct communication. In the *AR-CPR* condition, the dispatcher used the VR setup as described in Section 8.2. In the *Phone-CPR* condition, only headphone and microphone were used, without the VR setup.

The setup for the *AR-CPR* includes using the visual aids shown to the bystander in AR, including the dispatchers embodiment for hand placement, the visual metronome to indicate optimal compression frequency, and the video instructions for the chin lift. Additionally In *Phone-CPR*, the only additional modality besides the voice of the dispatcher is an audible metronome for the optimal compression frequency. Next, we will describe the procedure for the experiment including the protocol for dispatcher and how these aids were employed.

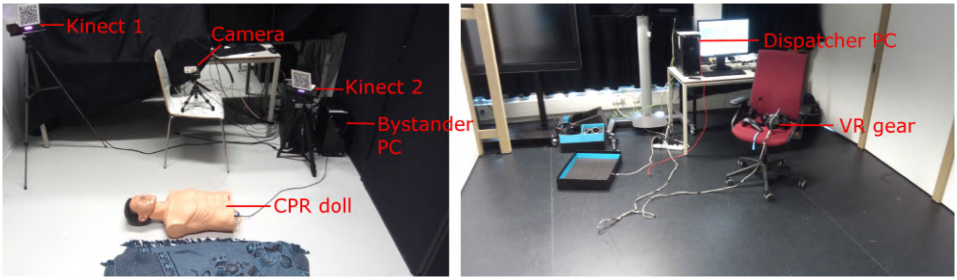


Figure 8.3. Experiment setup. Left; bystander room. Right; dispatcher room.

8.3.3 PROCEDURE

Participants signed a consent form and were then led into the *bystander room* where the CPR doll was located. They were then given the phone or AR-headset depending on the condition. The experimenter initiated the connection with the participant from the dispatcher room by calling on the phone or AR-headset. The dispatcher then gave a summary of the task, that is, that the participant would be guided through how to administer CPR to the doll as if it was an unconscious person, and that this would include compressions of the chest. To avoid tracking issues with the AR-headset, participants in both conditions were instructed not to administer the rescue breaths by mouth-to-mouth contact, but by merely blowing onto the doll's face from a distance.

The experiment started after this introduction, from here on following a protocol for the interaction between dispatcher and bystander inspired by the dispatcher protocol used in Dutch emergency call centers. This includes some basic information so that the dispatcher can send help, and to make sure that the bystander is safe (*Where* *are you?*, *Are your surroundings safe?*), followed by some checks on the victim (*Can you talk to the victim?*, *Can you see if the victim is still breathing*). The instructions for administering CPR follow, such as asking the bystander to kneel by the victim. The three main tasks are explained. First to perform the *chin lift*, which is needed to ensure the victim's airway is not occluded by their tongue. This is followed by instructions for the hand placement for the chest compressions. After 30 compressions, the bystander is told to administer two rescue breaths, before continuing with the chest compressions again.

In the AR condition, instructions for the chin lift were given with a virtual instruction video placed next to the victim's head in AR. Hand placement was guided using the virtual hands of the dispatcher. In the phone condition, both chin lift and hand placement were explained through verbal instructions ("*Put two fingers under the chin and lift the chin up [...]*", "*Place the heel of one hand in the center of the chest*"). In both the AR and Phone conditions, count and speed of the compressions was monitored by the dispatcher and bystander counting together, along an audible metronome. In the AR condition, the visual metronome was placed if the bystander deviated too far from the optimal rate, which was monitored in real-time by the dispatcher using the CPR doll sensors. No VR functionality was used by the dispatcher in the phone condition. This procedure of 30:2 compressions and breaths was repeated eight times. Once completed, the participant was asked to stop, and was lead to the dispatcher room to fill out the post-experiment questionnaire.

8.3.4 HYPOTHESES

For the quality of the administered CPR we expected an overall improvement in the *AR-CPR* condition, given the various AR aids shown to the participants in that condition, and the access that the dispatcher had to the doll sensor data. This is particularly beneficial for giving feedback on the compression depth. For compression frequency however, the audible metronome may be expected to act as a high floor for the phone condition.

While these comparisons on the task performance may seem to obviously favor the *AR-CPR* condition, there was also a concern that the complexity of the technology may actually counteract these benefits. The HoloLens headset has a limited field of view, and it's almost 600g in weight may also negatively affect the experience given the physical nature of the task. The aforementioned Raw-TLX instrument for measuring task load may reveal if that is the case. While the perceived task load may remain similarly high, we expected the overall mental demand to decrease.

Again, our main focus in this chapter however is less on the performance measure or usability, but rather on the effects our ARC intervention had on the subjective experience of the task, when compared to the voice-only condition. Although some of the differences we hypothesized for the On-Site Operator experience in the previous studies did not materialize, we consider the differences between conditions in this experiment to be larger. In particular, we expect perceived co-presence to be rated higher in the AR condition due to the complete absence of an embodiment in the phone condition. We expect this to manifest in a significant difference in the co-presence subscale. Given the use of other communication elements and the ability for the dispatcher to give more specific feedback on the bystander's performance, we also expect the communication-related measures of the Networked Minds instrument, such as perceived message understanding and behavioral interdependence, to score significantly higher in the experimental condition.

8.3.5 PARTICIPANTS

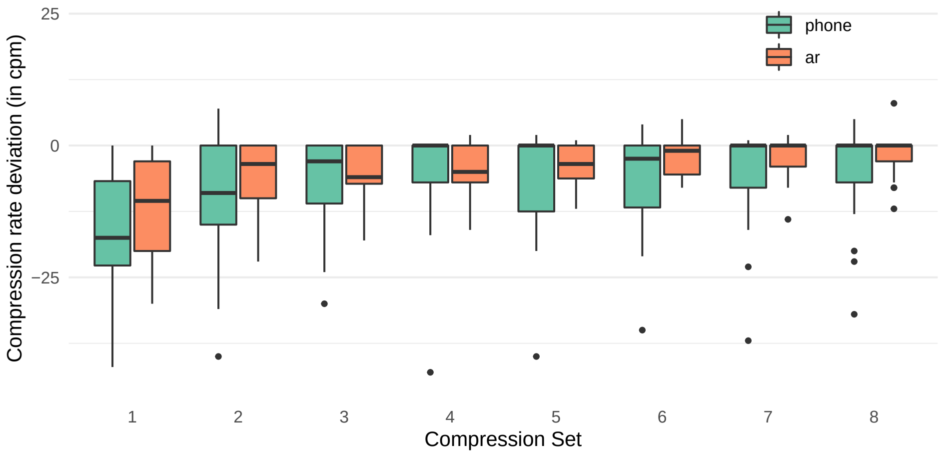
Using convenience sampling, n=60 participants were recruited. The mean age was 23.5 years (SD=5.7). Gender was almost balanced, with 13 females in each condition. The only selection criteria was that participants had not had any formal course or training on administering CPR in the past two years. One participant however indicated in a control question in the post-experiment questionnaire that they had recently participated in a CPR training. Due to a technical error, sensor data wasn't recorded for another participant. Both of these participants happened to be assigned to the *AR-CPR* condition.

8.3.6 DATA PROCESSING

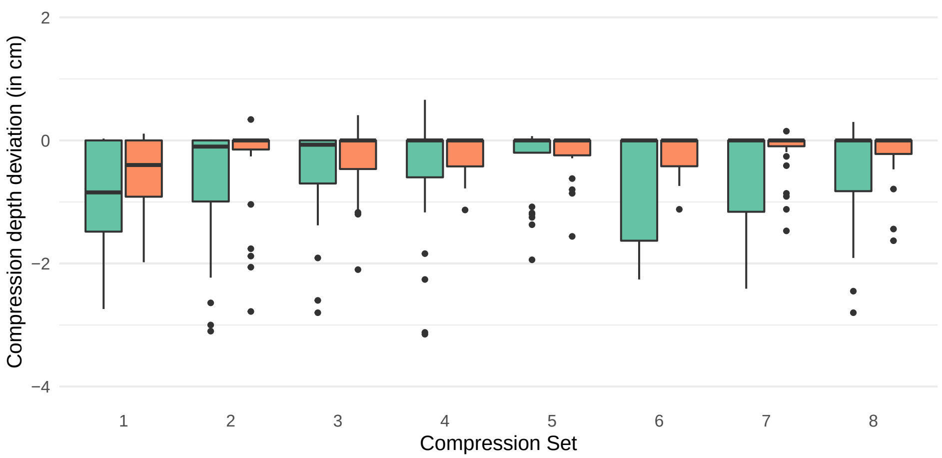
We post-processed the data from the force sensor to extract compression rate and depth deviations. After peak detection, windows for the eight sets were detected by looking at groups of consecutive compressions without pauses longer than three seconds. Since participants did not always perform exactly 30 compressions per set, we used a tolerance of ± 5 compressions. For the statistical analysis of the performance measures, we excluded data of participants where we were unable to detect the full eight sets of compression. After these exclusions, we are left with n = 23 participants in the phone condition and n = 25 in the AR condition.

8.4 RESULTS

The two performance measures, deviation from ideal compression depth and ideal compression frequency, are visualized in Figure 8.4. Shown are the deviations of the average compression rate and depth per compression set from the ideal ranges, namely 100-120 cpm and 5-6 cm respectively. For the subjective measures, in Figure 8.6 we visualize the scores on the subscales of the NASA-TLX and Networked Minds social presence instruments.



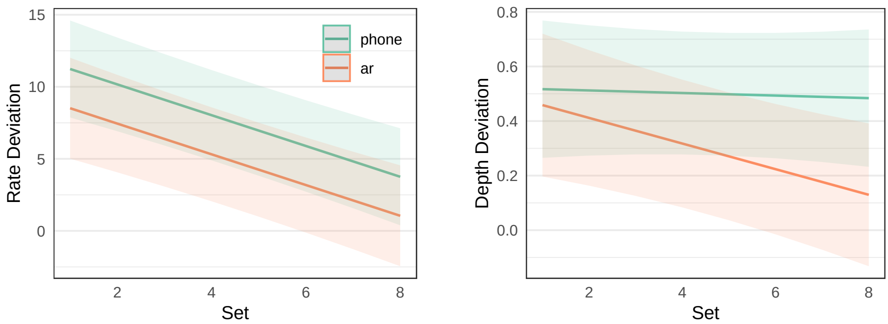
(a) Deviation from range of optimal compression frequency of 100-120 compressions per minute (cpm). Negative data points mean that the average rate during a set was slower than the optimum rate. Positives mean too fast.



(b) Deviation from range of optimal compression depth of 5-6 cm. Each datapoint is the average depth of compression. Negative data points mean that the average depth during a set was too shallow. Positives mean to deep.

Figure 8.4. Deviations from optimum compression rate (a) and depth (b) over the eight compression sets (n = 28 in AR, n = 30 in phone).

Performance Measures For statistical analysis of the performance measures, we fitted two linear mixed models using the restricted maximum likelihood (REML), one to predict deviation of compression depth, and one to predict deviation of compression rate, from the independent variables *condition* and *set*, with condition being either *phone* or *AR*, and set being the eight repeated sets of CPR. The models included *participant id* (pid) as random effect (formula: `~ 1 | pid`). Before fitting the models, we recoded deviation to use the absolute difference from the ideal ranges (i.e. too fast and too slow receive the same value).



(a) Predicted values of rate deviation **(b)** Predicted values of depth deviation

Figure 8.5. Predicted values of the models of average deviation over eight compression sets for a) compression rate b) compression depth

For deviation of compression rate (formula: `DepthDeviation ~ condition * set`), the model's total explanatory power is substantial (conditional $R^2 = 0.67$) and the part related to the fixed effects alone (marginal R^2) is of 0.11. The model's intercept is at 10.94 cpm (95% CI [8.87, 13.02], $t(378) = 10.36$, $p < .001$). For deviation of compression depth (formula: `RateDeviation ~ condition * set`), the model's total explanatory power is substantial (conditional $R^2 = 0.52$) and the part related to the fixed effects alone (marginal R^2) is of 0.04. The model's intercept is at 0.51 cm (95% CI [0.36, 0.67], $t(378) = 6.37$, $p < .001$).

The effect of *set* on the deviation of compression rate is statistically significant and negative ($\beta = -1.07$, 95% CI [-1.28, -0.86], $t(378) = -10.01$, $p < .001$; Std. $\beta = -0.30$, 95% CI [-0.35, -0.24]), meaning that particpants on average reduced the deviation of compression rate from the optimal range over time. The effect of *condition* was statistically non-significant ($\beta = -1.93$, 95% CI [-4.86, 1.01], $t(378) = -1.29$, $p = 0.198$; Std. $\beta = -0.23$, 95% CI [-0.55, 0.08]). There were also no significant interaction effects between *condition* and *set*. Visualizing the model's predicted values of rate deviation (see Figure 8.5a) indeed shows that the error reduces over time, independent of the condition. The slope offset shows a trend to slightly favor the *AR* condition, but again there was no significant main effect here.

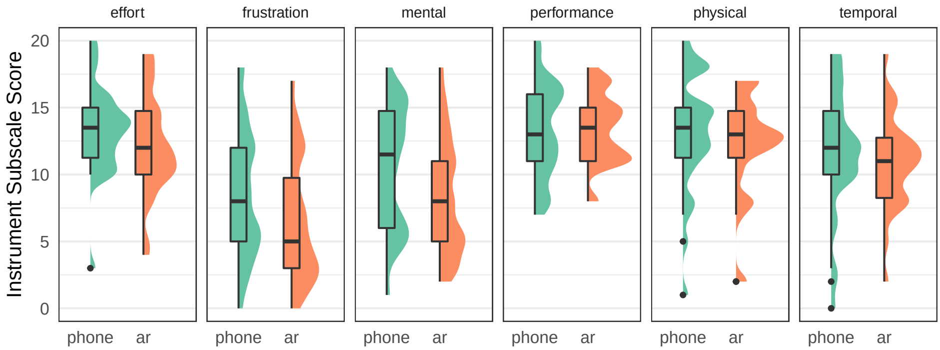
For compression depth, the model predictions look somewhat different. As with deviation of compression rate, we see a significant negative effect of *set* on deviation of compression depth (β = -0.03, 95% CI [-0.05, -6.61e-03], t(378) = -2.64, p = 0.009; Std. β = -0.09, 95% CI [-0.16, -0.02]), meaning participants reduced their deviation from the optimal range of compression depth over time. And again, there is no effect of *condition* (β = -0.01, 95% CI [-0.24, 0.21], t(378) = -0.10, p = 0.920; Std. β = -0.23, 95% CI [-0.53, 0.07]). However, here we do observe a statistically significant, negative interaction effect of *set* and *condition* (β = -0.03, 95% CI [-0.06, -2.71e-03], t(378) = -2.16, p = 0.031; Std. β = -0.11, 95% CI [-0.21, -9.90e-03]). Comparing the model's predictions by condition, the rate of improvement over time is estimated to almost constant (.005 cm per set, SE = .014) in the phone condition, but more pronounced for the AR condition (.05 cm per set, SE=.014). Again, these trends are visualized in the model's predicted depth deviation values shown Figure 8.5b.

Subjective Measures As for the two instruments measuring the subjective experience, we performed independent t-tests on the subscale scores of the NASA-TLX and Networked Minds instrument. On the NASA-TLX instrument, there were no significant differences. On the Networked Minds instrument, there was a significant difference on Attentional Allocation (t(58)=2.66, p=0.01) when comparing scores between conditions. In the AR condition, participants rated AA higher (M=4.36, SD=0.54) than participants in the phone condition (M=3.99, SD=0.53). There were no other subscales where scores were significantly different between conditions.

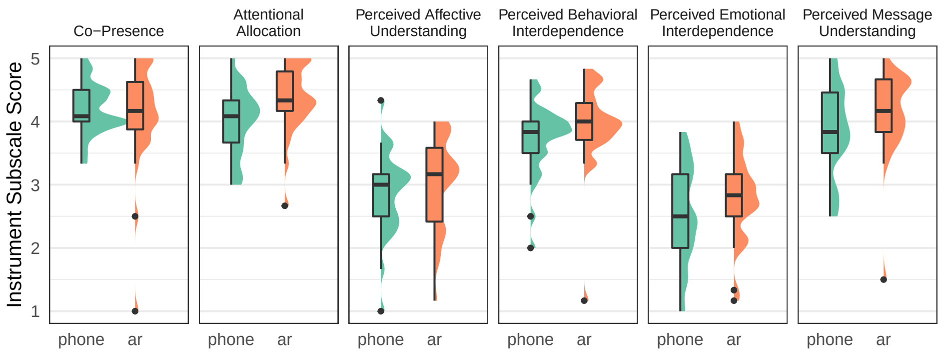
8.5 DISCUSSION

We evaluated ARC, our prototype for a telepresent emergency dispatcher that supports first responders with administering CPR in a lab setting. The ARC system allowed the dispatcher to be virtually present and see the bystander. ARC also allowed the dispatcher to make use of visual aids to support giving instructions and feedback. It was evaluated against the current state of the art, namely dispatcher support over a phone-based support.

In the wild, the immediate and correct administration of CPR can significantly reduce damage and increase survival rates. In our evaluation, we looked at performance measures that reflect two quality aspects of the administered CPR: optimal rate and depth of compressions. We have also looked at potential differences in subjective experience for the bystander, specifically task load and perceived social presence of the dispatcher.



(a) Responses to the 6 items in the NASA-TLX instrument. Higher scores mean more required effort, physical demand, etc.



(b) Average scores computed for each of the six constructs of the Networked minds instrument.

Figure 8.6. Responses to TLX (a) and Networked Minds (b) instruments by condition (n = 30 for both AR and phone.)

Difference in performance The main significant difference on performance measures was the difference in the change of deviation of compression depth throughout the several compression sets. As visible in Figure 8.4b, participants typically deviated on the side of too shallow compressions. Using the AR system, the dispatcher was able to reduce this deviation over time, while the deviation remained stagnant when using the phone system. For compression rate, participants typically erred on being too slow (see Figure 8.4a). However, we did not find a significant difference in CPR rate between the condition, nor in interaction with the progression over time. Looking at the trends, it seems that there is some rate of improvement that is similar across the conditions, although the slope for the AR condition is slightly offset in its favor.

These results largely fit our hypothesis, as we expected the audible metronome, the tool that helps to keep the right pace in the state of the art of phone assistance, to act as a high floor for compression rate, leaving little room for additional improvements through the feedback of the dispatcher or through the visual metronome. For compression depth however, we anticipated an improvement given the added ability for the AR dispatcher to give feedback when depth deviation was large. This is not possible in the state of the art, which explains why improvement is difficult over time.

Difference in perceived task load We hypothesized that having the dispatcher virtually present with access to visual aids for giving feedback and instructions would result in decreased task load, specifically in terms of mental demand. The results did not show any significant difference. Looking at the trends however, the median scores (see Figure 8.6a) are higher in the AR condition than in the phone condition in all but the *performance* subscale. However, the largest difference between the condition's median scores remains in the perceived required mental demand, in favor of the AR condition. While this is promising, it should be noted that no significant difference does not allow us to conclude the inverse, that the AR system is not perceived to be significantly *worse* than the less involved state-of-the-art approach, which was an alternative hypothesis we considered in Section 8.3.4, given some of the downsides of the prototype setup in the experimental condition, such as the weight of the headset.

Difference in social presence We expected that the AR condition would outperform the phone condition significantly on the co-presence subscale and on the communication related aspects, such as perceived message understanding. Given the substantially different interfaces, it is especially surprising that there was no effect on co-presence. In the previous study, where this difference also did not materialize, we suspected that a floor effect of the voice modality was responsible. This may also be the case here, as again, the main common modality that was used for communication during the experiment was voice in both conditions. Looking at the mean values of the co-presence score, they are indeed close towards the upper end of the scale for both conditions.

Instead, we found a significant increase in *attentional allocation* in the experimental condition. The attentional allocation is informed by questions regarding the focus participants received from and gave to the dispatcher. We haven't formulated a hypothesis regarding this measure. A possible post-hoc explanation could be that even though the AR technology did not strongly mediated co-presence, it still made the communication with the remote dispatcher more salient. This also raises the question whether the AR embodiment visible to the participants was sufficiently attributed to being part of the dispatcher's embodiment, or rather being perceived as one of the other visual aids of the system, such as the visual metronome.

Limitations & Future Work There are a number of limitations to this study. First of all, having one of the experimenters act as the dispatcher is not an ideal methodology. Given the high interactivity required in this context, it would be worthwhile to find an experiment design where a trained confederate acting as the dispatcher can remain blind to the experimental condition. What is more, for a true evaluation of the system, the user experience of the dispatcher needs to be included in the research as well. Another methodological concern is the sensitivity of the social presence instrument, as the floor for several subscales is already quite high by merely interacting via phone. We also had to exclude several participants from the performance analysis due to difficulties with detecting the exact sets and compressions from the sensor data. It is possible that the excluded data is biased towards being data of poor performance, however since slightly more data from the phone condition had been excluded, this potential bias is not favouring our hypothesis.

Even given some of the high social presence scores of the AR prototype in absolute terms, there are still improvements to be made if one goal of the system is to maximize perceived co-presence with the dispatcher. These are the technical limitations already encountered in the previous study, such as the limited field of view, but also the way in which the dispatcher embodiment is displayed in general. The dispatcher's hands for example are mostly salient towards the beginning of a session, when instructions for hand-placement are given.

Additionally, to address the concern raised in the discussion above regarding the attribution of the embodiment and virtual aids, we recommend in future to improve the design of the OpenIMPRESS embodiment appearance towards a full and connected body representation. It may be worthwhile to include control questions post-experiment to measure to what extent different elements, such as the embodiment or visual aids, are attributed to being part of the Remote Operator, being part of the system, or lacking ownership.

Finally, there is one implicit research question in the motivation for this work that we were unable to answer with the present evaluation. That is, how might improvements on perceived aspects of social presence - if achieved by an improved prototype - affect both the subjective experience in a real emergency situation, that is, by reducing negative emotions such as anxiety, and then indirectly also affect the quality of the CPR positively. Also, is this something where we find a similar high floor effect based on what the voice of a trained, empathetic dispatcher can achieve, or can the virtual presence of the dispatcher provide additional improvements?

8.6 CONCLUSION

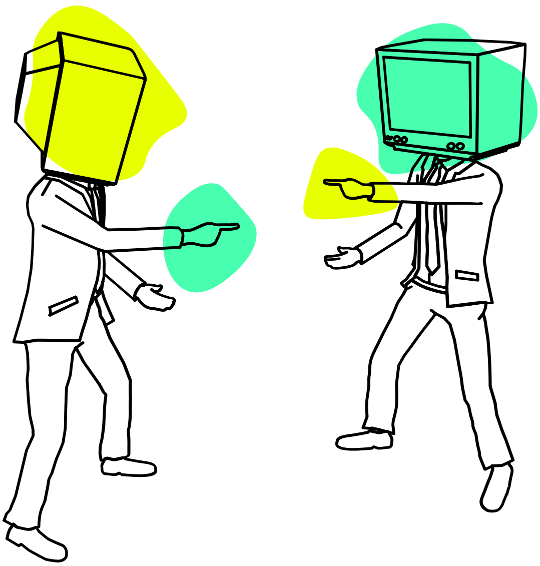
We have evaluated our prototype for a telepresent dispatcher in emergency situations using Augmented Reality technology. Our expert dispatcher taught participants how to administer CPR to a doll through either a phone or through our prototype, in both cases following a standardized protocol for emergency dispatchers. Our hypotheses on improved quality of the administered CPR and increases on certain measures of perceived task load and social presence of the remote dispatcher when using the AR prototype could only partially be supported by the results.

We found some measurable improvements on the rate at which the quality of CPR improved throughout sessions in terms of the optimal compression depth, but no main effect of our experimental condition on this or the other measure, namely deviations from optimal compression rate. In terms of perceived task load, we could not find significant improvements when using the AR prototype, although the trends are promising. For the social presence measures, it seems that the system did not mediate co-presence of the dispatcher more. Instead, an unexpected increase of the perceived attention received from and given to the dispatcher from the bystander was found.

Our main vision for such a system in a future where AR technology is more pervasive and wearable on a day-to-day basis is that it not only can improve the quality of the administered CPR through multimodal instructions from telepresent dispatchers. In addition, we imagine that the system may allow dispatchers to reduce negative emotions of the bystanders, such as anxiety or stress, by mediating the dispatcher's social presence to them. Ultimately thus, such a system could both improve the survival rates of the victims and improve the subjective experience of the bystander.

9

THIRD ORDER SOCIAL PRESENCE



In the previous two chapters we have employed mixed reality technology to mediate social interactions remotely. Besides performance measures, we have also looked at how users perceive the social presence of their remote interaction partners, depending on the communication medium - that is, different interface configurations in Chapter 7, or mediation through phone versus the *AR* prototype in the previous chapter. We have applied a measure of social presence to examine how the mediating technology affects the perceived level of co-presence and mental representation of interaction partners in the social interaction tasks that participants engaged in. This measure is an important indicator of how well our technology functions to mediate social interaction, and where it may be limiting or distorting.

For both studies, these tasks featured asymmetric roles for the interaction partners, where the person working on site (the On-Site Operator, or OSO) would deal with the task at hand, for example through manipulation of puzzle pieces or by administering CPR to a doll, while the remote person (the Remote Operator, or RO) would assist the OSO with expertise or knowledge only available to them. The OpenIMPRESS system that was used for both studies foresees different interfaces for its users to suit the different roles based on the access that the two users need to each other and to their respective environments. The OSO needs to interact with their physical environment, and thus uses an AR headset. The RO accesses that remote location through a VR setup, as their own physical environment is irrelevant for the interaction.

Given these differences in users' roles and interface, it was perhaps not surprising that individual social presence scores were affected differently by changes in the mediation technology. However, as we will discuss in the remainder of this chapter, the way in which Harms and Biocca's social presence instrument is typically employed, a third dimension is averaged away: The *intersubjective level*, a higher order concept and measure of *symmetry* in social presence embedded in the instrument by design.

On the intersubjective level of social presence, we consider the level at which perceived social presence is mutual between interactants, or whether interactants feel that presence might be skewed in some way to favor themselves or their partner. This level seems particularly relevant to highly asymmetric scenarios of mediated interaction, like the ones we have examined.

In the following, we will first revisit the design of Harms and Biocca's social presence instrument, in particular the three dimensions of social presence, including the *intersubjective level*. We will also provide a brief survey of some of previous works that have employed the instrument and also applied these symmetry measures in their analysis. Finally, we conduct and discuss an exploratory post-hoc analysis of the data collected in the two previous studies that takes the symmetry measures into account.

9.1 THE INTERSUBJECTIVE LEVEL OF SOCIAL PRESENCE

The development and validation of the Networked Minds measure of social presence can be traced through a number of publications, mainly by the authors Harms and Biocca (Biocca, Burgoon, et al., 2001; Biocca and Harms, 2002, 2003; Biocca, Harms, and Burgoon, 2003; Biocca, Harms, and Gregg, 2001; Harms and Biocca, 2004). Biocca and Harms conceptualize social presence in a model with three levels. Level 1 is the perception of spatial co-presence with others. Level 2 is the subjective level, concerned with how interactants perceive access to each other's bodies and minds, the *psycho-behavioral accessability* (Biocca and Harms, 2002). Levels 1 and

2 are captured by the six constructs of *co-presence* (Level 1), *attentional allocation*, perceived *message understanding*, *emotional understanding*, *behavioral interdependence* and *emotional interdependence* (Level 2) in the updated and validated version of the instrument (Harms and Biocca, 2004). We have previously discussed these constructs in Section 7.1.

The third level of social presence is the *intersubjective level*, which is meant to capture the extent of how *"one individual perceives the social presence to be mutual"*, and *"intersubjectively the degree to which a pair of interactants share this sense of social presence among each other"* (Biocca and Harms, 2002). To capture this level in their instrument, Harms and Biocca thus formulate each item in their instrument as a pair of *reciprocal items*. An example of such a pair is *My partner's thoughts were clear to me* and *My thoughts were clear to my partner*. The first evaluates one's own internal sense of the other's social presence, and the second evaluates how one perceives the other's sense of their presence. Each of the six subscales have three pairs of reciprocal items associated with them in the updated and validated version of the instrument (Harms and Biocca, 2004).

It should be noted that the way one finds the Networked Minds instrument applied typically is by computing the six subscale scores as averages, without further discrimination of the reciprocal items. In fact, when looking at the 19 papers from the recent literature survey on the role of social presence in AR and VR interventions for cooperative tasks (Osmers et al., 2021), even though twelve of these papers used a version of Harms and Biocca's instrument, none performed a symmetry analysis or an analysis that treats the reciprocal item pairs differently – this includes our own work that we reported on in the previous two Chapters.

However, if one is explicitly interested in the dynamics between different users of the same system, a more in-depth analysis can be performed, where these reciprocal items for each construct are grouped, and the *symmetry* of these pairs within and across paired users are computed. Harms and Biocca describe two symmetries as measures of the intersubjective level of social presence.

Within-interactant Symmetry The *within-interactant symmetry* reflects how an individual user perceives their interaction partner's and own presence comparatively. It can be computed as the correlation between their own responses to the reciprocal items, that is, A's responses to *My partner's (B) thoughts were clear to me (A)* and *My thoughts (A) were clear to my partner (B)*.

But how does (lack of) within-interactant symmetry shape the interaction between users? Biocca and Harms state that it is *"important from the individual's own construction of the interaction as it guides the successive interaction."* (Biocca and Harms, 2002). They illustrate it with the challenge of communicating sarcasm when using text-based communication. If a user expects that a text message may not be clearly received as being sarcastic due to the limitations of the textual medium, they might feel inclined to append a winking emoticon as a way of compensating for the lack of the speech modality, where the speakers tone of voice is used to mediate the sarcastic meaning of the message.

Cross-interactant Symmetry *Cross-interactant symmetry* reflects how similarly users perceive another user's social presence. To give an example from a dyadic interaction between users A and B, this would be the correlation between A's response to *My parter's (B) thoughts were clear to me (A)* and B's response to *My thoughts (B) were clear to my partner (A)*.

According to Biocca and Harms, cross-interactant symmetry is an important feature of social presence as it requires, *"framed from a theory of mind perspective[,] a level of access to another's emotional state and intentions. It is by decoding the content of the interaction, as it exists in the context of the other individual's affective state, that a level of connection or understanding [...] is reached"* (Biocca and Harms, 2002).

High cross-interactant symmetry means that users perceive that they can accurately model each other's *access* to the other. To continue with the example above, a user may adapt their communication based on their own perception of the other user's ability to perceive, that is, by inserting a smiley in a sarcastic message. High cross-interactant symmetry means that indeed, the other user might have not been able to perceive the sarcasm otherwise, and thus appreciates the inclusion of the smiley.

According to Biocca and Harms' guide (Biocca and Harms, 2003), these two scores are computed from the correlation between a user's ratings of their own and the other's experience (for within-interactant symmetry), and the correlations between the perceptions of the experience attributed to individual users by all users (including that user themselves, for cross-interactant symmetry). Unfortunately there is no more detailed description for the exact computations and methods for analysis in an experimental context.

We performed a non-exhaustive literature search for examples of symmetry analyses using the Networked Minds instrument. We searched for all papers referencing one of the three common references used for the instrument (Biocca and Harms, 2003; Biocca, Harms, and Burgoon, 2003; Harms and Biocca, 2004) in combination with the term *symmetry*. After scanning abstracts and methodologies, only three papers remained that performed a symmetry analysis on the social presence measures in some capacity: one from the field of social robotics, where pilots of a telepresence robot reported high within-interactant symmetry, meaning that they perceived remote interaction partners to share their state of social presence (Kristoffersson, Eklundh, and Loutfi, 2013) - unfortunately, further details of how this analysis was carried out are underreported in this publication. Next, in their study on dyadic meditation in VR, Järvelä et al. find increased social presence when using enabling interventions such as shared breathing and biofeedback mechanisms in the VR environment (Järvelä et al., 2021). They additionally computed cross-interactant symmetry scores – by computing the absolute difference and subtracting it from the scale maximum value – which were not affected by the experimental conditions. Finally, in their study on the effect of a hologram-based communication system on social presence when compared to video-based systems, Mazgaj et al. conduct a symmetry analysis by computing the correlation within participants reciprocal item pairs, and find a significant correlation for these scores on the co-presence measure in the hologram condition, that is not there for the video condition (Mazgaj et al., 2021).

Finally, it's worth mentioning that there are also studies that look at the two directions of perceived presence separately, but not relating them to each other in a symmetry analysis. These papers would likely not show up in our literature search unless they happen to also include the term *symmetry* or happen to cite the version of Biocca and Harms's guide that explicitly discusses this analysis (Biocca and Harms, 2003). An example of this is the study of Kontogiorgos et al., 2020. Here, the effects of social robot embodiment on users' perception of the robot when failure occurs during the interaction are investigated. The social presence constructs are split and analyzed by their direction (own and other's perceived experience), but a symmetry analysis is not performed.

To summarize, it was surprising to only find so few examples of third order social presence analysis using the Networked Minds instrument. What's more, the exact methods used vary somewhat, or are underreported. In particular, methods using absolute differences between scores, such as the one used by Järvelä et al. (2021) loose the information of the direction of the asymmetry, that is, which of the users perceived the other's presence more. This is important information especially when users have different roles in the interaction, or use different interfaces than the other. In the remainder of this chapter, we will re-examine the data from our studies in Chapters 7 and 8, to look at how social presence was perceived intersubjectively for the task performed with the OpenIMPRESS system.

9.2 EXPLORATIVE ANALYSIS

We will first look at intersubjective social presence across both studies. In this analysis we focus on the *within-interactant symmetry* measure as it can be computed from individual participant responses. *Cross-interactant symmetry* would require responses from all users, so it cannot be computed for the *CPR* study in Chapter 8, where the social presence instrument was not administered to the RO. Our approach is to first visualize the symmetry measure for each of the six lower-order social presence measures to identify potential patterns, separately for the two studies, participant roles and experimental condition. In the *CPR* study in Chapter 8, the experimental condition was to either use voice-only phone communication with the dispatcher, or to use the experimental system that would enable a virtual embodiment for the remote dispatcher in augmented reality.

To keep the comparisons across the two studies as fair as possible, we only select the data from the *puzzle* study (Chapter 7) that are most analogous to the conditions in the *CPR* study, that is, the *NoEmbodiment* configuration as analogue to the *phone* condition, and the *Baseline* configuration as an analogue to the *AR-CPR* condition. Since they were part of different experiments with different methodologies, we decided to describe these in the following as the different *experiences*, rather than conditions. We have the *WithEmbodiment* (formerly *Baseline* and *AR*) and the *WithoutEmbodiment* (formerly *NoEmbodiment* and *Phone*) experiences. See Table 9.1 for an overview of the samples.

experiment	experience	role	n
CPR	WithoutEmbodiment	OSO	30
CPR	WithEmbodiment	OSO	30
PUZ	WithoutEmbodiment	OSO	10
PUZ	WithoutEmbodiment	RO	10
PUZ	WithEmbodiment	OSO	30
PUZ	WithEmbodiment	RO	30

Table 9.1. Data samples included in the analysis. All data from the CPR study, and the analogous experiences from the Puzzle Study.

9.2.1 COMPUTING A SYMMETRY SCORE

To compute the *within-interactant*, or *subjective* symmetry, we first need to annotate the pairs of reciprocal items per subscale of the Networked Minds instrument. Each item is either has a *target*, either *self* or *other*. Items coded with *self* relate to one's own perception of the other, while *other* items ask the participant to think about the *other*'s perception of them. For example, *I noticed my partner* is coded as *self*, and *My partner noticed me* is coded as *other*. Each construct of the instrument features exactly three pairs of reciprocal items. For the full version of the instrument, including the annotated *target* as used in the remainder of this analysis, see Table 9.2. This table also features the common shortcuts (*aa*, *pei*, etc.) used in the following. For a detailed description of the different subscales of this instrument, refer to Section 7.1.

subscale	Q	Question	Invert	Target
co-presence (co)	Q1	I noticed [MY PARTNER]	0	self
	Q2	[MY PARTNER] noticed me	0	other
	Q3	[MY PARTNER]'s presence was obvious to me	0	self
	Q4	My presence was obvious to [MY PARTNER]	0	other
	Q5	[MY PARTNER] caught my attention	0	self
	Q6	I caught [MY PARTNER]'s attention	0	other
attentional allocation (aa)	Q7	I was easily distracted from [MY PARTNER] when other things were going on	1	self
	Q8	[MY PARTNER] was easily distracted from me when other things were going on	1	other
	Q9	I remained focused on [MY PARTNER] throughout our interaction	0	self
	Q10	[MY PARTNER] remained focused on me throughout our interaction	0	other
	Q11	[MY PARTNER] did not receive my full attention	1	self
	Q12	I did not receive [MY PARTNER]'s full attention	1	other
perceived mutual understanding (pmu)	Q13	My thoughts were clear to [MY PARTNER]	0	other
	Q14	[MY PARTNER]'s thoughts were clear to me	0	self
	Q15	It was easy to understand [MY PARTNER]	0	self
	Q16	[MY PARTNER] found it easy to understand me	0	other
	Q17	Understanding [MY PARTNER] was difficult	1	self
	Q18	[MY PARTNER] had difficulty understanding me	1	other
perceived affective understanding (pau)	Q19	I could tell how [MY PARTNER] felt	0	self
	Q20	[MY PARTNER] could tell how I felt	0	other
	Q21	[MY PARTNER]'s emotions were not clear to me	1	self
	Q22	My emotions were not clear to [MY PARTNER]	1	other
	Q23	I could describe [MY PARTNER]'s feelings accurately	0	self
	Q24	[MY PARTNER] could describe my feelings accurately	0	other
perceived emotional interdepend. (pei)	Q25	I was sometimes influenced by [MY PARTNER]'s moods	0	self
	Q26	[MY PARTNER] was sometimes influenced by my moods	0	other
	Q27	[MY PARTNER]'s feelings influenced the mood of our interaction	0	self
	Q28	My feelings influenced the mood of our interaction	0	other
	Q29	[MY PARTNER]'s attitudes influenced how I felt	0	self
	Q30	My attitudes influenced how [MY PARTNER] felt	0	other
perceived behavioral interdepend. (pbi)	Q31	My behavior was often in direct response to [MY PARTNER]'s behavior	0	self
	Q32	The behavior of [MY PARTNER] was often in direct response to my behavior	0	other
	Q33	I reciprocated [MY PARTNER]'s actions	0	self
	Q34	[MY PARTNER] reciprocated my actions	0	other
	Q35	[MY PARTNER]'s behavior was closely tied to my behavior	0	other
	Q36	My behavior was closely tied to [MY PARTNER]'s behavior	0	self

Table 9.2. Items from the Networked Mind instrument (Harms and Biocca, 2004) with the Target annotated.

The symmetry score for a participant is the average score of *self* items in a subscale minus the average score of *other*, normalized by the scale range. If a participant scores all *self* items 5 and all *other* items 1, the maximum positive asymmetry score of 1 will be obtained. If both are equal, the symmetry score is 0. If *other* items are favored, the asymmetry is negative.

To illustrate this with the example of the *attentional allocation* construct. If the ratio for this construct is near zero, users perceived that they extend the same amount of attention towards their interaction partner, as they feel that their interaction partner extended attention towards them. When score ratios are positive, participants felt stronger that they paid more attention to their interaction partner than they felt that their interaction partner paid to them (and vice versa for negative score ratios).

It should be noted that the symmetry score is not always equally intuitive to interpret, as some reciprocal item pairs for the social presence constructs use different semantics. Consider this item pair from the perceived emotional interdependence (*pei*) construct:

Q27 "[MY PARTNER]'s feelings influenced the mood of our interaction"

Q28 "My feelings influenced the mood of our interaction"

Here *the mood of our interaction* is not clearly differentiated if the other's perception of the self, or the own perception of the other is being rated. Following the pattern of the other item pairs in that construct, we can label Q27 as the *self*, and Q28 as the *other* item. However, it remains helpful to relate back to the exact semantics of the questions that load a construct, and to the exact coding of the reciprocal pairs, when interpreting the direction of a symmetry score.

9.2.2 SYMMETRY ACROSS STUDY, EXPERIENCE AND ROLE

The resulting distribution of within-interactant symmetry scores across studies (*puzzle* and *CPR* study) and role (OSO and RO, RO only for *puzzle* study) are shown in Figure 9.1. Note that for the *puzzle* study data, the datasets are not balanced across the *experiences*, as every participant in the study experienced the *Baseline* configuration, while only a third of the participants were assigned to the *NoEmbodiment* condition. We use box plots to get an impression in which directions the symmetry skews (i.e. asymmetric), or whether it is centered around the 0-value (i.e. symmetric). This visualization gives a first impression of an overall 'signature' of the within-interactant symmetry by social presence construct, across the different tasks that participants did in the respective study.

A first interesting observation in the data in Figure 9.1 is that *CPR* data, symmetry scores are skewed away from zero more than in the *puzzle* data. A possible explanation could be the different natures of the tasks and the status assigned to the different roles between the studies. In the CPR task, the RO was the experimenter and was instructing the participant. In the puzzle task on the other hand, the roles in the task were much more equal in terms of status due to the collaborative nature, with the main difference being in the access of information and access to the physical space between the collaborators.

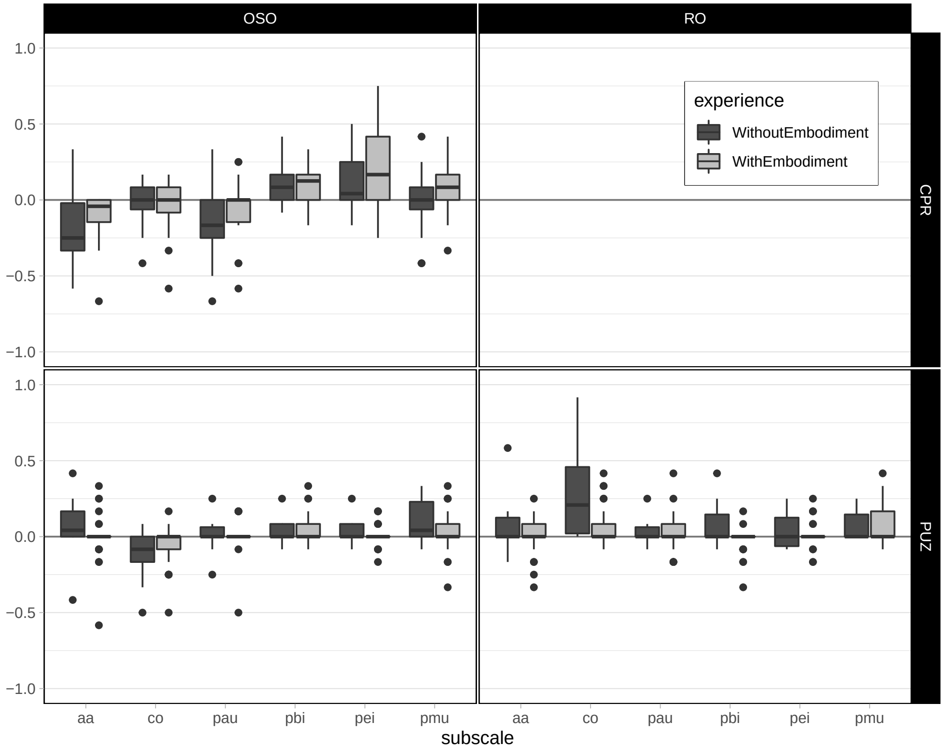


Figure 9.1. Within-interactant symmetry as difference between the average of reciprocal item pairs per social presence construct. Data from the *CPR* and *Puzzle* (PUZ) experiments, split by the role (OSO or RO) of the participant and whether they engaged in the experience with the RO virtually embodied (WithEmbodiment) or invisible (WithoutEmbodiment). See Table 9.2 for a key of the short hands used for the construct names.

What's more, the symmetry skew directions (if any) in both *CPR* and *puzzle* data are quite similar between the *WithEmbodiment* and *WithoutEmbodiment* experiences, suggesting that the task and relationship between interactants may bias the symmetry scores more than the technology used. There are, however, examples where the overall symmetry skew direction is similar, but the magnitude of the median noticeably differ between *experience*. One prominent example of this is the perceived symmetry of co-presence (*co*) for the RO in the puzzle task, where the asymmetry seems to be rated as more pronounced by participants in the *WithoutEmbodiment* experience.

For the *CPR* data, there are two similar examples, on the attentional allocation (*aa*) and perceived affective understanding (*pau*) constructs. Overall, these skew in favor of the *other*, that is to say, participants rated items like "[MY PARTNER] remained focused on me throughout our interaction" and "[MY PARTNER] could describe my feelings accurately" higher than their reciprocal item pairs ("I remained focused on [MY PARTNER] throughout our interaction" and "I could describe [MY PARTNER]'s feelings accurately"). Notably however, this skew is more pronounced in the *WithoutEmbodiment* experience. We might expect the overall symmetry skew to be more symmetric in situations where a novice needs to follow the instructions of the remote expert, as both need to pay attention to each other in turns: The novice pays attention when they observe how the experts demonstrates the execution of the task, and the expert pays attention to the novice while they attempt to perform the task.

For a task like performing *CPR* however, once the two actions are explained at the beginning, the interaction between novice and expert changes, where the novice focuses on repeating the task first and foremost, and the expert monitoring the performance, giving feedback only when necessary. While this explains the overall skew direction on the *aa* and *pau* constructs, it doesn't quite explain why this skew moderated towards more symmetry when the virtual embodiment was added. One hypothesis could be that the mediation technology afforded more access of the OSO to the RO, as the RO was not just voice-only, but was also visible and addressable through their virtual embodiment rendered in the OSO's AR display. We will address this hypothesis in more detail in the discussion in Section 9.3.

For the *CPR* data, overall symmetry scores across the social presence constructs do not all skew in the same direction: While the *aa* and *pau* construct favor the *other* items, perceived behavioral interdependence (*pbi*) and perceived emotional interdependence (*pei*) favor the *self* items. This apparent contradiction is best explained by formulating the symmetry directions in a more descriptive way: While participants feel that the instructor paid more attention to them (*aa*), and that the instructor had a better understanding of their emotions (*pau*) than the other way around, they felt that their mood and behavior were more affected by the instructor, than the instructor's mood and behavior were affected by theirs (*pei* and *pbi*). Overall, these asymmetries are not surprising to arise in *remote expert, local novice* interactions, when the export observes and steers the behavior of the novice.

In the *puzzle* data, the distributions of most subscales are centered closer around the zero-value, indicating an overall high perceived symmetry between the RO and OSO. For the *puzzle* data we can also already look at how skew directions compare between interactants, as a first indication of cross-interactant symmetry. Again, the *co* construct stands out, as it not only skews away from the zero mark for both RO and OSO, but also in opposite directions. To explain this more descriptively for the co-presence construct: OSOs rated questions like *"The RO noticed me"* higher than *"I noticed the RO"*, while ROs rated questions like *"I noticed the OSO"* higher than *"The*

OSO noticed me". So there seems to be agreement on this asymmetry, both perceive that the RO is more co-present with the OSO than the OSO is co-present with the RO.

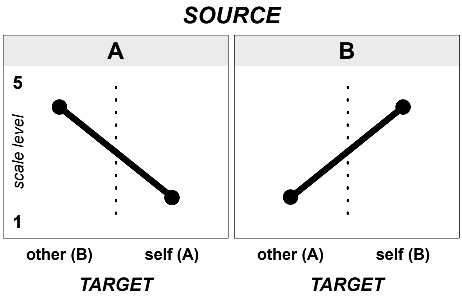
A final observation on the *puzzle* data is that most subscales slightly favor the *self* account, which perhaps reflects some tendency to more confidently describe one's own inner state rather than to describe the inner state of the other participant.

This visualization showed some interesting tendencies, especially since it allows us to contrast these within-interactant symmetry scores between different styles of tasks engaged in with similar technology. In the next step, we want to take a closer look at cross-interactant symmetry.

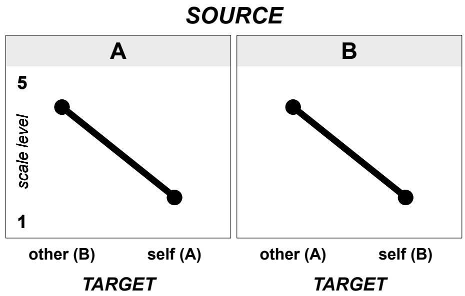
9.2.3 PERCEIVED CO-PRESENCE SYMMETRIES

For cross-interactant symmetry, we are limited to the data from the puzzle study, where we have responses from both participants, the Remote Operator and On-Site Operator. We will focus only on the measure that stood out the most in Figure 9.1 for the puzzle study data, the *co-presence* measure.

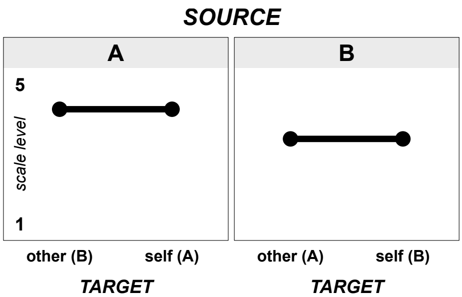
First however, we want to develop an intuition on how to visualize and interpret both within- and cross-interactant (a)symmetries. In Figure 9.2, we propose a visualization of a dyads' responses to the respective reciprocal item pairs for a given subscale of the Networked Minds instrument. For the *within-interactant symmetry score* computed in Section 9.2.1, we took the ratio between the *self* and *other* items of the item pairs. Here, instead, we plot the mean values of the *self* and *other* items and represent the ratio as a slope by connecting these data points. This results in two slopes for each social presence measure in the instrument, one representing the *within-interactant symmetry* for participant A, and one for participant B. The dotted lines in Figure 9.2a visualize the *symmetry axes* for the cross-interactant symmetry. Sorting the data points on the x-axis by *target* (i.e. *self* or *other* instead of by *A* and *B*) causes a third symmetry axis to appear between the panels for A and B, which represent the cross-interactant symmetry. If the within-interactant symmetry slopes are mirrored between these panels (such as in Figure 9.2a and Figure 9.2c), there is also mutual cross-interactant symmetry. In other words, both A and B's perceptions agree with each other completely. In Figure 9.2b, there is no cross-interactant symmetry, as there is no agreement on their reciprocal item scores. Visualizing the data this way allows us to identify both within-interactant symmetry by looking at the steepness and direction of the slopes, as well as the level of cross-interactant symmetry by checking whether the slope angles are mirrored across the symmetry axis between the panels. In contrast to the previous visualization, we can still see differences in absolute score levels (see Figure 9.2c). There are also examples where there cross-interactant symmetry is not mutual, such as in Figure 9.2d.



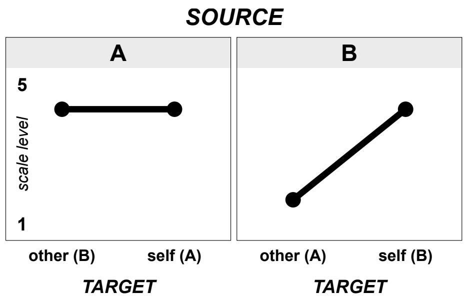
(a) within-interactant asymmetry, which is 'mirrored' intersubjectively. Both agree that A's access to B is higher than B's access to A.



(b) within-interactant asymmetry and cross-interactant asymmetry. Both A and B feel that they have better access to the other, than the other has to them.



(c) Subjective and cross-interactant symmetry. Both feel that they have similarly levels of access to each other.



(d) within-interactant symmetry perceived by A, but asymmetry perceived by B, so cross-interactant asymmetry.

Figure 9.2. Example patterns of within- and cross-interactant symmetries. Levels of *TARGET* represent the average score of the subscale items split by their reciprocal target, that is, the evaluated own sense of social presence vs their perception of their partner's (other) sense of social presence. Semantics of the examples in the sub-captions use the wording *access to X* as a shorthand stand-in for some social presence constructs, such as perceived affective understanding

Based on the tendencies in the exploratory analysis above, we will focus on the symmetry of co-presence construct scores. So besides *role* and *embodiment condition* used in the original analysis of the puzzle experiment, we introduce the third variable *target*, describing whether a user evaluates their perceived social presence (*self*), or how they perceive the *other*'s sense of social presence. In Figure 9.3, we visualize the data in using the same layout as in the examples shown in Figure 9.2.

We can see that the pattern resembles our example in Figure 9.2c when looking at the *WithEmbodiment* data, showing both within- and cross-interactant symmetry. When looking at the *WithoutEmbodiment* data, we see that within-interactant symmetry decreases for both, but much more so for the RO, resuling in a pattern somewhere between examples *a)* and *c)* in Figure 9.2.

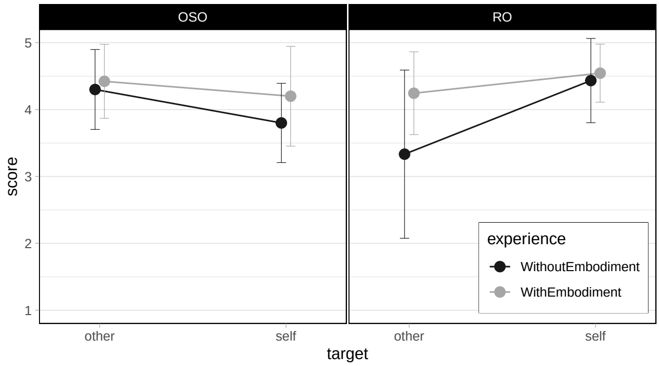


Figure 9.3. Mean co-presence score in the puzzle data per role, with separate lines for the two types of experience (WithEmbodiment and WithoutEmbodiment). Error bars show the standard deviation.

9.3 DISCUSSION

In our original analyses in Chapters 7 and 8, a single measure per social presence construct was used for statistical analysis: the average scores over all items associated with that subscale, which is the way in which this instrument is commonly employed. In this chapter, we have explored within- and cross-interactant symmetry, two additional dimensions that are implicitly included in Biocca and Harms's instrument for social presence.

Our visualizations of the *within-interactant symmetries* revealed that for some subscales, participants did rate their own sense of social presence differently than they rated the other's sense of social presence. The extent of this difference also differed between the two studies we have analyzed, suggesting that perceived symmetry in social presence is not only informed by the technology used to mediate the interaction, but also by the context of the task and the role of the interlocutor. The OSO perceived biases towards the Remote Operator in the *CPR* study being more aware of the OSO's presence than vice versa may be explained by the nature of instructing the task of CPR. The main interaction between RO and OSO on transferring the knowledge was comparatively brief, as once the OSO starts performing CPR themselves, they are focussed on that task for several iterations, with the RO giving feedback on the performance. On the other hand, the task performed by the team of RO and OSO in the *puzzle* experiment was much more collaborative in nature, with similar status of between the interactants, which might explain that social presence constructs were overall perceived to be as more *symmetric*.

However, there were also some instances where the *experience* afforded by the mediation technology seem to affect the direction and magnitude of the symmetry scores. For the *CPR*, the asymmetry magnitude of the *aa* and *pau* constructs seem to shrink when the RO's presence is being mediated through a virtual embodiment, as apposed to being voice only. As we have already suggested above, this may be seen as the technology moderating the aforementioned bias due to the differences in role between RO and OSO in the *CPR* task.

We also looked at *cross-interactant symmetry*, that is, the extent to which interlocutors' perceptions of each other's sense of social presence agree with each other.

In the original analysis of this data (in Chapter 7), we found that the RO reported significantly lower levels of co-presence when their embodiment at the remote location was removed – even without considering symmetry (i.e. when just averaging out all items in the *co* construct). At the same time, there was no difference for the OSO. Our speculation there was that, although the OSO has not perceived any difference in presence between the conditions, the OSO still behaved differently towards the RO in a way that was noticed by the RO.

When taking the reciprocal item pairs into account, we see that this lower score was mainly carried by the subset of items that had the RO report on how they perceived the OSO's level of co-presence, while, on average, their own perceived co-presence seem unchanged, meaning that the cross-interactant symmetry stopped being mutual.

It is also worthwhile to try and explain how the differences between the experiences afforded in the experimental conditions informed this outcome. Recall that the visibility of the RO's embodiment in the OSO's environment was the only design aspect changed in that experimental condition compared to the fully featured *Baseline* setup. The RO was still able to move around freely in the remote environment (unlike in the *FixedView* condition). The RO's perception of the OSO thus was not affected at all. As we already speculated, being invisible to the OSO means that the OSO would never be able, for example, to visually address the RO, even if the RO positioned themselves right in front of the OSO. On the other hand, when the RO is visible in the OSO's environment, this is possible. Even if the RO is positioned out of the limited field of view of the OSO's AR display, the RO can take action and reposition themselves in order to be visible to the OSO.

This does still not explain the cross-interactant asymmetry on the co-presence measure: Why did this difference did not manifest (as strongly) for the OSO? One possible explanation could be that the voice modality acted as a ceiling effect for co-presence, but at different levels for the RO than for the OSO, due to their different roles in the encountered tasks. That is, some tasks that may be easier to solve with only the voice of the RO (causing the ceiling effect for the OSO), but where only the voice of the OSO was not enough (to cause a ceiling effect for the RO).

Speculations like these may be suitable to improve the system in the future - but what are we seeking to improve upon? This is something that we haven't addressed with this analysis, but that could be interesting to incorporate in evaluations in the future: Are there correlations between disagreement (asymmetries) on presence constructs and performance measures on the collaborative tasks?

There are some limitations to this analysis, such as the low and imbalanced sample size for the puzzle data. We have not formulated rigorous criteria for when we qualify an outcome as *symmetric* or *asymmetric* beyond the example patterns. What's more, we used two different methods for the symmetry scores for the different analysis, first by computing ratios per participant for the within-interactant symmetry scores, second by computing the computation means and presenting the slopes as the symmetry scores. In the latter we do not lose the absolute distribution of social presence score levels from the population. We should stress that these are our own interpretations of how to approach and explain symmetry in social presence, as the guides for the Networked Minds (such as Biocca and Harms, 2003) do not define exact methods. Alternative methods may be more appropriate, such as correlation scores per-participant and per-dyad.

9.4 CONCLUSION

Higher order social presence constructs are included in the commonly used Networked Minds instrument for social presence, although rarely considered in research. Especially for social interactions mediated with technology that give asymmetric access to interactants, such as our *OpenIMPRESS* system, social presence symmetry may be used as a diagnostic tool to explain variances in user experience. However, besides mediation technology, also task context, such as nature of the task and relationship between interactants may color how participants perceived symmetry is rendered.

10

A PASSENGER OUTSIDE OF THE VEHICLE

It is easy to relate the OpenIMPRESS system to the science fiction concepts described in the introduction of this thesis. Telepresence systems have been described with terms like beaming (i.e. Steed et al., 2012), and at least from the RO point of view, the point-cloud representation of the remote capture site and of the OSO has certain similarities with how *Holograms* are depicted in the Star Wars movies. The work discussed in the previous hapters also relates to the science-fiction scenario of the SoIC-TV, described in Section 1.1.

The concept for the *Augmented Reality CPR* system (ARC) in Chapter 8 was already situated in its own science-fiction scenario, where head worn AR technology is a pervasive reality. We described ARC as a future version of the emergency call system, where the dispatcher can, whenever appropriate, be telepresent with a first responder, and get access to the scene of the incident.

We could also envision a scenario where a service similar to ARC could play out in a world of SoIC-TV s. A smart home may detect a medical emergency with one of its inhabitants, causing it to evaluate the best option for a first response. Besides calling the ambulance, a nearby SoIC-TV *driver* is called and connected with a suitable remote dispatcher as *passenger*. This could be a trained first responder, or even a machine that is capable of guiding the *driver* through the protocol for responding to the given emergency[1].

This scenario extends the SoIC-TV with a new aspect: So far, we envisioned the *passenger* to sit inside the figurative vehicle of the *driver*, giving instructions to the *driver* from the inside. Here, instead, the *passenger* (i.e. the emergency dispatcher) is stepping out of the vehicle, being virtually embodied in front of the *driver*, in the same shared space, to give directions and instructions that relate to that environment, such as the hand placement on the chest of a person receiving CPR. Similarly, the *passenger* now has access to the *driver* from an outside perspective, allowing to better

[1] We can only hope that the SoIC-TV *driver* does not need to wear a TV-sized system on their head when having to apply rescue breaths.

observe the performance and behavior of the *driver* in a certain task and adapt their instructions accordingly.

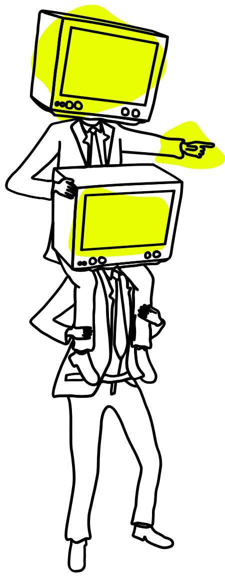
If we extrapolate the findings in this part of the thesis to the SoIC-TV, it is focussed on the social interaction between *driver* and *passenger* when engaging in the *remote expert, local novice* scenarios described in Section 5.1. These typically do not foresee any third interlocutors as part of the interaction. In the next part, we will look at interaction scenarios where a third interlocutor is involved. To summarize it briefly in terms of the SoIC-TV universe: A SoIC-TV *driver* is tasked to interact with a bystander, while a virtual *passenger* dictates the behavior and speech of the *driver*, unbeknownst to the bystander.

PART *IV*

THE VIRTUAL HOMUNCULUS: MULTIMODAL CYRANOIDS

11

EMBODIMENT & MIND



In the 1897 play *Cyrano de Bergerac* by Edmond Rostand, Cyrano, the main character, tries to woo his cousin Roxane. Cyrano however lacks in confidence due to his looks. He learns that Roxane fancies his colleague, the handsome but dim-witted Christian. Christian reciprocates, but requires help with wooing Roxane on his own, lacking the eloquence. The more educated and worldy Cyrano offers to help by writing letters for Christian, and whispers the right words into Christian's ears from a balcony while Christian is on a date with Roxane. Cyrano does the thinking, but Christian has got the looks. In the play, Roxane doesn't really figure this out until years later when Cyrano reads out a letter to her - supposedly from Christian - on his deathbed (he got hit on the head by a log that fell from a building). Why did Roxane not figure this out earlier? Taking a philosophical approach, from the experience of being human, we experience the world from the inside of our body. This experience will lead us to conclude that this is the same for others we communicate with. We believe the source of a thought communicated to us in a conversation is the mind of our conversation partner, which is somehow located in their body or head. Body and mind seem inseparable. However, as we have seen in the previous experiments, and as we have seen in Cyrano's story, interactions can be mediated – by technology or through whispering – and the source of a thought may actually be located elsewhere.

In the study reported on in Chapter 3, we used the Wizard of Oz approach to have a humanoid virtual agent conduct an interview with a participant. Although we suggested to the participants that this virtual agent was autonomous, the *mind* of the agent was, in part, located with the experimenter, who selects the utterances for the agent to say, and in part located inside the machine that was programmed to render the virtual character and to implement certain rules, for example for the gaze behavior.

In the previous chapter we have discussed social interactions in mixed reality, with other humans rather than with virtual agents. These experiences were strictly designed to create a "place illusion", that is, for a remote visitor to feel as if they were co-present with a local user. Arguably however, the embodiment shown to the person was a virtual one, reconstructed from the depth sensors on the remote location, thus part of the place illusion to make the social interaction work was for the remote visitor to accept that the rendered body's original source is coherent with a real, physical person on the other end, that they are interacting with in real time.

Technology-mediated or -simulated social interactions like the ones described in the previous chapters of this thesis show how elastic the coupling between a virtual embodiment and the source that controls it can be, without affecting users' ability to sustain social interactions.

One instance where this elasticity may have been overstretched was in the study on our system for remote assistance on Augmented Reality CPR in Chapter 8. Co-presence scores reported by the first responders were not significantly higher when compared to the condition without a virtual embodiment. This lead us to speculate that the main part of the dispatcher's virtual representation shown to the dispatcher, the virtual hands, may not have been (strongly) attributed to the dispatcher, but instead to other parts of the system. Of course there could also be another explanation: the hands could be neither attributed to the remote dispatcher nor the system, but to a third, albeit never introduced person.

We would intuitively say that this is an unlikely explanation. One would not expect some third, unknown agent to be involved without being given some reason. Now, in this chapter, we are explicitly looking at a scenario where there is such a third, unknown and never introduced entity, that is the true source of actions perceived through an embodiment, similar to what is illustrated in *Cyrano de Bergerac*.

11.1 A DYSTOPIAN SCENARIO

To give an example of this in a technology mediated interaction context, consider the SoIC-TV scenario that we sketched in the introduction of this Thesis: A physical body, the *driver*, offers to embody the mind of the remote person. While, to the bystander, it may be more or less obvious in the SoIC-TV case (with a TV on the *driver's* head), we can conceive technology that does this in a more subtle way - as we will see, an earpiece could be enough.

Now, what happens when sources of mind and the embodiment are mixed in social interactions? What if a virtual agent is controlled half by an algorithm, half acts on behalf of another human? Our impressions and beliefs of others are being formed from first impressions, based on the other's appearance, as well as on their responses. While one might be able to probe whether a virtual human is controlled by a machine or by another human if one sets out to do so, this might be more difficult the smarter machines become and the more realistic virtual or robotic agents start to look and move.

What about the decisively dystopian case where a human is employed to act and speak on behalf of a machine? Here, first impressions are on the side of the machine, as it appears like a human. How much of an effort does the machine need to make to keep up the illusion for others that they are interacting with a human instead?

In the remainder of this Chapter[1], we will discuss so-called *cyranic interfaces* that were used explore these questions. Specifically the apparatus called *Echoborg*, used to embody text-only chatbots through humans. This work inspired us to extrapolate this method to the domain of Embodied Conversational Agents (ECAs). We discuss similar methodologies from this domain where the true source of an ECA's actions are manipulated under deception of a participant for experimental reasons, and we discuss how the appearance and embodiment impact the perception of actors in social interactions.

In the next chapter, we will first present our own system that allows a human to serve as embodiment for multimodal ECAs. In Chapter 13, we will then present a first user study conducted with this system, exploring how the same ECA is perceived when embodied either by a virtual or real human.

[1]Based on our paper S. Falcone, J. Kolkmeier, et al. (Sept. 2022). "The multimodal EchoBorg: not as smart as it looks". In: *Journal on Multimodal User Interfaces* 16.3, pages 293–302. ISSN: 1783-8738. DOI: 10.1007/s12193-022-00389-z. URL: https://doi.org/10.1007/s12193-022-00389-z

11.2 CYRANOIDS & ECHOBORGS

Inspired by *Cyrano de Bergerac*, infamous social psychologist Milgram explored the idea of *cyranic illusions*, when a human - a *cyranoid* - is used as a mere proxy to relay or *shadow* speech from another human, the *speech source*, using a radio ear-piece. Those talking to the cyranoid, the *interactants*, would believe that they were interacting with the human in front of them, while they were truly getting the responses from the invisible speech source. Milgram was interested in how large the incongruency between the identity of the speech source and the cyranoid could be, while remaining undetected by the interactants, speculating that the larger the incongruency, the sooner interactants would detect that something was off. However, preliminary experiments showed that interactants are able to explain away perceived incongruencies. One prominent case that Milgram discussed was that of a child shadowing the responses of an invisible adult - as an example of two highly incongruent identities - while interacting with other adults. The adults did not suspect anything, but rather integrated the incongruency between the appearance of the child and the way it would talk into a plausible truth by attributing high intelligence and maturity to the child (some speculated the child would turn out a prodigy).

In the more recent work by Corti et al., this method is extended to not have cyranoids shadow the speech of another human, but rather to shadow the speech of chatbots - which in turn would receive the cyranoid's conversation partners (i.e. the participants) transcribed speech as input (Corti and Gillespie, 2015, 2016). This special case of a cyranoid is called *Echoborg*. Just like with cyranoids, a person who engages in interaction with an Echoborg has no reason to believe they would be interacting with anybody (or anything) except for the person they see in front of them.

In one experiment, Corti and Gillespie had participants engage in 10-minute interactions with different chatbot engines, either using the text-based interface or (under deception) through the Echoborg, and evaluate their conversation partner and the interaction afterwards. Results were that while participants described their conversation partners as *mechanical* and *robotic* after using the textual interface, the same conversation experienced through an Echoborg caused participants to describe their conversation partners with human characteristics, such as *shy*, *awkward* and *autistic* (Corti and Gillespie, 2015). Participants' ratings on how comfortable they felt during the interaction showed that interactions with the Echoborg was rated significantly less comfortable than with the textual interface for two of the three types of chatbot engines.

Corti and Gillespie draw two interesting conclusions from this. For one, the different groups of adjectives used between the different interfaces (textual/screen-based and Echoborg) demonstrate how the impressions and judgements of interaction partners is primarily colored by what participants saw in front of them, rather than by what the content of the interaction was. More technical adjectives were used for the textual interface, and the Echoborg was described with human-like characteristics - "*Participants who spoke with an echoborg based their personality judgements to what they saw: a human person sitting directly in front of them. Participants who spoke with a text interface based their personality judgements to what they saw: a computer screen*" (Corti and Gillespie, 2015). Second, the observation that participants feel less comfortable with the same conversation through an Echoborg than through the textual interface further underlines that "*face-to-face, in-the-flesh interactions place much higher intersubjective demands on the parties to an encounter*" (Corti and Gillespie, 2015).

In Milgram's experiment, participants explained away children's ability to talk like adults by assuming the child was gifted. Here, unfitting and delayed responses by the Echoborg are explained away by participants as the person being shy and awkward. So while these incongruencies still affect the perception, the interesting aspect of these methods is that judgements by participants are based on what is in front of them.

Corti proposes to make use of this in research and development of intelligent agents. The Echoborg methodology could be used as a benchmark condition of embodied agents, used for understanding how social perception is related to the interface or form of the embodiment, as well as forward-looking studies where the Echoborgs take the role of embodied agents for tasks that these agents are not yet equipped for (such as object manipulation) while following the protocols generated by a machine.

There is one aspect in which Corti's Echoborgs fall short for them to be useful in the context of embodied agent research: Echoborgs only shadow speech, not other modalities. Meanwhile, for embodied agent research, non-verbal behaviors such as gestures, facial expressions and gaze behaviors receive as much attention as the speech modality. So an Echoborg system that allows for shadowing additional modalities could be a valuable contribution, as it would allow for more experimental control.

In the remainder of this chapter, we will explore the potential role for Echoborgs in HCI research. First, we will provide more grounding and background on evaluation methods in HCI research, in particular for embodied conversational agent (ECA) research, that is, involving virtual or social robots to mediate social interaction. We will also provide the background on the role of the appearance of agent embodiment and perception in social interactions with embodied agents and avatars. Then we will present our prototype apparatus for an advanced Echoborg system, that also features multimodal behavior shadowing. We will finally report on a first user study performed with this apparatus, in which we administer instruments from the aforementioned field of HCI research, to get a first idea on how users perceive our

Echoborg, and the quality of the interaction with it.

11.3 BACKGROUND

In cyranoid and Echoborg experiments, the role of the target embodiment is to be deceptive about the true source of its actions. We will first look at the background from the ECA research domain, where the true source of an ECA's actions are manipulated under deception of a participant for experimental reasons (Section 11.3.1). In the case of cyranoids and echoborgs, this deception works as interactants will form first impressions and expectations based on the appearance of an embodiment. So in Section 11.3.2, we will look in more detail at the relationship between the appearance of an embodiment and the perceived mind and personality of these embodiments, as well as their effect on how social interactions are perceived.

11.3.1 DECEPTION METHODS IN ECA RESEARCH

Non-technical evaluation of ECA research contributions may involve users engaging in an interaction with the ECA system, often under a deception setting where the true source of some actions is not the system, but rather a human.

A popular method is *Wizard-of-Oz* (WOZ). If we are interested in the effect of a certain behavior on the perception of the ECA, the researcher may - typically under deception of the participant - take over the control of that behavior. For example, in Ter Maat, Truong, and Heylen (2010), researchers take over the utterance selections - following a strict experimental protocol - and act out different turn-taking strategies to investigate how these affect perception of the agent personality-related attributes.

Part of a modern ECA is also the ability to automatically sense certain behaviors in the user to adapt its own behavior. This may include the automatic recognition of speech, estimation of the user's emotional state based on features detected in speech or facial expressions. Often, the automated sensing capabilities available to perceive the behavior of the human conversation partner may not be sufficiently implemented or reliable. Here researchers may also act as wizards by not taking over the behaviour of the agent, but rather replace these sensing capabilities, manually indicating to the system information of the user's behavior. In the simplest case, this could simply be a single button to indicate to the ECA system when the participant starts and stops speaking such as described in Reidsma, Kok, et al. (2011).

More advanced versions of the WoZ allow for new possibilities in terms of the research design. The Switching Wizard-of-Oz method (SWOZ, Poppe, Maat, and Heylen, 2012) has participants interact with an ECA, while in the background a wizard, observing the behavior of another conversation partner or confederate controls the behavior of the ECA such that it reflects the behavior of the conversation partner or confederate. What's more, the SWOZ method can be used, on the fly, to switch between system generated behavior and wizarded behavior without the participant noticing. For evaluation purposes, participants can be instructed to indicate when the behavior of the ECA seems off. In Poppe, Maat, and Heylen (2012) for example, the confederate was instructed to simply listen to the participant as if they were sitting in front of them. Then, the wizard would observe whenever the confederate would exhibit a backchannel (nods), and in turn would trigger a backchannel animation on the ECA.

As we can see, with the WoZ approach the focus is on evaluating or at least complementing certain behaviors when presented on an ECA. As illustrated in the introduction, the Echoborg method allows us to replace the embodiment of the virtual agent with an actual human, keeping the backend of the ECA the same. In the next subsection we discuss in more detail the relationship between the shape and medium of an embodiment and how it affects how it and interaction with it is perceived.

11.3.2 AGENCY BELIEFS & IMPRESSION FORMING

When engaging in social interaction with others, humans quickly form impressions on the others' skills (Ambady and Rosenthal, 1993), personality and attitudes towards others (Argyle, 1988; Campbell and Rushton, 1978; Levesque and Kenny, 1993). These impressions can be formed very quickly based on a few seconds observing the other's non-verbal behavior such as facial expressions and gestures (Ambady and Rosenthal, 1993), but also based on static attributes such as clothes worn in photographs (Naumann et al., 2009).

Effects of virtual agent behavior on perception of agent personality and interpersonal attitudes have been investigated in perceptual studies; such as effects of properties of gestures (Neff, Toothman, et al., 2011; Smith and Neff, 2017) with language (Neff, Wang, et al., 2010) on personalty, posture (Normoyle et al., 2013) on emotion, as well as studies focussing on impression forming during first encounters with virtual characters (Cafaro, Vilhjálmsson, Bickmore, Heylen, Jóhannsdóttir, et al., 2012).

Not just appearance, but also shape and medium affect how humans perceive their interaction partners. For example, in Bailenson, Yee, et al. (2006), dyads engaged in technology mediated interactions at various levels of *behavioural and form realism*[2], that affected the level of perceived co-presence but also affected the level of self-disclosure.

[2] voice only, video conference, and simple virtual polygon-avatars

The conditions in Bailenson, Yee, et al. (2006) only include virtual and mediated interactions. An interesting question in our context however is whether there is any measurable effect on the agent being physically embodied or not. In a more recent survey, Li discusses works that investigate how the experience of interacting with phyically copresent social robots, telepresence robots and virtual agents (Li, 2015), concluding that *"robots were more persuasive and perceived more positively when physically present in a user's environment than when digitally-displayed on a screen either as a video feed of the same robot or as a virtual character analog"*.

A growing body of work that investigates how social (game) experiences are affected by beliefs about the agency of other players, and whether they are physically co-present consistently finds that the mere belief that another player is human (positively) affects various aspects of the experience, both in self-report (Gajadhar, De Kort, and IJsselsteijn, 2008; Lim and Reeves, 2010; Weibel et al., 2008) (i.e. increased presence and enjoyment) and physiological measures (Kätsyri et al., 2013; Lim and Reeves, 2010; Ravaja, 2009). For example, Kätsyri et al. (2013) found that in a first-person video game, the outcome of the game causes reward-areas of the brain to respond differently, depending on the belief about whether the opponent was controlled by a human or by a computer.

While these last works may be describing unconscious effects, others have looked at how visible interaction outcomes, i.e. how agents are treated, as benchmarks or measures for agent behavior. Melo and Gratch propose to benchmark agents based on whether people, "in a specific social situation, [...] act with the virtual agent in the same manner as they would with a real human" (Melo and Gratch, 2015), something that requires agents to be perceived as featuring higher levels of *mind* (ability to act and plan) and *experience* (ability to feel emotion). The authors raise that prior beliefs about humans' and virtual agents' levels of mind and experience will differ. Humans are treated more favorably in most contexts by default, while virtual agents need to make an extra effort of displaying behavioral strategies to actively demonstrate these capabilities to sway the users' perception, and treat the agent as if they were human.

While Melo and Gratch do not consider the special case of human-embodied machines in their work, another experiment by Corti and Gillespie evaluated Echoborgs according to a benchmark in the same spirit. They looked at how users treat and interact with the agent, rather than relying on self-report of agent perception. Specifically, they measured the amount of *repair behaviors* initiated by participants during conversations. These are conversational turns where the speaker indicates some need for clarification, that is, in references to a *trouble source* in the previous utterance by another speaker. Such repair activities are important to establish and maintain common ground (Clark and Brennan, 1991). This was measured under the conditions of the chatbot interacting through the text interface, or when embodied by an

Echoborg. As an additional factor, they varied whether participants were informed about the respective condition, that is, whether they knew that the Echoborg was only talking on behalf of the chatbot. Again, the results show that appearance and belief about agency interact with how interlocutors are treated: the likelihood for offering certain types of repair behaviors was highest when the agent was embodied by a human in the covert condition, that is, when the participant was unaware that they were talking to an agent. Meanwhile, in conditions where the same conversation was had using a text interface, less effort to offer repair was made (Corti and Gillespie, 2016).

The chatbot in Corti and Gillespie's study typically was not able to respond to the initiated repair behaviors in a meaningful way, so while humans may have offered to initiate these bilateral interactions, they would not be implemented by the agent. Meanwhile, such behaviors with bilateral connection that require constant exchange between agent and interactant are what, for example, Melo and Gratch refer to as the strategies an agent should employ to demonstrate their capacities of mind and experience.

An interesting example of such bilateral connection of behavior is demonstrated by Bevacqua et al., who equip virtual characters with the ability to show different levels of behavior *coupling* in their body movements, along with an instrument to measure the perceived levels of coupling via self-report (Bevacqua et al., 2014). The coupling construct represents the continuous perceived mutual influence between participant and agent. In other words, the perceived level of mutual flow. Other included constructs relate to *co-presence*, the feeling of the other *being with* the participant, to the *engagement* elicited, and the *believability*, that is, the extent to which the agent is perceived as human-like. Part of their findings was that engagement and co-present scores correlate with the perceived level of coupling. They recommend that agent cognitive architectures feature the ability to show coupling behavior as a means to improve the sense of co-presence with the agent.

11.4 CONCLUSION

In this chapter, besides discussing the methodology and prior works on Echoborgs, we discussed that there are already similar deception-based methods common in ECA research. These, however, are usually in the configurations where human experimenters dictate, unbeknownst to participants, some or all of the behaviors of the embodied agent, such as a social robot or virtual human. We also discussed some research on the role of participants' belief in what or who controls a given artificial embodiment in how interactions are experienced, as well as the role of the shape and appearance of the embodiment

Corti and Gillespie propose to explore the Echoborg method further in artificial agent research - the advantage that they see with evaluating machine generated conversation through an Echoborg is that this way, biases that humans have when interacting with machines are removed, and these machine generated behaviors can now be evaluated in the context of human-human interaction. In particular, we are curious how interactions play out when comparing virtual embodied conversational agents with those embodied by humans using the Echoborg method.

Beyond the lab setting, Echoborg-like systems could also help to answer questions raised by our science-fiction scenario of the SoIC-TV scenario we have introduced in Section 1.1.3. Imagining a *driver* can relay thoughts and messages from remote users to others without those people realizing, such as in the Milgram's experiments? What happens when the source is not human but a machine, such as a human-embodied *Google Assistant*? Will it still be perceived machine-like? Will interlocutors be more willing to comply with requests?

Corti and Gillespie answer some of these questions in their experiments, for the case of chatbots being interacted with through either text interfaces or humans acting as Echoborgs. As mentioned however, the Echoborg setup is limited to verbal behaviors only, and so far, their work has only been used to compare textual interfaces with humanoid embodiments.

To investigate the effects of virtual, physical embodiments, that is, doing comparisons with virtual and humanoid embodiments, and to investigate the effects of machine generated behaviors under a human-human interaction, we can advance a number of aspects in terms of an Echoborg-like apparatus. Even though works from ECA and VA research discussed in Section 11.3.1 were not intended to exhaustively cover the various modalities that are being investigated on ECAs, we have seen that they encompass much more than just the modality of speech: turn-taking, backchannels, gaze, gestures and facial expressions are all being implemented and evaluated on ECAs. In the following chapter we describe an apparatus that allows us to create an Echoborg-like system that allows the shadowing of both verbal and nonverbal behaviors.

12

A CYRANIC INTERFACE FOR MULTIMODAL ECHOBORGS

In this Chapter[1], we will describe our first version of an Echoborg-like apparatus that allows a trained user to shadow behaviours generated by our Embodied Conversational Agent (ECA) system. As discussed in the previous chapter, the Echoborg-like apparatuses existing so far focus exclusively on shadowing speech behaviors, typically through an earpiece. This makes sense when working with chatbots, which respond exclusively in terms of text messages.

However, as discussed in Section 11.3.1, modern embodied conversational agent research is interested in more modalities than just speech, and virtual agent (VA) embodiments are increasingly capable of more realistic behaviors in terms of their animation and rendering. In line with this, we find it crucial to also investigate whether we can have an Echoborg also shadow behaviors in those other modalities that are being produced by modern ECA systems. The apparatus that we describe in this chapter therefore is called the *Multimodal Echoborg* (MEB).

Another aspect that we will pay attention to for this apparatus are the experimental conditions we seek to support. In the prior works we discussed in the previous chapter, the comparisons were between conversations through textual interfaces for chatbots, and conversations with the same chatbot through an Echoborg apparatus. Instead, we are interested in comparisons between two different embodiments; ECAs embodied through their typical VA, versus the same ECAs embodied by a human using the MEB apparatus.

In the following we will first present our existing ECA platform and then how we extended it to feature a MEB embodiment. This will serve as our apparatus for the experiment presented in the next chapter.

---

[1]Based on our paper S. Falcone, J. Kolkmeier, et al. (Sept. 2022). "The multimodal EchoBorg: not as smart as it looks". In: *Journal on Multimodal User Interfaces* 16.3, pages 293–302. ISSN: 1783-8738. DOI: 10.1007/s12193-022-00389-z. URL: https://doi.org/10.1007/s12193-022-00389-z

12.1 THE EXISTING PLATFORM

Our system is built on a SAIBA (Vilhjálmsson et al., 2007) compliant base setup which has been successfully employed in a number of recent projects and research studies on ECAs. SAIBA is a reference architecture for modular virtual agent systems. The modules we employ are the ASAP (Articulated Social Agent Platform) behaviour planner and realizer (Van Welbergen, Reidsma, and Kopp, 2012), the Flipper Dialogue Manager (Maat and Heylen, 2011; Waterschoot et al., 2018) and the Dialogue Game Execution Platform (DGEP, Lawrence et al., 2017). For rendering, we employ our ASAP Unity bridge (Kolkmeier, Bruijnes, Reidsma, and Heylen, 2017), which allows us to render our virtual agents on a wide range of target platforms, also including VR and AR hardware (Kolkmeier, Bruijnes, and Reidsma, 2017).

The *Flipper* dialogue manager acts as a flexible state machine that allows the control and exchange between the other modules based on the state of the dialogue. In DGEP, dialogues are modelled in terms of possible moves in a given state with a given outcome. A higher level planning component can then select the best move to reach a certain goal for the given actor. Note that all participants in the dialogue, agents and human users alike, are modelled in the same way. User's dialogue-related utterances are mapped to their possible dialogue moves either by the experimenter in a WoZ fashion, or by a separate sensing module.

Based on the state of the dialogue, Flipper will plan different behaviors for the agents, such as listening behavior, or the execution of the next selected dialogue move. Dialogue moves typically encode only functional information (intents) (think of "greet user", or "say 'hello' in a friendly way", ideally formalized using the Functional Markup Language, FML (Cafaro, Vilhjálmsson, Bickmore, Heylen, and Pelachaud, 2014; Heylen et al., 2008)), which are then translated by third components into plans for actual surface behavior. These third components typically include NLG (Natural Language Generation) and NVBG (Non-verbal Behavior Generation) components, that build up the text to be spoken (if any) and the gestures and facial expressions to accompany it, these behavior plans are encoded in behavior markup language (BML) and sent to ASAP for realization. Especially the step from translating a formal functional description to planned surface behavior description can be complex and implementation often very project specific, and is out of the scope for this discussion. A diagram of the resulting system is shown in Figure 12.1. The design of the individual components for the cyranic inteface is discussed in the next section.

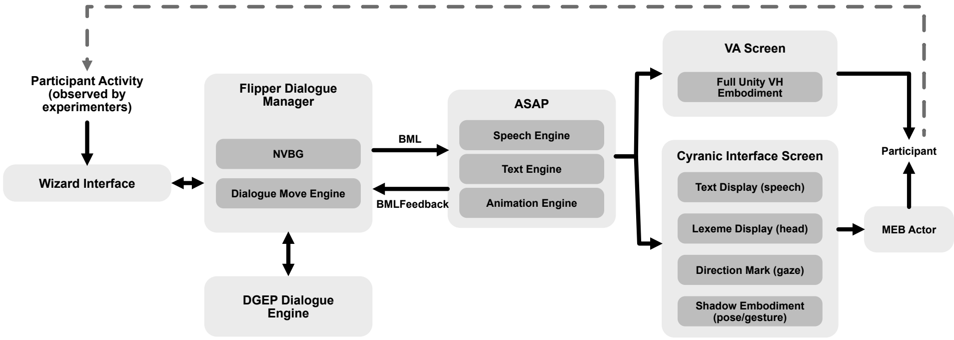


Figure 12.1. System Diagram of the MEB setup, showing both the regular VA display and display through a MEB using the Cyranic Interface.

12.2 EXPLORING ECA SHADOWING APPROACHES

One advantage of working with a modular agent architecture is that the different modules should be easily replaceable or extendable. This is true in particular for the ASAP realizer, where the behavior realization is composed of a number of engines to which instructions from the different modalities specified in the behavior markup can be routed (Reidsma and Welbergen, 2013). Previously, for example, we have used this to create hybrid robotic and virtual embodiments (Linssen and Theune, 2017). Following this approach, we explored some options as to how we can present the to-be-shadowed behavior across the various relevant modalities to the person acting as MEB. Finally, we will describe the resulting apparatus used for the remainder of this work.

12.2.1 SPEECH

In principle, to implement the classic cyranic interface requires no work at all: we can stream the speech synthesized by our Text-To-Speech (TTS) engine to an earpiece worn by the MEB. However, we have also explored alternative implementations for speech shadowing, such as a teleprompter-like setup. Here, we employed a second engine that would extract the surface text to be realized by the speech engine, streaming it to a remote client where it would be rendered on a screen. This allows us both to display the entire sentence at once, and incrementally spell out (or color in) the words to be spoken at the speed at which the TTS engine would have produced the speech. In this prototype of the teleprompter-like interface, additional speech markup was lost, but could potentially be added back by translating it to text markup. For example, when the TTS engine is instructed to speak louder or stress particular words, we could increase text size or change the font weight of certain words.

12.2.2 GESTURE

We considered a similar solution for gestures, where we would display a textual indication for the type of gestures. The gestures our virtual characters are able to display are typically coded by lexemes such as `refuse` or `deictic_self`. However, these lexemes are not always equally human-readable, descriptive or helpful in terms of their exact version. For example, it would not be obvious what the difference is between `beat1` and `beat2`, as at this stage, where the behavior may be planned, it is not clear how it will be realized by the responsible engine and embodiment. Alternatively, we could look to display the functional description that was obtained in the planning stage. While this may be more human-readable (i.e. *greet user*), it circumvents both the planning and realization stage. Both of these approaches leave, not just in terms of our architecture, but quite literally, both the planning and *realization* of the gestures up to the MEB.

Another option that we explored was to show the realized gestures animated on a character for the MEB to shadow. This approach makes it possible to shadow overall style, size and dynamics of the gestures . Here, of course, the disadvantage is that both the functional and semantic meaning may get lost (or may not be immediately evident to the MEB).

12.2.3 BACKCHANNELS

We can consider backchannels to be a subset of gestures. However, we elected to treat them separately, as they 1) mainly occur while listening, without any other (hand-) gestures at the same time, 2) because the set of head gestures that we currently employ for backchannels is limited to nodding. Instead, we elected to route the realization of *nod* behaviors to a discrete signal emitter that toggles the word *Nod* on the teleprompter display mentioned earlier. Alternatively, we considered having this signal trigger physical actuators, such as a small light in the field of view of the MEB or a tactile actuator that is placed in their neck.

12.2.4 GAZE

While gaze behavior is certainly one of the most complex human behaviors, for this present work we decided to limit gaze to the current gaze target at a given moment. As will be discussed later, the interaction in the user study would involve three interlocutors, and the simplified model for social gaze that we employ simply determines which of the other two interlocutors' faces will be looked at and for how long (before alternating to the only other possible target). We do not model averted gaze or environment gaze targets. To achieve this, we added another discrete element to the screen that we are now already using for the teleprompter function and discrete nod signals. This gaze element indicates to the MEB to either gaze at the dialogue parter to their left or to their right.

Of course, this solution works for our three-party dialogue setting, where no other gaze targets are required. However, for gaze in particular, we could envision more powerful cyranic interfaces in the future, using unobtrusive AR glasses that overlay instructions as to where (and how) to look into the AR overlay, potentially even enabling saccades and similar micro behaviors that constitute to gaze.

12.3 RESULTING APPARATUS FOR MULTIMODAL ECHOBORGS

Together with the confederate that would later act as our MEB, we explored some combinations for shadowing the different modalities as discussed above. We limited our work to these modalities, and not included facial expressions, touch and other modalities.

We found that shadowing speech, especially if the utterances are somewhat predictable, is easiest when read from a teleprompter-style interface. Similarly, for gestures, shadowing from a virtual puppet was easier than interpreting the higher-level functional description. Signals for giving backchannels (Nods) was easily shadowed from the lexeme-style interface (i.e. words popping up on the screen for the style of nod to be performed). Given the simple level of gaze behavior to be shadowed, the interface designed was found to be sufficient.

All information was shown on a single screen. In experiment settings, to be able to keep participants blind to the nature of the MEB required us to position this display behind the participant (see Figure 12.2). Additionally, we implemented an automatic procedure that would disable the screen unless there was an ongoing interaction (i.e. based on the presence of signals on the BML progress and BML feedback channel). This was to prevent the participant from seeing any part of the interface while they were not seated and facing away from the screen, that is, while entering or leaving the room.

Figure 12.2. This is a frame from the video recording from the user study, showing the cyranic interface on the screen behind the male participant. On the left is our confederate acting as the MEB. Visible in the center of the screen is the character used for gesture shadowing. The gaze direction indicator is currently activated on the right side of the screen, telling the actress to look at the debater sitting to her right side (not shown here). Speech is being shadowed from the teleprompter text (on the top of the screen). For backchannels, the letters 'NOD' would appear on the screen briefly (not shown here).

13

EVALUATING THE MEB

In this chapter[1] we will discuss the user study that we conducted with the apparatus for multimodal behavior shadowing for our Embodied Conversational Agent (ECA) platform described in the previous chapter. The goals were to evaluate our technical implementation, and to probe how participants perceive an interaction with our Multimodal EchoBorg (MEB), compared to the same interaction with a virtual embodiment of the ECA. We also look to form a first impression on how feasible the MEB methodology is for future ECA research.

The dialogue scenario that participants would engage in during the experiment was based on a previously recorded corpus of an ethical debate with two human debaters and a human moderator. The dialogue utterances as well as ECA behaviors were informed on what human debaters in this corpus did. We will describe in more detail the dialogue scenario, corpus recordings and the modelling of ECA behaviors in the next three sections.

In the user study we had the participant (acting as the moderator) lead a discussion between two debaters controlled by our dialogue system. Between subjects, we would replace one of the debaters with a trained MEB that would shadow the behaviors using our MEB interface. In a post-experiment questionnaire we included self-report measures on the overall experience of the interaction, as well as on a number of constructs of how participants would perceive the two debaters. For additional insights into how the experimental conditions relate to a comparable human-only interaction, we had also administered some of these measures to the participants of the human-human interaction corpus. The experiment design will be explained in more details in Section 13.2

[1] Based on our paper S. Falcone, J. Kolkmeier, et al. (Sept. 2022). "The multimodal EchoBorg: not as smart as it looks". In: *Journal on Multimodal User Interfaces* 16.3, pages 293–302. ISSN: 1783-8738. DOI: 10.1007/s12193-022-00389-z. URL: https://doi.org/10.1007/s12193-022-00389-z

13.1 DIALOGUE SCENARIO

ECA systems typically operate in a predefined context and do not typically afford open-ended dialogues such as the chatbot evaluated in the previous work by Gillespie and Corti (2016). We decided to create a simple dialogue scenario for these purposes, which we will describe first before going into detail on the experimental design. Since this present work was designed to be exploratory and not suited for a high-powered experiment, one requirement for this scenario was that it would be multi-party to allow for a potential within-subject design.

We elected to model an ethical-debate-like scenario, with a moderator and two opposing debaters, discussing different variations of the *Trolley Dilemma* (Thomson, 1976). The role of the moderator was important to give structure to the dialogue, and it would be a good role for a participant to enact, as it allows for fair post-hoc comparisons in a within-subject design. What's more, a moderator has comparatively predictable dialogue moves, allowing for a controlled flow of the experiment.

13.1.1 HUMAN-HUMAN-HUMAN CORPUS

In order to be able to model this scenario sufficiently and to inform the design of utterances and gestures to be used by our ECA, we first recorded a small corpus of three humans (HHH) engaging in such a debate. We recorded audio, video, and we transcribed the dialogues. In total, we recorded 6 triads (2 females, 16 males). Moreover, we administered some of the same questionnaires regarding the interaction experience and interlocutor perception that we planned to use in the user study later. Even though these interactions would be different from the interactions with a dialogue system, they may be interesting to compare in an exploratory fashion regardless.

13.1.2 MODELLING OF ECA

Using the dialogue argumentation framework, we created a simple dialogue game. We structured the arguments observed in the corpus according to topic (see Table 13.1). For each of these arguments, we selected verbatim responses used by debaters in the corpus data. These were to be used as surface text for our agents to speak out. To not favor one debater over the other based on how convincing the respective arguments were, we had the arguments ranked on their strength in a brief online study, and selected similarly ranked utterances for each topic for both the proponent and opponent..

Key Argument	Moral question
Fate	Can fate decide for the life of human beings?
Numeric	Human life is a qualitative or quantitative matter?
Economic	Is it better to save more lives because they are a greater resource for society?
Responsibility	If we make the choice of pulling the lever, do we become responsible for murder?
Inaction	Can 'inaction' be considered as 'action'?

Table 13.1. The key *Trolley Dilemma* arguments that we have identified based on the HHH corpus, and the moral questions which describe them.

For the non-verbal behavior of the agents, we consulted the video recordings from the corpus of the selected utterances and presented them to an actor. The actor would then act out these utterances wearing a motion capture suit. This yielded full-body gesture animations for each utterance. The motion capture recordings and selected arguments were then combined into behaviors[2], and linked to the corresponding dialogue move in the dialogue manager (the *Dialogue Move Engine* component in Figure 12.1). We also implemented a simple rule-based behavior tree for backchannels and gaze while listening (the *NVBG* component in Figure 12.1).

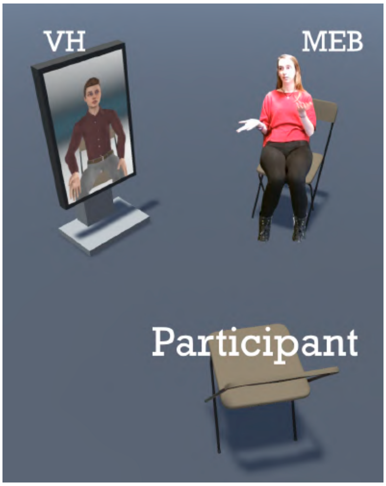
13.2 RESEARCH DESIGN

Participants were assigned to one of two conditions: *Human-Agent-Agent* (HAA) or *Human-Echoborg-Agent* (HEA).

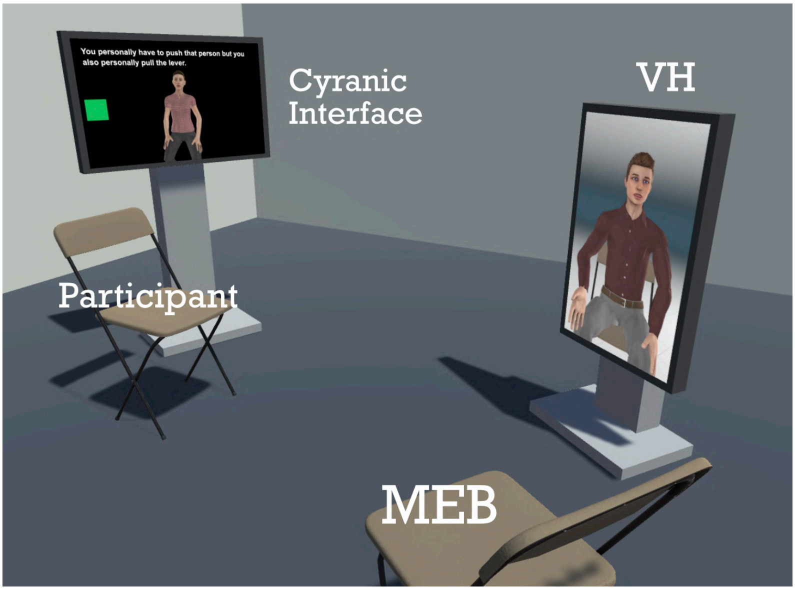
Participants were always assigned to the role of the moderator, while the debaters (*proponent* and *opponent*) were always acted out by our ECAs. Just to emphasize, when referring to ECA(s), we include both the MEB and the VA, as the same ECA but with different embodiments. In both HAA and HEA, the opponent ECA was always embodied by the virtual agent (VA). In HEA, the proponent ECA was embodied by the MEB, while in HAA, the proponent ECA was also embodied by a VA. We call this between-subject variable *proponent embodiment*. For those participants assigned to the HEA condition it is also interesting to compare their ratings of the VA embodiment of the opponent versus the MEB proponent embodiment. This is a within-subject variable which we refer to as *debater embodiment*.

[2]In the form of BML messages comprised of the motion captured animation as gesture behavior, and the verbatim argument text as speech behavior

(a) HAA



(b) HEA



(c) View of the MEB (in HEA)

Figure 13.1. 3D illustration of the debater placement in (a) HAA and (b) HEA conditions. The MEB view in (c) shows how the screen with the cyranic interface was positioned, behind the participant.

13.2.1 PARTICIPANTS

In total, 36 participants were sampled from the university staff and student body, between 19 and 46 years old, 16 females and 20 males, and the number of participants was equally distributed between conditions. The user study was approved by an ethics board.

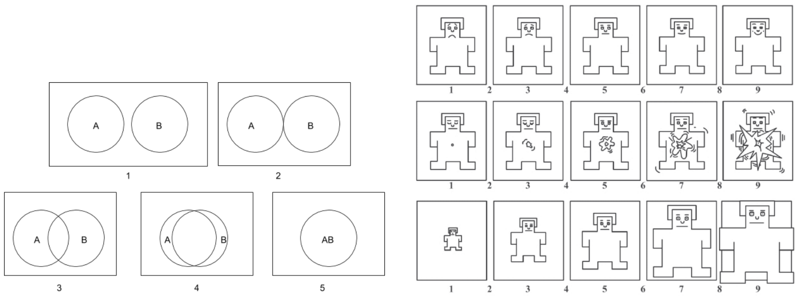
13.2.2 MEASURES

The main measures in this study were administered in a post-experiment questionnaire. To quantify perception of appearance and agency we included the Godspeed Questionnaire Series (GQS, Bartneck et al., 2009), as it covers both aspects of the appearance that are relevant to us, such as *anthropomorphism* and *animacy*, as well as constructs on how *intelligent* and *likeable* participants felt the debaters were. This questionnaire uses semantic differentials for its constructs. That is, participants were asked to rate their impression of the target on a scale from 1 to 5, with the extreme ends labeled with differential adjective pairs. Pairs for *animacy* include "dead"–"alive" and "artificial"–"lifelike". Pairs for *anthropomorphism* include "unconscious"–"conscious" and "moving elegantly"–"moving rigidly". Pairs for *likability* include "unkind"–"kind" and "unpleasant"–"pleasant". Pairs for *perceived intelligence* include "ignorant"–"knowledgeable" and "foolish"–"sensible".

We added measures on trust and closeness (using the non-verbal scales described in Figure 13.2a) to look at what relationship developed between the participant and the respective debater during the interaction.

In addition, we included the questionnaire used by (Bevacqua et al., 2014) that includes questions on constructs related to *coupling*, the continuous perceived mutual influence between the participant and another actor (i.e., "I had the feeling that the [other's] behavior was connected to mine" and "I had the impression that the [other] was following my movements."), as well as on *co-presence*(i.e., "I had the feeling that the [other] was aware of my presence" and "I had the impression that I was in the presence of another being"), as well as on *engagement* (i.e., "I enjoyed [debating] with the [other]"), and on *believability* (i.e., "The [other's] behavior made me think of human behavior"), that is, the extent to which an object or character (i.e., our debaters) are perceived as human-like. Just like the Godspeed and closeness scales, these items were administered twice, once for each of the two debaters.

To quantify the overall experience of the interaction during the debate, we included the core module of the Game Experience Questionnaire (IJsselsteijn, De Kort, and Poels, 2013). With it, the game experience (in our case 'debate experience') can be assessed through scores on seven components. Each component is loaded by at least five items. Participants indicate their agreement with the statement on a five-point scale (0: "not at all", 1: "slightly", 2: "moderately", 3: "fairly" and 4: "extremely"): Immersion (i.e. "I found it impressive"), Flow (i.e. "I lost track of time"), Competence (i.e. "I was good at it"), Positive and Negative Affect (i.e. "I enjoyed it"), Tension (i.e. "I felt annoyed"), and Challenge (i.e. "I had to put a lot of effort into it").



(a) Inclusion of Other in the Self **(b)** Self-Assessment Manikins

Figure 13.2. (a) Inclusion of Other in the Self scale (Aron, Aron, and Smollan, 1992). Participants were asked to indicate which picture they felt best represented the relationship between themselves and the target (i.e. debater). (b) The Self-Assessment Manikins (Bradley and Lang, 1994). Participants for each of the three columns (valence, arousal & dominance), participants were asked to indicate (on the number scale under the pictures) how they had felt during the interaction.

Finally, we included the Self-Assessment Manikin (Bradley and Lang, 1994), again another non-verbal pictorial assessment technique, to measure pleasure, arousal and dominance in the interaction (see Figure 13.2b).

13.2.3 HYPOTHESES

Although this work was exploratory in nature, we did try to formulate a number of hypotheses on the outcome of the quantitative measures based on our current understanding on how our MEB may be perceived.

The scale-ends on the GQS instrument are mostly contrasts of machine-like qualities (low) and human-like qualities (high). Even though the MEB shares the same machine-based mind with the Virtual Agent embodiment, we still expected the MEB to be rated more favorably (human-like), based on their appearance alone:

Hypothesis 1 Participants will rate a MEB higher than a VA embodied conversational agent on the key concepts: anthropomorphism, animacy, likeability, and perceived intelligence.

The discussed literature demonstrates that experiences are more engaging when participants believe they are interacting with a human than when they are interacting with a machine, even if the behavior of the other players are otherwise equal.

We further discussed works in the background section that found that the mere belief about agency affects the overall perceived quality of the experience (i.e. Kätsyri et al., 2013; Lim and Reeves, 2010; Weibel et al., 2008.). We expected this to manifest in the Game Experience Questionnaire, which measures flow and engagement, as well as on the Self-Assessment Manikins for emotional response.

Hypothesis 2 The quality of the interaction with the ECA, as reflected in constructs such as flow, arousal and engagement (as measured by the GEQ and SAM), will be rated more positively by participants when the ECA is embodied by the MEB.

Another dimension in which we evaluated our ECAs was in how interactive they were perceived. Here we used the *coupling* instrument from Bevacqua et al. (2014). Of course, all ECAs in our experiment were equally (non-)interactive with regards to their conversational activities. Non-verbal behaviours like nodding and gaze direction were not generated based on the behavior of the participant, but fully deterministic. However, we again follow the same intuition as for the previous two hypotheses: We expected the human to be favored based on the mere belief that the human was capable of those behaviors, while the VA would have to prove these capabilities (i.e. Melo and Gratch, 2015).

Hypothesis 3 On measures regarding the bilateral relationship between the ECA and participant during the interaction (as reflected by the coupling instrument Bevacqua et al., 2014), the ECA will score higher when embodied by the MEB.

13.2.4 MATERIALS & APPARATUS

For this study we used the apparatus discussed in the previous chapter, meaning the dialogue system configured to run the ethical debate scenario. The two debaters and the moderator were located on the corners of an equilateral triangle facing the center of the space (see Figures 13.1a and 13.1b).

VAs were shown on large TV screens in portrait mode. When the proponent was embodied by the MEB, that screen was replaced by a chair for the MEB to sit on. For the MEB's cyranic interface, a large screen was placed behind and out of sight of the participant, facing the MEB (see Figure 13.1c). Due to the fact that there were other screens in the experiment room, participants did not get suspicious in seeing the screen behind their chair while entering in the room. Moreover, all the screens were, or appeared as, turned off when participants entered the room. Therefore, they could not see the agent on the screen.

	HHH Corpus	HAA	HEA
Sessions	6	18	18
Participant	18 (16 male)	18 (9 male)	18 (11 male)
Estimated duration	30 minutes	30 minutes	30 minutes
Roles:			
Moderator	Human	Human	Human
Opponent	Human	Agent 1	Agent 1
Proponent	Human	Agent 2	Echoborg

Table 13.2. Configuration of the corpus recording session (HHH) and the two experiment conditions (HAA and HEA).

13.2.5 MULTIMODAL ECHOBORG TRAINING

We recruited an experienced actress from the student body to act as the MEB in this user study (see Figure 13.1b). Following a number of training sessions of the debate with the researchers, she became familiar with the scenario and behaviours. While not systematically quantifying the accuracy of shadowing, comparing recordings of MEB behaviors with the VA behaviours showed that the actress was able to shadow the speech and gestures reliably.

13.2.6 PROCEDURE

Participants entered the room, were provided with the instructions for the experiment and were asked to sign the consent form. For the instructions, participants were told that they and two other participants[3] would engage in an ethical debate about the trolley dilemma. They were told that they got assigned the role of the moderator, who would lead the debate. For this, they were provided with queue cards to structure the debate around. The cue cards contained variation of the dilemma, which were to be introduced whenever debaters exhausted their arguments. The cue card for the original version of the dilemma had to be chosen first, but the participant was free to choose the order of the variations after that. Participants were then asked to be seated in the experiment area (see Figure 13.1). In the condition where the

[3]The nature of the other debaters was intentionally kept ambiguous

proponent ECA was embodied by our confederate through the MEB apparatus, we had both the confederate and the participant enter at the same time, going through the same procedure and being seated in the experiment area at the same time. During the debate, the experimenters used the *Wizard Interface* (see Figure 12.1) to indicate participant utterances (i.e. the selected cues). Debates would end when the all variations of the dilemma on the cue cards were addressed by the debaters, which would take around 10 minutes. After the debate was completed, participants were asked to fill out the post-experiment questionnaire[4]. After that participants were asked about their impression of the interaction, and then they were debriefed about the nature of the experiment.

13.3 RESULTS

We will first report on the quantitative analysis to be able to test our initial hypotheses described in Section 13.2.3. We will then discuss the results also qualitatively, also drawing from the comparisons with how participants in the corpus sessions reported on the interaction that the experimental conditions were modeled on.

13.3.1 QUANTITATIVE ANALYSIS

We conducted one-way ANOVA on the effect of the embodiment on our measures for perception of the debaters and interaction experience overall. This was done first for the between subject variable *proponent embodiment*. Since participants assigned to the Human-Echoborg-Agent (HEA) condition (n = 18) saw both a virtual and MEB embodiment, we have done an additional one-way ANOVA for this group on the within-subject variable *debater embodiment*.

In the between-subject analysis, two sub-scales of the Godspeed instrument showed significant effects: animacy ($F(1,34) = 5.834$, $p = 0.021$, $\eta^2 p = 0.146$) and anthropomorphism ($F(1,34) = 20.061$, $p < 0.001$, $\eta^2 p = 0.371$). Post-hoc tests revealed that the proponent was rated higher on both those sub-scales when embodied by the MEB rather than by the Virtual Agent. One sub-scale of the coupling instrument showed a significant effect between subjects: coupling ($F(1,34) = 16.920$, $p < 0.001$, $\eta^2 p = 0.332$). The post-hoc test reveals the same directionality as before: the proponent received higher coupling scores when embodied by the MEB.

The within-subject analysis revealed a significant effect on the anthropomorphism of the debater embodiment on the anthropomorphism subscale ($F(1,17) = 12.190$, $p = 0.003$, $\eta^2 p = 0.418$) and on the perceived intelligence ($F(1,17) = 4.322$, $p = 0.053$,

[4] Also the MEB actor in the HEA condition

Table 13.3. Statistics of pairwise comparisons. Top half showing the comparisons of scores attributed to the proponent debater embodiment (VA or MEB) between subject. Bottom half showing the comparisons of scores attributed to the opponent (MEB) and the proponent (VA) within the HEA condition.

	Scale	Subscale	contrast	Estimate	SE	df	t.ratio	p.value
between	godspeed	animacy	VA - MEB	-0.722	0.299	34	-2.415	0.021
	godspeed	anthropomorphism	VA - MEB	-1.233	0.275	34	-4.479	0.000
	coupling	copresence	VA - MEB	-0.639	0.155	34	-4.113	0.000
within	godspeed	anthropomorphism	VA - MEB	-1.022	0.293	17	-3.491	0.003
	godspeed	intelligence	VA - MEB	-0.537	0.258	17	-2.079	0.053
	coupling	copresence	VA - MEB	-0.667	0.133	17	-5.030	0.000

$\eta^2 p = 0.203$). Here as well, post-hoc tests reveal that debaters embodied by the MEB were perceived more favorably.

There were no other significant effects of proponent embodiment or debater embodiment found on any other instrument. Statistics for the between and within post-hoc tests are reported in Table 13.3.

13.4 QUALITATIVE ANALYSIS

Beyond the quantitative analysis, given the exploratory nature of the study, we take a closer qualitative look at the distributions of the responses shown and the tendencies they might indicate. Although speculations about these tendencies may not be supported statistically by the data, they may be helpful to discuss and to formulate new hypotheses about the complex nature of how MEBs are perceived, and how they may be used in ECA research in the future. This also allows us to include into our analysis the data from the recording sessions for the corpus on which the interaction in the experiment was modeled. Finally, we also include some feedback received from the participants after the experiment, and a brief description of our confederate's experience using the apparatus and acting as MEB.

13.4.1 EXPERIENCE OF THE INTERACTION

Figures 13.3 and 13.4 show the distribution of participant responses to the different constructs on the SAM and Game Experience Questionnaire instruments. In addition to the experiment data, we added the responses to the same questionnaires from the corpus sessions, where three non-confederate human participants conducted the debate (labeled HHH, see Section 13.1.1). To keep these somewhat comparable, we are only showing the responses from the moderator in the HHH corpus sessions, excluding the responses of the two debaters (from whom we do not have any responses in the two experiment conditions).

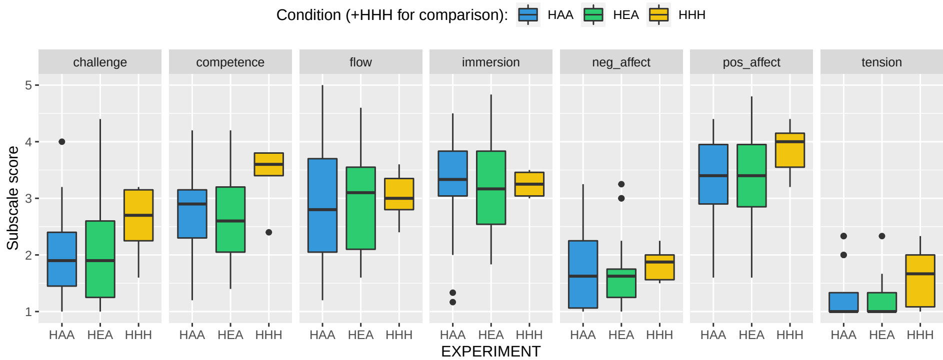


Figure 13.3. Game Experience Questionnaire. Note that for this visualization, response values are remapped to 1-5 from the original 0-4 range.

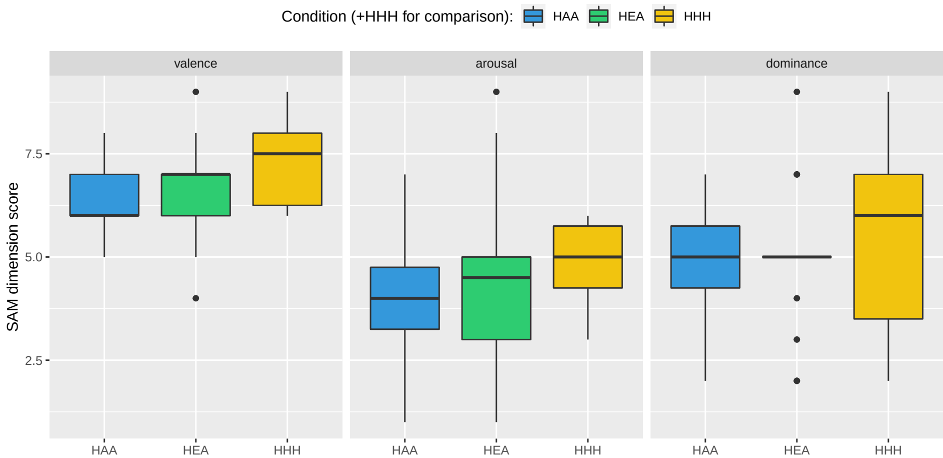


Figure 13.4. SAM scores

For the SAM scores shown in Figure 13.4, there only seem to be small tendencies when comparing the two experiment conditions (HAA and HEA). The medians on valence and arousal are somewhat higher in the MEB condition than in the agent only condition, that is, increased reported feelings of valence and arousal. Responses to the same instrument in the human-only corpus indicate that a similar interaction with only humans elicits higher feelings of valence, arousal and dominance.

The Game Experience Questionnaire (see Figure 13.3) measures constructs on the experience overall rather than the individual emotions. Again, we find that the main differences manifest between the human only corpus and the two experiment conditions overall. The tendency seems to be that moderators in HHH felt that the experience of the debate had more tension, more positive affect, was invoking of their competence more. Between the MEB and VA conditions, again there seem to be none or only subtle differences. On tension, there is a strong floor effect for the experimental conditions, while the HHH are low but not centered around the minimum value. Notable is also how in a number of constructs, the variance differs wildly between HHH and the two experimental conditions. Participants in HHH rate the level of competence, flow immersion and positive affect similarly, while these constructs were rated with much more variance by participants in the experimental conditions. One explanation could be that for the MEB embodiment, participants felt unclear about its identity or personality, causing individuals overall vary also in their interpretation of their feelings towards the experience. However, since variance is also larger in responses from the condition without the MEB, perhaps it's about the identity of the other debaters in general, whose nature were kept equally ambivalent (all debaters were introduced as other participants). A third explanation could be regarding the overall difference in the experience. Participants might feel more confident about how they appraise an experience familiar to them (such as a debate with other humans), compared to the more unusual experience of debating with agents (or a mix of agents and humans).

From the related work on the experience of interaction while (not) co-present with other players, difference between the human-only experiences and the experiences in the experiment were to be expected. However, previous work we discussed also has consistently found that the mere belief of agency is sufficient to elicit these responses. The premise of the MEB condition was that they would elicit this belief of a human debater with high agency merely by the embodiment being a human one. Consequently, we had hypothesized that the overall experience would be perceived more positively when at least one human was involved, compared to the condition that exclusively features artificial agents. Instead, these two conditions seem to perform at the same level. In the next section we look in more detail at the perception that moderators had of the different embodiments.

13.4.2 QUALITATIVE ANALYSIS: PERCEPTION OF DEBATERS

In Figures 13.5 to 13.7 we show moderator responses to the instruments measuring attributes of the individual debaters. As above, for comparison, we also display the responses that moderators gave when the same instruments were administered after the human-only corpus recordings. Note that for these visualizations, we grouped all scores attributed to agent & embodiments together, regardless of their role in the debate.

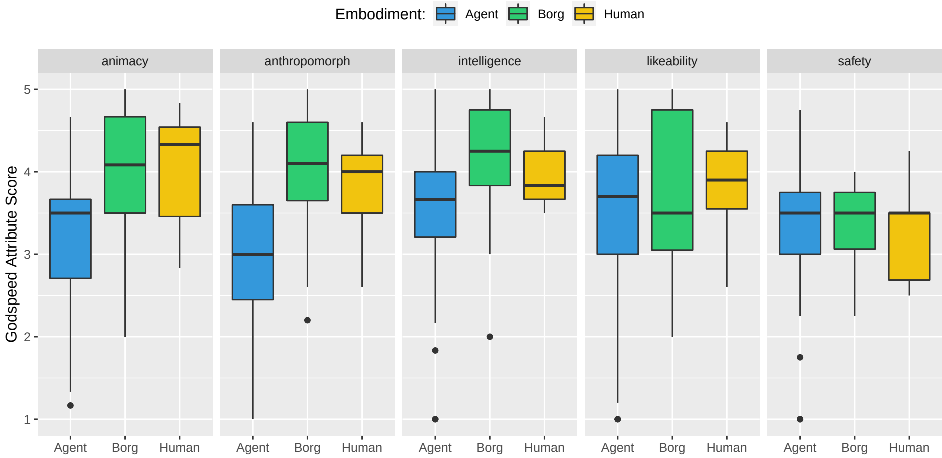


Figure 13.5. Perceived Godspeed attributes

When asked about how moderators felt about the different debaters individually, the pattern we have seen above disappears, where both experiment conditions received very similar scores. Instead, starting with the constructs measured in the GQS (Figure 13.5), the MEB in the experiment received similar ratings as the other human participants in the corpus recording sessions on *animacy* and *anthropomorphism*, while the agent received lower scores on these two constructs, which were also two of the main differences found in the quantitative analysis.

Interestingly, for intelligence, the MEB received higher scores than other human debaters received from the moderators in the corpus recordings. This also points to a weakness of our mixed experiment design, as the MEB condition features two different embodiments to be rated (VA and MEB), possibly causing participants to answer the questionnaires more differentiated.

Looking at Figure 13.6, overall the lowest scores on closeness were given to the agents, followed closely by the MEB. Again, responses in the human corpus centered mostly around the center of the scale, suggesting that more differentiated scores where elicited in the MEB condition, where two different embodiments were displayed.

The *coupling* instrument (see Figure 13.7) was not administered to participants of the corpus recording sessions, so we can only discuss differences between the experiment conditions. Again, we only found a significant difference on the *co-presence* subscale in the quantitative analysis, where the MEB was favored over the VA.

Looking at the distribution of the responses however, there is an interesting tendency in the *believability* scores, where moderators seem to rate debaters with a virtual

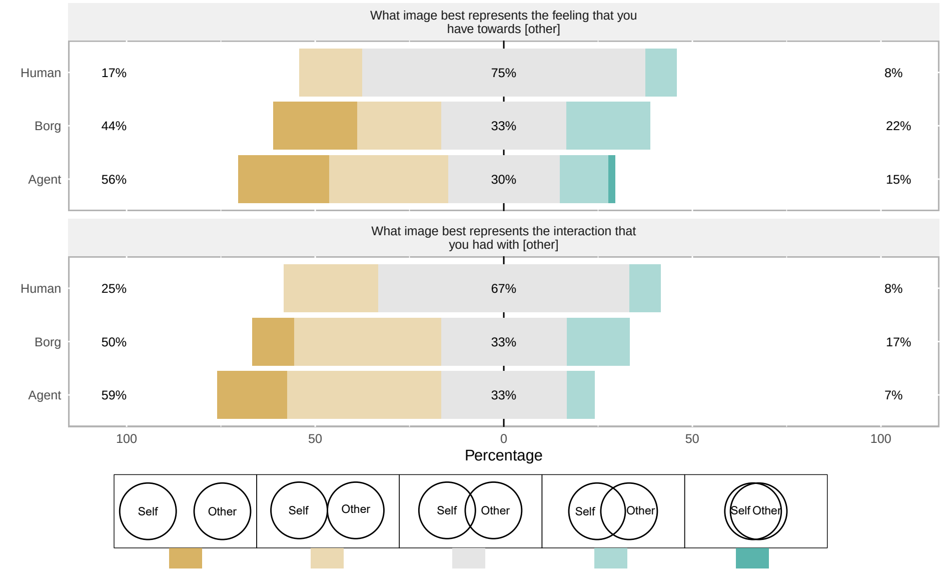


Figure 13.6. Inclusion of Other in the Self instrument responses per type of interaction partner.

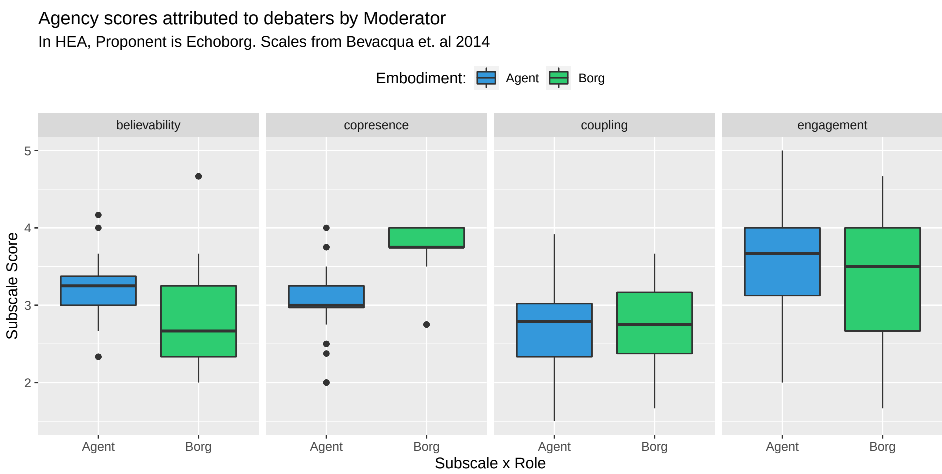


Figure 13.7. Perceived coupling (this instrument was not administered in the HHH corpus)

agent embodiment higher than scores given to debaters with MEB embodiments. On *coupling*, both embodiments score similarly low. Both embodiments were also found to be similarly *engaging*, although with notably higher variance in the responses compared to the other subscales.

So looking at these distributions, tendencies may point towards the MEB being not just perceived as a *normal* human debater in all accounts. The difference in *believability* suggests that participants felt that the MEB did not live up to their expectations as a human debater - the agent scored higher on believability, as the agent behavior could be better integrated with participant's expectations of a virtual agent.

This did not affect participants feeling of co-presence, where the MEB clearly scores better than the agent embodiment, as with the closeness instrument discussed above, this may be mainly driven by the physical manifestation of the MEB, compared to the screen-based display of the VA.

13.4.3 PARTICIPANT FEEDBACK AND CONFEDERATE EXPERIENCE

A contribution of this work is the first implementation of a Multimodal Echoborg apparatus for ECA systems. For evaluation purposes we gathered insights from participants on their experience with the MEB interlocutor (prior to revealing that she was a confederate). We also asked our MEB actress to keep logs and notes on her own experience and performance after each session.

All participants in the HEA condition reported that they noticed something off about the human interlocutor, often using the term *awkward*. Some thought that she was shy, others thought that she may have some mild form of a mental disorder. Only two participants however guessed that she was acting and involved in the experiment.

The actress noted a number of instances in her logs where she remembered deviating from the behavior on the MEB interface she was supposed to shadow. Specifically in the speech and gesture modalities. Most deviations were in speech. Some notes read: "I thought that was the sentence I had to say but instead I said it faster.", "I tried not to look at the screen because I felt that the participant might notice something is happening behind him.", and "The participant wanted to speak and I had to speak over him.". Regarding gestures, she noted some instances of having difficulty following the movements on the MEB interface: "I had the impulse to follow my own reaction to what I was saying".

Based on spot-checks in the video recordings, we confirmed that the majority of deviations occurred during gesture shadowing. Meanwhile, deviations in speech shadowing were small. Regarding the listening behavior (timed head nods), there were little variations in the actresses shadowing performance.

13.5 DISCUSSION

In this section, we will first revisit the hypothesis about perception of the ECAs and overall experience. Then we will reflect on the implications of these findings, in terms of the use of an MEB-like apparatus for ECA research. We will finally discuss some limitations and future work, and reflect on the concept of humans embodying machines in general.

13.5.1 DOES AN MEB HAVE THE BENEFIT OF THE DOUBT?

We can only partially support our first hypothesis. We expected the MEB to outperform the VA embodiment in the subscales of the Godspeed instrument. We could only find evidence for this on the anthropomorphism and animacy subscale. There was only near-significance on the perceived intelligence subscale (although the tendency was in favor of the MEB). No difference was found on the likability ratings. Interestingly, the MEB seems to outperform only on measures regarding the outer appearance of the embodiment, that is, humanoid looking and moving. At the same time, participants did not falsely overestimate traits related to the behavior of the ECA, such as intelligence and likability.

The second hypothesis cannot be accepted. We hypothesized that the interaction would be more enjoyable when a human was involved, based on research that showed positive effects on how an experience is perceived, mediated by mere belief of true agency of an interactant. These effects did not seem to translate to our context, as no difference was measured on any of the Game Experience Questionnaire subscales, nor on the emotional response measured using SAM. After also considering data from the HHH corpus in the qualitative analysis, it seems that the limitations of the conversation afforded by our ECA system simply were not as engaging as those afforded in the corpus recordings that the system was modeled on. The human embodiment was not able to overshadow its limited mind.

This lack of an engaging experience may also be explained by the lack of our ECAs' ability to engage in bilateral connection with the other dialogue partners (including the participant). For our third hypothesis, we had expected again that the human embodiment would be sufficient to involve the belief that the MEB was at least more capable of such bilateral behavior than the VA. This however was not the case. The only subscale of the coupling instrument that showed a significant difference was the co-presence subscale. As already discussed above, this may be best explained by the MEBs physical form, while the VA was limited to a screen-based display.

In absolute terms, the coupling scores in fact were overall low, reflecting the limitations of our ECA system for both embodiments equally. From the insights shared post experiment, participants noticed that the MEB was not behaving normally, explaining it away with different potential personality attributes or even disorders.

The finding that in tendency, the agent was actually perceived as more believable than the MEB, may also point towards an explanation of betrayed expectations. Just as described in the related work on ways to benchmark agent believability (i.e. (Melo and Gratch, 2015)), participants have lower prior beliefs about virtual agents' ability to have mind and experience than about humans. Lower believability scores for the MEB may show negative effects if such a prior high belief turns out to be unwarranted. This also connects to prior work by Nowak and Biocca on agents with highly anthropomorphic avatars, which "*set up higher expectations that lead to reduced [co-]presence when these expectations were not met*" (Nowak and Biocca, 2003).

13.5.2 LIMITATIONS & FUTURE WORK

There are a number of limitations to acknowledge regarding the present work. A higher powered study is required to be able to test the hypotheses properly. There was also a lack of counterbalancing in the debater role and gender. Additionally, the chosen mixed design is problematic, and comparing both between subject on all data and again within subject on a subset of the data may have artificially inflated the power of those tests. A dyadic interaction with strict between subject design would allow for a more rigorous investigation of MEB perception. What is more, we rely entirely on self-report measures for this study. Instead, methodologies such as the ones used by Corti and Gillespie that use benchmarks based on behavioral measures are preferred.

We also have to consider a general limitation of Echoborg-like methods, which is the chance of nonverbal leakage, or other display of behavior by the confederate that was not intended by the source of the system. Based on some spot checks with the confederate's self report, there were also deviations in how gestures were shadowed. Also, due to the implementation of the MEB interface, there may have been systematic biases in gaze direction of the MEB towards the screen that we have not accounted for. For larger studies, the shadowing performance should be monitored more rigorously, and potentially modeled in the analysis.

In the future, we should also evaluate an MEB in a more dynamic setting, with an ECA that has a more advanced cognitive system, for example integrating Social Signal Processing (SSP) that allows for display of behaviors contingent on how other interactants behave.

13.5.3 MEB FOR ECA RESEARCH

One motivation for this work was to investigate whether Echborg-like methodologies could be a useful contribution to the tool set of methodologies used in ECA research, next to for example the popular Wizard-of-Oz method. We have proposed that the ability to shadow multimodal behaviors is an important requirement for the context of ECA research. Based on our experience designing the apparatus and conducting this experiment, we can reflect on how well multimodal shadowing worked in our experiment, using our first implementation of the MEB.

As we have already discussed extensively in the limitations above, spot checking self-report data from the confederate with the recording, we notice that deviations were especially noticeable in the gesture modality. To be more practical for ECA research, and to generally achieve a more integrated MEB system, improvements to the apparatus are required.

We may expect innovations on wearable AR glasses in the future, that allow for rich instructions on gaze behavior or head movement in a covert way. Haptic displays certainly could play a role in providing encoded information about gestures in various body parts. Vibrotactile actuators placed in the MEB's neck could encode nodding intensity and frequency, while actuators on forearms or hands could encode information about the use of hand shapes and gestures. Potentially, when integrating an ECA system with external sensors for SSP, not just other interactant's behavior could be observed, but also the MEB's behavior. In such a fully integrated version of an MEB, this could help not only to monitor deviations in the behavior shadowing, but also serve as real-time feedback to correct and improve the MEB performance during training or in-situ.

13.5.4 MEB AND VA BENCHMARKS

We may conclude from this that with some more improvements to the apparatus, ECA research with the MEB could be possible, but is it also useful? The intuition as to why Echoborg-like methodologies could benefit artificial agent research was that when embodied by a human, they would evoke higher (human-like) beliefs in interactants about the agent's level of mind and experience, which in turn also causes interactants to treat them and interact with them more like with humans. In the words of Corti and Gillespie: *"Artificial agents cannot enter the world of human intersubjectivity without the support of their human interactants, and this support is contingent upon interactants' supposition that complex intersubjectivity is achievable."* (Corti and Gillespie, 2016). If we want to learn whether novel ECAs cognitive models' can keep up with those of a human, we need to be able to evaluate interactions under a human-human condition, which is something that the MEB could afford in the future.

However, we should also acknowledge a nuanced view on this. As Corti and Gillespie also raise, *"As more advanced artificial intelligence develops and as people are raised in a world in which socially advanced artificial agents are ubiquitous, the expectations people will place on the intersubjective capacities of their machine interlocutors may increase"* (Corti and Gillespie, 2016). Thus, being used to more capable artificial agents in the future may make it normal to interact with machines as with humans, diminishing benefits from MEBs in these experimental settings.

Another important aspect to remember is that, especially in the field of ECA research, it is not always the goal to fully mimic human-human interaction with an ECA. In certain contexts, users recognizing the machine for what it is, gives value to the ECA in the first place, such as for virtual therapists (Lucas et al., 2014).

13.5.5 MEB AND SOIC-TV

Let's review how this research relates to the SoIC-TV system that we described in the introduction of this thesis. The original use cases of the device is to allow owners of the SoIC-TV to offer their services as a telepresence system as a *driver*, similar to how Uber drivers rent out their time and car as a mobility service. Remote users, or *passengers*, would be able to call in and have the *driver* serve as their human embodiment, while the *passenger's* face would be visible on the SoIC-TV.

While the MEB apparatus designed for this research was not intended to be used for telepresence applications, it still has a lot in common with the SoIC-TV scenario described in Section 1.1.3. Here, we proposed that with our system, *drivers* could also be rented by machines instead of just human *passengers*.

As the MEB apparatus is not wearable, let's assume an advanced version as proposed in Section 13.5.2, with a set of covert AR glasses used as interface for multimodal behavior shadowing. Of course, the same limitations still apply as identified above; a *driver* will require certain training, especially the more modalities they are expected to shadow from a machine *passenger*.

The uncertainty about the quality or deviations of shadowing performance also translates from concerns in lab-setting to possible requirements for field application of such a system. We have already mentioned that an MEB-directed SSP software may help evaluate shadowing performance in a lab setting. To elaborate on this point further, we can also take a closer look at how behavior realization with artificial embodiments usually deals with this very practical problem of deviations between requested behavior and deviations from it in the embodiment. Here, a feedback loop is typically established between the behavior planning engine and the actual embodiment on which these behaviors play out. This is true for any embodiment crossing into the physical domain, such as for joints on robots. In physical joints, exact reproduction of a requested motion may vary depending on environmental factors such as changes in temperature or physical load. Using a feedback loop that measures actual joint position and feeds this back to the behavior planning system helps to update plans for the motion dynamically, such that it meets other requirements. As an example, a joint lagging behind too far might cause the planning system to decide to abort continuing of the current gesture, in order to try and meet the time requirements for the next gesture.

So for a proper wearable MEB system for machine *passengers*, not only do we need to consider more actuators to allow for better signaling the desired shadowing performance, but also we should consider wearable sensors that give feedback on the *driver's* performance to the *passenger*.

Finally, we should also address the non-technical implications of this work for the SoIC-TV scenarios. We had reiterated in Section 13.5.4 prior motivation as to why the use of a human embodiment for evaluating machine behaviors may be useful in the context of ECA research. The main reason being the experimental consequences of observing or evaluating human-machine interaction when due to prior expectations, machines are treated differently to begin with. Although we have not discussed related work from, for example, the domain of persuasive systems, it is easy to imagine scenarios where humans are more likely to dismiss a request coming from a machine than from another human. So, beyond the better access to the physical domain, another reason for machines to employ a SoIC-TV *driver* (instead of sending an email) would be to benefit from an increased ability to persuade other humans.

13.6 CONCLUSION

Summarizing, we have evaluated our Echoborg apparatus for multimodal behavior shadowing (MEB) that is integrated with our existing Embodied Conversational Agent (ECA) framework. Our research questions were about how the perception of the interaction and perception of the individual agents would differ when the ECA was embodied either by the MEB or by our virtual agent (VA). To summarize our findings, let's go back to Rostand's play *Cyrano de Bergerac* that inspired Milgram to develop the cyranic method, and ultimately inspired the Echoborg method.

The moral of the story was that Roxane was attracted to Christian's body, but ultimately fell in love with Cyrano's mind. This is a feat not likely repeated by our MEB. In our experiment, body and mind of the MEB are inversely incongruent - a human body with the limited mind of a machine. Our MEB was perceived more human-like on the surface, despite its limited mind. However, the qualitative analysis showed that some of these limitations shone through on other measures, and a higher powered study may reveal that without a socially more capable ECA system, the initial favor a human embodiment may receive will turn into a negative in the course of the interaction with an MEB.

What's more, we can say that (luckily), the titular *Mind Transplant* was not fully achieved with our system. The EB actor remained in control of their own facilities and required their own facilities to follow the events on the cyranic interface and translate them into their own behavior. As we have discussed, this translation will remain a source of error between the desired behavior of the ECA system and the true realization performed by a human embodiment. Future improvements to the cyranic interface as well as a potential feedback loop between ECA and embodiment may be worthwhile to explore for mitigation of these errors.

The concept remains useful for future research, where participants should overlook their existing biases when interacting with machines, so that the MEB helps with benchmarking improved artificial systems that Roxane could end up falling in love with.

PART V

CONCLUSION

14

DISCUSSION AND FUTURE WORK

In this final part, I will reflect on the work presented in this thesis, how it contributes to the overall field of mediated social interaction technology, and what directions it could take in the future. Where appropriate, I'll also discuss ethical implications.

Taken side by side, the primary research questions investigated in the studies discussed in this thesis do not strictly build on each other. They do not aim to solve a single problem, and thus do not allow a singular conclusion, evaluating the extent to which some problem was solved. However, the studies presented in this thesis do share a commonality: They employ novel technological artifacts or research apparatuses to explore social interactions beyond what we experience in our everyday context of co-located and mediated social interaction. As such, the contribution of this research can perhaps best be examined in terms of how it contributes to understanding the evolution of Mixed and Virtual Reality media. Slater summarized the early research with VR with the focus on simulating known experiences - much as television was originally used to mediate theater, before other experiences were invented:

"*Yet it is a medium that has the potential to go far beyond anything that has been experienced before in terms of transcending the bounds of physical reality, through transforming your sense of place and through non-invasive alterations of the sense of our own body*" (Slater, 2009)

Again, the primary research questions of the studies presented in this thesis are not explicitly about exploring or transcending the *bounds of physical reality* with the apparatuses built for these studies. However, most provide some novel aspects in the context in which they are applied, in the experience that they afford and the way they were designed. In the following, we will discuss these aspects along three themes. First, the experience of sharing mixed reality spaces in different ways with other humans and machines, and second, the experience of perceiving those actors we share space with when their agency is not always clearly defined. Finally, third, we reflect on the role technology plays in affording these interactions, and reflecting on the technology and software developed in the course of this thesis.

14.1 THEME 1: SHARING SPACE IN MIXED REALITY

One theme in this thesis is that mixed reality technology can be used to mediate social interactions in both virtual and physical shared spaces. In Chapter 2, we investigated humans sharing space with virtual humans in a virtual environment. In Chapters 7 and 8, we used mixed reality technology to allow remote users to navigate a physical space, while at the same time showing an avatar of the remote user to local users using augmented reality. The remote user shared the virtual representation of the physical space with the local user in the same way that the local user shared the physical space with the virtual representation of the remote user.

We learned that virtual humans violating social norms regarding the use of virtual space doesn't go unnoticed by human interlocutors, causing them to both adjust their own use of the virtual space, and also affecting how they perceive the virtual human's personality. The way virtual embodied actors behave in virtual spaces matters. In theory, the space in VR can be infinite, but in social interactions, use of space is a social signal that matters.

This is also the case when we look at scenarios such as in the two experiments with the OpenIMPRESS telepresence system. Here, to afford communication around objects in the physical space of one user, we aligned the (representations of) physical space and (representations of) user embodiments in such a way that they were consistent with each other both for the local and remote users. This is required for certain types of communication to work properly, such as establishing eye contact, as well as for gaze direction and pointing gestures that reference other objects in the physical space – or, for that matter, shared virtual objects, such as the shared widgets used in the *CPR* study used to provide feedback on the local user's performance to both users.

However, note that the two apparatuses in Chapter 2 and Chapter 7 are also quite different from each other in how they deal with (artificial) physical constraints and realism of the experience. In the study on proxemics, we sought to *simulate* a co-located interaction with VHs in VR. Human-like appearance of the VHs was a priority, including an animation system that featured realistic locomotion for (re-)establishing social distance. The participants did exclusively use their own locomotion to reposition themselves in the room-scale virtual environment. With the OpenIMPRESS system, we did not adhere to the same requirements. The remote user was able to manipulate their position in the representation of the remote environment using the hand-controllers, even while physically standing still. This could be done without regards for obstacles, and without regard for constraints that would be imposed by a human body. A human body needs to stand, sit or crouch to adjust their point of view in space. Using the OpenIMPRESS interface, the RO can position themselves close to the floor, or inside a tight space in the remote environment, while physically remaining in whatever position is comfortable[1]. Of course, these physical constraints cannot be overcome by the local user that is engaged with the physical space, but also for them, the experience will be different from a co-located interaction, as the RO's embodiment will still float around freely.

So, between the apparatuses in Chapter 2 and Chapter 7, the ways of how space is shared are different. At the same time, for different use cases – simulating co-located social interactions versus facilitating remote collaborations – other requirements will apply. The choice to disregard these physical constraints for the more application oriented scenarios of the OpenIMPRESS seemed successful, as we found that remote user's experience in terms of social presence and usability, as well as team task performance have improved when the remote user's presence was not limited to physical constraints[2].

One of the obvious areas of improvement of the OpenIMPRESS system would be in the capabilities of the RO embodiment in terms of communication modalities, as it is currently limited to voice and position and rotation of head and hands. Adding a face for stronger cues through eye-gaze and facial expressions may allow for better communication. However, we can also speculate that the possibility for users to violate social norms may come to play again once we also introduce such a more capable, human-like embodiment for the RO. While this may allow for better communication, affordances and expectations might also change and conflict with the lack of physical constraints discussed above. A high-quality human head with realistic skin and eyes floating through walls may feel more *uncanny* than our current, simplistic avatar.

We've highlighted *social presence* as a measure for how the presence of others and the interaction with them is perceived between interactants in our telepresence system. In Chapter 9, we provide a reflection on how we employ one of the most popular measures of social presence in evaluations of (collaborative) telepresence systems. We have highlighted the third, intersubjective dimension of social presence that is implicitly included in the Networked Minds instrument, allowing for analysis of perceived within- and cross-interactant symmetry. These indicate how interactants perceive each other's presence differently, and the extent to which this feeling is mutual. Especially in the assymetric experiences such as the ones we address with our OpenIMPRESS system, these symmetries may prove as a useful diagnostic tool and deserve more attention in the future.

[1]Head movement through space are still tracked, and the virtual viewpoint reflects these movements. The controller-based movements manipulate the root transform of the tracking area.

[2]Although it should be noted that the physical constraints in the *FixedPosition* condition were different from the ones we described for the *proxemics* study

14.2 THEME 2: SOURCES OF MIND

Another theme in this thesis is how humans perceive mediated social interactions with other interactants of unusual or uncertain sources of agency. For example, in the *social conflict* study in Chapter 3, we allowed a virtual human to engage in a conversation with teachers by having the VH claim that they have a complaint received from a real student, and are discussing it with the participant on behalf of the exam committee. Naturally, there was no prior information on the participant or prior students available to the agent. What's more, we used the *Wizard-of-Oz* method to control the VH utterances. Albeit in a different context, in the *Multimodal Echoborg* (MEB) study in Chapter 13, the opposite happens: A human talks on behalf of a machine. An Embodied Conversational Agent (ECA) system produced verbal and non-verbal behavior which our confederate was asked to *shadow* using the MEB apparatus.

In both cases, these deceptions, that is to say, the true level of agency or source of mind of participants' interlocutors were only disclosed to participants after the experiment was completed. We investigated how participants experienced these unusual types of mediated interactions.

From the *social conflict* study, we did gather some insights into what participants thought about possible scenarios where machines are used to mediate these types of conflicts on behalf of others. Some explicitly noted that there was uncertainty about *who* the VH was, and consequently it was difficult to understand what the result or possible outcome of the interaction would be. Others found it easier to talk about the conflict with a machine compared to talking about it with a human, an effect often observed in VH literature. Others again mentioned that they simply felt it was not appropriate for a machine or Virtual Human to judge someone. In summary, the unclarity or individual interpretations of the VH's agency or source of mind may have colored these interactions significantly on an individual level.

In the *Multimodal Echoborg* study, our analysis of the experience with the machine-driven human focussed on more constructs from HCI instruments, such as game experience and agent perception. Our initial hypothesis was that overall, interactions with the ECA would be rated more favorably when it was embodied by a human (using the MEB apparatus) than when it was embodied by a VH. These hypotheses were mostly not supported by the results. Our follow-up analysis suggests that instead, the human embodiment raised expectations about the interaction that the human embodiment could not fulfil, as it was limited to the capabilities of the ECA system. Although not significant, the VH embodiment with the same capabilities received higher average ratings on the *believability* measure.

Variations in the true source of mind is also a theme that was developed throughout this thesis when discussing the variations of our SoIC-TV telepresence services. We explored scenarios both for how a human could rent a SoIC-TV driver for conflict situations in Chapter 4, and how a machine may employ a human embodiment to achieve goals requiring social interactions in the physical domain in Section 13.5.5.

Taking together both the experiences investigated in the user studies and the SoIC-TV scenarios, we are painting a fictive, but plausible picture for how in a future with increasingly pervasive technology that affords mixed reality mediated social interactions, there are also opportunities for social interactions with agents that may not exactly fit into patterns that we are familiar with today from our daily co-located social interactions. Machines acting on behalf of humans, and humans acting on behalf of machines are perhaps only two extreme ends on a continuous scale of social interlocutors we might encounter with mixed-agency. We learned that being transparent about the true sources of agency may lead to less confusion on the part of the user, while at the same time we speculated how intentional deception may empower machines to achieve goals in the physical domain that they were not able to achieve before.

Designers of these mixed-agency systems will also need to consider the complex ethical aspects of these systems. Being deceived about who it is that we are talking to is already a pervasive problem in *normal* interactions on the web. In future environments that feature mixed reality technology, such as the *Metaverse*, interactions with increasingly smart machines employing increasingly persuasive embodiments for interactions will take place. Similarly, humans with malicious intentions may be able to exploit the ability to appear in such environments with embodiments that help them to appear more trustworthy than they would be in a co-located interaction.

A common ethical concern about the internet is that malicious players can exploit the reach they gain through the internet in a way that would not be possible in the physical domain. For example, sending out one million scam-letters may not be profitable, while sending out tens of millions of scam emails comes at almost no cost. Although we may be verging into a science-fiction scenario, we can conceive how with a SoIC-TV telepresence system, a malicious actor could now exploit a network of human actors *at scale* for a coordinated attack.

14.3 THEME 3: TECHNOLOGY

Finally, I want to reflect on the contributions made around the technology and software developed in this thesis. As mentioned in the introduction, using existing technology in a way to achieve new experiences has always been an important motivation for this work. While some experiences may seem more mundane or common by the time this thesis is published, there are still some interesting aspects to be discussed.

Starting with the studies on the OpenIMPRESS system for asymmetric telepresence presented in Chapter 6, the implemented design aspects of this system were based on existing work on mixed reality telepresence. In the study on the different implementations of the interface and telepresence experience in Chapter 7, we were able to support a number of the initial design rationales. What's more, we have integrated a novel feature with the marker-based alignment of depth-sensors at the capture sites (see Section 6.2.1). The resulting system is a framework that is both easy to use and easy to extend to other settings, as was done for the AR-CPR study in Chapter 8.

Next, we have made some contributions to Virtual Human and Embodied Conversational Agent technologies in the course of this thesis. The ASAP Realizer to Unity bridge developed for and used in the studies involving ECAs (i.e. Chapter 3 and Chapter 12) has since also been employed in other projects involving ECAs and multimodal social robots.

We have also raised some questions on how ECA systems might need to be adapted in the future in Section 13.5.5, if they need to be integrated in hybrid mixed reality settings such as the SoIC-TV service. Here, classic ECA systems for virtual humans or social robots typically implement a feedback system between the behavior planner, and the embodiment that is realizing the requested behaviors. In mixed settings such as the MEB setup, this means extending the system with real-time feedback on the MEB's (or SoIC-TV *driver*'s) performance, that is, on whether utterances are completed, were interrupted, or spoken incorrectly, as well as monitoring the progress on gestures and locomotion behaviors, to name but a few.

14.4 CONCLUSION

I have discussed a number of studies on different mediated interaction experiences using mixed reality technology. Although these studies are varying in their primary research questions, they paint an overall picture of how MR technology will, in the future, not just allow experiences that transcend what's familiar in terms of *physical experiences*, but also transcend what is familiar to us in *social experiences*. Hybrid interactions with humans acting on behalf of machines and machines acting on behalf of humans, represented by various physical or virtual embodiments in a mixed reality will pose both challenges to designers and users of these systems.

ACKNOWLEDGMENTS

When you enjoy something very much, you really take your time with it because you don't really want it to end. I have certainly taken my time with this thesis, and most of it in the amazing family that is the HMI research group. I want to address and thank all of those that supported and influenced me, and made this time so enjoyable. It might be best to just start at the beginning. Maybe a little before that...

First, I really want to thank those that were involved in organizing and teaching the CREATE bachelor program. Especially Alfred, Anton, Angelika, Chris, Dennis, Edwin and Gerrit. Being a *proefkonijn* in the pilot group was crazy and fun, inspiring and formative. I learned about the creative process and how not to fear failure (as much), about humans and technology, how to make things, and how make everything talk to each other – regardless of boundaries between physical and virtual. I will always get a kick out of this stuff. Thank you!

After CREATE, I started my master at HMI. The way I remember it, on the first face-to-face meeting with Dirk, I expressed my desire to help with ongoing projects at the department. I also remember that Dirk's reaction wasn't quite satisfying to me. Instead of open enthusiasm, there was just a sober, sonorous "okay... hmmmm". So typical. Still, soon after, I was involved in all kinds of projects, making *Touching Virtual Agents* with Gijs, dialogue managers with Merijn and Rieks, tracking social robots and people using MoCap with Jered. I got a huge kick out of that. Thank you all for getting me involved.

There were two experiences that stood out from this time, and that really made me want to commit and stay at HMI for doing a PhD. First, the FROG EU project. Going to Seville several times with Vanessa, Daphne, Lynn and Randy to get that green robot to tell people about the Alcázar, getting it to point its snorkel at those moorish facades... I got a HELL of a kick out of that! Thank you. Also thank you to that reviewer during the practice review, who completely tore apart all of that work. That taught me to not be too defensive when getting constructive criticism. But damn, that hurt.

Randy and Lynn, thank you most of all. The friendship we formed during FROG and kept since means a lot to me, you have helped me so much with so many things since (including comforting me after that practice review). Let's take a trip back to Sevilla?

Lynn, without your support, I don't think I would've ever finished this thesis – at the very least, it would read a lot worse.

The second experience that really made me want to become a PhD candidate at HMI was the infamous eNTERFACE workshop in Lisbon. I don't know where to start or stop, I felt right at home with you guys between Super Bock, virtual human arms, Daft Punk, people hanging from the ceiling, greasy food with potato chips, haptic sleeves, the *Lux*, sleeping in the hallway, preparing user studies, and faux pas[3]. Gijs, Merijn, Robby and Ronald, the ultimate role models for a young researcher? Thank you guys for this amazing time (there and after), I got a real kick out of that.

I was so motivated to stay at HMI. Also, since you already had me on tape saying that out loud, there was no way out anyway. I remember, it was a little more scary and less fun at first. That's because suddenly you have to do stuff on your own, instead of together with all the other nice people. Luckily, there was the Happiness Committee always there to support. Merel, Jeroen & CZ, keeping things Kawaii (and a little Metal) at HMI for years. Special thanks to CZ, the voice of ambition and compassion, both motivating and supportive. All the love to you guys and to little ZAGA!

One tradition at the HMI Lab are the *Thunder Thursday* hacking evenings. Who doesn't remember the Santa's Lap Simulator using Aura Interactors? Thanks to everybody who joined in on these, especially to the fanatics that stayed long enough to get stuff to actually work, like Daniel! I really got a kick out of these!

Another tradition that I'll fondly remember are the (board) game nights. Special thanks go to Jelte for keeping these (and other activities) alive, and for loving polar bears, and apologies for all those jokes. You sometimes point out how different we two are, manifested in the dispute about our favored style of games: collaborative VS competitive. What's better? We should fight it out… I mean figure that out together some day!

Speaking about games – Simon, Roelof, Camille, I kinda want this HMI CS clan to work out after all. Maybe we can get some sponsorship, any ideas? Oh, Lorenzo, there you are! Why yes, I am interested in more flag facts and dog pics! Thank you! Also, will we ever finish that duet-singing-agent? I would definitely get a kick out of that! Also, thanks to all the other foosball fanatics. I really got a kick out of beating you, especially when that got harder and harder.

Thanks to Charlotte and Alice for your help and support. Dennis, thank you for the mentoring on the fundamentals of interactions with VAs, on behavior realization, on ASAP and Flipper (also Daniel). Do you ever feel like having some sort of Stockholm Syndrome with the ASAP codebase? I feel it's holding me hostage, but I also really get

[3]https://www.youtube.com/watch?v=BZVBTtX92Nc

a kick out of it. Thank you, Bob. I really missed having someone to spar with in the home office. Thank you Mariët for the best feedback on paper drafts anyone could wish for.

Many thanks to my colleagues on the 4TU project, my fellow (former) PhD students Minha, Fran, Sima, Bernd and Michel. So many fond memories from the write-a-thons and winterschools (thank you Khiet and others for organizing those!). Especially thank you Minha, no need to share the many fun and inappropriate memories here, but they were always profound. Those made me grow as a researcher and as a person. I really enjoyed our collaboration on the social conflict study. No thanks to that lab supervisor guy in Eindhoven though for not allowing us to stay late.

Also thanks to the students I got to supervise, especially the research topics and master thesis projects. Thanks Emiel, Sander (who worked on the OpenIMPRESS system and studies), Joost, Karolina, Sanne and Wessel. Also special thanks to Sara. Our collaboration on your master Thesis with Merijn and the resulting paper on the Echoborg research contributed a lot to the last chapter of this thesis.

Dirk, thank you for being my supervisor and friend, for forming HMI into the loving, fun and interesting family that it is, creating the space for all these (and many more) experiences. And for being reliably calming, for somehow making things work out just fine in the end. Whenever we spoke I always felt better about everything afterwards. Perhaps sometimes I should have felt some urgency about finishing my thesis as well? Oh well, it was done when it was done.

Also thank you for making it possible for me to go abroad for the internships. Thanks to Mohamed and the ISIR team for having me there. Special thanks to Michael and the entire Sausalito team for the great time I had at FB (twice!).

Many thanks to my friends Emiel, David and Thomas for so many things. So many good memories from WorkSPACE to Fraksl, Sweden, to BYOB, to the Design Lab, keyboards and projects projects projects. Thank you for always having an open ear, I'm glad we're there for each other. I can't wait for us to find some time for some interesting stuff in the future, hopefully with all of us physically colocated again.

Finally, special thanks to my family for their support, especially to my parents for making it possible for me to pursue life in this way. Thank you Nele for hosting so many of the people mentioned above on our farm during that HMI outing on Osnabrück, which is still fondly remembered by everybody who was there!

Eternal thanks to Sarah, for the love, patience, support, understanding, critical view and advise, for your help with the illustrations, and for not getting too angry when I ruin the bread. Looking forward to some much-needed vacation together. I will reaaaallyyy get a kick out of that ♡

BIBLIOGRAPHY

Abella, B. S., N. Sandbo, P. Vassilatos, J. P. Alvarado, N. O'Hearn, H. N. Wigder, P. Hoffman, K. Tynus, T. L. V. Hoek, and L. B. Becker (2005). "Chest compression rates during cardiopulmonary resuscitation are suboptimal". In: *Circulation* 111.4, pages 428–434.

Adcock, M., S. Anderson, and B. Thomas (2013). "Remote{F}usion: real time depth camera fusion for remote collaboration on physical tasks". In: *Proc. 12th ACM SIGGRAPH Int. Conf. Virtual-Reality Contin. Its Appl. Ind. (VRCAI '13)*. New York, NY, USA: ACM, pages 235–242. DOI: http://dx.doi.org/10.1145/2534329.2534331.

Alem, L. and J. Li (2011). "A study of gestures in a video-mediated collaborative assembly task". In: *Adv. Human-Computer Interact.*, page 1.

Ambady, N. and R. Rosenthal (1993). "Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness." In: *Journal of personality and social psychology* 64.3, page 431.

Amores, J., X. Benavides, and P. Maes (2015). "Showme: A remote collaboration system that supports immersive gestural communication". In: *Proc. 33rd Annu. ACM Conf. Ext. Abstr. Hum. Factors Comput. Syst.* ACM, pages 1343–1348.

Argyle, M. (1988). *Bodily communication*. Methuen.

Argyle, M. and J. Dean (1965). "Eye-Contact, Distance and Affiliation". In: *Sociometry* 28.3, pages 289–304. ISSN: 00380431. DOI: 10.2307/2786027.

Aron, A., E. N. Aron, and D. Smollan (1992). "Inclusion of Other in the Self Scale and the structure of interpersonal closeness". In: *Journal of personality and social psychology* 63.4, page 596.

Bailenson, J. N., N. Yee, D. Merget, and R. Schroeder (2006). "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". In: *Presence: Teleoperators and Virtual Environments* 15.4, pages 359–372.

Bailenson, J. N., J. Blascovich, A. C. Beall, and J. M. Loomis (2001). *Equilibrium Theory Revisited: Mutual Gaze and Personal Space in Virtual Environments*. DOI: 10.1162/105474601753272844.

Baños, R. M., C. Botella, M. Alcañiz, V. Liaño, B. Guerrero, and B. Rey (2004). "Immersion and emotion: their impact on the sense of presence". In: *CyberPsychology Behav.* 7.6, pages 734–741.

Bartneck, C., D. Kulić, E. Croft, and S. Zoghbi (2009). "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International journal of social robotics* 1.1, pages 71–81.

Benedek, M. and C. Kaernbach (2010). "A continuous measure of phasic electrodermal activity". In: *Journal of neuroscience methods* 190.1, pages 80–91.

Bevacqua, E., I. Stanković, A. Maatallaoui, A. Nédélec, and P. De Loor (2014). "Effects of coupling in human-virtual agent body interaction". In: *International Conference on Intelligent Virtual Agents*. Springer, pages 54–63.

Biocca, F., J. Burgoon, C. Harms, M. Stoner, et al. (2001). "Criteria and scope conditions for a theory and measure of social presence". In: *Presence: Teleoperators and virtual environments*.

Biocca, F. and C. Harms (2002). "Defining and measuring social presence: Contribution to the networked minds theory and measure". In: *Proceedings of PRESENCE* 2002, pages 1–36.

Biocca, F. and C. Harms (2003). "Guide to the networked minds social presence inventory v. 1.2". In.

Biocca, F., C. Harms, and J. K. Burgoon (2003). "Toward a more robust theory and measure of social presence: Review and suggested criteria". In: *Presence: Teleoperators & virtual environments* 12.5, pages 456–480.

Biocca, F., C. Harms, and J. Gregg (2001). "The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity". In: *4th annual international workshop on presence, Philadelphia, PA*, pages 1–9.

Blankendaal, R., T. Bosse, C. Gerritsen, T. De Jong, and J. De Man (2015). "Are aggressive agents as scary as aggressive humans?" In: *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, pages 553–561.

Bosse, T., T. Hartmann, R. A. Blankendaal, N. Dokter, M. Otte, and L. Goedschalk (2018). "Virtually bad: A study on virtual agents that physically threaten human beings". In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, pages 1258–1266.

Bradley, M. M. and P. J. Lang (1994). "Measuring emotion: the self-assessment manikin and the semantic differential". In: *Journal of behavior therapy and experimental psychiatry* 25.1, pages 49–59.

Brooke, J. (1996). "SUS-A quick and dirty usability scale". In: *Usability Eval. Ind.* 189.194, pages 4–7.

Cafaro, A., H. H. Vilhjálmsson, T. Bickmore, D. Heylen, K. R. Jóhannsdóttir, and G. S. Valgardsson (2012). "First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters". In: *International conference on intelligent virtual agents*. Springer, pages 67–80.

Cafaro, A., H. H. Vilhjálmsson, T. Bickmore, D. Heylen, and C. Pelachaud (2014). "Representing communicative functions in saiba with a unified function markup language". In: *International Conference on Intelligent Virtual Agents*. Springer, pages 81–94.

Campbell, A. and J. P. Rushton (1978). "Bodily communication and personality". In: *British Journal of Social and Clinical Psychology* 17.1, pages 31–36.

Carless, D. (2006). "Differing perceptions in the feedback process". In: *Studies in higher education* 31.2, pages 219–233. ISSN: 0307-5079. DOI: 10.1080/03075070600572132. arXiv: arXiv:1011.1669v3.

Chénéchal, M. L., T. Duval, V. Gouranton, J. Royan, and B. Arnaldi (2016). "Vishnu: virtual immersive support for Helpi{N}g users: an interaction paradigm for collaborative remote guiding in mixed reality". In: *Proc. IEEE Third VR Int. Work. Collab. Virtual Environ.* IEEE, pages 9–12.

Clark, H. H. and S. E. Brennan (1991). "Grounding in communication." In.

Corti, K. (2015). "Developing the cyranoid method of mediated interpersonal communication in a social psychological context: applications in person perception, human-computer interaction, and first-person research". PhD thesis. The London School of Economics and Political Science (LSE).

Corti, K. and A. Gillespie (2015). "Offscreen and in the chair next to your: conversational agents speaking through actual human bodies". In: *Intelligent Virtual Agents*. Springer, pages 405–417.

Corti, K. and A. Gillespie (2016). "Co-constructing intersubjectivity with artificial conversational agents: people are more likely to initiate repairs of misunderstandings with agents represented as human". In: *Computers in Human Behavior* 58, pages 431–442.

Coutts, L. M. and F. W. Schneider (1976). "Affiliative conflict theory: An investigation of the intimacy equilibrium and compensation hypothesis." In: *J. Pers. Soc. Psychol.* 34.6, pages 1135–1142. ISSN: 0022-3514. DOI: 10.1037/0022-3514.34.6.1135.

Creswell, J., V. Clark, M. Gutmann, and W. Hanson (2008). "An expanded Typology for classifying mixed methods research into designs". In: *The mixed methods reader. Sage publications*, pages 159–96.

Dahlbäck, N., A. Jönsson, and L. Ahrenberg (1993). "Wizard of Oz studies - why and how". In: *Knowledge-based systems* 6.4, pages 258–266. ISSN: 09507051. DOI: 10.1016/0950-7051(93)90017-N.

De Dreu, C. K., D. Van Dierendonck, and M. T. Dijkstra (2004). "Conflict at work and individual well-being". In: *International journal of conflict management* 15.1, pages 6–26.

Diaz, G., J. Cooper, D. Kit, and M. Hayhoe (2013). "Real-time recording and classification of eye movements in an immersive virtual environment". In: *J. Vis.* 13.12, page 5.

Fairchild, A., S. Campion, A. García, R. Wolff, T. Fernando, and D. Roberts (2017). "A Mixed Reality Telepresence System for Collaborative Space Operation". In: *IEEE Trans. Circuits Syst. Video Technol.* 27.4, pages 814–827.

Falcone, S., J. Kolkmeier, M. Bruijnes, and D. Heylen (Sept. 2022). "The multimodal EchoBorg: not as smart as it looks". In: *Journal on Multimodal User Interfaces* 16.3, pages 293–302. ISSN: 1783-8738. DOI: 10.1007/s12193-022-00389-z. URL: https://doi.org/10.1007/s12193-022-00389-z.

Feldmann, I., W. Waizenegger, N. Atzpadin, and O. Schreer (2010). "Real-time depth estimation for immersive {3D} videoconferencing". In: *3DTV-Conference True Vision-Capture, Transm. Disp. 3D Video (3DTV-CON), 2010*. IEEE, pages 1–4.

Findlen, R. A. (2000). "Conflict: The skeleton in academe's closet". In: *New Directions for Community Colleges* 109, pages 41–49.

Ford, B., P. Srisuresh, and D. Kegel (2005). *Peer-to-Peer Communication Across Network Address Translators*. URL: http://www.brynosaurus.com/pub/net/p2pnat/.

Fussell, S. R., L. D. Setlock, and R. E. Kraut (2003). "Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks". In: *Proc. SIGCHI Conf. Hum. factors Comput. Syst.* ACM, pages 513–520.

Gajadhar, B. J., Y. A. De Kort, and W. A. IJsselsteijn (2008). "Shared fun is doubled fun: player enjoyment as a function of social setting". In: *International Conference on Fun and Games*. Springer, pages 106–117.

Gao, L., H. Bai, G. Lee, and M. Billinghurst (2016). "An oriented point-cloud view for MR remote collaboration". In: *SIGGRAPH ASIA 2016 Mob. Graph. Interact. Appl.* ACM, page 8.

Garau, M., M. Slater, D.-P. Pertaub, and S. Razzaque (2005). "The responses of people to virtual humans in an immersive virtual environment". In: *Presence: Teleoperators & Virtual Environments* 14.1, pages 104–116.

Giesselink, S. (2018). "Teach me CPR now! Augmented learning for one-shot teaching of cardiopulmonary resuscitation in out-of-hospital cardiac arrest". Master's thesis. University of Twente.

Gillespie, A. and K. Corti (2016). "The body that speaks: recombining bodies and speech sources in unscripted face-to-face communication". In: *Frontiers in psychology* 7, page 1300.

Groen, E. L. and J. E. Bos (2008). "Simulator sickness depends on frequency of the simulator motion mismatch: An observation". In: *Presence* 17.6, pages 584–593.

Guadagno, R. E., J. Blascovich, J. N. Bailenson, and C. McCall (2007). "Virtual humans and persuasion: The effects of agency and behavioral realism". In: *Media Psychology* 10.1, pages 1–22.

Guadagno, R. E. and R. B. Cialdini (2002). "Online persuasion: An examination of gender differences in computer-mediated interpersonal influence." In: *Group Dynamics: Theory, Research, and Practice* 6.1, page 38.

Hall, E. T. (1969). *The hidden dimension*. Volume 1990. Anchor Books New York.

Hardy, J. D. and T. W. Smith (1988). "Cynical hostility and vulnerability to disease: Social support, life stress, and physiological response to conflict." In: *Health Psychology* 7.5, page 447.

Harms, C. and F. Biocca (2004). "Internal consistency and reliability of the networked minds measure of social presence". In.

Harmsen, E. (2018). "OpenIMPRESS: an open immersive telepresence system". Master's thesis. University of Twente.

Hart, S. G. (2006). "NASA-task load index (NASA-TLX); 20 years later". In: *Proc. Hum. factors Ergon. Soc. Annu. Meet.* Volume 50. 9. Sage Publications Sage CA: Los Angeles, CA, pages 904–908.

Hashimoto, T., T. Mojaverian, and H. S. Kim (2012). "Culture, interpersonal stress, and psychological distress". In: *Journal of Cross-Cultural Psychology* 43.4, pages 527–532.

Heylen, D., S. Kopp, S. C. Marsella, C. Pelachaud, and H. Vilhjálmsson (2008). "The next step towards a function markup language". In: *International Workshop on Intelligent Virtual Agents*. Springer, pages 270–280.

Holmberg, M., S. Holmberg, and J. Herlitz (2001). "Factors modifying the effect of bystander cardiopulmonary resuscitation on survival in out-of-hospital cardiac arrest patients in Sweden". In: *Eur. Heart J.* 22.6, pages 511–519.

Holmberg, M., S. Holmberg, J. Herlitz, B. Gårdelöv, et al. (1998). "Survival after cardiac arrest outside hospital in Sweden". In: *Resuscitation* 36.1, pages 29–36.

Huang, W. and L. Alem (2011). "Supporting hand gestures in mobile remote collaboration: a usability evaluation". In: *Proc. 25th BCS Conf. Human-Computer Interact.* British Computer Society, pages 211–216.

Huisman, G., J. Kolkmeier, and D. Heylen (2014a). "With Us or Against Us: Simulated Social Touch by Virtual Agents in a Cooperative or Competitive Setting". In: *Intell. Virtual Agents*. Springer, pages 204–213. ISBN: 3319097660.

Huisman, G., J. Kolkmeier, and D. Heylen (2014b). "With us or against us: simulated social touch by virtual agents in a cooperative or competitive setting". In: *International Conference on Intelligent Virtual Agents*. Springer, pages 204–213.

Ibayashi, H., Y. Sugiura, D. Sakamoto, N. Miyata, M. Tada, T. Okuma, T. Kurata, M. Mochimaru, and T. Igarashi (2015). "Dollhouse VR: A multi-view, multi-user collaborative design workspace with VR technology". In: *SIGGRAPH Asia 2015 Posters, SA 2015*, pages 2–3. ISSN: 1882-7764. DOI: 10.1145/2820926.2820948.

IJsselsteijn, W., Y. De Kort, and K. Poels (2013). "The game experience questionnaire". In.

Järvelä, S., B. Cowley, M. Salminen, G. Jacucci, J. Hamari, and N. Ravaja (2021). "Augmented Virtual Reality Meditation: Shared Dyadic Biofeedback Increases Social Presence Via Respiratory Synchrony". In: *ACM Transactions on Social Computing* 4.2, pages 1–19.

Jordan, P., O. Mubin, M. Obaid, and P. A. Silva (2018). “Exploring the referral and usage of science fiction in HCI literature”. In: *International Conference of Design, User Experience, and Usability*. Springer, pages 19–38.

Kasahara, S. and J. Rekimoto (2014). “JackIn: integrating first-person view with out-of-body vision generation for human-human augmentation”. In: *Proc. 5th Augment. Hum. Int. Conf.* ACM, page 46.

Kasahara, S. and J. Rekimoto (2015). “JackIn head: immersive visual telepresence system with omnidirectional wearable camera for remote collaboration”. In: *Proc. 21st ACM Symp. Virtual Real. Softw. Technol.* ACM, pages 217–225.

Kastanis, I. and M. Slater (2012). “Reinforcement Learning Utilizes Proxemics : An Avatar Learns to Manipulate the Position of People in Immersive Virtual Reality”. In: *Trans. Appl. Percept.* 9, 3:1–3:15. ISSN: 1544-3558. DOI: 10.1145/2134203.2134206.

Kätsyri, J., R. Hari, N. Ravaja, and L. Nummenmaa (2013). “The opponent matters: elevated fMRI reward responses to winning against a human versus a computer opponent during interactive video game playing”. In: *Cerebral Cortex* 23.12, pages 2829–2839.

Koenker, R. (2019). *quantreg: Quantile Regression*. R package version 5.42. URL: https://CRAN.R-project.org/package=quantreg.

Kolkmeier, J., M. Bruijnes, and D. Reidsma (2017). “A demonstration of the ASAP Realizer-Unity3D bridge for virtual and mixed reality applications”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 223–226.

Kolkmeier, J., M. Bruijnes, D. Reidsma, and D. Heylen (2017). “An ASAP realizer-unity3d bridge for virtual and mixed reality applications”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 227–230.

Kolkmeier, J., E. Harmsen, S. Giesselink, D. Reidsma, M. Theune, and D. Heylen (2018). “With a little help from a holographic friend: The OpenIMPRESS mixed reality telepresence toolkit for remote collaboration systems”. In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–11.

Kolkmeier, J., M. Lee, and D. Heylen (2017). “Moral conflicts in VR: Addressing grade disputes with a virtual trainer”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 231–234.

Kolkmeier, J., J. Vroon, and D. Heylen (2016). “Interacting with virtual agents in shared space: Single and joint effects of gaze and proxemics”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 1–14.

Komiyama, R., T. Miyaki, and J. Rekimoto (2017). “JackIn space: designing a seamless transition between first and third person view for effective telepresence collaborations”. In: *Proc. 8th Augment. Hum. Int. Conf.* ACM, page 14.

Kontogiorgos, D., S. van Waveren, O. Wallberg, A. Pereira, I. Leite, and J. Gustafson (2020). “Embodiment effects in interactions with failing robots”. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14.

Koruth, K. J., A. Lang, R. F. Potter, and R. L. Bailey (2015). “A comparative analysis of dynamic and static indicators of parasympathetic and sympathetic nervous system activation during TV viewing”. In: *Communication Methods and Measures* 9.1-2, pages 78–100.

Kowalski, M., J. Naruniec, and M. Daniluk (2015). “Live Scan3D: A Fast and Inexpensive 3D Data Acquisition System for Multiple Kinect v2 Sensors”. In: *3D Vis. (3DV), 2015 Int. Conf.* IEEE, pages 318–325.

Kreibig, S. D. (2010). “Autonomic nervous system activity in emotion: A review”. In: *Biological psychology* 84.3, pages 394–421.

Kristoffersson, A., S. Coradeschi, and A. Loutfi (2013). “A review of mobile robotic telepresence”. In: *Adv. Human-Computer Interact.* 2013, page 3.

Kristoffersson, A., K. S. Eklundh, and A. Loutfi (2013). “Measuring the quality of interaction in mobile robotic telepresence: A pilot’s perspective”. In: *International Journal of Social Robotics* 5.1, pages 89–101.

Lang, P. J., M. K. Greenwald, M. M. Bradley, and A. O. Hamm (1993). “Looking at pictures: Affective, facial, visceral, and behavioral reactions”. In: *Psychophysiology* 30.3, pages 261–273.

Lawrence, J., M. Snaith, B. Konat, K. Budzynska, and C. Reed (2017). “Debating technology for dialogical argument: Sensemaking, engagement, and analytics”. In: *ACM Transactions on Internet Technology (TOIT)* 17.3, pages 1–23.

Lee, G., T. Teo, S. Kim, and M. Billinghurst (2017). “Mixed reality collaboration through sharing a live panorama”. In: *SIGGRAPH Asia 2017 Mob. Graph. Interact. Appl.* New York, NY, USA: ACM. DOI: https://doi.org/10.1145/3132787.3139203.

Lee, M., J. Kolkmeier, D. Heylen, and W. IJsselsteijn (2021). “Who Makes Your Heart Beat? What Makes You Sweat? Social Conflict in Virtual Reality for Educators”. In: *Frontiers in Psychology* 12.

Levesque, M. J. and D. A. Kenny (1993). “Accuracy of behavioral predictions at zero acquaintance: A social relations analysis.” In: *Journal of Personality and Social Psychology* 65.6, page 1178.

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

Lim, S. and B. Reeves (2010). “Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player”. In: *International Journal of Human-Computer Studies* 68.1-2, pages 57–68.

Linssen, J. and M. Theune (2017). “R3d3: The rolling receptionist robot with double dutch dialogue”. In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. Springer, pages 189–190.

Lucas, G. M., J. Gratch, A. King, and L.-P. Morency (2014). “It’s only a computer: Virtual humans increase willingness to disclose”. In: *Computers in Human Behavior* 37, pages 94–100.

Maat, M., K. P. Truong, D. Heylen, and M. Ter Maat (2010). “How turn-taking strategies influence users’ impressions of an agent”. In: *Intell. Virtual Agents*. Springer, pages 441–453. ISBN: 3642158919.

Maat, M. t. and D. Heylen (2011). “Flipper: An information state component for spoken dialogue systems”. In: *International Workshop on Intelligent Virtual Agents*. Springer, pages 470–472.

Maimone, A. and H. Fuchs (2011). “Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras”. In: *Mix. Augment. Real. (ISMAR), 2011 10th IEEE Int. Symp.* IEEE, pages 137–146.

Mazgaj, M., A. D’Amato, J. Elson, and D. Derrick (2021). “Exploring the Effects of Real-Time Hologram Communication on Social Presence, Novelty, and Affect”. In: *Proceedings of the 54th Hawaii International Conference on System Sciences*, page 471.

Meehan, M., B. Insko, M. Whitton, and F. P. Brooks Jr (2002). “Physiological measures of presence in stressful virtual environments”. In: *Acm transactions on graphics (tog)*. Volume 21. 3. ACM, pages 645–652.

Melo, C. M. de and J. Gratch (2015). “Beyond Believability: Quantifying the Differences Between Real and Virtual Humans”. In: *Intelligent Virtual Agents*. Edited by W.-P. Brinkman, J. Broekens, and D. Heylen. Cham: Springer International Publishing, pages 109–118. ISBN: 978-3-319-21996-7.

Mezo, P. G., R. E. McCabe, M. M. Antony, and K. Burns (2005). “Psychometric validation of a monitoring-blunting measure for social anxiety disorder: the coping styles questionnaire for social situations (CSQSS)”. In: *Depression and anxiety* 22.1, pages 20–27.

Naumann, L. P., S. Vazire, P. J. Rentfrow, and S. D. Gosling (2009). “Personality judgments based on physical appearance”. In: *Personality and social psychology bulletin* 35.12, pages 1661–1671.

Neff, M., N. Toothman, R. Bowmani, J. E. F. Tree, and M. A. Walker (2011). “Don’t scratch! Self-adaptors reflect emotional stability”. In: *International Workshop on Intelligent Virtual Agents*. Springer, pages 398–411.

Neff, M., Y. Wang, R. Abbott, and M. Walker (2010). “Evaluating the effect of gesture and language on personality perception in conversational agents”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 222–235.

Nicholson, S. (2016). “The State of Escape: Escape Room Design and Facilities”. In: *Meaningful Play 2016. Lansing, Michigan.*

Normoyle, A., F. Liu, M. Kapadia, N. I. Badler, and S. Jörg (2013). “The effect of posture and dynamics on the perception of emotion”. In: *Proceedings of the ACM Symposium on Applied Perception*. ACM, pages 91–98.

Nowak, K. L. and F. Biocca (2003). “The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments”. In: *Presence: Teleoperators & Virtual Environments* 12.5, pages 481–494.

Oda, O., C. Elvezio, M. Sukan, S. Feiner, and B. Tversky (2015). "Virtual Replicas for Remote Assistance in Virtual and Augmented Reality". In: *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol.* New York, NY, USA: ACM, pages 405–415.

Oda, O. and S. Feiner (2012). "{3D} referencing techniques for physical objects in shared augmented reality". In: *Proc. 2012 IEEE Int. Symp. Mix. Augment. Real.* Pages 207–215.

Orts-Escolano, S., C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, et al. (2016). "Holoportation: Virtual {3D} teleportation in real-time". In: *Proc. 29th Annu. Symp. User Interface Softw. Technol.* ACM, pages 741–754.

Osmers, N., M. Prilla, O. Blunk, G. George Brown, M. Janßen, and N. Kahrl (2021). "The Role of Social Presence for Cooperation in Augmented Reality on Head Mounted Devices: A Literature Review". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17.

Patterson, M. L. (1977). "Interpersonal distance, affect, and equilibrium theory". In: *J. Soc. Psychol.* 101.2, pages 205–214. ISSN: 0022-4545. DOI: 10.1080/00224545.1977.9924008.

Piumsomboon, T., A. Dey, B. Ens, Y. Lee, G. Lee, and M. Billinghurst (2017). "Exploring enhancements for remote mixed reality collaboration". In: *SIGGRAPH Asia 2017 Mob. Graph. Interact. Appl.* New York, NY, USA: ACM.

Piumsomboon, T., G. Lee, J. Hart, B. Ens, R. Lindeman, B. Thomas, and M. Billinghurst (2018). "Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration". In: *Proc. 2018 CHI Conf. Hum. Factors Comput. Syst.* New York, NY, USA: ACM.

Poelman, R., O. Akman, S. Lukosch, and P. Jonker (2012). "As if being there: mediated reality for crime scene investigation". In: *Proc. ACM 2012 Conf. Comput. Support. Coop. Work.* ACM, pages 1267–1276.

Poggi, I., F. D'Errico, and L. Vincze (2013). "Comments by words, face and body". In: *Journal on Multimodal User Interfaces* 7.1, pages 67–78.

Poppe, R., M. ter Maat, and D. Heylen (2012). "Online behavior evaluation with the Switching Wizard of Oz". In: *International Conference on Intelligent Virtual Agents*. Springer, pages 486–488.

Rae, I., B. Mutlu, and L. Takayama (2014). "Bodies in motion: mobility, presence, and task awareness in telepresence". In: *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst.* ACM, pages 2153–2162.

Ravaja, N. (2009). "The psychophysiology of digital gaming: The effect of a non co-located opponent". In: *Media Psychology* 12.3, pages 268–294.

Regenbrecht, H., T. Lum, P. Kohler, C. Ott, M. Wagner, W. Wilke, and E. Mueller (2004). "Using Augmented Virtuality for Remote Collaboration". In: *Presence Teleoperators Virtual Environ.* 13.3, pages 338–354. ISSN: 1054-7460. DOI: 10 . 1162 / 1054746041422334.

Reidsma, D., I. de Kok, D. Neiberg, S. C. Pammi, B. van Straalen, K. Truong, and H. van Welbergen (2011). "Continuous interaction with a virtual human". In: *Journal on Multimodal User Interfaces* 4.2, pages 97–118.

Reidsma, D. and H. van Welbergen (2013). "AsapRealizer in practice–A modular and extensible architecture for a BML Realizer". In: *Entertainment computing* 4.3, pages 157–169.

Ribeiro, F. (2015). *UMA - Unity Multipiurpose Avatar System*. URL: https://github.com/huika/UMA.

Rogers, B. and M. Graham (1979). "Motion parallax as an independent cue for depth perception". In: *Perception* 8.2, pages 125–134.

Rosenfeld, H. M., B. E. Breck, S. H. Smith, and S. Kehoe (1984). "Intimacy-mediators of the proximity-gaze compensation effect: Movement, conversational role, acquaintance, and gender". In: *J. Nonverbal Behav.* 8.4, pages 235–249. ISSN: 01915886. DOI: 10.1007/BF00985981.

Schubert, T., F. Friedmann, and H. Regenbrecht (2001). "The experience of presence: Factor analytic insights". In: *Presence Teleoperators Virtual Environ.* 10.3, pages 266–281.

Slater, M. (2009). "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1535, pages 3549–3557.

Smith, H. J. and M. Neff (2017). "Understanding the impact of animated gesture performance on personality perceptions". In: *ACM Transactions on Graphics (TOG)* 36.4, page 49.

Sodhi, R. S., B. R. Jones, D. Forsyth, B. P. Bailey, and G. Maciocci (2013). "BeThere: 3D mobile collaboration with spatial input". In: *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.* ACM, pages 179–188.

Steed, A., W. Steptoe, W. Oyekoya, F. Pece, T. Weyrich, J. Kautz, D. Friedman, A. Peer, M. Solazzi, F. Tecchia, et al. (2012). "Beaming: an asymmetric telepresence system". In: *IEEE Comput. Graph. Appl.* 32.6, pages 10–17.

Stiell, I. G., S. P. Brown, J. Christenson, S. Cheskes, G. Nichol, J. Powell, B. Bigham, L. J. Morrison, J. Larsen, E. Hess, et al. (2012). "What is the role of chest compression depth during out-of-hospital cardiac arrest resuscitation?" In: *Crit. Care Med.* 40.4, page 1192.

Suzuki, S.-I. (1997). "Development of a new psychological stress response scale (SRS-18) and investigation of the reliability and the validity". In: *Jpn J Behav Med* 4, pages 22–29.

Tait, M. and M. Billinghurst (2015). "The effect of view independence in a collaborative ar system". In: *Comput. Support. Coop. Work* 24.6, pages 563–589.

Tarvainen, M. P., J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen (2014). "Kubios HRV–heart rate variability analysis software". In: *Computer methods and programs in biomedicine* 113.1, pages 210–220.

Tecchia, F., L. Alem, and W. Huang (2012). “3D helping hands: a gesture based MR system for remote collaboration”. In: *Proc. 11th ACM SIGGRAPH Int. Conf. Virtual-Reality Contin. its Appl. Ind.* ACM, pages 323–328.

Ter Maat, M., K. P. Truong, and D. Heylen (2010). “How turn-taking strategies influence users’ impressions of an agent”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 441–453.

Thomson, J. J. (1976). “Killing, letting die, and the trolley problem”. In: *The Monist* 59.2, pages 204–217.

Trahanias, P., W. Burgard, A. Argyros, D. Hahnel, H. Baltzakis, P. Pfaff, and C. Stachniss (2005). “TOURBOT and WebFAIR: Web-operated mobile robots for tele-presence in populated exhibitions”. In: *IEEE Robot. Autom. Mag.* 12.2, pages 77–89.

Valenzuela, T. D., D. J. Roe, S. Cretin, D. W. Spaite, and M. P. Larsen (1997). “Estimating effectiveness of cardiac arrest interventions”. In: *Circulation* 96.10, pages 3308–3313.

Van Welbergen, H., D. Reidsma, and S. Kopp (2012). “An incremental multimodal realizer for behavior co-articulation and coordination”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 175–188.

Van Welbergen, H., R. Yaghoubzadeh, and S. Kopp (2014). “AsapRealizer 2.0: The next steps in fluent behavior realization for ECAs”. In: *International Conference on Intelligent Virtual Agents*. Springer, pages 449–462.

Vilhjálmsson, H., N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, et al. (2007). “The behavior markup language: Recent developments and challenges”. In: *International Workshop on Intelligent Virtual Agents*. Springer, pages 99–111.

Waalewijn, R. A., J. G. P. Tijssen, and R. W. Koster (2001). “Bystander initiated actions in out-of-hospital cardiopulmonary resuscitation: results from the Amsterdam Resuscitation Study (ARRESUST)”. In: *Resuscitation* 50.3, pages 273–279.

Waterschoot, J. van, M. Bruijnes, J. Flokstra, D. Reidsma, D. Davison, M. Theune, and D. Heylen (2018). “Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents”. In: *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. ACM, pages 43–50.

Weibel, D., B. Wissmath, S. Habegger, Y. Steiner, and R. Groner (2008). “Playing online games against computer-vs. human-controlled opponents: Effects on presence, flow, and enjoyment”. In: *Computers in Human Behavior* 24.5, pages 2274–2291.

SIKS DISSERTATION SERIES

2011

01 Botond Cseke (RUN), Variational Algorithms for Bayesian Inference in Latent Gaussian Models
02 Nick Tinnemeier (UU), Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language
03 Jan Martijn van der Werf (TUE), Compositional Design and Verification of Component-Based Information Systems
04 Hado van Hasselt (UU), Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference
05 Bas van der Raadt (VU), Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.
06 Yiwen Wang (TUE), Semantically-Enhanced Recommendations in Cultural Heritage
07 Yujia Cao (UT), Multimodal Information Presentation for High Load Human Computer Interaction
08 Nieske Vergunst (UU), BDI-based Generation of Robust Task-Oriented Dialogues
09 Tim de Jong (OU), Contextualised Mobile Media for Learning
10 Bart Bogaert (UvT), Cloud Content Contention
11 Dhaval Vyas (UT), Designing for Awareness: An Experience-focused HCI Perspective
12 Carmen Bratosin (TUE), Grid Architecture for Distributed Process Mining
13 Xiaoyu Mao (UvT), Airport under Control. Multiagent Scheduling for Airport Ground Handling
14 Milan Lovric (EUR), Behavioral Finance and Agent-Based Artificial Markets
15 Marijn Koolen (UvA), The Meaning of Structure: the Value of Link Evidence for Information Retrieval
16 Maarten Schadd (UM), Selective Search in Games of Different Complexity
17 Jiyin He (UVA), Exploring Topic Structure: Coherence, Diversity and Relatedness
18 Mark Ponsen (UM), Strategic Decision-Making in complex games
19 Ellen Rusman (OU), The Mind's Eye on Personal Profiles
20 Qing Gu (VU), Guiding service-oriented software engineering - A view-based approach
21 Linda Terlouw (TUD), Modularization and Specification of Service-Oriented Systems
22 Junte Zhang (UVA), System Evaluation of Archival Description and Access
23 Wouter Weerkamp (UVA), Finding People and their Utterances in Social Media
24 Herwin van Welbergen (UT), Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior
25 Syed Waqar ul Qounain Jaffry (VU), Analysis and Validation of Models for Trust Dynamics

26 Matthijs Aart Pontier (VU), Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots
27 Aniel Bhulai (VU), Dynamic website optimization through autonomous management of design patterns
28 Rianne Kaptein (UVA), Effective Focused Retrieval by Exploiting Query Context and Document Structure
29 Faisal Kamiran (TUE), Discrimination-aware Classification
30 Egon van den Broek (UT), Affective Signal Processing (ASP): Unraveling the mystery of emotions
31 Ludo Waltman (EUR), Computational and Game-Theoretic Approaches for Modeling Bounded Rationality
32 Nees-Jan van Eck (EUR), Methodological Advances in Bibliometric Mapping of Science
33 Tom van der Weide (UU), Arguing to Motivate Decisions
34 Paolo Turrini (UU), Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations
35 Maaike Harbers (UU), Explaining Agent Behavior in Virtual Training
36 Erik van der Spek (UU), Experiments in serious game design: a cognitive approach
37 Adriana Burlutiu (RUN), Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference
38 Nyree Lemmens (UM), Bee-inspired Distributed Optimization
39 Joost Westra (UU), Organizing Adaptation using Agents in Serious Games
40 Viktor Clerc (VU), Architectural Knowledge Management in Global Software Development
41 Luan Ibraimi (UT), Cryptographically Enforced Distributed Data Access Control
42 Michal Sindlar (UU), Explaining Behavior through Mental State Attribution
43 Henk van der Schuur (UU), Process Improvement through Software Operation Knowledge
44 Boris Reuderink (UT), Robust Brain-Computer Interfaces
45 Herman Stehouwer (UvT), Statistical Language Models for Alternative Sequence Selection
46 Beibei Hu (TUD), Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work
47 Azizi Bin Ab Aziz (VU), Exploring Computational Models for Intelligent Support of Persons with Depression
48 Mark Ter Maat (UT), Response Selection and Turn-taking for a Sensitive Artificial Listening Agent
49 Andreea Niculescu (UT), Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012
01 Terry Kakeeto (UvT), Relationship Marketing for SMEs in Uganda
02 Muhammad Umair (VU), Adaptivity, emotion, and Rationality in Human and Ambient Agent Models
03 Adam Vanya (VU), Supporting Architecture Evolution by Mining Software Repositories

04 Jurriaan Souer (UU), Development of Content Management System-based Web Applications
05 Marijn Plomp (UU), Maturing Interorganisational Information Systems
06 Wolfgang Reinhardt (OU), Awareness Support for Knowledge Workers in Research Networks
07 Rianne van Lambalgen (VU), When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions
08 Gerben de Vries (UVA), Kernel Methods for Vessel Trajectories
09 Ricardo Neisse (UT), Trust and Privacy Management Support for Context-Aware Service Platforms
10 David Smits (TUE), Towards a Generic Distributed Adaptive Hypermedia Environment
11 J.C.B. Rantham Prabhakara (TUE), Process Mining in the Large: Preprocessing, Discovery, and Diagnostics
12 Kees van der Sluijs (TUE), Model Driven Design and Data Integration in Semantic Web Information Systems
13 Suleman Shahid (UvT), Fun and Face: Exploring non-verbal expressions of emotion during playful interactions
14 Evgeny Knutov (TUE), Generic Adaptation Framework for Unifying Adaptive Web-based Systems
15 Natalie van der Wal (VU), Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.
16 Fiemke Both (VU), Helping people by understanding them - Ambient Agents supporting task execution and depression treatment
17 Amal Elgammal (UvT), Towards a Comprehensive Framework for Business Process Compliance
18 Eltjo Poort (VU), Improving Solution Architecting Practices
19 Helen Schonenberg (TUE), What's Next? Operational Support for Business Process Execution
20 Ali Bahramisharif (RUN), Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing
21 Roberto Cornacchia (TUD), Querying Sparse Matrices for Information Retrieval
22 Thijs Vis (UvT), Intelligence, politie en veiligheidsdienst: verenigbare grootheden?
23 Christian Muehl (UT), Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction
24 Laurens van der Werff (UT), Evaluation of Noisy Transcripts for Spoken Document Retrieval
25 Silja Eckartz (UT), Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application
26 Emile de Maat (UVA), Making Sense of Legal Text
27 Hayrettin Gurkok (UT), Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games
28 Nancy Pascall (UvT), Engendering Technology Empowering Women
29 Almer Tigelaar (UT), Peer-to-Peer Information Retrieval
30 Alina Pommeranz (TUD), Designing Human-Centered Systems for Reflective Decision Making

31 Emily Bagarukayo (RUN), A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure
32 Wietske Visser (TUD), Qualitative multi-criteria preference representation and reasoning
33 Rory Sie (OUN), Coalitions in Cooperation Networks (COCOON)
34 Pavol Jancura (RUN), Evolutionary analysis in PPI networks and applications
35 Evert Haasdijk (VU), Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics
36 Denis Ssebugwawo (RUN), Analysis and Evaluation of Collaborative Modeling Processes
37 Agnes Nakakawa (RUN), A Collaboration Process for Enterprise Architecture Creation
38 Selmar Smit (VU), Parameter Tuning and Scientific Testing in Evolutionary Algorithms
39 Hassan Fatemi (UT), Risk-aware design of value and coordination networks
40 Agus Gunawan (UvT), Information Access for SMEs in Indonesia
41 Sebastian Kelle (OU), Game Design Patterns for Learning
42 Dominique Verpoorten (OU), Reflection Amplifiers in self-regulated Learning
43 Withdrawn
44 Anna Tordai (VU), On Combining Alignment Techniques
45 Benedikt Kratz (UvT), A Model and Language for Business-aware Transactions
46 Simon Carter (UVA), Exploration and Exploitation of Multilingual Data for Statistical Machine Translation
47 Manos Tsagkias (UVA), Mining Social Media: Tracking Content and Predicting Behavior
48 Jorn Bakker (TUE), Handling Abrupt Changes in Evolving Time-series Data
49 Michael Kaisers (UM), Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions
50 Steven van Kervel (TUD), Ontologogy driven Enterprise Information Systems Engineering
51 Jeroen de Jong (TUD), Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching

2013
01 Viorel Milea (EUR), News Analytics for Financial Decision Support
02 Erietta Liarou (CWI), MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing
03 Szymon Klarman (VU), Reasoning with Contexts in Description Logics
04 Chetan Yadati (TUD), Coordinating autonomous planning and scheduling
05 Dulce Pumareja (UT), Groupware Requirements Evolutions Patterns
06 Romulo Goncalves (CWI), The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience
07 Giel van Lankveld (UvT), Quantifying Individual Player Differences
08 Robbert-Jan Merk (VU), Making enemies: cognitive modeling for opponent agents in fighter pilot simulators
09 Fabio Gori (RUN), Metagenomic Data Analysis: Computational Methods and Applications

10 Jeewanie Jayasinghe Arachchige (UvT), A Unified Modeling Framework for Service Design.
11 Evangelos Pournaras (TUD), Multi-level Reconfigurable Self-organization in Overlay Services
12 Marian Razavian (VU), Knowledge-driven Migration to Services
13 Mohammad Safiri (UT), Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly
14 Jafar Tanha (UVA), Ensemble Approaches to Semi-Supervised Learning Learning
15 Daniel Hennes (UM), Multiagent Learning - Dynamic Games and Applications
16 Eric Kok (UU), Exploring the practical benefits of argumentation in multi-agent deliberation
17 Koen Kok (VU), The PowerMatcher: Smart Coordination for the Smart Electricity Grid
18 Jeroen Janssens (UvT), Outlier Selection and One-Class Classification
19 Renze Steenhuizen (TUD), Coordinated Multi-Agent Planning and Scheduling
20 Katja Hofmann (UvA), Fast and Reliable Online Learning to Rank for Information Retrieval
21 Sander Wubben (UvT), Text-to-text generation by monolingual machine translation
22 Tom Claassen (RUN), Causal Discovery and Logic
23 Patricio de Alencar Silva (UvT), Value Activity Monitoring
24 Haitham Bou Ammar (UM), Automated Transfer in Reinforcement Learning
25 Agnieszka Anna Latoszek-Berendsen (UM), Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System
26 Alireza Zarghami (UT), Architectural Support for Dynamic Homecare Service Provisioning
27 Mohammad Huq (UT), Inference-based Framework Managing Data Provenance
28 Frans van der Sluis (UT), When Complexity becomes Interesting: An Inquiry into the Information eXperience
29 Iwan de Kok (UT), Listening Heads
30 Joyce Nakatumba (TUE), Resource-Aware Business Process Management: Analysis and Support
31 Dinh Khoa Nguyen (UvT), Blueprint Model and Language for Engineering Cloud Applications
32 Kamakshi Rajagopal (OUN), Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development
33 Qi Gao (TUD), User Modeling and Personalization in the Microblogging Sphere
34 Kien Tjin-Kam-Jet (UT), Distributed Deep Web Search
35 Abdallah El Ali (UvA), Minimal Mobile Human Computer Interaction
36 Than Lam Hoang (TUe), Pattern Mining in Data Streams
37 Dirk Börner (OUN), Ambient Learning Displays
38 Eelco den Heijer (VU), Autonomous Evolutionary Art
39 Joop de Jong (TUD), A Method for Enterprise Ontology based Design of Enterprise Information Systems
40 Pim Nijssen (UM), Monte-Carlo Tree Search for Multi-Player Games
41 Jochem Liem (UVA), Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning
42 Léon Planken (TUD), Algorithms for Simple Temporal Reasoning

43 Marc Bron (UVA), Exploration and Contextualization through Interaction and Concepts

2014

01 Nicola Barile (UU), Studies in Learning Monotone Models from Data
02 Fiona Tuliyano (RUN), Combining System Dynamics with a Domain Modeling Method
03 Sergio Raul Duarte Torres (UT), Information Retrieval for Children: Search Behavior and Solutions
04 Hanna Jochmann-Mannak (UT), Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation
05 Jurriaan van Reijsen (UU), Knowledge Perspectives on Advancing Dynamic Capability
06 Damian Tamburri (VU), Supporting Networked Software Development
07 Arya Adriansyah (TUE), Aligning Observed and Modeled Behavior
08 Samur Araujo (TUD), Data Integration over Distributed and Heterogeneous Data Endpoints
09 Philip Jackson (UvT), Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language
10 Ivan Salvador Razo Zapata (VU), Service Value Networks
11 Janneke van der Zwaan (TUD), An Empathic Virtual Buddy for Social Support
12 Willem van Willigen (VU), Look Ma, No Hands: Aspects of Autonomous Vehicle Control
13 Arlette van Wissen (VU), Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains
14 Yangyang Shi (TUD), Language Models With Meta-information
15 Natalya Mogles (VU), Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare
16 Krystyna Milian (VU), Supporting trial recruitment and design by automatically interpreting eligibility criteria
17 Kathrin Dentler (VU), Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability
18 Mattijs Ghijsen (UVA), Methods and Models for the Design and Study of Dynamic Agent Organizations
19 Vinicius Ramos (TUE), Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support
20 Mena Habib (UT), Named Entity Extraction and Disambiguation for Informal Text: The Missing Link
21 Kassidy Clark (TUD), Negotiation and Monitoring in Open Environments
22 Marieke Peeters (UU), Personalized Educational Games - Developing agent-supported scenario-based training
23 Eleftherios Sidirourgos (UvA/CWI), Space Efficient Indexes for the Big Data Era
24 Davide Ceolin (VU), Trusting Semi-structured Web Data
25 Martijn Lappenschaar (RUN), New network models for the analysis of disease interaction
26 Tim Baarslag (TUD), What to Bid and When to Stop
27 Rui Jorge Almeida (EUR), Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty
28 Anna Chmielowiec (VU), Decentralized k-Clique Matching

29 Jaap Kabbedijk (UU), Variability in Multi-Tenant Enterprise Software
30 Peter de Cock (UvT), Anticipating Criminal Behaviour
31 Leo van Moergestel (UU), Agent Technology in Agile Multiparallel Manufacturing and Product Support
32 Naser Ayat (UvA), On Entity Resolution in Probabilistic Data
33 Tesfa Tegegne (RUN), Service Discovery in eHealth
34 Christina Manteli (VU), The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.
35 Joost van Ooijen (UU), Cognitive Agents in Virtual Worlds: A Middleware Design Approach
36 Joos Buijs (TUE), Flexible Evolutionary Algorithms for Mining Structured Process Models
37 Maral Dadvar (UT), Experts and Machines United Against Cyberbullying
38 Danny Plass-Oude Bos (UT), Making brain-computer interfaces better: improving usability through post-processing.
39 Jasmina Maric (UvT), Web Communities, Immigration, and Social Capital
40 Walter Omona (RUN), A Framework for Knowledge Management Using ICT in Higher Education
41 Frederic Hogenboom (EUR), Automated Detection of Financial Events in News Text
42 Carsten Eijckhof (CWI/TUD), Contextual Multidimensional Relevance Models
43 Kevin Vlaanderen (UU), Supporting Process Improvement using Method Increments
44 Paulien Meesters (UvT), Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.
45 Birgit Schmitz (OUN), Mobile Games for Learning: A Pattern-Based Approach
46 Ke Tao (TUD), Social Web Data Analytics: Relevance, Redundancy, Diversity
47 Shangsong Liang (UVA), Fusion and Diversification in Information Retrieval

2015

01 Niels Netten (UvA), Machine Learning for Relevance of Information in Crisis Response
02 Faiza Bukhsh (UvT), Smart auditing: Innovative Compliance Checking in Customs Controls
03 Twan van Laarhoven (RUN), Machine learning for network data
04 Howard Spoelstra (OUN), Collaborations in Open Learning Environments
05 Christoph Bösch (UT), Cryptographically Enforced Search Pattern Hiding
06 Farideh Heidari (TUD), Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes
07 Maria-Hendrike Peetz (UvA), Time-Aware Online Reputation Analysis
08 Jie Jiang (TUD), Organizational Compliance: An agent-based model for designing and evaluating organizational interactions
09 Randy Klaassen (UT), HCI Perspectives on Behavior Change Support Systems
10 Henry Hermans (OUN), OpenU: design of an integrated system to support lifelong learning
11 Yongming Luo (TUE), Designing algorithms for big graph datasets: A study of computing bisimulation and joins
12 Julie M. Birkholz (VU), Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks

13 Giuseppe Procaccianti (VU), Energy-Efficient Software
14 Bart van Straalen (UT), A cognitive approach to modeling bad news conversations
15 Klaas Andries de Graaf (VU), Ontology-based Software Architecture Documentation
16 Changyun Wei (UT), Cognitive Coordination for Cooperative Multi-Robot Teamwork
17 André van Cleeff (UT), Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs
18 Holger Pirk (CWI), Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories
19 Bernardo Tabuenca (OUN), Ubiquitous Technology for Lifelong Learners
20 Lois Vanhée (UU), Using Culture and Values to Support Flexible Coordination
21 Sibren Fetter (OUN), Using Peer-Support to Expand and Stabilize Online Learning
22 Zhemin Zhu (UT), Co-occurrence Rate Networks
23 Luit Gazendam (VU), Cataloguer Support in Cultural Heritage
24 Richard Berendsen (UVA), Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation
25 Steven Woudenberg (UU), Bayesian Tools for Early Disease Detection
26 Alexander Hogenboom (EUR), Sentiment Analysis of Text Guided by Semantics and Structure
27 Sándor Héman (CWI), Updating compressed colomn stores
28 Janet Bagorogoza (TiU), Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO
29 Hendrik Baier (UM), Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains
30 Kiavash Bahreini (OU), Real-time Multimodal Emotion Recognition in E-Learning
31 Yakup Koç (TUD), On the robustness of Power Grids
32 Jerome Gard (UL), Corporate Venture Management in SMEs
33 Frederik Schadd (TUD), Ontology Mapping with Auxiliary Resources
34 Victor de Graaf (UT), Gesocial Recommender Systems
35 Jungxao Xu (TUD), Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction

2016
01 Syed Saiden Abbas (RUN), Recognition of Shapes by Humans and Machines
02 Michiel Christiaan Meulendijk (UU), Optimizing medication reviews through decision support: prescribing a better pill to swallow
03 Maya Sappelli (RUN), Knowledge Work in Context: User Centered Knowledge Worker Support
04 Laurens Rietveld (VU), Publishing and Consuming Linked Data
05 Evgeny Sherkhonov (UVA), Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers
06 Michel Wilson (TUD), Robust scheduling in an uncertain environment
07 Jeroen de Man (VU), Measuring and modeling negative emotions for virtual training
08 Matje van de Camp (TiU), A Link to the Past: Constructing Historical Social Networks from Unstructured Data
09 Archana Nottamkandath (VU), Trusting Crowdsourced Information on Cultural Artefacts

10 George Karafotias (VUA), Parameter Control for Evolutionary Algorithms
11 Anne Schuth (UVA), Search Engines that Learn from Their Users
12 Max Knobbout (UU), Logics for Modelling and Verifying Normative Multi-Agent Systems
13 Nana Baah Gyan (VU), The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach
14 Ravi Khadka (UU), Revisiting Legacy Software System Modernization
15 Steffen Michels (RUN), Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments
16 Guangliang Li (UVA), Socially Intelligent Autonomous Agents that Learn from Human Reward
17 Berend Weel (VU), Towards Embodied Evolution of Robot Organisms
18 Albert Meroño Peñuela (VU), Refining Statistical Data on the Web
19 Julia Efremova (Tu/e), Mining Social Structures from Genealogical Data
20 Daan Odijk (UVA), Context & Semantics in News & Web Search
21 Alejandro Moreno Célleri (UT), From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground
22 Grace Lewis (VU), Software Architecture Strategies for Cyber-Foraging Systems
23 Fei Cai (UVA), Query Auto Completion in Information Retrieval
24 Brend Wanders (UT), Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach
25 Julia Kiseleva (TU/e), Using Contextual Information to Understand Searching and Browsing Behavior
26 Dilhan Thilakarathne (VU), In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains
27 Wen Li (TUD), Understanding Geo-spatial Information on Social Media
28 Mingxin Zhang (TUD), Large-scale Agent-based Social Simulation - A study on epidemic prediction and control
29 Nicolas Höning (TUD), Peak reduction in decentralised electricity systems - Markets and prices for flexible planning
30 Ruud Mattheij (UvT), The Eyes Have It
31 Mohammad Khelghati (UT), Deep web content monitoring
32 Eelco Vriezekolk (UT), Assessing Telecommunication Service Availability Risks for Crisis Organisations
33 Peter Bloem (UVA), Single Sample Statistics, exercises in learning from just one example
34 Dennis Schunselaar (TUE), Configurable Process Trees: Elicitation, Analysis, and Enactment
35 Zhaochun Ren (UVA), Monitoring Social Media: Summarization, Classification and Recommendation
36 Daphne Karreman (UT), Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies
37 Giovanni Sileno (UvA), Aligning Law and Action - a conceptual and computational inquiry

38 Andrea Minuto (UT), Materials that Matter - Smart Materials meet Art & Interaction Design
39 Merijn Bruijnes (UT), Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect
40 Christian Detweiler (TUD), Accounting for Values in Design
41 Thomas King (TUD), Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance
42 Spyros Martzoukos (UVA), Combinatorial and Compositional Aspects of Bilingual Aligned Corpora
43 Saskia Koldijk (RUN), Context-Aware Support for Stress Self-Management: From Theory to Practice
44 Thibault Sellam (UVA), Automatic Assistants for Database Exploration
45 Bram van de Laar (UT), Experiencing Brain-Computer Interface Control
46 Jorge Gallego Perez (UT), Robots to Make you Happy
47 Christina Weber (UL), Real-time foresight - Preparedness for dynamic innovation networks
48 Tanja Buttler (TUD), Collecting Lessons Learned
49 Gleb Polevoy (TUD), Participation and Interaction in Projects. A Game-Theoretic Analysis
50 Yan Wang (UVT), The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains

2017
01 Jan-Jaap Oerlemans (UL), Investigating Cybercrime
02 Sjoerd Timmer (UU), Designing and Understanding Forensic Bayesian Networks using Argumentation
03 Daniël Harold Telgen (UU), Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines
04 Mrunal Gawade (CWI), Multi-core Parallelism in a Column-store
05 Mahdieh Shadi (UVA), Collaboration Behavior
06 Damir Vandic (EUR), Intelligent Information Systems for Web Product Search
07 Roel Bertens (UU), Insight in Information: from Abstract to Anomaly
08 Rob Konijn (VU) , Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery
09 Dong Nguyen (UT), Text as Social and Cultural Data: A Computational Perspective on Variation in Text
10 Robby van Delden (UT), (Steering) Interactive Play Behavior
11 Florian Kunneman (RUN), Modelling patterns of time and emotion in Twitter #anticipointment
12 Sander Leemans (TUE), Robust Process Mining with Guarantees
13 Gijs Huisman (UT), Social Touch Technology - Extending the reach of social touch through haptic technology
14 Shoshannah Tekofsky (UvT), You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior
15 Peter Berck (RUN), Memory-Based Text Correction

16 Aleksandr Chuklin (UVA), Understanding and Modeling Users of Modern Search Engines
17 Daniel Dimov (UL), Crowdsourced Online Dispute Resolution
18 Ridho Reinanda (UVA), Entity Associations for Search
19 Jeroen Vuurens (UT), Proximity of Terms, Texts and Semantic Vectors in Information Retrieval
20 Mohammadbashir Sedighi (TUD), Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility
21 Jeroen Linssen (UT), Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)
22 Sara Magliacane (VU), Logics for causal inference under uncertainty
23 David Graus (UVA), Entities of Interest — Discovery in Digital Traces
24 Chang Wang (TUD), Use of Affordances for Efficient Robot Learning
25 Veruska Zamborlini (VU), Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search
26 Merel Jung (UT), Socially intelligent robots that understand and respond to human touch
27 Michiel Joosse (UT), Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors
28 John Klein (VU), Architecture Practices for Complex Contexts
29 Adel Alhuraibi (UvT), From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"
30 Wilma Latuny (UvT), The Power of Facial Expressions
31 Ben Ruijl (UL), Advances in computational methods for QFT calculations
32 Thaer Samar (RUN), Access to and Retrievability of Content in Web Archives
33 Brigit van Loggem (OU), Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity
34 Maren Scheffel (OU), The Evaluation Framework for Learning Analytics
35 Martine de Vos (VU), Interpreting natural science spreadsheets
36 Yuanhao Guo (UL), Shape Analysis for Phenotype Characterisation from High-throughput Imaging
37 Alejandro Montes Garcia (TUE), WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy
38 Alex Kayal (TUD), Normative Social Applications
39 Sara Ahmadi (RUN), Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR
40 Altaf Hussain Abro (VUA), Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems
41 Adnan Manzoor (VUA), Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle
42 Elena Sokolova (RUN), Causal discovery from mixed and missing data with applications on ADHD datasets
43 Maaike de Boer (RUN), Semantic Mapping in Video Retrieval
44 Garm Lucassen (UU), Understanding User Stories - Computational Linguistics in Agile Requirements Engineering

45 Bas Testerink (UU), Decentralized Runtime Norm Enforcement
46 Jan Schneider (OU), Sensor-based Learning Support
47 Jie Yang (TUD), Crowd Knowledge Creation Acceleration
48 Angel Suarez (OU), Collaborative inquiry-based learning

2018

01 Han van der Aa (VUA), Comparing and Aligning Process Representations
02 Felix Mannhardt (TUE), Multi-perspective Process Mining
03 Steven Bosems (UT), Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction
04 Jordan Janeiro (TUD), Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks
05 Hugo Huurdeman (UVA), Supporting the Complex Dynamics of the Information Seeking Process
06 Dan Ionita (UT), Model-Driven Information Security Risk Assessment of Socio-Technical Systems
07 Jieting Luo (UU), A formal account of opportunism in multi-agent systems
08 Rick Smetsers (RUN), Advances in Model Learning for Software Systems
09 Xu Xie (TUD), Data Assimilation in Discrete Event Simulations
10 Julienka Mollee (VUA), Moving forward: supporting physical activity behavior change through intelligent technology
11 Mahdi Sargolzaei (UVA), Enabling Framework for Service-oriented Collaborative Networks
12 Xixi Lu (TUE), Using behavioral context in process mining
13 Seyed Amin Tabatabaei (VUA), Computing a Sustainable Future
14 Bart Joosten (UVT), Detecting Social Signals with Spatiotemporal Gabor Filters
15 Naser Davarzani (UM), Biomarker discovery in heart failure
16 Jaebok Kim (UT), Automatic recognition of engagement and emotion in a group of children
17 Jianpeng Zhang (TUE), On Graph Sample Clustering
18 Henriette Nakad (UL), De Notaris en Private Rechtspraak
19 Minh Duc Pham (VUA), Emergent relational schemas for RDF
20 Manxia Liu (RUN), Time and Bayesian Networks
21 Aad Slootmaker (OUN), EMERGO: a generic platform for authoring and playing scenario-based serious games
22 Eric Fernandes de Mello Araujo (VUA), Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks
23 Kim Schouten (EUR), Semantics-driven Aspect-Based Sentiment Analysis
24 Jered Vroon (UT), Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots
25 Riste Gligorov (VUA), Serious Games in Audio-Visual Collections
26 Roelof Anne Jelle de Vries (UT),Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology
27 Maikel Leemans (TUE), Hierarchical Process Mining for Scalable Software Analysis
28 Christian Willemse (UT), Social Touch Technologies: How they feel and how they make you feel

29 Yu Gu (UVT), Emotion Recognition from Mandarin Speech
30 Wouter Beek, The "K" in "semantic web" stands for "knowledge": scaling semantics to the web

2019

01 Rob van Eijk (UL),Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification
02 Emmanuelle Beauxis Aussalet (CWI, UU), Statistics and Visualizations for Assessing Class Size Uncertainty
03 Eduardo Gonzalez Lopez de Murillas (TUE), Process Mining on Databases: Extracting Event Data from Real Life Data Sources
04 Ridho Rahmadi (RUN), Finding stable causal structures from clinical data
05 Sebastiaan van Zelst (TUE), Process Mining with Streaming Data
06 Chris Dijkshoorn (VU), Nichesourcing for Improving Access to Linked Cultural Heritage Datasets
07 Soude Fazeli (TUD), Recommender Systems in Social Learning Platforms
08 Frits de Nijs (TUD), Resource-constrained Multi-agent Markov Decision Processes
09 Fahimeh Alizadeh Moghaddam (UVA), Self-adaptation for energy efficiency in software systems
10 Qing Chuan Ye (EUR), Multi-objective Optimization Methods for Allocation and Prediction
11 Yue Zhao (TUD), Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs
12 Jacqueline Heinerman (VU), Better Together
13 Guanliang Chen (TUD), MOOC Analytics: Learner Modeling and Content Generation
14 Daniel Davis (TUD), Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses
15 Erwin Walraven (TUD), Planning under Uncertainty in Constrained and Partially Observable Environments
16 Guangming Li (TUE), Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models
17 Ali Hurriyetoglu (RUN),Extracting actionable information from microtexts
18 Gerard Wagenaar (UU), Artefacts in Agile Team Communication
19 Vincent Koeman (TUD), Tools for Developing Cognitive Agents
20 Chide Groenouwe (UU), Fostering technically augmented human collective intelligence
21 Cong Liu (TUE), Software Data Analytics: Architectural Model Discovery and Design Pattern Detection
22 Martin van den Berg (VU),Improving IT Decisions with Enterprise Architecture
23 Qin Liu (TUD), Intelligent Control Systems: Learning, Interpreting, Verification
24 Anca Dumitrache (VU), Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing
25 Emiel van Miltenburg (VU), Pragmatic factors in (automatic) image description
26 Prince Singh (UT), An Integration Platform for Synchromodal Transport
27 Alessandra Antonaci (OUN), The Gamification Design Process applied to (Massive) Open Online Courses

28 Esther Kuindersma (UL), Cleared for take-off: Game-based learning to prepare airline pilots for critical situations
29 Daniel Formolo (VU), Using virtual agents for simulation and training of social skills in safety-critical circumstances
30 Vahid Yazdanpanah (UT), Multiagent Industrial Symbiosis Systems
31 Milan Jelisavcic (VU), Alive and Kicking: Baby Steps in Robotics
32 Chiara Sironi (UM), Monte-Carlo Tree Search for Artificial General Intelligence in Games
33 Anil Yaman (TUE), Evolution of Biologically Inspired Learning in Artificial Neural Networks
34 Negar Ahmadi (TUE), EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES
35 Lisa Facey-Shaw (OUN), Gamification with digital badges in learning programming
36 Kevin Ackermans (OUN), Designing Video-Enhanced Rubrics to Master Complex Skills
37 Jian Fang (TUD), Database Acceleration on FPGAs
38 Akos Kadar (OUN), Learning visually grounded and multilingual representations

2020
01 Armon Toubman (UL), Calculated Moves: Generating Air Combat Behaviour
02 Marcos de Paula Bueno (UL), Unraveling Temporal Processes using Probabilistic Graphical Models
03 Mostafa Deghani (UvA), Learning with Imperfect Supervision for Language Understanding
04 Maarten van Gompel (RUN), Context as Linguistic Bridges
05 Yulong Pei (TUE), On local and global structure mining
06 Preethu Rose Anish (UT), Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support
07 Wim van der Vegt (OUN), Towards a software architecture for reusable game components
08 Ali Mirsoleimani (UL),Structured Parallel Programming for Monte Carlo Tree Search
09 Myriam Traub (UU), Measuring Tool Bias and Improving Data Quality for Digital Humanities Research
10 Alifah Syamsiyah (TUE), In-database Preprocessing for Process Mining
11 Sepideh Mesbah (TUD), Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models
12 Ward van Breda (VU), Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment
13 Marco Virgolin (CWI), Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming
14 Mark Raasveldt (CWI/UL), Integrating Analytics with Relational Databases
15 Konstantinos Georgiadis (OUN), Smart CAT: Machine Learning for Configurable Assessments in Serious Games
16 Ilona Wilmont (RUN), Cognitive Aspects of Conceptual Modelling
17 Daniele Di Mitri (OUN), The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences

18 Georgios Methenitis (TUD), Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems
19 Guido van Capelleveen (UT), Industrial Symbiosis Recommender Systems
20 Albert Hankel (VU), Embedding Green ICT Maturity in Organisations
21 Karine da Silva Miras de Araujo (VU), Where is the robot?: Life as it could be
22 Maryam Masoud Khamis (RUN), Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar
23 Rianne Conijn (UT), The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging
24 Lenin da Nobrega Medeiros (VUA/RUN), How are you feeling, human? Towards emotionally supportive chatbots
25 Xin Du (TUE), The Uncertainty in Exceptional Model Mining
26 Krzysztof Leszek Sadowski (UU), GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization
27 Ekaterina Muravyeva (TUD), Personal data and informed consent in an educational context
28 Bibeg Limbu (TUD), Multimodal interaction for deliberate practice: Training complex skills with augmented reality
29 Ioan Gabriel Bucur (RUN), Being Bayesian about Causal Inference
30 Bob Zadok Blok (UL), Creatief, Creatieve, Creatiefst
31 Gongjin Lan (VU), Learning better – From Baby to Better
32 Jason Rhuggenaath (TUE), Revenue management in online markets: pricing and online advertising
33 Rick Gilsing (TUE), Supporting service-dominant business model evaluation in the context of business model innovation
34 Anna Bon (MU), Intervention or Collaboration? Redesigning Information and Communication Technologies for Development
35 Siamak Farshidi (UU), Multi-Criteria Decision-Making in Software Production

2021

01 Francisco Xavier Dos Santos Fonseca (TUD),Location-based Games for Social Interaction in Public Space
02 Rijk Mercuur (TUD), Simulating Human Routines:Integrating Social Practice Theory in Agent-Based Models
03 Seyyed Hadi Hashemi (UVA), Modeling Users Interacting with Smart Devices
04 Ioana Jivet (OU), The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning
05 Davide Dell'Anna (UU), Data-Driven Supervision of Autonomous Systems
06 Daniel Davison (UT), "Hey robot, what do you think?" How children learn with a social robot
07 Armel Lefebvre (UU), Research data management for open science
08 Nardie Fanchamps (OU), The Influence of Sense-Reason-Act Programming on Computational Thinking
09 Cristina Zaga (UT), The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play
10 Quinten Meertens (UvA), Misclassification Bias in Statistical Learning

11 Anne van Rossum (UL), Nonparametric Bayesian Methods in Robotic Vision
12 Lei Pi (UL), External Knowledge Absorption in Chinese SMEs
13 Bob R. Schadenberg (UT), Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning
14 Negin Samaeemofrad (UL), Business Incubators: The Impact of Their Support
15 Onat Ege Adali (TU/e), Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm
16 Esam A. H. Ghaleb (UM), BIMODAL EMOTION RECOGNITION FROM AUDIO-VISUAL CUES
17 Dario Dotti (UM), Human Behavior Understanding from motion and bodily cues using deep neural networks
18 Remi Wieten (UU), Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks
19 Roberto Verdecchia (VU), Architectural Technical Debt: Identification and Management
20 Masoud Mansoury (TU/e), Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems
21 Pedro Thiago Timb□ Holanda (CWI), Progressive Indexes
22 Sihang Qiu (TUD), Conversational Crowdsourcing
23 Hugo Manuel Proen□a (LIACS), Robust rules for prediction and description
24 Kaijie Zhu (TUE), On Efficient Temporal Subgraph Query Processing
25 Eoin Martino Grua (VUA), The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications
26 Benno Kruit (CWI & VUA), Reading the Grid: Extending Knowledge Bases from Human-readable Tables
27 Jelte van Waterschoot (UT), Personalized and Personal Conversations: Designing Agents Who Want to Connect With You
28 Christoph Selig (UL), Understanding the Heterogeneity of Corporate Entrepreneurship Programs

2022

1 Judith van Stegeren (UT), Flavor text generation for role-playing video games
2 Paulo da Costa (TU/e), Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey
3 Ali el Hassouni (VUA), A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare
4 Ünal Aksu (UU), A Cross-Organizational Process Mining Framework
5 Shiwei Liu (TU/e), Sparse Neural Network Training with In-Time Over-Parameterization
6 Reza Refaei Afshar (TU/e), Machine Learning for Ad Publishers in Real Time Bidding
7 Sambit Praharaj (OU), Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics
8 Maikel L. van Eck (TU/e), Process Mining for Smart Product Design
9 Oana Andreea Inel (VUA), Understanding Events: A Diversity-driven Human-Machine Approach

10 Felipe Moraes Gomes (TUD), Examining the Effectiveness of Collaborative Search Engines
11 Mirjam de Haas (UvT), Staying engaged in child-robot interaction, a quantitative approach to studying preschoolers' engagement with robots and tasks during second-language tutoring
12 Guanyi Chen (UU), Computational Generation of Chinese Noun Phrases
13 Xander Wilcke (VUA), Machine Learning on Multimodal Knowledge Graphs: Opportunities, Challenges, and Methods for Learning on Real-World Heterogeneous and Spatially-Oriented Knowledge
14 Michiel Overeem (UU), Evolution of Low-Code Platforms
15 Jelmer Jan Koorn (UU), Work in Process: Unearthing Meaning using Process Mining
16 Pieter Gijsbers (TU/e), Systems for AutoML Research
17 Laura van der Lubbe (VUA), Empowering vulnerable people with serious games and gamification