Daniel Roth

Intrapersonal, Interpersonal, and Hybrid Interactions in Virtual Reality

Dissertation submitted for the degree of Doctor rerum naturalium (Dr. rer. nat.)
Faculty of Mathematics and Computer Science, University of Würzburg, Germany
2019
Advisor: Prof. Dr. Marc Erich Latoschik
Reviewer: Prof. Dr. Gary Bente, Michigan State University
Reviewer: Prof. Jeremy Bailenson, Ph.D., Stanford University

https://doi.org/10.25972/OPUS-18862

Abstract

Virtual reality and related media and communication technologies have a growing impact on professional application fields and our daily life. Virtual environments have the potential to change the way we perceive ourselves and how we interact with others. In comparison to other technologies, virtual reality allows for the convincing display of a virtual self-representation, an avatar, to oneself and also to others. This is referred to as user embodiment. Avatars can be of varying realism and abstraction in their appearance and in the behaviors they convey. Such user-embodying interfaces, in turn, can impact the perception of the self as well as the perception of interactions. For researchers, designers, and developers it is of particular interest to understand these perceptual impacts, to apply them to therapy, assistive applications, social platforms, or games, for example. The present thesis investigates and relates these impacts with regard to three areas: intrapersonal effects, interpersonal effects, and effects of social augmentations provided by the simulation.

With regard to *intrapersonal* effects, we specifically explore which simulation properties impact the illusion of owning and controlling a virtual body, as well as a perceived change in body schema. Our studies lead to the construction of an instrument to measure these dimensions and our results indicate that these dimensions are especially affected by the level of immersion, the simulation latency, as well as the level of personalization of the avatar.

With regard to *interpersonal* effects we compare physical and user-embodied social interactions, as well as different degrees of freedom in the replication of nonverbal behavior. Our results suggest that functional levels of interaction are maintained, whereas aspects of presence can be affected by avatar-mediated interactions, and collaborative motor coordination can be disturbed by immersive simulations.

Social interaction is composed of many unknown symbols and harmonic patterns that define our understanding and interpersonal rapport. For successful virtual social interactions, a mere replication of physical world behaviors to virtual environments may seem feasible. However, the potential of mediated social interactions goes beyond this mere replication. In a third vein of research, we propose and evaluate alternative concepts on how computers can be used to actively engage in mediating social interactions, namely *hybrid avatar-agent technologies*. Specifically, we investigated the possibilities to augment social behaviors by modifying and transforming user input according to social phenomena and behavior, such as non-verbal mimicry, directed gaze, joint attention, and grouping. Based on our results we argue that such technologies could be beneficial for computer-mediated social interactions such as to compensate for lacking sensory input and disturbances in data transmission or to increase aspects of social presence by visual substitution or amplification of social behaviors.

Based on related work and presented findings, the present thesis proposes the perspective of considering *computers as social mediators*. Concluding from prototypes and empirical studies, the potential of technology to be an active mediator of social perception with regard to the perception of the self, as well as the perception of social interactions may benefit our society by enabling further methods for diagnosis, treatment, and training, as well as the inclusion of individuals with social disorders. To this regard, we discuss implications for our society and ethical aspects. This thesis extends previous empirical work and further presents novel instruments, concepts, and implications to open up new perspectives for the development of virtual reality, mixed reality, and augmented reality applications.

Zusammenfassung

Virtual Reality und weitere Medien- und Kommunikationstechnologien haben einen wachsenden Einfluss auf professionelle Anwendungsbereiche und unseren Alltag. Virtuelle Umgebungen haben das Potenzial, Einfluss darauf zu nehmen, wie Mensche sich selbst wahrnehmen und wie sie mit anderen umgehen. Im Vergleich zu anderen Technologien ermöglicht Virtual Reality die überzeugende Visualisierung einer virtuellen Selbstdarstellung, eines Avatars, sichtbar für den Nutzer/die Nutzerin selbst aber auch für andere. Dies bezeichnet man als Nutzerverkörperung. Avatare können von unterschiedlichem Realismus und Abstraktion in Bezug auf ihr Aussehen sowie der Darstellung von Verhaltensweisen geprägt sein. Solche nutzerverkörpernde Schnittstellen wiederum können die Wahrnehmung des Selbst sowie die Wahrnehmung von Interaktionen beeinflussen. Für Forscher/-innen, Designer/-innen und Entwickler/-innen ist es von besonderem Interesse, diese Wahrnehmungseffekte zu verstehen, um sie beispielsweise auf Therapie, assistive Anwendungen, soziale Plattformen oder Spiele anzuwenden. Die vorliegende Arbeit untersucht und bezieht sich auf diese Auswirkungen in drei Bereichen: intrapersonelle Effekte, zwischenmenschliche Effekte sowie Effekte durch soziale Augmentierungen, die durch die Simulation bereitgestellt werden.

Im Hinblick auf *intrapersonelle* Effekte widmet sich die vorliegende Arbeit insbesondere der Frage, welche Simulationseigenschaften die Illusion des Besitzens/Innehabens und der Kontrolle eines virtuellen Körpers sowie eine wahrgenommene Veränderung des Körperschemas beeinflussen. Die vorgestellten Studien führen zur Konstruktion eines Instruments zur Erfassung dieser Dimensionen und die Ergebnisse zeigen, dass die empfundene Verkörperung besonders von dem Grad der Immersion, der Simulationslatenz sowie dem Grad der Personalisierung des Avatars abhängt.

Im Hinblick auf *zwischenmenschliche* Effekte vergleicht diese Dissertation physische (realweltliche) und virtuelle soziale Interaktionen sowie unterschiedliche Freiheitsgrade in der Replikation nonverbalen Verhaltens. Die Ergebnisse deuten darauf hin, dass die funktionalen Ebenen der Interaktion aufrechterhalten werden, während Aspekte der Präsenz durch avatarvermittelte Interaktionen beeinflusst werden und die kollaborative motorische Koordination durch immersive Simulationen gestört werden kann.

Die soziale Interaktion besteht aus vielen unbekannten Symbolen und harmonischen Mustern, die das menschliche Verständnis und zwischenmenschliche Beziehungen definieren. Für erfolgreiche virtuelle soziale Interaktionen mag eine bloße Replikation von physikalischen Weltverhaltensweisen auf virtuelle Umgebungen möglich erscheinen. Das Potenzial computervermittelter sozialer Interaktionen geht jedoch über diese bloße Replikation hinaus. Im dritten Bereich dieser Arbeit werden alternative Konzepte vorgeschlagen und evaluiert, wie Computer genutzt werden können, um eine aktive Rolle in sozialen Interaktionen einzunehmen. Diese Technologien werden als *hybride Avatar-Agenten-Technologien* definiert. Insbesondere wird untersucht, welche Möglichkeiten das soziale Verhalten zu erweitern emtstehen, indem die Verhaltensweisen der Benutzer/-innen entsprechend sozialer Phänomene und Verhaltensweisen modifiziert und transformiert werden. Beispiele sind die nonverbale Spiegelung, der Fokus des Blicks, eine gemeinsame Aufmerksamkeit und die Gruppenbildung. Basierend auf den Ergebnissen argumentiert diese Arbeit, dass solche Technologien für computervermittelte soziale Interaktionen von Vorteil sein könnten, beispielsweise zum Ausgleich fehlender Sensorik, Störungen bei der Datenübertragung oder zur Verbesserung sozialer Präsenz durch visuelle Substitution oder Verstärkung des sozialen Verhaltens. Basierend auf verwandten Arbeiten und präsentierten Ergebnissen wird abgeleitet, dass Computer als soziale Mediatoren fungieren können. Ausgehend von Prototypen und empirischen Studien kann das Potenzial der Technologie, ein aktiver Vermittler in Bezug auf die Wahrnehmung des Selbst sowie die Wahrnehmung sozialer Interaktionen zu sein, unserer Gesellschaft zugutekommen. Dadurch können beispielsweise weitere Methoden zur Diagnose, der Behandlung und Ausbildung sowie der Inklusion von Menschen mit sozialen Störungen ermöglicht werden. In diesem Zusammenhang werden die Auswirkungen auf unsere Gesellschaft und ethische Aspekte diskutiert. Diese Arbeit erweitert frühere empirische Arbeiten und präsentiert darüber hinaus neue Instrumente, Konzepte und Implikationen, um neue Perspektiven für die Entwicklung von Virtual Reality, Mixed Reality und Augmented Reality Anwendungen zu beleuchten.

To my wife, the most wonderful person on earth,
and my family, to whom I owe everything.

Acknowledgments

"It is the supreme art of the teacher to awaken joy in creative expression and knowledge." –Albert Einstein

Words cannot express my sincere gratitude to my advisor, Prof. Dr. Marc Erich Latoschik, for his continuous support, his guidance, and motivation. I thank him for giving me the opportunity to progress with my research, his openness towards the topic, and for keeping his patience at times when I lost mine. I could not have imagined a better advisor.

I am in deepest gratitude to my secondary reviewer, Prof. Dr. Gary Bente, who initially provided me with the opportunity to pursue in academia. His inspiration, and continuous personal and scientific guidance, support, and many mind-opening discussions were of priceless value for me.

I am in debt to Prof. Jeremy Bailenson, Ph.D., who provided a review of my thesis. I am deeply honored, given that I learned so much from his work and vision.

With greatest gratitude I thank Dr. Dr. Kai Vogeley, who had open words and advice continuously along the way, supported my progress, and tremendously contributed to the blue-sky ideas I pursued in my work with superior wisdom.

With greatest gratitude I thank Prof. Dr. Sebastian von Mammen for providing academic advice and a second shelter. His valuable comments to my works and inspiring guidance in teaching strongly supported my progress.

Following my scientific advisors I thank my mentor in life, my friend, soulmate, and wife, Prof. Dr. Diana Rieger. This thesis would not have been possible without her support. I hope to make up for many occasions where I could not prioritize the most important things in life.

I thank my whole family, especially my parents Gabriele & Otto Roth, my grandmother Barbara Stysch, my brother Jr. Prof. Dr. Martin Roth, his wife Kanako Roth-Kurashima, Otto Roth Jr., and Benjamin Roth. I grew up in a happy, protected and loving environment and I cannot thank them enough for the strength, support and love they provided me with. I thank my family in law, Cornelia Rieger, Carina Rieger, and Alex Langstrof, and my closest friends who stuck with me since school or college and had open ears in many situations, Jan Reinhardt, Johannes Schmidt, Pauline Mantell, Daniel Ahrens, Daniel Erpilev, Roman & Lena Breitwieser, Hans Fischer, Felix and Teresa Zaiser.

Further, I would like to thank and wish the best of luck for their successful graduation to my "brothers in crime", my colleagues that went through this process with me. I thank Dominik Gall, for becoming a true friend during my time in Würzburg, for sharing his thoughts with me during ups and downs and for the feedback to the projects and to initial drafts of this thesis. I thank Jan-Philipp Stauffert in similar manner, and for continuous technical support, feedback and support during the projects, and for his feedback to final drafts. I thank Tobias Feigl from Fraunhofer IIS for his open ear to many of my thoughts (and problems), and for becoming a friend, not only in scientific context. I thank Sebastian Oberdörfer for the mutual thought exchange and healing discussions.

I express my sincere thankfullness to the HCI Group, Sandra Schubert, Prof Dr. Birgit Lugrin, Assistant Prof. Dr. Jean-Luc Lugrin, Dr. Florian Niebling, Dr. Martin Fischbach, Andrea Bartl, Benjamin Eckstein, Negin Hamzeheinejad, Florian Kern, Andreas Knote, Kristof Korwisi, Maximilian Landeck, Samantha Monty, Andreas Müller, David Obremski, Dr. Philipp Schaper, Christian Schell, (formerly) Hendrik Striepe, Susanne Stenglin, Fabian Unruh, Erik Wolf, and Chris Zimmerer. They welcomed (and tolerated) me, each of them contributing through help and favor, advise, or personal conversations.

I thank my VR community colleagues, Dr. Iana Podkosova, Assistant Prof. Dr. Gerd Bruder, Assistant Prof. Niels Christian Nilsson (Ph.D.), and Dr. Alexander Kulik for their general guidance and feedback on drafts of the present thesis, and finally for the many joyful moments at VR conferences. I am in dept for the help they provided with no ulterior motives to reclaim. I hope our paths continue to cross for a long time.

I thank the former department of Media Psychology at the University of Cologne (Dr. Lena Frischlich, Roman Frischlich-Rybak, Dr. Thomas Dratsch, Björn Günther, Dr. Johannes Breuer, Dr. Tim Wulf, Olivia Rutkowski, Bastian Jarczewsky, Arvid Hoffman, and Monika Wegener-Hergarden), where my academic path started. I especially thank them for accepting and respecting the fact that they had to work with a technologist.

I thank the Michigan State University (Department of Communication, Arts, and Science) most importantly Prof. Brandon VanDerHeide (Ph.D.) for his thoughts and feedback on drafts of this thesis. The visit was of great inspiration which I soaked in during lively discussions. I will continue to appreciate the additional pounds I added during every visit thanks to "Ron salads".

I thank Prof. Dr. Arnulph Fuhrmann, Dr. Christopher Mutschler, Prof. Mario Botsch, Dr. Thomas Waltemate, Jascha Achenbach, Dr. Alexandra Georgescu, Mathis Jording, Eric Novotny, and Reed Reynolds, for their collaboration and fruitful interaction.

Empirical studies are not possible without voluntary participation. I therefore thank all participants that participated in our studies.

Last but certainly not least, I would thank exceptional (former) students and student co-workers that accompanied me and worked intensely with me along the way, Julia Büser, Dmitri Galakhov, Roman Rybak, Carola Bloch, Arvid Hofmann, Kristoffer Waldow, Felix Stetter, Johanna Behrendt, David Zilch, Sebastian Lammers, Constantin Kleinbeck, David Fernes, Peter Kullmann, David Mal, Chris Felix Purps, Daniel Götz, Jonathan Stapf, Christopher Göttfert, Franziska Westermeier, Larissa Brübach, and Elisabeth Ganal. Without their help, many projects would not have been possible, and I thank them greatly for that. I wish all of them the best for their future.

List of Publications

Related Publications and External Contributions

In addition to unpublished work (at the time of writing), the findings presented in this thesis resulted in several publications. Parts of the papers are used directly or indirectly in this thesis. The following lists the publications which stand in direct relation to the present writing for each study. I had the honor to work with exceptional colleagues and students that all contributed to these publications from different perspectives. The following therefore includes a description of these contributions (multiple persons are ordered by seniority) for each study.

Chapter 2

The Hybrid Avatar-Agent Technology Concept The concept presented in Section 2.4.5 was initially designed by the author, Kai Vogeley (University Hospital Cologne/ Research Center Jülich), and Gary Bente (Michigan State University), and further refined together with Marc Erich Latoschik.

Daniel Roth, Marc Erich Latoschik, Kai Vogeley, and Gary Bente. 2015. Hybrid Avatar-Agent Technology – A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com*, 14, 2, 107–114. ISSN: 2196-6826. DOI: 10.1515/icom-2015-0030

Background and Related Work in Chapter 2 and consecutive chapters are also discussed in two book chapters, and derive from the publications related to Chapter 3 and following (see listings below).

Daniel Roth, Jean-Luc Lugrin, Sebastian von Mammen, and Marc Erich Latoschik. 2017. Controllers & inputs: Masters of puppets. In *Avatar, Assembled: The Social and Technical Anatomy of Digital Bodies, Jaime Banks (Ed.)* 281–290. ISBN: 978-1-4331-4326-7

Daniel Roth, Jan-Philipp Stauffert, and Marc Erich Latoschik. 2019. Avatar Embodiment, Behavior Replication, and Kinematics in Virtual Reality (in press). In *VR Developer Gems, William R Sherman (Ed.)* VR Developer Gems. Volume 1. Taylor & Francis. ISBN: 978-1-138-03012-1

Chapter 3

Study 1 was initiated by the author, Jean-Luc Lugrin, and Marc Erich Latoschik. Based on previous work, the apparatus was developed with the help of David Zilch as part of a Master thesis. David Zilch conducted the user study, based on the study design by the author, Jean-Luc Lugrin, Marc Erich Latoschik, and David Zilch. The scale designed for and derived from the Study 1 was initiated by the author. The initial items were constructed together with Jean-Luc Lugrin. The data was analyzed by the author.

Jean-Luc Lugrin, David Zilch, Daniel Roth, Gary Bente, and Marc Erich Latoschik. 2016. Facebo: Real-time face and body tracking for faithful avatar synthesis. In *2016 IEEE Virtual Reality*. IEEE VR 2016. IEEE, Greenville, SC, USA, 225–226. DOI: 10.1109/VR.2016.7504735

Marc Erich Latoschik, Jean-Luc Lugrin, and Daniel Roth. 2016. FakeMi: a fake mirror system for avatar embodiment studies. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. ACM, Munich, Germany, 73–76. ISBN: 978-1-4503-4491-3. DOI: 10.1145/2993369.2993399

Daniel Roth, Jean-Luc Lugrin, Marc Erich Latoschik, and Stephan Huber. 2017. Alpha IVBO - Construction of a Scale to Measure the Illusion of Virtual Body Ownership. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. ACM, Denver, CO, USA, 2875–2883. ISBN: 978-1-4503-4656-6. DOI: 10.1145/3027063.3053272

Study 2 was initiated by Marc Erich Latoschik and Mario Botsch (University of Bielefeld). The experiment was designed by Dominik Gall, the author, and Marc Erich Latoschik in collaboration with Thomas Waltemate and Mario Botsch. The Apparatus was developed by Thomas Waltemate and Mario Botsch. Thomas Waltemate and colleagues (University of Bielefeld) conducted the experiment.

Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24, 4, 1643–1652. ISSN: 1077-2626. DOI: 10.1109/TVCG.2018.2794629

Study 3 was initiated by Marc Erich Latoschik. The study design was constructed and discussed by the author, Dominik Gall, and Marc Erich Latoschik. The apparatus was built by the author with help of David Fernes as part of a Bachelor project. The experiment was conducted by the author and Dominik Gall with help of David Fernes, Samantha Straka (University of Würzburg), and Carola Bloch (Ludwig-Maximilians-University, Munich). The photogrammetric avatars were created by Jascha Achenbach, Thomas Waltemate, and Mario Botsch (University of

Bielefeld). The analysis was conducted by the author. All authors contributed to the paper. Marc Erich Latoschik was the lead author.

Marc Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The Effect of Avatar Realism in Immersive Social Virtual Realities. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology*. VRST '17. Gothenburg, Sweden, 39:1–39:10. DOI: 10.1145/3139131.3139156

Studies 4, 5, and 6 were initiated by the author. The apparatus was developed by the author with help of David Fernes, Christopher Göttfert, and Patrick Schulz as part of their Bachelor thesis and Bachelor projects, who also helped in conducting the study. The scale construction and analysis was performed by the author.

Study 7 was initiated by the author. The apparatus was developed based on previous works of Study 4, 5, and 6. The jitter injection was developed with help of Jan-Philipp Stauffert. Peter Kullmann and Christopher Göttfert helped in conducting the study.

The resulting scale and experiment summary of Studies 4, 5, 6, and 7 has been submitted for publication.

Chapter 4

Study 8 was initiated by the author, Arnulph Fuhrmann (TH Köln), and Gary Bente (Michigan State University). The apparatus was developed with help of Dmitri Galakhov (TH Köln) as part of a Bachelor thesis, who also helped in conducting the study.

Daniel Roth, Jean-Luc Lugrin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. Avatar Realism and Social Interaction Quality in Virtual Reality. In *2016 IEEE Virtual Reality*. IEEE VR 2016. IEEE, Greenville, SC, USA, 277–278. DOI: 10.1109/VR.2016.7504761

Study 9 was initiated by the author, Arnulph Fuhrmann, and Gary Bente. The apparatus was developed based on previous work with help of Kristoffer Waldow (TH Köln) and Felix Stetter (TH Köln), who also helped in conducting the study.

Study 10 was initiated by the author. The apparatus was developed based on previous work with help of Kristoffer Waldow as part of a Bachelor thesis. Kristoffer Waldow, Carola Bloch, Arvid Hoffman and Bastian Jarczewsky helped in conducting the study. The apparatus was discussed in the following publications.

Daniel Roth, Kristoffer Waldow, Felix Stetter, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. SIAMC: A Socially Immersive Avatar Mediated

Communication Platform. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. ACM, Munich, Germany, 357–358. ISBN: 978-1-4503-4491-3. DOI: 10.1145/2993369.2996302

Daniel Roth, Kristoffer Waldow, Marc Erich Latoschik, Arnulph Fuhrmann, and Gary Bente. 2017. Socially Immersive Avatar-based Communication. In *2017 IEEE Virtual Reality (VR)*. IEEE VR 2017. IEEE, Los Angeles, USA, 259–260. DOI: 10.1109/VR.2017.7892275

Chapter 5

Study 11 was initiated by the author. The apparatus was developed with help of David Mal, Ivan Polyschev, Maximilian Wiedemann, Christoph Klöffel, Christian Purps, and Jens To as part of a seminar project. The study was conducted with help of David Mal, Peter Kullmann, and Christian Felix Purps.

Daniel Roth, David Mal, Christian Felix Purps, Peter Kullmann, and Marc Erich Latoschik. 2018. Injecting Nonverbal Mimicry with Hybrid Avatar-Agent Technologies: A Naive Approach. In *Proceedings of the Symposium on Spatial User Interaction*. SUI '18. ACM, Berlin, Germany, 69–73. DOI: 10.1145/3267782.3267791

Study 12 was initiated by the author. The apparatus was developed with help of Roman Eyck and Peter Kullmann as part of a Bachelor and Master project. The study was conducted with help of Peter Kullmann and Dominik Gall.

Daniel Roth, Peter Kullmann, Gary Bente, Dominik Gall, and Marc Erich Latoschik. 2018. Effects of Hybrid and Synthetic Social Gaze in Avatar-Mediated Interactions. In *Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018*. ISMAR '18. IEEE, Munich, Germany. DOI: 10.1109/ISMAR-Adjunct.2018.00044

Study 13 was initiated by the author. The apparatus was developed with help of Constantin Kleinbeck as part of a Master thesis as well as Tobias Feigl (Fraunhofer IIS Nürnberg) who also helped in conducting the study.

Daniel Roth, Constantin Kleinbeck, Tobias Feigl, and Christopher Mutschler. 2017. Social Augmentations in Multi-User Virtual Reality: A Virtual Museum Experience. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. ISMAR '17. Nantes, France, 42–43. ISBN: 978-0-7695-6327-5. DOI: 10.1109/ismar-adjunct.2017.28

Daniel Roth, Constantin Kleinbeck, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. 2018. Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities. In *2018 IEEE Conference on Virtual Reality and 3D*

User Interfaces (VR). IEEE VR 2018. IEEE, Reutlingen, Germany, 215–222. DOI: 10.1109/VR.2018.8447550

Technologies from the developments of this chapter have been summarized and further extended, and have been accepted for publication.

Daniel Roth, Gary Bente, Peter Kullmann, David Mal, Christian Felix Purps, Kai Vogeley, and Marc Erich Latoschik. 2019, in press. Technologies for social augmentations in user-embodied virtual reality. In *25th ACM Symposium on Virtual Reality Software and Technology* (VRST '19). DOI: https://doi.org/10.1145/3359996.3364269

Further Publications

In addition to papers that directly relate to the presented studies, the following publications inform the body of this work.

Daniel Roth, Sebastian von Mammen, Julian Keil, and Marc Erich Latoschik. 2019. Approaching Difficult Terrain with Sensitivity: A Virtual Reality Game on the Five Stages of Grief (in press). In *11th International Conference on Virtual Worlds and Games for Serious Applications (VS Games)*. Vienna, Austria

Tobias Feigl, Daniel Roth, Stefan Gradl, Markus Wirth, Marc Erich Latoschik, Michael Philippsen, and Christopher Mutschler. 2019. Sick moves! motion parameters as indicators of simulator sickness. *IEEE Transactions on Visualization and Computer Graphics*, 25, 11, (November 2019), 3146–3157. DOI: 10.1109/TVCG.2019.2932224

Alexander Geiger, Gary Bente, Sebastian Lammers, Ralf Tepest, Daniel Roth, Danilo Bzdok, and Kai Vogeley. 2019. Distinct functional roles of the mirror neuron system and the mentalizing system. *NeuroImage*, 202, 116102. ISSN: 1053-8119. DOI: 10/gf9j2n

Sebastian Lammers, Gary Bente, Ralf Tepest, Mathis Jording, Daniel Roth, and Kai Vogeley. 2019. Introducing acass: an annotated character animation stimulus set for controlled (e)motion perception studies. *Frontiers in Robotics and AI*, 6, 94. ISSN: 2296-9144. DOI: 10.3389/frobt.2019.00094

Daniel Roth, Carola Bloch, Josephine Schmitt, Lena Frischlich, Marc Erich Latoschik, and Gary Bente. 2019. Perceived authenticity, empathy, and pro-social intentions evoked through avatar-mediated self-disclosures. In *Proceedings of Mensch Und Computer 2019* (MuC'19). ACM, Hamburg, Germany, 21–30. ISBN: 978-1-4503-7198-8. DOI: 10.1145/3340764.3340797

R. M. Reynolds, E. Novotny, J. Lee, D. Roth, and G. Bente. 2019. Ambiguous bodies: the role of displayed arousal in emotion [Mis]Perception. *Journal of Nonverbal Behavior*, (July 2019). ISSN: 1573-3653. DOI: 10/gf9jxq

Negin Hamzeheinejad, Daniel Roth, Daniel Götz, Franz Weilbach, and Marc Erich Latoschik. 2019. Physiological Effectivity and User Experience of Immersive Gait Rehabilitation. In *The First IEEE VR Workshop on Applied VR for Enhanced Healthcare (AVEH)*. IEEE

Sebastian von Mammen, Andreas Knote, Daniel Roth, and Marc Erich Latoschik. 2018. Games Engineering (Beschreibung der Disziplin). In *In: Games Studieren-Was, Wie, Wo?: Staatliche Studienangebote Im Bereich Digitaler Spiele. Bartholdy, Björn, Breitlauch, Linda, Czauderna, André, Freyermuth, Gundolf S. (Eds).* Volume 6. Transcript Verlag, 67ff

Daniel Roth and Carolin Wienrich. 2018. Effects of Media Immersiveness on the Perception of Virtual Characters. In *Ralf Dörner, Paul Grimm, Christian Geiger (Eds.), Proceedings of the 15th Workshop on Virtual and Augmented Reality of the GI Special Interest Group VR/AR. 2018.* Düsseldorf, Germany, 133–144

Jean-Luc Lugrin, Florian Kern, Ruben Schmidt, Constantin Kleinbeck, Daniel Roth, Christian Daxer, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. 2018. A Location-Based VR Museum. In *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. IEEE, Wuerzburg, Germany, 1–2. DOI: 10.1109/VS-Games.2018.8493404

Samuel Truman, Nicolas Rapp, Daniel Roth, and Sebastian von Mammen. 2018. Rethinking Real-Time Strategy Games for Virtual Reality. In *Proceedings of the 13th International Conference on the Foundations of Digital Games* (FDG'18). ACM, 31:1–31:6. DOI: 10.1145/3235765.3235801

Peter Ziegler, Daniel Roth, Andreas Knots, Michael Kreuzer, and Sebastian von Mammen. 2018. Simulator Sick but still Immersed: A Comparison of Head-Object Collision Handling and their Impact on Fun, Immersion, and Simulator Sickness. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 743–744. DOI: 10.1109/VR.2018.8446221

Marc Erich Latoschik, Jean-Luc Lugrin, Michael Habel, Daniel Roth, Christian Seufert, and Silke Grafe. 2016. Breaking Bad Behavior: Immersive Training of Classroom Management. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 317–318. DOI: 10.1145/2993369.2996308

Jean-Luc Lugrin, Marc Erich Latoschik, Michael Habel, Daniel Roth, Christian Seufert, and Silke Grafe. 2016. Breaking Bad Behaviours: A New Tool for Learning Classroom Management using Virtual Reality. *Frontiers in ICT*, 3, 26. DOI: 10.3389/fict.2016.00026

Gary Bente, Daniel Roth, Thomas Dratsch, and Kai Kaspar. 2016. Emotions of my Kin: Disambiguating Expressive Body Movement in Minimal Groups. *Culture and Brain*, 4, 1, 51–71. DOI: 10.1007/s40167-016-0038-3

Jean-Luc Lugrin, Ivan Polyschev, Daniel Roth, and Marc Erich Latoschik. 2016. Avatar anthropomorphism and acrophobia. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 315–316. DOI: 10.1145/2993369.2996313

Daniel Roth, Jean-Luc Lugrin, Julia Büser, Gary Bente, Arnulph Fuhrmann, and Marc Erich Latoschik. 2016. A Simplified Inverse Kinematic Approach for Embodied VR Applications. In *Proceedings of the IEEE Virtual Reality (IEEE VR) Conference 2016*. IEEE VR 2016. Greenville, SC, USA. DOI: 10.1109/VR.2016.7504760

Daniel Roth, Carola Bloch, Anne-Kathrin Wilbers, Kai Kaspar, Marc Erich Latoschik, and Gary Bente. 2015. Quantification of Signal Carriers for Emotion Recognition from Body Movement and Facial Affects. In *Abstracts of the 18th European Conference on Eye Movements, 2015, Vienna. Journal of Eye Movement Research, 8(4):1.* 192

Alexandra Livia Georgescu, Bojana Kuzmanovic, Daniel Roth, Gary Bente, and Kai Vogeley. 2014. The Use of Virtual Characters to Assess and Train Non-Verbal Communication in High-Functioning Autism. *Frontiers in Human Neuroscience*, 8. ISSN: 1662-5161. DOI: 10.3389/fnhum.2014.00807

Contents

Acronyms

AI artificial intelligence
ANOVA analysis of variance
AR augmented reality
BOI body ownership illusion
CASA computers as social actors
CASM computers as social mediators
CFA confirmatory factor analysis
CMC computer-mediated communication
DLP digital light processing
DOF degrees of freedom
DSLR digital single lens reflex
FOV field of view
HAAT hybrid avatar–agent technology
HCI human-computer interaction
HMD head-mounted display
HSV hue, saturation, value
IMU inertial measurement unit
IPQ igroup presence questionnaire
IPS interpersonal synchronization
IVBO illusion of virtual body ownership
MR mixed reality
MTP motion-to-photon
PCA principle component analysis
PI place illusion
PR physical reality
PSI plausibility illusion
RHI rubber hand illusion
RIS real-time interactive system
SMCs sensorimotor contingencies
SIPT social information processing theory
SVE shared virtual environment
SVR social virtual reality
TSI transformed social interaction
VBO virtual body ownership
VE virtual environment
VR virtual reality

Chapter 1

Introduction

1.1 Motivation

Research in the context of virtual reality (VR), mixed reality (MR), augmented reality (AR), and human-computer interaction (HCI) in general aims at improving and understanding the interaction between the human user and the computer system. Such interactions may be user-embodied interactions, that is, interactions in which users are represented to themselves and/or to others by avatars [28, 27]. Avatars are defined as virtual characters that are driven by the user's movements [13]. The design and development of HCI systems that support virtual embodiment, in turn, can affect both, the way we perceive ourselves, our own body scheme, as well as the way we perceive others in social interactions and the interaction itself. The simulation medium is therefore an interference in the natural cycle of perception and action [102] of a human user and plays an active role in stimulating our senses in different levels of vividness, respectively representational richness [321]. The present thesis aims at exploring the impacts of this interference with regard to individual and social interactions, and how it can be designed to actively moderate user-embodied interactions to foster future developments. It explores potentials in both, the replication of interactions closer to the physical reality (PR) as well as the abstraction and augmentation of interactions beyond the PR. We present empirical findings and guidelines for future developments of HCI systems, that may leverage computers as social mediators. The use cases for the presented results and implications are manifold. A better understanding of user-embodied interfaces can be beneficial to extend VR-based therapeutic, rehabilitation, and training applications, for example targeting post-traumatic stress disorders [257], phobia [212], social anxiety disorder [5], eating disorders [256], or motor rehabilitation [71]. Related research in psychiatry argues that many psycho-pathological conditions (i.e., mental disorders or generally norm-deviances) relate to communication disturbances (see e.g., [341]). Thus, especially therapy, training as well as research in this area can benefit from insights of how technologies can have an active part not only in the rendering of appearance, but also of behaviors in social interactions [231, 248, 111]. Collecting knowledge about the effects of user embodiment is important for game design [251, 161] and information visualization [178]. Finally, knowledge gained by the present research could inform interfaces for computer-supported collaborative work [28], and, more generally, any collaborative and communicative VR, MR, and AR application [289].

1.2 Structure, Research Questions and Contributions

The remainder of Chapter 2 will put the objectives of the present work into context. Chapter 2 further contributes a technological concept on the social augmentation of interactions in shared virtual environment (SVE) and presents a model that defines computers as social mediators. The empirical body of this thesis is divided into three parts that evaluate user-embodied interactions: *Intrapersonal effects*, *interpersonal effects*, and effects of *hybrid social interactions* that evaluate the proposed model of viewing computers as social mediators. Chapter 3 investigates intrapersonal effects of user-embodied interfaces as well as the latent variables that lead to the illusion of virtual body ownership.

RQ1a: How do simulation properties affect virtual embodiment?

RQ1b: What latent variables are responsible for the adaptation of a virtual body?

Chapter 3 contributes with findings on how realism, personalization, immersion and latency structure and affect the perception of virtual body ownership and agency. We show that immersion, realism, and personalization foster the acceptance (ownership) of a virtual body, and that both, constant and dynamic (jitter) latency in the reproduction of behavior hinders agency. We contribute a measure to assess the acceptance (ownership) of and control (agency) over a virtual body and the perceived change towards the own body schema. We further assess the relation of these dimensions to related constructs. The presented results guide further development and allow the application of the measurement instrument in future studies. Chapter 4 investigates how technologies for user-embodied interactions affect social interaction between two users.

RQ2a: How do virtual social interactions with user embodiment compare to physical social interactions?

RQ2b: How do technological properties affect user-embodied virtual social interactions?

Chapter 4 presents two prototype developments for immersive and semi-immersive user-embodied interaction in a SVE. Prototype 1 allows for the repliction of body movement whereas prototype 2 allows for the systematic variation in replicating body movement, facial expression, and gaze in a semi-immersive simulation. In three studies we find that compared to PR, social virtual interactions can be affected with regard to aspects of social presence and affect. The performance of a collaborative motor task performed in VR was reduced compared to the performance of participants in PR. Different levels of behavior realism can further affect telepresence and, indicated by our findings, the eeriness perception of the partner's avatar. We thus conclude that for social interactions in VR, simulation aspects such as behavior and appearance realism or latency can lead to changes in the perception of the interactions, and we discuss our findings. Chapter 5 contributes by conceptualizing different forms of social augmentations in virtual interactions and investigates resulting impacts.

RQ3a: What possible modifications can be designed to augment user-embodied social interactions?

RQ3b: What impacts on the perception of interactions arise from augmented social interactions?

Chapter 5 offers three contributions: hybrid and synthesized social gaze based on a behavior model, artificial nonverbal mimicry, and the visually augmentation of known nonverbal phenomena (eye contact, joint attention, grouping) based on abstraction from available sensor data. We present a prototype and study for each augmentation form and empirical findings on how social augmentations affect aspects of presence and the resulting user behavior. Based on our findings we conclude that artificial gaze models can be beneficial, for example in the case of lacking sensor data or transmission deficiencies. Artificial mimicry did not decrease the overall experience of embodied interactions to a significant degree. Further, visual transformations of social phenomena can increase aspects of presence and impact the users' social behavior. Technological implications are drawn and an architecture prototype is presented. The overarching empirical research goals are summarized in the model depicted in Fig. 1.1.

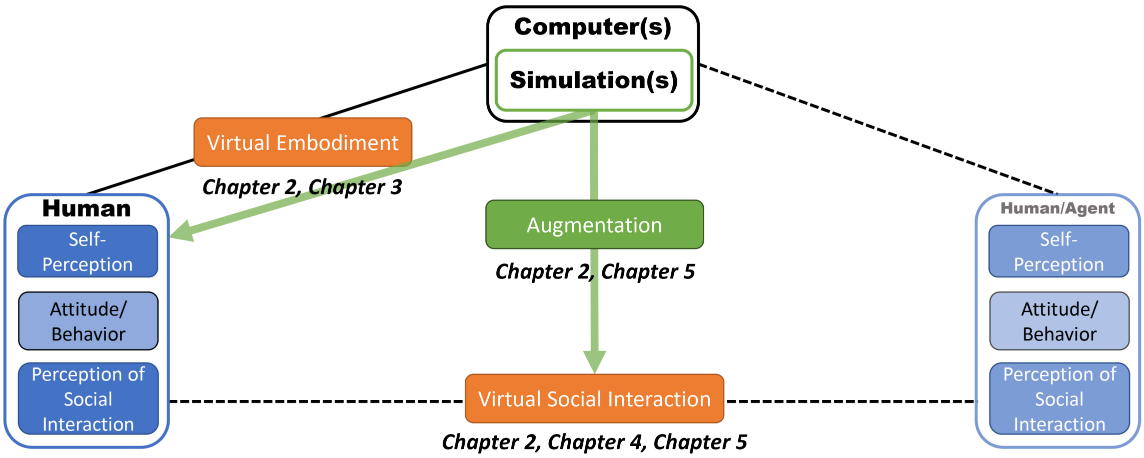

Figure 1.1: Model of interdependencies and overarching research themes. The simulation and its characteristics (green) impact the perception of the self through virtual embodiment and the perception of others through virtual social interactions (orange). Chapter 3 investigates the impacts on self-perception of virtual embodiment. Chapter 4 investigates user embodied social interactions. Chapter 5 explores the augmentation of social behaviors in SVEs.

Chapter 6 summarizes the findings presented in this thesis. We further present an ethical reflection on considering computers as social mediators. We describe how artificial intelligence (AI) could impact autonomy and privacy, and provide proposals how to counteract negative impact.

The remainder of this section will present the backgrounds of the present work with regard to both, the perceptual as well as technological relations, before deriving a model that summarizes previous work and derived hypothetical constraints.

1.3 Context: Interaction in Virtual Environments

VR systems are essentially communication systems in form of the "transfer of energy between two entities" [146, p. 10]. From this viewpoint, the user is included in the system description and interacts with the system through sensors and actuators. As with all communication systems, Shannon's basic mathematical description of communication applies; An information source selects a message to be transmitted. A transmitter produces a signal for the transmission which is transmitted through a channel and during this process subject to noise. A receiver processes the signal to deliver the initial message to the destination. Thus, noise–any communication-system generated signal that has the capacity to alter the reception of any message–is introduced during message exchange [291]. For example, in typical user-system interactions, the translation from intention to motor action and virtual feedback suffers from the delay of the processing time used by the machine (see Section 2.2.6). The accuracy of such interactions suffers from the precision of the sensor and post-processing. In turn, this noise affects the *perception–action cycle* of human information processing (see Section 2.1.1). Hence the general experience in a virtual environment (VE) is affected by the simulation may it be a single-user interacting with a system or multiple users interacting with the system and each other.

The terms VE and VR are often used interchangeably [175]. While previous work on VEs is often based on semi-immersive systems, such as desktop systems without a stereoscopic display or head-related tracking, most VR research today considers simulations that are based on the stereoscopic visual display of a camera perspective that is rendered in relation to the user's head movement. The level of virtuality of a system can be described on the basis of the *reality-virtuality continuum* described by Milgram, Takemura, Utsumi, and Kishino [206], that depicts levels of virtuality from real environments to AR to augmented virtuality to completely VEs. All in-between states are summarized under the umbrella term of MR.

When the reality-virtuality continuum is strictly applied, all current systems are MR systems. However, in the following, we use the more specific term VR for interactive graphical systems that render a three-dimensional environment considering the user's head movement, and the term VE as a broader description of an interactive system that renders a graphical environment. VEs and VR systems can be further described by a number of properties.

The level of immersion is a way to classify VR systems based on technological properties: "The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities" [297, p. 1]. Immersion in turn can foster the perception of *presence*, or *'being there'* [135] in a virtual environment (see Section 2.2).

Systems that support multiple (two or more) users, are often referred to as SVEs [289, 287], collaborative virtual environments [28], multi-user virtual environments [59, 99, 100], multi-user virtual worlds [74, 27] or multi-user VR [58, 165]. Users may be colocated or separated in the PR, as they can be colocated or seperated in the VE — see also Section 2.3 and [241] for a taxonomy.

Interfaces for three-dimensional interactions with VEs are often described using the term 3D user interface [175] which can be technically differentiated by the degrees of freedom (DOF) of the interaction. Among others, 3D user interfaces may for example make use of head motion, hand motion, or full body motion to provide useful interactions [97]. All non-autonomous interaction in VEs is dependent on user input and feedback provided by technology [50]. In consequence, no interactive system is independent of the user, which is why the user is constantly "in the loop" and human actions and senses have to be considered.

A special case of 3D user interfaces are interfaces that embody the user. By tracking human body motion and mapping the respective motion to avatars, the user virtually exists in the environment as part of the virtual projection. The user literally takes the perspective of a virtual representation. A final categorization of VEs is therefore the fidelity of user embodiment.

The resulting phenomenon of the sensual stimulation, the illusion of virtual body ownership, triggers a chain of cognitive processes that can contradict previously learned information and physical context. It is therefore essential to understand the principles of information processing when engaging the topic of user-embodied interfaces.

Chapter 2

Background and Related Work

2.1 Mechanisms of Embodiment

2.1.1 The Body and the Mind: Information Processing

Humans are fascinating biological machines with regard to the potential to perceive and process information. Our senses allow for a tremendous information throughput. Stimuli are captured through somatovisceral receptors and processed by the medulla, the brain stem, the thalamus, and the cortex followed by integrative and efferent systems that results in our behavior: the perceptions and reactions with cognitive, affective, motivational, motor, and vegetative components [375]. The approximate bandwidths of our eyes (10000000 bits/s), ears (100000 bits/s), skin (1000000 bits/s), taste buds (1000 bit/s), and nose (100000 bits/s) result in an overflow of information for our brain to consciously process [219]. Thus, incoming signals are processed and filtered for irrelevant and redundant information, reducing the information flow consciously perceived to 40, 30, and 5 bits/s for eyes, ears, and skin, respectively [376]. To handle this information, the necessary magnitude of the channel capacity of the brain, the human CPU, is estimated at 10000000000 bits/s [219].

In the physical world, our neural mechanisms process sensory information such as the visual information taken from social cues and behaviors. Based on *bottom-up* processes like stimulation and sensation, processed through short-term memory [116], as well as higher level *top-down* processes such as expectations, preacquired knowledge (long-term memory), and the use of contextual information [124], we make decisions and form intentions, and consecutively execute response actions — which is also referred to as the *perception–action cycle* [102].

Considering the above stated numbers are approximate, it is important to understand the challenge of the development of user-embodied interfaces, that is, to respect for the sensory and motor system [233], and to understand how technology impacts the flow of information processing that relates perception, cognition, and action [360].

2.1.2 Cogito Ergo Sum - Embodiment and its Subcomponents

Embodiment can be described as a part of self-consciousnes and throughout the literature, different phenomena were addressed to be components of embodiment.

Embodiment has for example be described as "the experience that the self is localized at the position of our body at a certain position in space" [186, p. 150] following [181]. Blanke and Metzinger [43] and Blanke [42] argue that through multisensory information processing, humans can perceive a conscious experience of self-identification (or body ownership), self-location, and the perspective one perceives the world with (first-person perspective). Further works investigated body ownership as a part of self-consciousness, and found that it can be disrupted by conflicting visual-somatosensory input [181].

Body ownership can be described as the experience and allocation of a bodily self as one's own body, as "my body," the particular perception of the own body as the source of bodily sensations, unique to oneself so that the own body is ever present in one's mental life [105, 335, 332]. A key instrument in the strong focus on investigating body ownership was the rubber hand illusion (RHI) [47]). The RHI is a psychological experiment that induces ownership of an artificial body part in the form of a rubber hand by simultaneously inducing (visually hidden) tactile stimulation of the physical hand combined with a visible tactile stimulation of the rubber hand. Caused by the stimulation, participants start to perceive the rubber hand as part of their body. This was substantiated through subjective questioning as well as a displacement measure where participants were asked to judge the location of their "hand" which indicated a drift in their perception. Throughout previous research, it has become salient that bottom-up accounts (multisensory processing and integration) drive the perception of body ownership and might be sufficient for an illusory body ownership experience. However, further findings suggest that top-down processes that include internal body maps (for example form and appearance matching) at least modulate this illusion in the PR [127, 332, 52] (see Chapter 3 for further details).

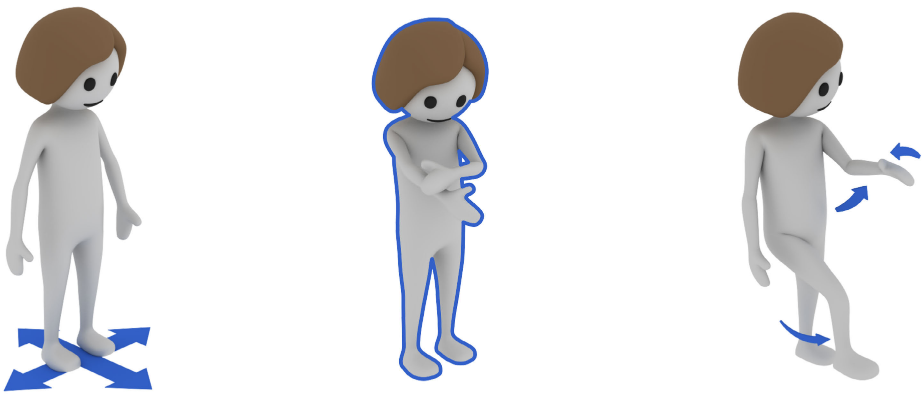

Figure 2.1: Three components of embodiment: self-location, body ownership, and agency. *Left:* Self-location refers to the experience of the self being located at the position of our body [186]. *Center:* Ownership relates to perceiving the body as the own, as the source of sensations. *Right:* Agency relates to the feeling of control over one's own actions [335].

It has been argued that agency (also referred to as the sense of agency [52]), meaning the "experience of oneself as the agent of one's own actions - and not of others' actions" ([75, p. 523,] following [105]) strongly relates to the construct of body ownership [335, 52]. In stronger relation to the context of this thesis, Tsakiris, Prabhu, and Haggard define agency as “the sense of intending and executing actions, including the feeling of controlling one’s own body movements, and, through them, events in the external environment” [335, p. 424]. While the exact relationship remains subject to further research, an RHI-based experiment analyzing a metric for assessing levels of embodiment of an artificial body part revealed the latent variables of *ownership*, *agency*, and *location* as factors where location was comprised of items focusing on the coherence between sensation and causation, as well as locational similarities of artificial and physical body part [184]. Figure 2.1 illustrates the differentiation of the three different perceptual concepts.

Kilteni, Groten, and Slater [157] summarize previous findings from physical and virtual experiments and argue to distinguish three components the sense of self-location (self-localizaion), the sense of agency, and the sense of body ownership. For the sake of consistency, we will stick to their differentiation of **self-location, agency, and ownership** for the remainder of this thesis. The findings from research considering the physical world can be extended to the virtual context.

2.1.3 Video Ergo Sum - Virtual Embodiment

Embodiment and especially body ownership receive ongoing attention in VR research. Following video-based approaches to manipulate bodily self-consciousness, which Lenggenhager, Tadi, Metzinger, and Blanke [181] entitled “video ergo sum” (abstracting René Descartes quote “cogito, ergo sum”–I think and therefore I am - referenced from [2, X, p. 535]), research found that the concept of the RHI also applies to virtual body parts [302] and even to entire virtual bodies [301, 189] to achieve an illusion of virtual body ownership (VBO), which in this context is also referred to as body transfer illusion [304]. According to Slater, Pérez Marcos, Ehrsson, and Sanchez-Vives’s review, the induction time for perception of (virtual) body ownership varies, depending on the method, between about 11 seconds up to 30 minutes [301]. Maselli and Slater [199] investigated factors driving VBO and also concluded that bottom-up factors like sensorimotor coherence and, in particular relevant to VEs, a first-person perspective, are driving factors. According to the authors, appearance moderates the experience insofar as for example realistic humanoid textures foster body ownership compared to abstract textures.

To investigate and alternate body perception, VR-based experimentation has also made use of virtual mirrors [119], and fake “physical” semi-immersive mirrors, that rely on *fish tank VR* projections [353] to evoke VBO, but keep a reference frame to the physical world [173]. Figure 2.2 depicts two approaches for experimental paradigms inducing and studying the VBO.

Summarizing previous research, the key aspects seem to be the degree and the precision with which *appearance*, *behavior*, and *stimulation* are rendered by a user-embodying interface, and which perspective (first vs. third person perspective) is presented [199, 158, 189, 119]. It remains essential to shared embodied VR to investigate preception and the technological impact of VBO as intrapersonal factors. We further discuss related research to VBO in more depth in chapter 3.

Figure 2.2: Examples of experimental platforms to study virtual embodiment and body ownership illusions. Top: A fake mirror system using a screen in combination with body and facial tracking (adapted from [173]). Bottom: Immersive HMD-based virtual mirror simulation based on passive optical marker tracking, HMD-based gaze tracking, and lower facial expression tracking.

Considering the current developments of VR technologies such as motion controllers and body tracking, it becomes salient that ownership and agency are crucial factors to consider when developing user-embodied interfaces. Design choices, for example character type, or the realism of replicated appearance and behavior may strongly influence these perceptual phenomena. Further, designers and developers could utilize distinct effects to create a non-realistic remapping of physical and virtual world. In this regard, avatars are of special concern.

2.1.4 Avatars - Means for Virtual Embodiment

The term avatar describes a digital representation of a human (user). The word as such derives from the Sanskrit word avatāra ("descent") which describes the incarnation of a deity in human or animal form in Hinduism [324]. In literacy avatars were for example topiced by Stephenson [318] in his popular science fiction novel "Snow Crash." Throughout the history of computer graphics and video

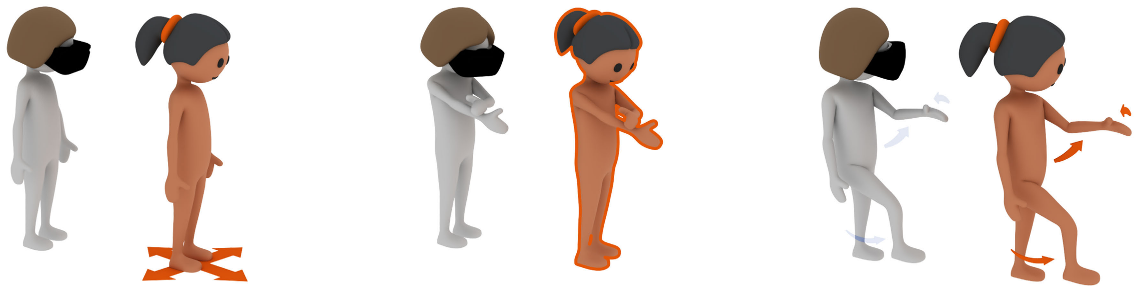

Figure 2.3: Three components of virtual embodiment according to the description of Kilteni, Groten, and Slater [157]: self-location, virtual body ownership, and agency. *Left:* Taking the perspective of an avatar (orange), a user (grey) can feel self-located in a VE. *Center:* On the basis of congruent stimulations, ownership over the virtual body can be percieved. *Right:* Agency over the virtual body is possible when the user's movements are tracked and retargeted to the virtual character's behavior.

games, the word avatar was used to describe a textual or graphical representation of the user [26]. However, whereas one might set a photo or pictographic avatar in a social network, forum or versioning system, Bailenson and Blascovich [13] define an avatar to be "a perceptible digital representation, whose behaviors reflect those executed, typically in real time by a specific human being." [p. 65]. In similar manner, Bell [26] describes: "An avatar is any digital representation (graphical or textual), beyond a simple label or name, that has agency (an ability to perform actions) and is controlled by a human agent in real time." [p. 3]. Both descriptions relate to VEs rather than a general description.

Two aspects of the above stated definitions can be further defined in the context of the present thesis. First, there are numerous options on how an avatar can be controlled, such as through gamepad or joysticks using certain control schemes, or, as also presented in this thesis, through motion tracking and retargeting, meaning the behavior data to avatar behavior, of physical movements [273]. Furthermore, the aspect of timeliness is of importance, as the delay between physical motion and virtual feedback is most likely to affect agency (see also Chapter 3 and Study 3.7). While there are no catastrophic outcomes if deadlines of behavior reproductions are not met (hard real time), outcomes of a behavior reproduction might be useless (firm real time) or decrease the value (soft real time) degrade the value of the result [294]. We can therefore describe *avatars as graphical user representations in a VE, which are animated by user behavior in firm or soft real time constraints*. While approaches of video-based reconstruction exist (see Section 2.3.1), in the present thesis we refer to avatars as rigged virtual characters that reproduce human behaviors [275]. Figure 2.4 describes the basic components of such characters. As appearance, structure and form of such virtual characters can be manifold, the question arises if users can adapt to altered self-representations.

2.1.5 Homuncular Flexibility

How far the flexibility in the adaptation of and embodiment through virtual representations (avatars) can reach was already examined in the 1980s and 1990s. In very

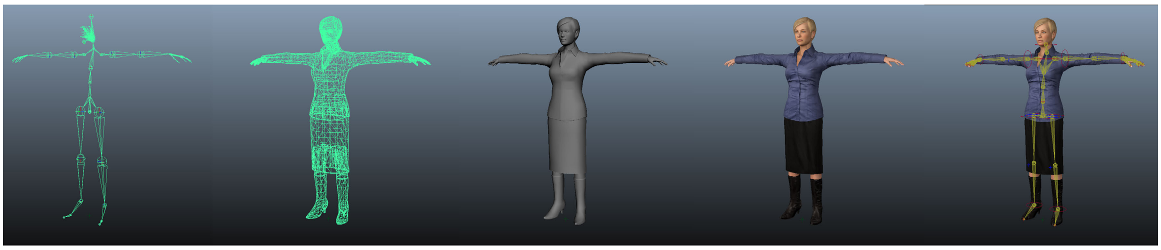

Figure 2.4: The building blocks of a character/avatar used for animation. *From left to right:* 1) A hierarchical skeleton defines the DOF joints can be restricted in their angular behavior. 2) A polygon mesh defines the surface hull of the avatar. 3) The surface material defines shading properties. 4) The texture defines the textural appearance. 5) A control rig can be used for pose solving in keyframe or inverse kinematik driven behaviors, based on a solver model.

early and visionary works, Jaron Lanier and colleagues investigated what was discovered by coincidence in a collaborative project by Tom Fourness and Jaron Lanier, namely the neural plasticity of humans in being able to adapt to non-human avatars that do not match any top-down expectations with regard to their appearance or behavioral control [365, 366]. In particular, the review by Won, Bailenson, and Lanier [365] describes published and unpublished explorations of adaptation to alternate anatomies (such as animal anatomies) and the flexibility of the somatosensory and motor cortex to adapt to such alterations in accepting and controlling the alternative visual representation. Their examples include remapping limbs in such a way that a participant's arm movement would control the avatar's foot, and vice versa, as well as remapping hip movements to animal body parts [320] or the adaptation to different gender appearances. This non-human remapping, which could be described as a form of virtual human augmentation, may also include additional virtual limbs ("ipsimodal remapping"), a remapping of sensory information to different channels (such as visual information presented through audio feedback, called "sensory substitution"), or a remapping of tracked input that is not consciously processed to alternative feedback, such as an avatar changing its color based on the emotions, which are tracked through neurophysiological sensors ("para-synthetic expression") [366]. In consequence to adapting altered body representations in VR, the users may perceive those as their own body and perceive VBO. In consequence, this illusion may change the perception of the own body scheme and as research shows, this can impact the user's attitude and behavior.

2.1.6 Impacts of Altered Virtual Bodies

The fact that, by utilizing VBO, humans can accept altered virtual bodies was exploited by Yee, Bailenson, and colleagues. Coined as the *Proteus Effect* [368, 367, 367, 369], they found a change in behavior, self-perception and identity of participants by taking the perspective of an avatar with altered appearance. In other words, participants changed their behavior and attitude according to behavior they attributed to their virtual representation.

For example, white skinned participants embodying a black skinned avatar showed a reduced racial bias [232, 19] or being placed in the perspective of an old looking avatar reduced negative stereotyping of the elderly [368]. Further confirmation of related effects of altered embodiment have been found with child-avatars and the overestimation of size [18]. Similarly, You and Sundar [370] found indirect impacts from avatar representation (with or without a backpack) on the perception of environmental steepness. In VR simulations for individual and shared experiences, it is therefore important to keep in mind that the adaptation of virtual bodies is not a one-way street; rather, it is a symbiotic effect in which control over a body goes along with top-down processing (see Section 2.1.1) and attributions based on what users have experienced throughout their lives, which can have a strong impact on perception and behavior.

2.1.7 Summary

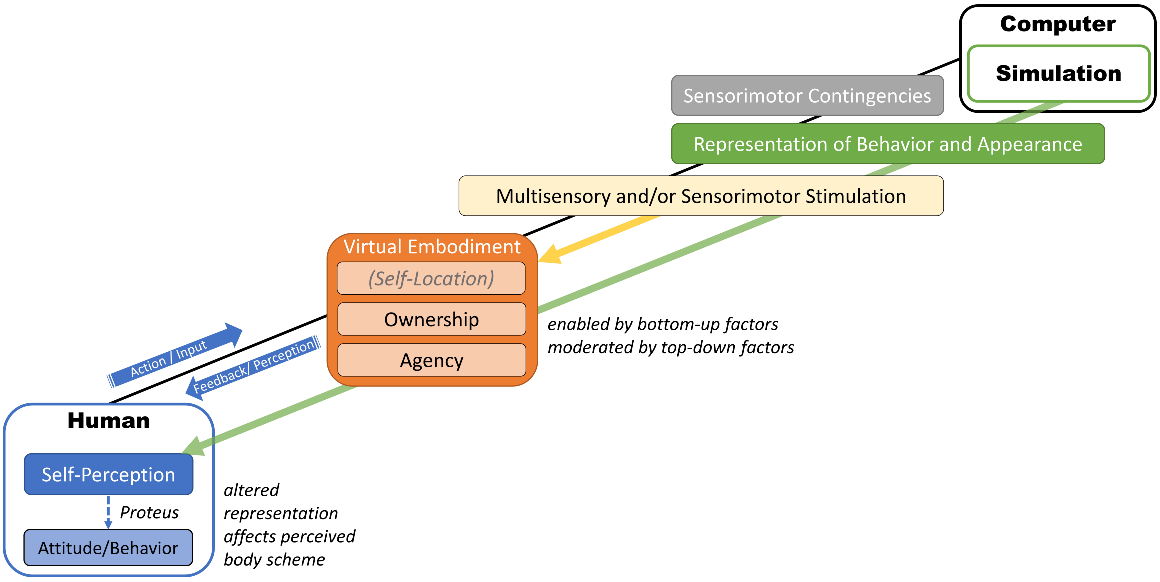

Figure 2.5: Model of the relations of interaction, virtual embodiment and self-perception. The user performs actions and is stimulated by (multimodal) feedback of the simulation. This loop interferes with the sensorimotor contingencies (SMCs) of human cognition (i.e., motion to predicted natural response). Through visual feedback of the simulation, avatars allow for virtual embodiment. It's components, such as ownership and agency are elicited by multisensory and/or sensorimotor stimulation [158], providing for the necessary bottom-up processing. Top down factors such as the humanness of the body (-part) representation were found to moderate the virtual body ownership [199]. Altered self-representations in the virtual world change the self-perception, and the user behaviors and attitudes can change according to what they attribute to their representation (Proteus effect).

In everyday life, healthy humans are self-aware, they perceive self-localization in the world through their body, their body as source of all sensations (ownership), and feel in control of actions and resulting events (agency) [43, 52, 186, 105, 335, 332, 75, 334, 332]. Previous research found that the human brain has the capability/the plasticity to adapt an artificial body (part) when sufficient bottom-up and top-down processing is triggered, such as with simultaneous visual and tactile stimulation [47, 333]. Through user-embodying interfaces that allow for the stimulation of similar processes, persons can adapt virtual body parts and whole virtual bodies [181, 302, 301, 304], even ones with abstract form or behavior representation [365, 366]. Figure 2.5 describes these processes. The adaptation of such virtual bodies can lead to the adaptation of user behavior. The users adapt their body scheme, behavior and attitude on the basis of their virtual representations [367, 367, 369, 368, 125, 232, 19].

While self-location (self-localization) was stated to be part of self-consciousness [186, 181, 43, 42, 157], especially in the context of VEs may seem related to another aspect of perception: presence or the sense of "being there" [135]. In contrast to self-location, presence must not necessarily involve to visualize or render a virtual body to the user [157].

2.2 Presence in Virtual Environments: Being there

Early on, Goffman [117] described presence as an aspect of everyday social and communicative life, in which individuals enter the presence of others to acquire and exchange information through communicative signals. In a different context, Minsky [207] used the term telepresence to describe the potentials achieved by versatile teleoperation systems that support sensory feedback in high quality, and thus allow for a remote presence that "possesses the strength of a giant or the delicacy of a surgeon" [p.1]. IJsselsteijn [141] interprets Minsky's description of telepresence as the "phenomenon that a human operator develops a sense of being physically present at a remote location through interaction with the system's human interface, that is,through the user's actions and the subsequent perceptual feedback he/she receives via the appropriate teleoperation technology" [p.1].

The general concept of presence, also referred to as the perception of "being in an environment" [321], or simply "being there" [135], was prominently investigated. Previous work proposed to distinguish different aspects of presence, for example, conceptually into physical, social, and self-related aspects [179] or on a psychometric basis into the factors involvement, sensory fidelity, realism, interface quality, adaptation, and immersion [363].

Previous research identified many aspects that affect the perception of presence as well as strategies to measure presence, which have been broadly discussed and argued upon in interesting debates [363, 364, 308, 321, 183, 179, 249]. In the context of the present thesis and in addition to telepresence, three previously identified aspects of presence are of special importance: Self-presence, social presence, and copresence.

2.2.1 Self-presence

An especially important aspect for interfaces embodying the user is self-presence, or "the extent to which the self is present (relevant) during media use" [249, p. 323], see also [179]. It is related to the self-awareness and self-similarity of the virtual representation in a virtual environment [252], or to "the extent to which a user feels that her avatar, or virtual representation of self inside a virtual world, is an extension of herself" [252, p.167]. Biocca describes self-presence as "the effect of virtual environment on the perception of ones body (i.e., body schema or body image), physiological states, emotional states, perceived traits, and identity", and thus "refers the effect of the sensory environment on mental models of the self" [39, p. 22f]. Putting these descriptions into context, self-presence seems to be naturally related to above noted components of embodiment.

2.2.2 Copresence

As noted above, in his book *The Presentation of Self in Everyday Life* Goffman [117] describes early on that presence is strongly related to the perceptions and impressions of others.

> *"When an individual enters the presence of others, they commonly seek to acquire information about him or to bring into play information about him already possessed. They will be interested in his general socio-economic status, his conception of self, his attitude toward them, his competence, his trustworthiness, etc."* [117, p.136]

In a communicative context, the concept of presence is therefore strongly bound to the concept of copresence, which can be described as the perception of "togetherness" or being colocated in a shared virtual world with others [82, 179, 288]. Nowak and Biocca [220] argue that this connection with others can be further distinguished with regard to level of copresence initiated by oneself, as well as the level of copresence initiated by the communication partner, which emphasizes an interaction.

2.2.3 Social Presence

In a similar sense, the perception of a social exchange or social presence, defined early on as "the degree of salience of the other people in the interaction" [295, p.65], or "a medium's ability to connect people" [220] influences the communication in mediated environments. Biocca [39] describes: "The minimum level of social presence occurs when users feel that a form, behavior, or sensory experience indicates the presence of another intelligence. The amount of social presence is the degree to which a user feels access to the intelligence, intentions, and sensory impressions of another." [p. 22]. Oh, Bailenson, and Welch [223] conducted a recent systematic review on social presence and summarize that immersion and context have a positive impact on social presence. They also find that a visual representation and a higher behavioral realism increase social presence. They argue that despite several findings of a positive impact of social presence on communication, depending on the situation, attempts to increase social presence may not always result in positive effects. For example, it may have negative impact on people with social anxiety. It seems clear that social presence stands in relation to copresence, but through the above mentioned description by Biocca, we may distinguish social presence (interpreting feelings, intentions) from copresence (perceiving connectedness, togetherness). As one component of presence perceived in SVEs, social presence is affected by the affordances a simulation presents to the user. To this regard, previous research identified characteristics that foster presence (in general), to which regard Slater [299] introduced helpful notions.

2.2.4 Simulation Characteristics that Foster Presence

To identify factors that affect presence it is helpful to refer to alternative descriptions. Slater argues presence to be a qualia (phenomenal awareness, consciousness) and describes presence as place illusion (PI), "constrained by the SMCs afforded by the virtual reality system" [299, p. 3549]. According to sensorimotor contingency theory, "action is constitutive for perception" and "actions and associated sensory stimulations are tied together by lawful relations", namely SMCs; [154, p.47] following [222, 217, 221]. For example, in the physical life these are implicit rules that guide action and perception. Looking behind an object is something we do not need to think about, we perform motion to change the visual perspective. In VR however, these SMCs are constrained, for example by the tracking fidelity and reproduction of perspective. While Slater [299] argues that there are no high fidelty VR systems (first-order systems) available and that PI cannot be directly assessed, he argues that PI, strongly bound to immersion, may also occur in lower level second-order systems when physical actions are accurately mapped by the system, meaning the system provides SMCs mapped between real and virtual world. Thus "People can report a feeling of 'being there' to the extent that they engage in *additional mental recreation* that transforms their actions for perception into the feeling of being in a space which they are already *not located* according to the rules of real-world SCs" [299, p. 3552]. In other words, if the system stimulates a contingent interaction between physical behavior and virtual feedback, PI can occur. In consequence, PI is limited to the boundaries of the immersion a system provides.

In his 2003 essay on presence terminology, Slater refers to immersion as follows: "The more that a system delivers displays (in all sensory modalities) and tracking that preserves fidelity in relation to their equivalent real-world sensory modalities, the more that it is 'immersive' " [297, p. 1]. Immersion therefore describes technological aspects that foster a simulation's realism or "vividness" (i.e., the richness in which the simulation is able to represent information to the senses; [321]) which can be assessed in a technical sense. Thus, in contrast to presence or PI, immersion is a system property that "can be objectively assessed, and relates to different issues than how it is perceived by humans" [297, p. 1]. Immersion is therefore a characteristic that can foster presence or PI, yet presence is not solely dependent on immersion. Some driving factors for presence identified by previous work (see [308] for a summary) are of importance because they strongly relate to the present thesis. They were identified as external factors, that is, factors supplied by the simulation technology as such.

Sensorial Richness Referring to works from Steuer [321] and others, Slater, Usoh, and Steed [308] concluded that the richness or vividness (in other words the quality and resolution) of information presented to the user, activating sensory organs, impacts the perception of presence in VEs. Their interpretation points out a multi-sensory richness in terms of the activation of different senses, whereas avoiding the *Cyborg's dilemma* [39], referring to a change of body schema based on the interface of the physical body with technology. However, one could also argue that the vividness of social and perceptual senses activated by simulations based on a reflection of the virtual self and of virtual others. As a result, in the context of this work, the present interpretation stresses the importance of the richness of user behavior transmitted and replicated in the VEs.

Self-Representation Among the factors fostering presence, Slater, Usoh, and Steed [308, p. 1] identified that the virtual self-representation of the user "should be similar in appearance to the participant's own body, respond correctly, and be seen to correlate with the movements of the participant." In consequence, one could argue that both, accuracy in appearance as well as in behavior replication is of importance for the vividness of VEs.

Consistency Across Displays In their summary and research Slater, Usoh, and Steed [308] and Slater, Steed, and Usoh [307] argued that a VE presented to a user needs to be consistent across all displays. Interpreting "displays" more generally as feedback, then it would be important to achieve similar levels of realism and vividness of the simulation across modalities. Therefore, a similar and consistent realism in the retargeting (i.e., mapping input to visual output) and replication of motor movements as well in the appearance may foster presence.

In conclusion, we identified and considered the factors of *sensory richness*, *self-representation*, and *consistency* may also describe key properties to potentially improve the state of present user-embodying immersive simulations.

2.2.5 Plausibility Illusion

Slater differentiated another dimension that impacts perception and behavior in VR, that is, the plausibility illusion (PSI), which is determined "by the extent to which the system can produce events that directly relate to the participant" [299, p. 3549]. The PSI, in contrast to PI/presence, depicts stimulation from external events beyond the control of the participant, such as reactions of other agents/avatars or simulation events, and thus "*correlations between external events not directly caused by the participant and his/her own sensations (both exteroceptive and interoceptive)*" [299, p. 3553]. While it is argued that PSI is not affected by physical realism, PSI may be affected by social and behavioral reactions of interaction partners and the plausibility of the overall context. Slater further argues that both PI and PSI have to occur in order for users to behave realistically in VR.

While differentiated from presence and PI, PSI may therefore be affected by the way a simulation is presented, in the context of the present thesis, specifically by the way interaction partners and their behavior is presented. For example, hearing a communication partner's voice and seeing an avatar that does not respond with respective mouth movements may hinder PSI. In turn, PSI and the credibility of the simulation may therefore affect aspects of social presence or copresence as a predecessor [223, 118].

2.2.6 Not Being There

As with factors that foster presence and vividenss of the experience, a vast number of factors have been identified that hinder presence and degrade the overall quality of the experience. Among those, two factors are especially important with regard to the present context: breaks in presence as well as the *timeliness* of the simulation, which are both related to the ability of the simulation to render persistently in real time. Both factors can be linked to the principles of human cognition so that we can refer to the way humans process information, as discussed in 2.1.1, and how further actions in the physical or virtual world are performed. The execution of actions on the basis of information, the *perception–action cycle* can be described as "the circular flow of *information* between an organism and its environment in the course of a sensory guided sequence of actions towards a goal" [328, p. 601], see also [102, 101]. According to the overview by Fuster [102], the prefrontal cortex of the human brain is mostly dedicated to memory, planning, and execution of actions. Reciprocal connections between the posterior cortex and the frontal cortex interface the sensory system with the motor system. These recurrent connections initiate a sensor-motor cycle of interaction and link the organism with the environment [102]. Rationally, when either expectations (top-down processes) or stimulation (bottom-up processes, see Section 2.1.1) and in consequence sensorimotor contingency is disturbed, the quality of the virtual experience suffers.

Breaks in Presence Breaks in presence can be broadly defined as discontinuities in a user's perception of presence between two environments, in other words, when the user switches between two alternate environmental gestalts ("I am in the environment the simulation depicts" vs. "I am in the laboratory doing an experiment") [305]. Thus, breaks in presence are situational variations based on specific events [182] which can for example occur due to unreal collision with virtual objects [374]. With respect to the topic of the present thesis, they may also occur due to tracking jitter or incorrect body poses, such as false calculation of the kinematics of the virtual character or an implausible collision of limbs of the virtual character. For example, when a markers of a marker-based tracking system are falsely interpreted (marker switch), rapid joint recalculations can for example lead to snapped off arm, which in turn can hinder the illusion of virtual body ownership (IVBO) and agency due to implausible physical conditions [158]. Rationally, such artifacts may also affect the credibility of the simulation and thus the related PSI.

Timeliness and Latency *Timeliness* refers to the ability of a real-time interactive system (RIS) to process, deliver, and display a simulation's computed information according to a user's actions and with no perceivable delay which is a fundamental challenge in the development of VR systems. All interactive VEs suffer from latency, which can be defined as "the time lag between a user's action and the system's response to this action" [96, p. 1], following Papadakis, Mania, and Koutroulis [229, p. 1]. Latency interferes in the natural human perception–action cycle and disrupts SMCs. Both, a higher constant motion-to-photon (MTP) latency as well as a dynamic latency (*latency jitter*) lead to mapping inaccuracies between the user's head orientation/position and the virtual camera's perspective. In consequence, both degrade the overall virtual experience and have negative effects such as reduced user performance, higher *simulator sickness*[1], and decreased perception of presence [92, 156, 313, 229, 96]. Figure 2.6 shows an RIS system that considers the user to be part of the loop.

Humans can detect *motion* at speeds as fast as 10 ms after stimulus [155], and *meaning* that around 13–40 ms (results vary) [195, 242]. Despite the fact that there are approaches to extremely low sensing and display latencies [95], the most problematic factor is that the data transmission in any distributed system is restricted by the laws of physics. Elbamby, Perfecto, Bennis, and Doppler [87] argue that while sensor delays could be negliglible in the future (around 1 ms) and display delays are likely to drop to about 5 ms, the remaining 14 ms for computation and communication of the data is most likely the bottleneck for distributed systems. However, local systems can achieve much lower latencies than currently examined, which roughly vary between 20 ms and 200 ms, depending on the sensor and display technologies. With regard to embodiment and behavior reconstruction, a full reconstruction of body poses and body motion was achieved in 38 ms for a local system using current technologies [349].

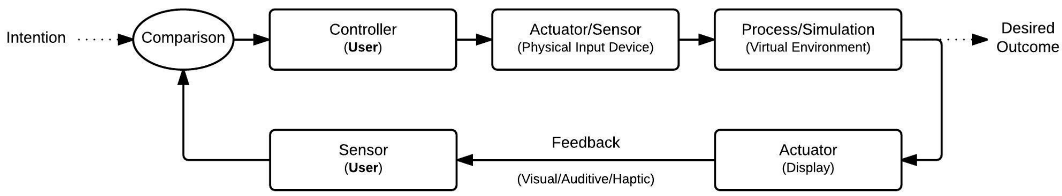

Figure 2.6: RIS loop considering the user as a part of the system (adapted from [273] following Englander and Englander [88] and Dorf and Bishop [79]). An initial intention results in action that is processed by the simulation. The simulation provides feedback to the user, who compares the outcome with the initial intention.

2.2.7 Summary

Among others, self-presence, copresence and social presence are aspects of presence that were identified by previous research. By the description of self-presence there seem to be parallels to virtual embodiment. Yet, self-presence may not necessarily be dependent upon embodiment, agency, or ownership. For example, an avatar controlled through keyboard and mouse, visualized from a third person perspective may extend the self, but may not necessarily evoke high levels of ownership and agency. Presence is strongly associated with the perception and impression of others [117]. Therefore, in computer-mediated interactions, aspects of social presence and copresence may be affected by the simulation's characteristic. In previous works

[1] It has been argued that a distinction should be made between simulator sickness and "cyber sickness" [312]. Because current VR research mostly assesses simulator sickness rather than cyber sickness (using the simulator sickness questionnaire [156]), we use the term universally to describe sickness resulting from virtual simulations.

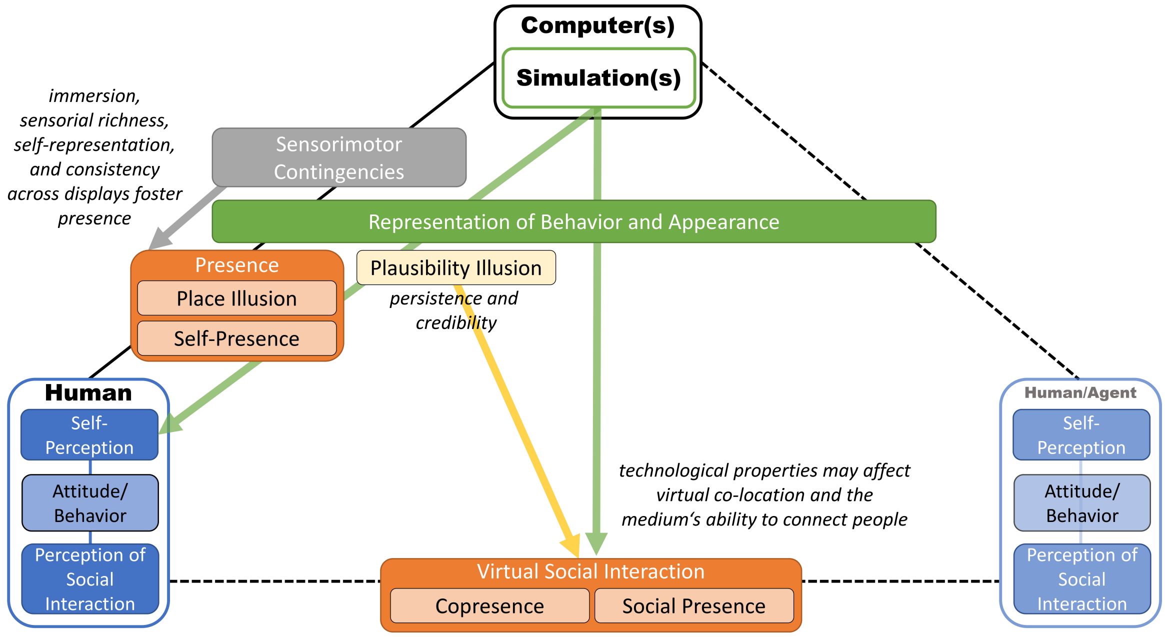

Figure 2.7: Model of the relations of interaction and aspects of presence. Properties of the simulation (sensorial richness, self-representation form, consistency of displays) affect presence. Simulation properties (e.g., context, modalities) can also affect copresence and social presence. As a related concept, PSI relates to the credibility and may affect social perception as well as self-perception.

Steuer [321], Slater, Usoh, and Steed [308], and Slater, Steed, and Usoh [307] argue that *sensorial richness*, the *self-representation*, and the *consistency across displays* may foster presence in VEs. PSI, as a related concept, can affect the perception of the self and the perception of others. Figure 2.7 depicts the aspects of presence and the relations identified by the literature. Social presence and copresence depend on interactions in shared environments.

2.3 Social Interaction in Shared Virtual Environments

In 1990, Blanchard, Burgess, Harvill, Lanier, Lasko, Oberman, and Teitel [41] were among the first to create a system for immersive networked VR rendered through "eyephones" which today are named head-mounted display (HMD). The RB2 ("reality built for 2") supported gestural interaction via tracking gloves, and the lab developed body sensing techniques. Further iterations of systems that allow for multi-user interaction can be categorized into two major approaches: 1) *Telepresence systems* that build upon Minsky's idea of telepresence [207] in the context of communication and collaboration [25], which typically involve RGB-based real-time 3D capturing and reconstruction of human users, and 2) *avatar-mediated systems* that drive computer generated virtual characters (virtual "surrogates"), based on human behavior input.

Figure 2.8: An example of a telepresence system from Beck, Kunert, Kulik, and Froehlich [25]. Users are sensed by RGB-depth cameras to reconstruct their appearance and behavior in virtual interactions. Reprinted with permission. Copyright ©2013, IEEE.

2.3.1 Telepresence Systems

In early works, Fuchs, Bishop, Arthur, McMillan, Bajcsy, Lee, Farid, and Kanade [98] conceptualized realistic telepresence systems based on a multi-camera (a passive "sea-of-cameras") approach, which later was essentially built, on the basis of multiple Kinect cameras [196]. In a similar approach, the system proposed by Beck, Kunert, Kulik, and Froehlich [25] enables projection-based multi-user interaction by tracking the users behaviors and appearances through multiple RGB-depth cameras and rendering their respective virtual images to the communication partners. They found similarities in the mutual understanding of pointing and tracing gestures in local and remote collaborating groups. Their application supports collaborative interaction and exploration [165]. Billinghurst and Kato [37] demonstrated that these principles can be adapted to MR and AR applications. Otsuka [227] created MMSpace, a system supporting kinetic displays mirroring the remote user's head position in group-to-group conversations, which they find to be superior to a static avatar. Exploiting recent sensing advances, Holoportation [226] allows the projection of users, tracked through multi-camera systems realistically and reliably onto AR devices.

Recent research demonstrates that telepresence systems based on RBG/RGB-depth sensing can create high-quality replications of appearance and behavior. One limitation is that, unless an alternative sensing system is used, these systems depend on a clean optical path between the camera sensor and the user for features such as replicating gaze, which would be hindered by an HMD. Second, and most important, they are limited to replicating the real or PR, contrary to avatar-based systems that allow for a variety of alterations.

2.3.2 Avatar-Mediated Systems

Avatar-mediated systems are an alternative to telepresence systems. Avatars, virtual characters driven by human movements [13], are a means of representing the user. In contrast to telepresence systems, avatar-mediated systems use a virtual character with a hierarchical anatomy to retarget human behavior to avatar behavior (see also

Figure 2.9: Examples of avatar mediated-systems. *Left:* An immersive system that reproduces behavior to avatars and renders to HMDs. Avatars transmitted can be of different form or gender. *Right:* A fish tank VR system that reproduces behaviors to avatars and renders to a projection display.

Section 2.1.4). Among early works, Slater and Steed [306] and Tromp, Bullock, Steed, Sadagic, Slater, and Frécon [330] demonstrated the feasibility and investigated the impact of multi-user avatar-mediated communication using the DIVE software [58] for distributed virtual environments [93] to gain first insights. Because behavior replication was limited to keyboard input, participants reported a lack of feedback due to missing body movement and body language [330]. In later iterations, Steptoe, Oyekoya, Murgia, Wolff, Rae, Guimaraes, Roberts, and Steed [319] and Roberts, Wolff, Rae, Steed, Aspin, McIntyre, Pena, Oyekoya, and Steptoe [260] developed systems to investigate additional modalities for social interaction by including eye gaze. Bente, Rüggenberg, Krämer, and Eschenburg [35] included gaze and hand tracking in a desktop-based avatar-mediated communication platform. With the improvements in body motion tracking and the improvement of pose recognition and kinematic solving (e.g., [296]), full-body motion could be reliably replicated to avatars by using RGB-depth sensing [322], or passive marker systems [271]. Figure 2.9 displays two avatar-mediated systems. An integration of body motion and facial expression in platforms using a fish tank VR metaphor (Figure 2.9 right) as well as more immersive approaches of embodied interactions will be discussed further in Section 4.3. Regarding the flexibility of avatar-based systems, Piumsomboon, Lee, Hart, Ens, Lindeman, Thomas, and Billinghurst [240] demonstrated the dynamic potential in the usage of avatars with respect to avatar size in symmetric and asymmetric remote collaboration in MR with redirected gaze and gestures. Steed, Steptoe, Oyekoya, Pece, Weyrich, Kautz, Friedman, Peer, Solazzi, Tecchia, et al. [314] combined telepresence and avatar-mediated approaches for asymetric communication.

Both telepresence systems and avatar-mediated systems are capable of being displayed via various technologies. Due to the flexibility of the dynamics of avatars, which are, by nature, exchange media for human behavior, we based the approaches and prototypes presented in this thesis on avatar-mediated systems. While the realism of appearance and behavior replication can currently be seen as less sophisticated than telepresence systems such as Holoportation [226], the systems from Beck and colleagues [25], or Maimone and colleagues [196], novel approaches to the personalization of avatars by means of photogrammetric scanning may close the gap of reduced appearance realism [1, 348] — the latter is discussed as part of Chapter 3.

2.3.3 The Elevation of Social VR and Embodied Interfaces

While consumer applications of SVEs date back to the first chat systems or multi-user dungeons or dimensions (see e.g., [288, 287, 45] for reviews) the improvement of game engines and network protocols contributed to further developments of more immersive and embodied SVEs. Most prominently, Second Life[2] was arguably the most successful consumer SVE application to date, gathering millions of users to explore a virtual world with real-time avatar animations. The release of affordable HMDs allowed SVEs to become more immersive and embodied, today often termed social virtual reality (SVR) applications.

The major difference between previous applications and SVR applications such as Oculus Rooms[3], AltspaceVR [4], High Fidelty[5] and others[6] is the ability to replicate a first person perspective and more natural behavior, as well as to allow for a higher level of immersion. All of the above are avatar-mediated systems, with different degrees of behavioral and appearance realism.

Despite the fact that behaviors are not necessarily replicated to full fidelity and today's versions are limited in terms of realistic behaviors and appearance, these applications extend previous technological foundations (see Section 2.3.2). For example, current applications include sophisticated frameworks for networking, transmission, environment creation, and animation, to achieve an overall increased realism and, thus, generally to allow for the *sensorial richness*, *realistic self-representation*, and *consistency across displays*.

2.3.4 Nonverbal Communication and its Relevance

A vast amount of psychological, philosophical, and cognitive research dedicated specifically to nonverbal communication underscores the importance of nonverbal behaviors for interpreting others and expressing oneself in social encounters. Thus it is an important dimension to consider when designing and developing shared virtual environments. The present work describes the most relevant phenomena to consider (see also Chapters 4 and 5).

Communication is one of the characteristics that decisively define human nature [329]. Social interaction through adaptive nonverbal behavior [55] is a crucial element of our everyday life that develops in early childhood [56]. Despite being the "hidden dimension" of communication [129], it is estimated that between 65% and 95% of information is communicated nonverbally in social interactions [200, 54]. Even for abstract SVE with little degree of replicated nonverbal behaviors of

[2] 2003, Linden Lab, San Francisco, USA, `https://secondlife.com/`
[3] 2019, Facebook Technologies, LLC USA, `https://www.oculus.com/experiences/gear-vr/1101959559889232/`
[4] 2019, Microsoft Corporation, USA, `https://altvr.com/`
[5] 2019, High Fidelity Inc., USA, `https://highfidelity.com/`
[6] `https://virtualrealitytimes.com/2017/04/16/social-vr/` (list of applications)

the users it may hold true that "*one cannot* not *communicate*" [356, p. 51] and even more important, one "cannot *not* respond to these communications" [355, p. 49]. For example, in everyday life, we show facial expression and body postures that even though not consciously executed allow for communicative interpretation. With regard to virtual social interaction this further implies that any visual representation, dynamic or static, implies a communicative interpretation from the partner.

The cause of this is the vast number of motor activation and expression possibilities—and in turn nonverbal communication channels—humans have are equipped with [72, 73]. Argyle [6] explained human social behavior and our nonverbal system, naming the signals: "facial expression; gaze (and pupil dilation); gestures, and other bodily movements; posture; bodily contact; spatial behaviour; clothes, and other aspects of appearance; non-verbal vocalizations; and smell" [p. 1]. As a dimensonal example, the entire human body consists of approximately 630 muscles, of which an average of 2.6 (i.e., a muscle redundancy) control one of the 240 kinematic DOF in the musculoskeletal system [372].

Bente and others [32, 33, 31, 342] distinguish the following functionalities of nonverbal behavior with relation to communication: 1) *modeling and coordination functions* that relate to the coordinated learning of motor behavior from others and groups and adapting future behaviors accordingly; 2) *discourse functions* that relate to speech understanding and production in interactions, such as the interpretation and production of beat gestures, emblemic gestures, and illustrations; 3) *dialogue functions* that include the interpretation and production of specific behavioral phenomena and back-channeling, such as eye contact, turn-taking, and head nods; and finally, 4) *socio-emotional functions* that relate to impression formation and the communication of emotions and interpersonal attitudes (see [342] for a recent description).

While specific aspects are discussed more detailed in Sections 4 and 5 of this thesis in the context of empirical work, it is important to understand the basic paradigm of intentional (or unintentional) nonverbal communication and, thus, the influence of person A on person B, as depicted in [6]; see Figure 2.10.

Figure 2.10: Encoding and decoding of communcative information. A conscious or unconscious state by person A results in the intentional or unintentional encoding of communicative information as a nonverbal signal, which is then decoded by person B (adapted from [6]).

This rationale is not particularly striking, but what is interesting in this notion from Argyle is what wording is used: the current "state" of A influencing the "state" of B in indirect manner. If we continue the format, adding the dimension of time (see Figure 2.11), we see that, logically, over the time, the state of A affects the state of B, and vice versa (see [285] for an engagement approach). Continuing this paradigm over time eventually leads to two major results: 1) the state of each individual affects the state of the other individual over time and 2) throughout the dimension of time, message sending and responding concludes in an adaptive process, that needs to be discussed in a little more detail.

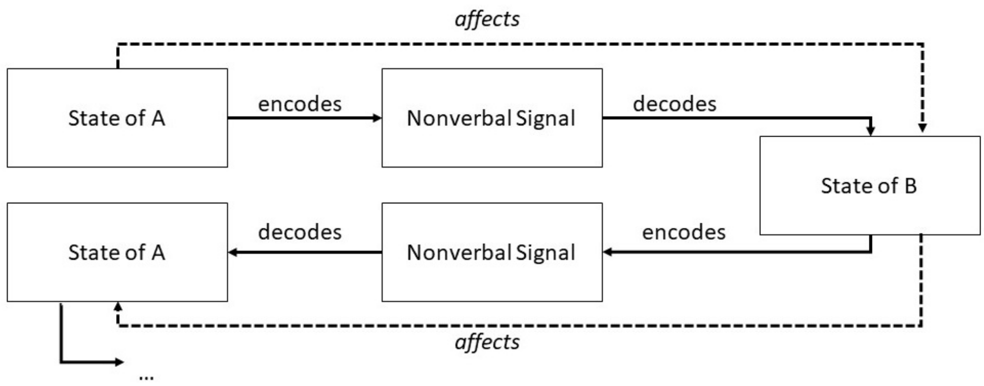

Figure 2.11: State adaptation of two interactants over time. Through the exchange of communicative messages over time, the state of A affects the state of B and vice versa. In turn, an adaptation over time occurs.

Interpersonal Synchronization, Adaptation, and Coupling

During social encounters, humans coordinate their feelings, intentions, and actions with others [107]. They shake hands, establish eye contact, move closer to each other, or even unconciously mimic interaction partners to create liking, rapport, and affiliation [168]. To express and perceive nonverbal behaviors, humans are equipped with multiple sensors and actuators, coupled with a neural processor [233] — see Section 2.1.1.

Humans perceive and respond to social behaviors such as gestures "...in accordance with an elaborate and secret code that is written nowhere, known by none, and understood by all" [284, p. 556]. These coordination processes are and have been previously investigated under the terminologies of interpersonal adaptation [55], interpersonal synchrony [68], interpersonal alignment [238], and interpersonal coordination [36], that all describe phenomena of *intentional or unintentional spatial and/or temporal movement coordination during joint social interaction*. We summarize these phenomena under the term interpersonal synchronization (IPS) for the remaining sections. There are a number of theories and empirical findings that follow different angles on how to explain these phenomena, but the following are of most importance to the context of this thesis.

The theory of mind [244] and theory theory [211] both constitute or relate to the idea that humans are capable of ascribing mental states to other persons and are using this information to explain and predict the actions of others, as in a mentalistic ability, a mental representation, or a theory by which we explain behaviors and infer other people's mental states, intentions, emotions, and personal traits. [197, 359]

The simulation theory most prominently underlined by the discovery of the mirror neuron system [163, 258] in apes, follows the understanding that certain visuomotor neurons responding to visual input in the (specific) form of an object-directed actions can induce motor activation [259]. Thus humans read out their own reactive embodiment, and the motor cortex can become active without any physical movement by that specific human. In turn, the theory relates this system to imitation and learning.

Finally, the embodiment or embodied cognition model of cognitive science evolved through the elevation of non-linear dynamic systems and opposition to a simplistic description of the mind based on discrete and syntactic rules [80], leading to a stronger inclusion of *the body* [338] and, finally, the inclusion of the dimension of intersubjective interaction [326, 106] in the model [104, 77] that influences the interpretation and understanding of virtual social interactions in the context of this thesis. To some degree, this model interfaces the importance of the body as the source of sensations, the strict interconnectivity and interdependency of body and mind in the form of sensorimotor coherence, and the processing of social interactions in a contextual role, accounting for processes of IPS. In consequence, this stresses the importance of time and the timeliness of the simulation, and whether or not the simulation, for example a SVE that supports user embodiment, is capable of transmitting nonverbal information to a degree that suffices to allow for the emergence of IPS. With regard to digital communication media, especially media that is asynchronous or text based, the question arises what strategies of human users were identified to uphold these patterns of interaction when the media offers less vividness in transmitting social behavior.

Social Information Processing in Mediated Communication

With regard to social information transmitted through computer-mediated communication and reduced levels of social cues that accompany communication absent of nonverbal markers, it has been argued in the cues filtered-out theory [69] that the reduced transmission of nonverbal cues in text-based or chat communications degrades and impairs the communication because nonverbal cues facilitate social functions such as identification, expression of emotion, and interpreting personalities. Yet many recent findings contradict this theory.

For instance, the social information processing theory (SIPT) [351] posits that users can creatively compensate for missing behavioral channels and social cues by encoding social information into other available channels and alternative cues [351]. This functional approach to communication states that, in addition to distinct verbal and nonverbal cues, any combination of social/nonverbal behaviors can convey information and contribute to communicative functions (see the recent chapter of Walther, Van Der Heide, Ramirez Jr, Burgoon, and Peña [352], for a review on theoretical directions).

There is a lack of research on VR and immersive environments that supports user embodiment, as most prior studies investigated text-based (chat) communication or multi-user video games [352]. Despite findings of positive impacts for higher modalities, it is yet unclear how and to what degree the activation and multimodal replication of nonverbal behaviors affects social judgements and the perception of social presence and copresence [223].

2.3.5 Summary

Systems that support user representations can be grouped into two main categories: i) telepresence systems, and ii) avatar-mediated systems. Commercial social VR systems are often avatar-mediated systems. Systems for virtual social interaction can transmit different levels of behavioral realism and reproduce different levels of realism in the user appearance. With regard to communication, the degree of user behavior reproduced by the simulation is of special importance, as nonverbal behavior is related to adaptation and synchronization processes in social interactions (nonverbal rapport). One could argue that a reduced realism and quality in the transmitted behavior disturbs the emergence of adaptation processes. SIPT [351, 352] argues that users can compensate to some degree for missing cues and behaviors, adapting their strategies of encoding social messages. Yet SIPT does not generally deny that differences in media may impact the communication [223]. Figure 2.12 describes the derived relations. Simulation characteristics may affect social interactions with regard to copresence, social presence and (virtual) rapport.

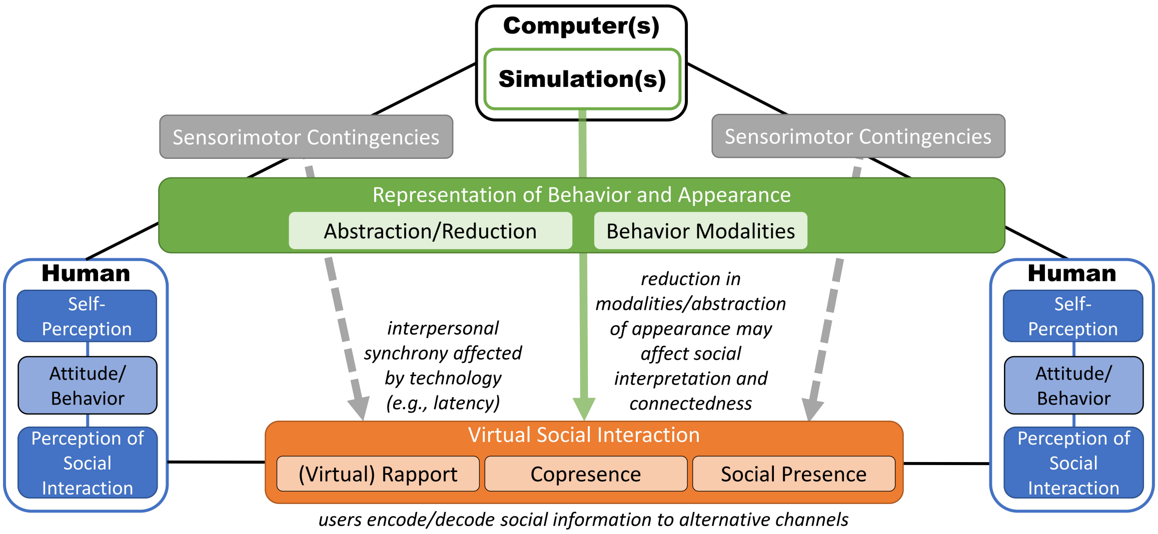

Figure 2.12: Potential impacts of simulation properties on virtual social interactions. Embodied virtual social interactions could be affected by the abstraction of the user representation, such as stylizations of avatars, as well as the abstraction of behaviors. The degree of behavior modalities transmitted is a further factor, that may reduce the ability to interpret and convey social signals and thus the possibilities to adapt nonverbally. However, according to SIPT, users may employ compensatory mechanisms to (partially) compensate for missing behavioral displays.

2.4 Augmenting Social Interactions

2.4.1 Computers as Social Actors

When describing the role and engagement of computers in mediated communications, it is important to consider the interaction and social rules between humans and computers, or, more general, humans and media. In 1994, Nass, Steuer, and

Tauber [216] defined computers as social actors (CASA). In a series of studies, they investigated social responses to entities they defined as human or computer. They investigated the impact of different terminologies and phrasing when users were exposed to a computer-supported learning scenario. The five studies found that 1) "social norms are applied to computers," 2) "notions of 'self' and 'other' are applied to computers," 3) "computer users respond socially to the computer itself," 4) "social responses are automatic and unconcious," and 5) "Human-computer interaction is socio-psychological" [p. 77]. Their findings suggest that human-computer relationships can be social, and that users apply social rules to their interactions with computers and other media. These and related findings were summarized by the media equation [253]. Their paradigm was replicated in numerous studies and Nass and Moon [215] concluded that mindless social responses [170] that are based on contextual cues are triggered when humans interact with computers. In other words, users socially respond to computers on the basis of learned social scripts for human-human interactions. These responses, similar to natural interactions, seem to be automatic, and thus without conscious effort. They further argued that anthropomorphism is *not* the central trigger for these responses in their experiments. They do discuss how perfect versions of technologies mimicking human characteristics may result in stronger social responses responses; however, they note as well that a lack of realism may, conversely, hinder those social responses [215, 253].

In the context of the present thesis one can consider two major entity types of social encounters in VEs: avatars and (embodied) agents. In contrast to human-driven avatars, (embodied) agents are driven by algorithms [13] that can, for example, support contextual and environmental awareness or the interpretation and expression of communicative signals [61, 60, 112]. To drive reactive agents, the social communication signals of human users must be interpreted through social signal processing [347, 340] and then mapped to agent reactions. In turn, agents can react to user behavior for example to achieve virtual rapport [121]. Recently, Daher et al. [70] found that priming observers to believe that an agent was intelligent increased social presence. Blascovich [44] concluded that "social influence can occur within digital virtual and immersive virtual environments, whether the 'others' present are computer agents or human avatars." [p. 127], which can be seen to extend CASA theorem to the context of SVEs and immersive interactions.

2.4.2 Response: Computers as Social Mediators

Technological factors identified by previous research point to the importance of the medium, the simulation, or in our case, simply the computer and its capabilities of sensing and replicating when considering computer-mediated communication (CMC). The mediating role of computers with regard to intrapersonal effects (self-perception) was described in Section 2.1.6 (Proteus Effect). In Section 2.3 we discussed that technological properties affect the perception of social interactions in VEs. This implies that a computer's social engagement is not limited to interaction with an agent entity [216] but is also apparent in mediating scenarios and/or replicating scenarios through user-embodied interfaces. Through virtual environments, they interfere with our social information processing, and thus can be described as social mediators with respect to both, self-perception as well as social interaction.

Hollan and Stornetta [138] described their view of the guiding principle of communication media development: “A belief in the efficacy of imitating face-to-face communication is an unquestioned presupposition of most current work on supporting communications in electronic media” [p. 119], and they then go further, stating that “looking at nonimitative approaches that focus on underlying requirements and the distinctive characteristics of the electronic media rather than on imitation of the mechanisms of face-to-face might lead to even better solutions” [p. 124].

Interpreting their statement, it can thus be concluded that computers could be used to actively control this mediation process, with regard to both, intrapersonal and interpersonal aspects. Therefore, this thesis argues for considering computers as social mediators. Considering computers as social mediators re-conceptualizes the role of computers in the context of user-embodying interfaces. With regard to an active mediating role by augmenting social behaviors, we argue that computer systems could be used to display artificial behavior derived from analyzing social phenomena and thus merging avatar and agent entity, see Figure 2.13.

Figure 2.13: Human-Computer Continuum. Continuum between human and computer entities and their embodied representations.

We argue that there are benefits in engaging actively in social mediation, by designing simulations that amplify, substitute, and transform social behaviors to alternative feedback, introducing hybrid forms of social interactions. Reviewing the SIPT and its rationale [351, 352], one could turn the argumentation: when users can creatively alternate their strategies to encode and decode social information to other channels, vice versa, this could be utilized to design communicative systems.

Studies of virtual environments have just begun to investigate such hybrid models of social interaction and some previous works exist that consider related rationals and hybrid representations.

2.4.3 Continuous Presence

A previous hybrid avatar/agent model has been introduced by Gerhard and colleagues [113, 114, 112, 115]. They base their argument on the vision of continuous presence [113], meaning the permanent user representation in an (educational) SVE. In the absence of the user (meaning: when being offline), they propose a continous embodied representation based on an intelligent agent. Such permanent embodiment could for example foster a community feeling and support educational SVEs [112]. Among other insights, Gerhard [112] concludes that when a user is offline, “The agent should be able to follow the conventions of natural human dialogues using verbal and non-verbal channels" [p. 170]. Consequently, when computers actively mediate virtual social interactions between human users, they may transform human behaviors to altered visual information by similar conventions.

2.4.4 Transformed Social Interaction

Bailenson and colleagues investigated the flexibility of communication in VEs and presented the concept of transformed social interaction (TSI) [12, 11, 17], which describes the decoupling of feedback information from actual physical behavior in CMC. These researchers argued that nonverbal behavior can be transformed strategically transformed to modify communication by, for example, directing the attention of a speaker to each of multiple listeners by individually rendering her/his gaze for each listener. In their studies, they found that this gaze augmentation resulted in significantly more agreement than with reduced-gaze or non-augmented gaze. However, participants in the augmented condition perceived less social presence [11]. Bailenson and colleagues further explored mapping behaviors to changes in the appearance and the form of a "emotibox" avatar, using facial feature tracking to deform a simplified box avatar, for example, changing its size and color [15]. The co-presence ratings and emotion transmission were lower in the avatar condition compared to video or voice only. Conversely, the emotibox condition led to more verbal and nonverbal disclosure [15].

In a similar manner, Boker and colleagues were the first exploring the modification of head movements and facial expressions by means of modifying active appearance models in real-time interactions [46]. In doing so, they found that these manipulations in turn impacted the behavior of the communication partner. More recently, Oh, Bailenson, Krämer, and Li [224] modified the facial expressions of avatars in desktop-based dyadic interactions and found that increasing an avatar's smile behavior increased the positivity of the evaluation of the conversation. Hart, Piumsomboon, Lawrence, Lee, Smith, and Billinghurst [133] recently presented the dynamic appearance modification of avatars as method of shaping and augmenting emotions in collaborative interactions, by modifying the appearance of an avatar based on emotional cues. In conclusion, multiple previous works explored the plasticity of social interactions by modifying and augmenting communicative aspects through active engagement of the medium. Yet, a general technological concept, especially considering the importance of bidirectional adaptation processes, is missing.

2.4.5 Hybrid Avatar–Agent Technologies

In order to address this gap, we introduce the technological concept of hybrid avatar–agent technology (HAAT) to utilize nonverbal phenomena of social interaction to augment virtual encounters [269]. Where previous approaches applied hybrid models to achieve continuous presence [112], to unidirectional modify social interaction (see Section 2.4.4) or for linear modifications of appearance [15], we propose a bidirectional approach that takes into account the previously mentioned processes of IPS and everyday social phenomena (see Section 2.3.4).

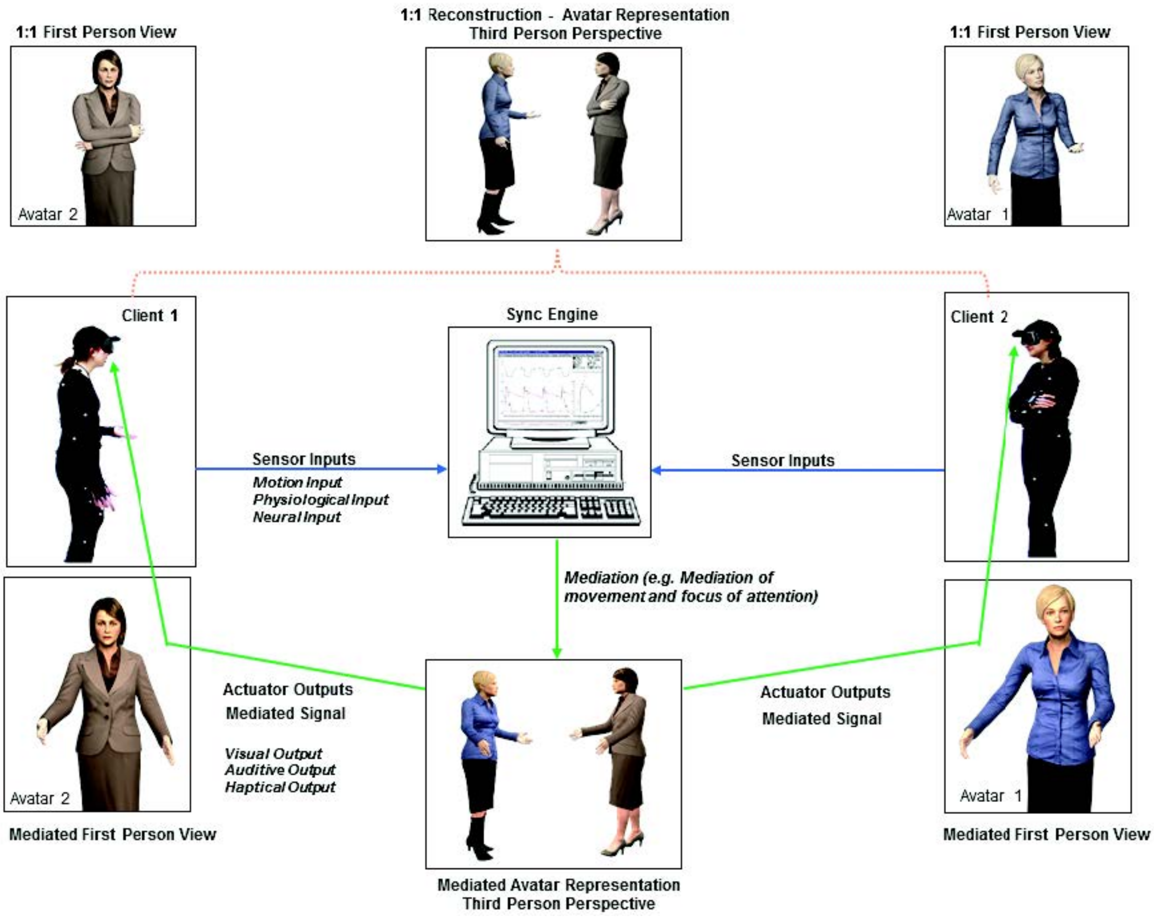

Figure 2.14: The initial hybrid avatar–agent technology concept: A functionality example. Users are embodied through avatars. Behavior is sensed and replicated, but in addition analyzed by a social AI. In the case of inadequate behavioral matching, rendered behavior is visually transformed by the social engine (sync engine) [269] ©2015, Walter de Gruyter and Company.

HAAT takes into account both human and technological aspects in social interactions to actively mediate virtual communication in the social dimension. In this respect, HAAT differs from TSI (see Section 2.4.4) as it considers the dyad or group, rather than the individual user, to be the layer of analysis and modification (see Figure 2.14). In that regard, HAAT is combining the capabilities of avatars that replicate human behavior with the potential of agents that display reactions to social phenomena and contextual rules.

Along with capturing and synthesizing nonverbal behavior, HAAT therefore targets the construction of an underlying social AI that interprets behaviors and modifies the respective virtual display according to what is adequate and fosters the interaction. To do so, the underlying model must identify phenomena of social interactions, analyze the current status, and modify the respective user behavior (and thus feedback) for the interlocutor. We propose this approach as an active engagement in the mediation process. We argue that HAAT technologies are useful for example in the context of intercultural communication, the inclusion of individuals with deficits in interpreting or expressing nonverbal behavior (e.g., social disorders), in the case of lacking sensory input, or transmission disturbances. For example, well known phenomena such as nonverbal mimicry [317] could be augmented to achieve higher levels of liking and affinity. Hybrid gaze models could be introduced to support people that have difficulties in expressing social gaze. Furthermore, the HAAT principle can be used to substitute or augment social phenomena in the case of missing sensory input, such as in large scale mobile applications that do not allow for multi modal behavioral sensing. The principles of intervention are further discussed in Chapter 5 along with exemplary prototypes and empirical studies.

In summary, we argue to consider computers as social mediators as they have been identified to affect self-perception, the perception of social interaction, and can further be used to actively augment social interactions. Previous works have proposed hybrid avatar-agent models to achieve continuous presence [113, 114, 112, 115] as well as to strategically transform social interaction by decoupling visual representation from physical behavior [12, 11, 17, 46, 224, 133]. While previous approaches rather consider intra-individual processes, we propose HAAT, which aims at considering the dyad or group as level of analyses.

2.5 Summary and Conclusion

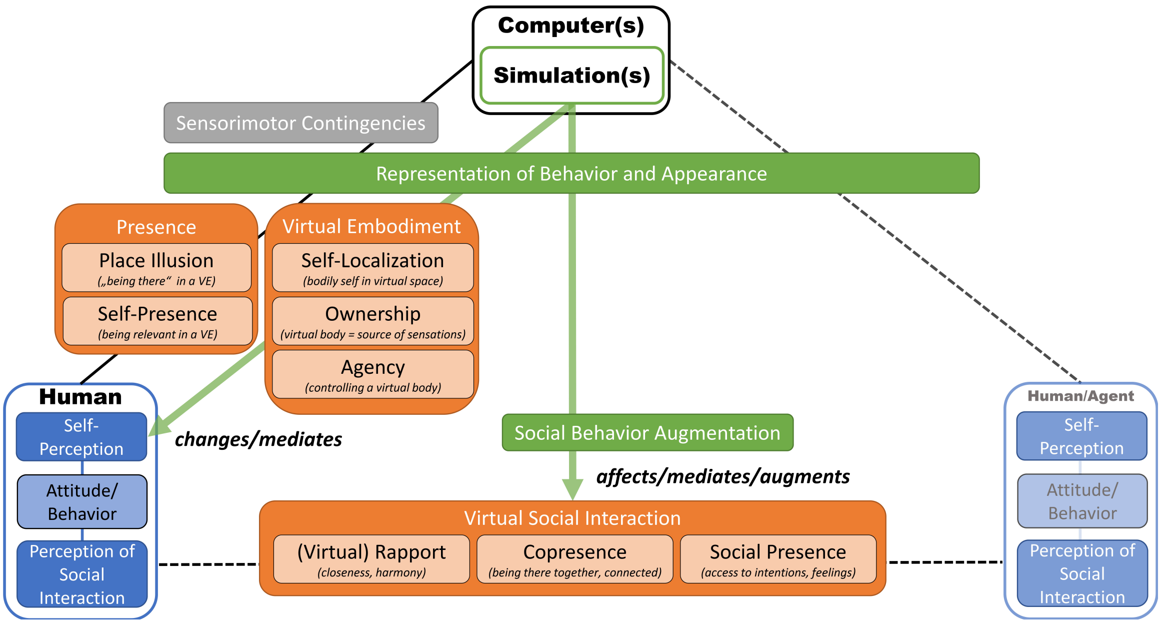

Figure 2.15: Model summary of the most important identified concepts. SMCs and vividness of the simulation fosters presence. Embodiment, ownership and agency are affected by bottom-up and top-down factors. Embodying an altered visual representation can lead to an adaptation of behaviors and attitudes. The simulation properties and replication of appearance and behavior can impact virtual social interactions. Social augmentations can actively engage in virtual social interaction.

We summarized basic context and related findings that represent the underlying principles of technological impacts on the self and on social interactions in VEs.

Figure 2.15 summarizes the findings. Intrapersonal interactions which describe the impact of the simulation technology on the self and self-perception are affected by top-down processes, which relate to expectations and previous knowledge, as well as bottom-up processes, which relate to sensorimotor stimulation. Previous literature describes two major concepts that are affected by simulation characteristics: Presence and embodiment. Both concepts share common ground. Similar to virtual embodiment, presence or PI describe the perception of being located in a virtual environment. Further, presence and agency rely on SMCs. Previous research identified that presence is fostered if the virtual environment is immersive and supports sensorial richness, a similar self-representation, and the consistency across displays. Self-representation in turn is the fundament for virtual embodiment. However, while previous works distinguished ownership and agency as concepts of (virtual) embodiment, a systematic investigation and relation to presence is missing. Further, throughout the literature, the assessment of virtual embodiment varies with regards to the measures applied in empirical studies. Both problems are therefore targeted in Chapter 3, specifically asking:

RQ1a: How do simulation properties affect virtual embodiment?

RQ1b: What latent variables are responsible for the adaptation of a virtual body?

Reviewing previous work and abstracting from the impacts on self-perception one could argue that characteristics that affect presence and embodiment may also affect interpersonal interactions. For example, facial expressions are known to be drivers of the perception and expression of emotion [86]. If those are not presented to communicators in VEs, one may assume that social impression formation and the interpretation of feelings are inhibited or disturbed. Yet, SIPT takes a countering position and argues that humans creatively adapt their communication on the basis of available channels and signals [351, 352]. Results investigating the impacts of technological factors vary [223] which is why we contribute to the further investigation of embodied social interaction in Chapter 4. We specifically asked two questions:

RQ2a: How do virtual social interactions with user embodiment compare to physical social interactions?

RQ2b: How do technological properties affect user-embodied virtual social interactions?

Effects such as the Proteus effect [367] already underline the mediating role of simulations with regard to the self-perception through embodied interfaces as well as a subsequent impact on attitudes and behavior. With regard to interpersonal effects, the majority of previous work has considered mediation in terms of a passive role that a medium plays in transmitting communicative information and how this passive mediating role affects the perception and execution of interactions. Motivated by theoretical proposals and earlier investigations [138, 216, 112, 12] we proposed computers as social mediators (CASM) and presented the technological concept of HAAT that aims at propagating an active role to an underlying social intelligence designed to intervene and augment social interactions. The investigation of prototypical interventions is the topic of Chapter 5, which addresses the following research questions:

> *RQ3a: What possible modifications can be designed to augment user-embodied social interactions?*
>
> *RQ3b: What impacts on the perception of interactions arise from augmented social interactions?*

The presented research themes describe aspects that potentially affect user-embodied interactions in VEs, summarized in Figure 2.15, which are: 1) the perception of embodiment, ownership, and agency, 2) presence, including social and copresence, 3) communicative behavior and how its execution and interpretation are affected by CMCs, and 4) the augmentation of social interactions.

Chapter 3

Intrapersonal Effects of User Embodiment

The present chapter focuses on the perception of virtual embodiment, and more specifically the VBO. The main purpose of the presented studies and derived knowledge was the validation and exploration of simulation aspects that impact virtual embodiment, with a specific focus on VBO and agency, along with the development of an instrument to assess the latent variables and the resulting impacts of virtual embodiment in the broader sense. Furthermore, we investigated relating constructs, such as presence, affect, in addition to pre-existing ownership and agency measures to gain insights about the relationships between the phenomena. The chapter is guided by the first two research questions:

RQ1a: How do simulation properties affect virtual embodiment?

RQ1b: What latent variables are responsible for the adaptation of a virtual body?

Section 3 gives a more detailed overview of previous findings with regard to body ownership in general and argues for the construction of a standardized measurement. Section 3.3 introduces an empirical study that led to a first instrument, the Alpha-IVBO scale, that reveals latent variables of virtual embodiment. Section 3.4 and 3.5 empirically explores novel aspects that may affect those latent variables, namely avatar personalization, immersion, and social interaction. Section 3.6 describes the generalization of the constructed measure which we termed the Beta-IVBO, along with its validation and test of reliability that led to the final Gold-IVBO.The results of four empirical studies on the impacts of simulation characteristics are reported, before the findings are summarized. As we specifically focused on body ownership and a respective measure, the following will present closely related work that considered the investigation of media and scenario properties that affect such illusions.

3.1 Triggers and Preventers of Body Ownership Illusions

Virutal simulations allow for a flexibility in context, sensory stimulation and behavior transmission. Previous findings that identified aspects that foster embodiment and ownership illusions (triggers), and aspects that hinder or prevent these illusions (preventers) [158].

Conditions that affect body ownership illusion (BOI)

- Crossmodal Stimulation
  - Visuomotor
    - *Congruent stimulation* → can elict BOI (e.g. [283])
    - *Delayed stimulation*→ can modulate or hinder BOI (e.g. [151])
    - *Manipulated stimulation* → inhibits BOI (e.g. [283])
    - *Anatomically incongruent stimulation* → can hinder BOI (e.g. [255])
  - Visuotactile
    - *Congruent stimulation* → can elict BOI (e.g. [235])
    - *Mismatched stimulation* → inhibits BOI (e.g. [302])
    - *Delayed/temporally decoupled* → can negatively affect/hinder BOI (e.g. [292])
  - Visuoproprioceptive
    - *Modified distance* → may affect BOI - not conclusive (see [158])
    - *Modified angle* → may not strictly affect BOI but not conclusive ([158])
    - *Slightly shifted perspective* → may still allow for BOI [198]
- Semantic Modulation (Shape/Structure)
  - Human
    - *Texture realism* → greater realism can positively modulate BOI (e.g. [199])
    - *Anatomical plausibility*
      - Implausible spatial configuration (rotated hand) → hinders BOI (e.g. [151])
      - Implausible perspective (3PP) → hinders BOI (e.g. [304])
      - Implausible structure (different body scale) → modulates BOI → ([158])
    - *Stimulation congruence* → no distinct findings
  - *Non-human* → weakens BOI (e.g. [333, 140])

Figure 3.1: Taxonomy of previous findings for factors that affect (virtual) embodiment (see [158] for an extended explanation and summary). Crossmodal congruent stimulations can elicit BOI, whereas (body) shape or texture moderate the effect. Regarding an implausible perspective, one could account a plausible perspective (i.e., 1PP) to a congruent visuomotor stimulation, rather than a semantic modulation.

Kilteni, Maselli, Kording, and Slater [158] present a constructive review of methods to induce and modify body ownership (and more general: embodiment) along with the related historical and theoretical grounding. They identify three crossmodal stimuli criteria, *visuotactile triggers*, *visuomotor triggers*, and *visuoproprioceptive modulations and triggers* (contributing to locating the body and body parts to a given time in space) that were found to impact what is in general referred to as BOIs, may they be evoked through the physical or the virtual world. All of which could be present in a VE or immersive VR simulation. For example, texture realism and spatial configuration are fundamental characteristics of a virtual simulation. A congruent visuomotor stimulation is key to elicit place illusion [299]. It is important to note that many previous findings, such as the RHI [47], base on the explicit crossmodal coupling of the triggering stimuli and therefore a single cue or sensation may not suffice to induce a BOI. Furthermore, they identify semantic constraints which are of importance for the present work: *shape*, *texture realism* (i.e., biological plausibility), *anatomical plausibility* of spatial configuration, and *stimulation congruence*, each of which may tolerate or not tolerate violations [158], see Figure 3.1. Reviewing the taxonomy presented in the Figure, one can see that the identified processes, to most degree, align with what was discussed in Section 2.1, namely bottom-up processes (here: congruent crossmodal stimulation and integration) facilitate embodiment to large degree, whereas top down processes (here: semantic modulation) moderate the effects.

While the findings presented in Figure 3.1 partly result from physical world experiments and we could argue about how they affect other dimensions of embodiment, such as self-location and agency, Maselli and Slater [199] reported empirical findings of VR experiments that match the above taxonomy with similar findings for the IVBO. Visuotactile stimulation did likely receive the most attention with regard to the induction of a VBO, which may be historically due to the original RHI experiment. Yet, aside from force feedback from hand controllers, it is the least present in current consumer VR simulations. Visuomotor stimulation in contrast is present in many VR applications, and, rationally, in many applications that support visual user embodiment. Furthermore, realism and abstraction of virtual bodies is a necessary choice for the design of every simulation supporting user embodiment, which is why it is of specific interest and will be adressed particularly in this thesis. Once designed, it is useful and of interest evaluate embodiment effects achieved through the simulation. Yet, a standardized instrument to do so is missing.

3.2 The Demand for a Standardized Measure

While many previous studies have adapted measures from originating experiments such as the RHI, for example displacement measures, the assessment of VBO is not consistent across previous work. A standardized and validated scale does not exist. Effects are often assessed with single items, which was argued to be problematic when analyzed individually [57]. Despite various approaches to cross-validate different measures, the validation often failed (see [272] for a discussion, and for a summary of applied measures, see [157] for a summary of applied measures). More recently, Gonzalez-Franco and Peck [118] underlined the request for a standardized assessment to measure VBO in avatar embodiment scenarios along with an extensive review.

The following sections therefore aim at investigating the effects of typical *stimulation properties* and semantic modulations of semi-immersive and immersive avatar embodiment scenarios, along with the *construction of a standardized measurement instrument*.

3.3 Study 1: Impacts of Avatar Appearance and Scale Construction

The first study collected initial data to investigate the impact of appearance on the perception of virtual embodiment [173]. We were specifically interested in i) how anthropomorphic characters compare to humanoid characters, and, ii) how a certain degree of personalization or individualization affects body ownership. Therefore, we incorporated additional aspects of self-representation to investigate their effects on body ownership and to assess latent variables. Two factors that have been widely neglected in previous research have been specifically addressed: 1) the employment of a *higher level of behavioral realism* through the replication of facial expressions to avatars, and 2) a variation of *personalization* and *stylization* (see Sections 2.1.3 and 3, respectively). Further, we addressed the fact of a lack of available measures by collecting previously used items to assess components of embodiment, more specifically, those that are related to ownership illusions and agency, along with further questions.

3.3.1 Method

Design

This experiment was conducted in a one-factor (avatar type) within-subject design. The exposure consisted of four different conditions that were presented in random order. In each condition, the participants were exposed to a fake mirror system (see Figure 3.2) and embodied as one of the four avatar types depicted in Figure 3.3 and instructed to perform certain movements and expressions, followed by a questionnaire assessment.

Apparatus

The FakeMi system (see Figure 3.2 and [191, 173] for further details) replicates human behavior and facial appearance to avatar behavior and appearance. Unity3D[7] was used to implement the virtual environment and render the simulation to a 55″ (1920px × 1080px) LG passive stereoscopic screen (55UB850V) mounted in portrait mode.

[7] Unity Technologies, `https://unity3d.com/`

Figure 3.2: The principle of the fake mirror system (adapted from [173], Copyright ©2016 ACM). The user's body movement and facial expression are tracked by sensing technology (Microsoft Kinect V2, PrimeSense Carmine V 1.09a). The simulation retargets these behaviors to avatars. The perspective is rendered according to a fish tank VR principle [353]. A stereoscopic screen displays passive steroscopic images of the simulation.

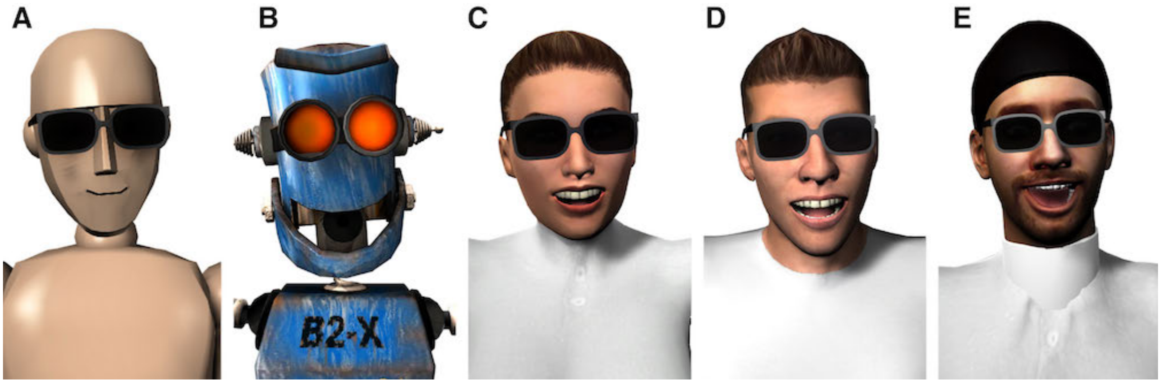

Figure 3.3: The avatars used in the study: A) The wooden mannequin, B) The robot, C) and D) a male and female generic humans, and E) an example of a personalized avatar based on the participant's facial scans were used. To not bias the level of similarity, the avatars were rendered with black glasses during the study to adapt to the passive stereoscopy glasses worn by the participant during the study. Adapted from [173], Copyright ©2016 ACM.

To replicate the participants' body movements, a Kinect V2[8] time-of-flight sensor tracked their movements. We used Brekel Pro Body[9] to acquire the skeletal information used to drive the avatar's body motion behavior in the simulation.

Facial deformations were tracked by a Carmine 1.09 short range RGB-depth camera[10] utilizing structured light coding to acquire a depth image and then processed by Faceshift Studio[11]. The acquired facial expression data that was used to drive the avatar's facial animations by generating a precomputed user template and respective deformation data in blendshape form [358, 357]. We could approximate the overall motion-to-photon latency by frame counting [134] between 150 ms - 200 ms. Figure 3.4 depicts example expressions mapped onto humanoid avatars.

With respect to the tracking frustums of both sensors, the approximate robust tracking volume estimated to 1.3 m × 1.3 m × 2 m for the body tracking and 0.6 m

[8] Microsoft Corporation, https://developer.microsoft.com/de-de/windows/kinect
[9] Brekel 3D, https://brekel.com/brekel-pro-body-v2/
[10] PrimeSense, Israel, acquired by Apple Inc. in 2013, https://www.apple.com/
[11] Faceshift AG, Switzerland, acquired by Apple Inc. in 2015, https://www.apple.com/

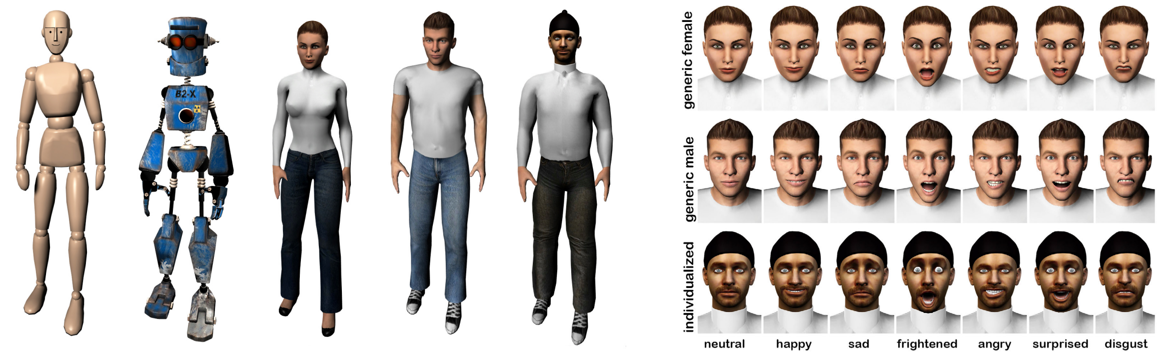

Figure 3.4: Full body views of the avatars used in Study 1 and exemplary facial expressions. *Left:* full body views of avatar examples (excluding glasses). All humanoid avatars had similar clothing (reprinted from [272], Copyright ©2017 Roth et al./David Zilch). *Right:* Exemplary expressions that were requested by participants during the procedure and, in turn, tracked by facial tracking to be replicated to the virtual environment. Note that during the experiment, the avatars were displayed with rendered stereo glasses for the sake of consistency, see Figure 3.3 top. Reprinted with permission of David Zilch.

$\times$ 0.6 $m \times$ 0.5 m for the facial tracking. The virtual characters were generated and modified using Poser[12] and Maya[13]. The wooden mannequin was derived from a related study[14] [270] and modified to be able to display a subset of facial expressions. The robot was acquired through an online shop and modified to be able to display of a subset of facial expression. Figure 3.3 depicts the avatar types used in the study.

In summary, the system allows for non-invasive behavioral tracking, avoiding a potential *Cyborg's dilemma* [39], meaning a third variable bias through a change of body scheme caused by invasive and wearable VR technology. This allows for the integration of different avatar types, including custom or respectively personalized avatars that incorporate the user's facial behavior and body animation. Furthermore, the apparatus let the users keep a physical frame of reference.

Procedure

We welcomed participants and informed them about the experiment. After they provided their written consent for participation, we prepared the personalized avatar. We created a model of the participant's head and took two full body pictures in order to combine the personalized head with an appropriate body that matched the participants measures.

A facial expression template was generated from the 21 facial expressions the participants performed for the calibration procedure. We provided long-sleeve shirts for the participants, to match the generic and personalized avatars' clothing. Participants were then equipped with polarization glasses, and calibrated for the body motion tracking. During the calibration processes, the fake mirror was covered. To calibrate appropriate scaling of the avatar, the mirror was uncovered, and the

[12] Smith Micro Software, Inc., https://my.smithmicro.com/poser-3d-animation-software.html
[13] Autodesk Inc., https://www.autodesk.com/products/maya/overview
[14] The avatar was initially created at the University of Cologne, Department for Media Psychology

avatar's scale was adjusted together with the participant using a temporary avatar template. The participant then stepped in front of the mirror and performed a pose calibration before the first trial was executed.

During each trial, prerecorded oral instructions were played back using an audio speaker asking the participant to 1) look at specific body parts (e.g., the right foot, left upper arm, belly), 2) make certain facial expressions that correspond with the basic emotions (anger, disgust, fear, joy, sadness, and surprise [85]), and 3) move distinct body parts (e.g., turn the head up, raise the right arm). The participants were given three seconds to execute each instruction, and a sound then reminded the participant to focus back on the screen followed by a pause of three seconds before the next instruction followed. This process ensured continuous interaction and execution of behaviors, as well as the focus on the virtual counterpart. We randomized the sequence of action instructions (i.e., type of behavior to be executed), as such, and each complete trial lasted for one minute. After each trial, we asked the participant to fill out a post-experimental questionnaire in digital form before continuing to the next trial for the next condition/avatar representation type. Following the final trial, the participants were exposed to the appearance of a virtual spider. The results of the latter are not subject to the present reporting.

Measures

We performed a literature review to extract previously used measures and extend these with additional missing items. After reviewing related work [47, 220, 243, 22, 184, 21, 23, 119, 136, 304, 151, 157, 199, 9, 189], 15 questions were derived and designed for a consecutive factor identification to explore latent variables (see Table 3.1).

In the evaluation, the items were assessed using a 7-point Likert type response format (strongly disagree - neither agree or disagree - strongly agree), in a random order with the following instructions: "Please answer the following questions according to your gut feeling, spontaneously and intuitively." The questions were collected to capture the influential variables previously found, such as the sensorimotor coherence (i.e., agency), the realism of appearance and personalization, and potential aftereffects or changes in the perception of the body scheme. In addition to the VBO questionnaire, avatar preference was determined and qualitative questions were assessed.

Participants

The final sample consisted of $N = 43$ participants (26 female, 17 male, $M_{age} = 23.49$, $SD_{age} = 7.64$), of which 39 were students at the time. Sixteen had previous experience with VR or tracking technologies. All reported speaking German for five years or longer, which ensured accurate understanding of the audio instructions and questions. The study was conducted at the University of Würzburg.

Table 3.1: List of Questions Assessed and Considered for the IVBO Scale.

1. myBody I felt as if the body I saw in the virtual mirror might be my body. *Adapted from* [304, 119]
2. myBodyParts I felt as if the body parts I looked upon were my body parts. *Adapted from* [151]
3. humanness The virtual body I saw was human-like. *Adapted from* [243]
4. myMovements The movements I saw in the virtual mirror seemed to be my own movements.
5. controlEnjoyment I enjoyed controlling the virtual body I saw in the virtual mirror.
6. controlMovements I felt as if I was controlling the movement I saw in the virtual mirror. *Adapted from* [151]
7. causeMovements I felt as if I was causing the movement I saw in the virtual mirror. *Adapted from* [151]
8. ownOtherbody The illusion of owning a different body than my real one was very strong during the experience.
9. myBodyChange At a time during the experiment I felt as if my real body changed in its shape, and/or texture.
10. myBodyCheck During or after the task, I felt the need to check that my body still looked like what I had in mind.
11. echoHeavyLight I felt an after-effect as if my body had become lighter/heavier.
12. echoTallSmall I felt an after-effect as if my body had become taller/smaller.
13. echoLargeThin I felt an after-effect as if my body had become larger/thinner.
14. *otherPerson* The body I saw in the mirror was another person. *Adapted from* [119]
15. *enjoyment* How did you like the overall experience in the virtual world?

3.3.2 Results

Prior to the main analyses, we excluded two items from the set of questions. The *enjoyment* item was excluded as we observed that it captured the general interest and enjoyment of the application rather than a construct related to the VBO.

Furthermore, we discarded the item *otherPerson* ("The body I saw in the mirror was another person") for the component analysis. The first reason for this was instability of the item during preliminary analysis. Second, and more important: the phrasing of the item was potentially misleading, as the construct of "another person" is not necessarily in relation to "another body," and could be denoted to having another self, and therefore, the interpretation may have been confusing for participants.

Principle Component Analysis

In a similar fashion to the work from Longo, Schüür, Kammers, Tsakiris, and Haggard [184], we used a principle component analysis (PCA) to identify latent variables, following the procedure described by Field [90]. A Varimax orthogonal rotation was applied, and an individual PCA was conducted for each score-set (i.e., each condition). The criterion from Kaiser [150] was applied due to the small sample size, and components with eigenvalues < 1 were discarded. Kaiser-Mayer-Olkin measurements of sample adequacy for the four calculations were $> .641$, and all Bartlett's tests for sphericity were significant ($p < .05$). The results of the exploratory PCA are depicted in Table 3.2. The analysis resulted in three latent components that could be consistently detected in three of the four conditions. In the fourth analysis of the personalized avatar condition, an additional fourth component remained but its items also loaded on the other components and the fourth component by itself did not explain a large amount of variance and was therefore discarded. The *humanness* item showed a cross loading in one condition (one PCA, respectively) and the *ownOtherBody* item showed crossloadings in two conditions.

The first derived component related to questions that were stated with regards to the acceptance of the body as the own, the feeling of ownership, and the humanness of the appearance, which is why we describe this component as *acceptance*, and it strongly relates to questions that target ownership perception. The second component gathered items that relate to agency, meaning the feeling of control over movements and their causation, which is why we describe this component as *control*. The third component we could identify from the analysis is related to changes in the perception of the body from what was previously memorized, such as a change in feeling taller or smaller, so we termed this component *change*.

We pursued analyses of the respective reliabilities for each component and each trial, which are depicted in Table 3.3. Assuming these are the latent components of the item collection, their reliabilities are good. The proposed final scale is provided in Table 3.4

Experimental Results

After establishing the 3-component scale, we performed analyses of variance for each resulting factor. We found a significant main effect for *acceptance*; $F(3,126) = 18.283, p < .001, \eta_p^2 = .303$. Pairwise comparisons showed that *acceptance* was rated significantly higher for human-looking avatars (generic, personalized) than for the rather cartonish/artifical avatars ($p < .011$), whereas there were no significant differences between both human-looking avatars or between the cartonish/artificial avatars. The analysis for the *change* component revealed a significant main effect of smaller size $F(3,126) = 2.76$, $p = .045$, $\eta_p^2 = .062$. Individual comparisons showed that despite a slight linear trend, only the difference between the individual avatar ($M = 2.35, SE = .17$) and the wooden mannequin ($M = 1.97, SE = .17$) was significant ($p = .022$). There were no significant effects for the *control* factor. This seems rational, as all conditions had similar properties with regard to the control of the avatars. Figure 3.5 depicts the results.

Table 3.2: PCA Results for the Conditions.

	Woodie			Robot		
Item	CH	CO	AC	CH	CO	AC
myBody			.908			.869
ownOtherBody	.497			.522		.477
myBodyParts			.875			.790
myMovements		.850			.864	
controlEnjoyment		.841			.776	
controlMovements		.918			.912	
causeMovements		.873			.866	
myBodyChange	.864			.829		
myBodyCheck	.565			.741		
echoHeavyLight	.928			.941		
humanness			.725			.707
echoTallSmall	.837			.771		
echoLargeThin	.839			.843		
Var. Init. (%)	35.6	26.5	11.7	36.4	23.8	11.5
Var. Rot.(%)	29.1	24.7	20.3	29.1	24.0	18.5
Eigenvalue Init.	4.67	3.45	1.52	4.73	3.09	1.49
Eigenvalue Rot.	3.79	3.21	2.64	3.79	3.12	2.40

	Human			Personalized			
Item	CH	CO	AC	CH	CO	AC	misc
myBody			.848			.922	
ownOtherBody	.628		.486	.427			-.641
myBodyParts			.895			.905	
myMovements		.737			.820		
controlEnjoyment		.819			.819		
controlMovements		.858			.879		
causeMovements		.851			.868		
myBodyChange	.699			.854			
myBodyCheck	.683			.422			.775
echoHeavyLight	.796			.865			
humanness		.612	.428			.739	
echoTallSmall	.867			.870			
echoLargeThin	.923			.911			
Var. Init. (%)	35.3	25.8	10.2	34.3	26.5	10.6	8.0
Var. Rot.(%)	28.8	24.8	17.7	26.8	24.7	19.3	8.7
Eigenvalue Init.	4.59	3.35	1.33	4.47	3.44	1.38	1.04
Eigenvalue Rot.	3.74	3.22	2.30	3.48	3.21	2.51	1.13

Notes. Loadings < .4 are not displayed. The bottom rows indicate the amount of variance explained before and after rotation, the initial eigenvalue, and the eigenvalue after the performed rotation.

Table 3.3: Component Reliability.

	Woodie	Robot	Human	Personalized
Acceptance	.857	.773	.787	.886
Control	.905	.899	.839	.887
Change	.859	.853	.872	.817

Note. Reliabilities of the components, as measured by Cronbach's α, tested against the scores of the individual trials in the experiment.

Table 3.4: The Initial Scale Proposal (Alpha-IVBO).

Component	Question
Acceptance	**1. myBody** I felt as if the body I saw in the virtual mirror might be my body.
	2. myBodyParts I felt as if the body parts I looked upon where my body parts.
	3. humanness The virtual body I saw was human-like.
Control	**4. myMovements** The movements I saw in the virtual mirror seemed to be my own movements.
	5. controlEnjoyment I enjoyed controlling the virtual body I saw in the virtual mirror.
	6. controlMovements I felt as if I was controlling the movement I saw in the virtual mirror.
	7. causeMovements I felt as if I was causing the movement I saw in the virtual mirror.
Change	**8. ownOtherbody** The illusion of owning a different body than my real one was very strong during the experience.
	9. myBodyChange At a time during the experiment I felt as if my real body changed in its shape, and/or texture.
	10. myBodyCheck During or after the task, I felt the need to check that my body really still looked like what I had in mind.
	11. echoHeavyLight I felt an after-effect as if my body had become lighter/heavier.
	12. echoTallSmall I felt an after-effect as if my body had become taller/smaller.
	13. echoLargeThin I felt an after-effect as if my body had become larger/thinner.

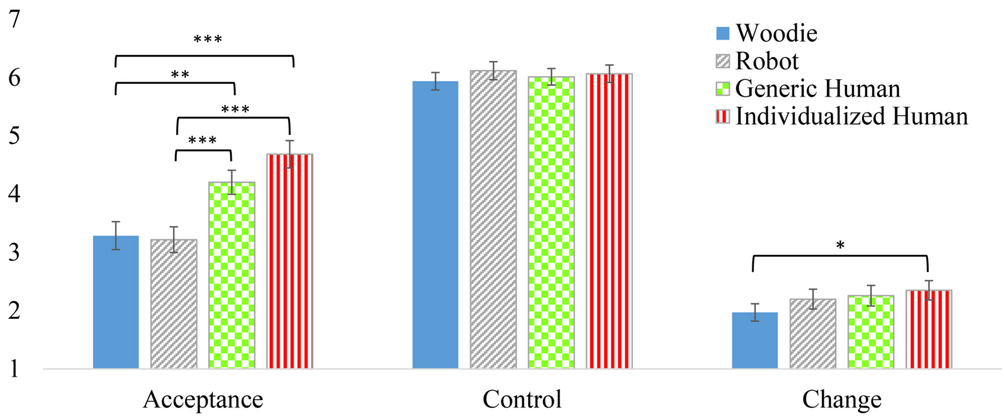

Figure 3.5: Descriptive results for acceptance, control, and change. *Note.* *** $p < .001$; ** $p < .01$; * $p < .05$; *bars denote the mean value; error bars denote the standard error.*

3.3.3 Discussion

In the first study, we explored the underlying components of VBO on the basis of a fake mirror metaphor. Through PCA, we defined three components. *Acceptance*, which was mostly related to top-down processes [199], the attribution of the avatar as a representation of the self. The items ask for the level of acceptance of the virtual body as one's own representation and thus related to previous research describing the sensation of body ownership [151, 158, 199, 198]. The second component identified was the *control* component. Items ask, for example, for the level of control and whether movements were caused by the participant. The component is therefore strongly related to agency, preliminary researched to be in relation but separable from ownership [335, 336, 151]. The third component identified was the component we termed *change*. This third component is associated with the feeling of transformation from what was previously known about the physical body, and might imply indications that the self-attribution was transformed [304] and, consequently, may indicate preceding effects related to the *Proteus Effect* [368].

Applying these measures to our study showed a higher acceptance of the human-like character appearances compared to more artistic/artificial body forms of the robot and the wooden mannequin. The personalized avatar showed higher ratings for acceptance, but not to a significant difference compared to the generic avatar. Interestingly, the change component showed highest values for the personalized avatar. This could for example imply that through a higher acceptance of the virtual body, the perceived body scheme is impacted. In other words, a higher sensation of ownership may moderate the perception of a change in body scheme.

Limitations

Some limitations have to be considered for the experiment. First, the fake mirror metaphor precludes generalization to more immersive scenarios, such as HMD-driven simulations, and we did not include specific questioning with regards to the perspective of the user. Second, the personalization was still partly generic, as the personalized avatars were created on the basis of an appearance template for the body. Special attention was paid only to the user's head (excluding the hair, which was covered by a virtual toque) for creation of the avatar. This implies further research, which will be described in the following section. Finally, the acceptance component inherited an assessment item that asked for human-likeness which may have had an impact on the acceptance measure as such. However, as described in Section 3, human-like appearance and body structure foster ownership [199, 158]. Two items resulted in cross-loadings and even evolved in an artificial fourth component, which shows that further validation is of need. In addition, we considered a (fake) virtual mirror paradigm, which implies the need for a further generalization to less restricted scenarios and additional validation, which will be described in Section 3.6.

Conclusion

In Study 1 we constructed a first measurement instrument to assess for components of virtual embodiment, which were found to be *acceptance* (ownership), *control* (agency) and *change* by a principle component analysis, based on the responses to the constructed scale. The results of the study indicate higher acceptance for characters with human appearance. Yet, we could not find a significant difference for a generic vs. personalized human avatar. Partly this might be due to the limitations of the study. Thus, another approach could provide more insights, namely the accurate photogrammetric capturing of users to provide personalized avatars. Further, more immersive setups could impact these results. These aspects were tackled in Study 2.

3.4 Study 2: Photogrammetric Personalization and Immersion

In order to investigate the impact of personalization and the level of immersion in more detail, this section presents the second study that was conducted in collaboration with the University of Bielefeld [348]. The study used a photogrammetry approach [1] to further improve the level of similarity, and thus personalization or individualization, and assess its impacts on VBO. A semi-immersive system was compared to an immersive system and multiple aspects that potentially affect VBO, such as the participants clothing, were examined. We further assessed related constructs (presence, similarity, emotional response) to gain insights about their relations to components of virtual embodiment. The scale proposed in the previous section (Section 3.3) was applied. The goal of the study was to further evaluate how personalization with a high fidelity reconstruction system affects the VBO and to further substantiate previously suggested effects of immersion (see [348] for further details).

3.4.1 Method

Design

The experiment followed a three-factor (medium, personalization, clothing) mixed design. The factor of *medium* was measured within subjects and consisted of two levels, a semi-immersive simulation that was displayed to the participant on an *L-shape* projection, as well as more immersive simulation that was displayed through an HMD (see Figure 3.6). The factor *personalization* was also measured within subjects and distinguished three levels of personalization types: a generic avatar created by a character generator, a generic avatar created by a photogrammetry approach [1], and a personalized avatar created using the same photogrammetry approach individually for each participant (see Figure 3.7). As a third between-subjects factor, we assessed the *clothing* types, which were split into groups and were used to compare whether personalization effects were stable for avatars that were created from persons wearing motion-capture clothing or casual dress. The factor was introduced in order to investigate the bias that can result from additional technological constraints the users were facing.

Apparatus

Two apparatuses were constructed for the experiment at the University of Bielefeld, the photogrammetric avatar generator and the virtual simulation system, depicted in Figure 3.6.

The photogrammetry procedure utilized two camera rigs to create the participants' personalized avatars. The body was photographed using 40 synchronously triggered digital single lens reflex (DSLR) cameras. The user's face was photographed by eight synchronously triggerd DSLR cameras. Based on the camera images, two dense textured point clouds were created through multi-view stereo reconstruction using Agisoft PhotoScan[15]. In order to create an accurate geometry and the respective rigged avatar mesh, a human template model[16] was fit to the two point clouds using registration, inverse kinematics and a statistical body model. Using the geometric mesh, a 4000px × 4000px texture and UV texture layout was computed using the geometry and image data. Finally, the high-resolution facial mesh and body mesh were merged using Poisson-based blending [234] (see [1, 348] for further details). The simulation and respective avatar embodiment consisted of an optical infrared passive-marker tracking system (OptiTrack Prime13W, 10 cameras, 120 Hz[17]), an L-shaped projection that covered a wall ($3\ m \times 2.3\ m$) and the floor ($3\ m \times 2.3\ m$), were used. The projection was driven by two stereoscopic projectors (2100px × 1200px, 60 Hz) using the INFITEC[18] spectral separation and filtering technique. In the immersive conditions, an HTC Vive[19] HMD (1080px × 1200px, 90 Hz) was

[15] Agisoft `https://www.agisoft.com/`
[16] Autodesk Character Generator, Autodesk, `https://charactergenerator.autodesk.com/`
[17] NaturalPoint, Inc. DBA OptiTrack, `https://optitrack.com/products/prime-13w/`
[18] Infitec GmbH, `http://infitec.net/?lang=de`
[19] HTC Corporation, `https://www.vive.com/`

Figure 3.6: Scanning rig and apparatus for the simulation. *Left.* Separate photogrammetry scans from the face and the body are combined for the creation of the avatar. *Right.* The user is tracked through body motion tracking (OptiTrack Prime Sense 13W) technology. In the L-shape condition, a stereoscopic L-shaped screen (front, floor) displays passive steroscopic images. In the HMD condition, the HMD displays the simulation to the user. The two top figures are reprinted from [174] © 2017 Latoschik, Achenbach, Botsch et al.; the two bottom figures are reprinted from [348] © 2018 IEEE.

used to display the simulation. To acquire skeletal transformations, the tracking system was interfaced with the rendering machines via network. The simulation was developed using C++ and OpenGL 4. Motion-to-photon latency was approximated to 62ms (projection) and 67ms (HMD approximated with desktop measurement) via frame counting. The simulation was rendered with high-performance machines (see [348] for further details).

Measures

After each trial, the above (Section 3.3.1) scale was employed using a 7-point Likert-style response format (0 – strongly disagree, 3 – neither agree nor disagree, 6 – strongly agree). In the concept of the present work we were especially interested

Figure 3.7: Avatar types used in the study. *Top.* Motion capture clothing. *Bottom.* Casual clothing. *Left to right.* CG generic avatars, photogrammetry scanned generic avatars, demonstration example of scanned personalized avatars of the participants. Reprinted from [348] © 2018 IEEE.

Table 3.5: Reliabilities Assessed by Cronbach's α.

Condition	Acceptance	Control	Change
HMD CG Generic	.701	.679	.891
HMD P Generic	.713	.877	.910
HMD P Personalized	.839	.801	.802
L-Shape CG Gen.	.775	.764	.909
L-Shape P Gen.	.781	.821	.934
L-Shape P Pers.	.767	.783	.837

Note. CG Character Generator; P Photogrammetry

in how the scale proposed in Section 3.3 would perform in alternative scenarios, meaning more immersive setups. To determine this, we analyzed Cronbach's Alpha for each individual subscale and each trial separately. The reliabilities are depicted in Table 3.5. Except from the *control* component assessment in the HMD character creator avatar condition, all reliabilities performed above .7.

In addition to the Alpha-IVBO components, three oral "in-situation" questions adapted from [151, 48] were used to identify the level of ownership during the exposure: 1) Ownership: "To what extent do you have the feeling that the virtual body is your body?" 2) Agency: "To what extent do you have the feeling that the virtual body moves just like you want it to, as if it is obeying your will?" 3) Subjective presence: "To what extent do you feel present in the virtual environment right now?" (0 - not at all to 10 - totally). Before the experiment, participants were informed about the concept of presence: "Presence is defined as the subjective impression of really being there in the virtual environment." In addition, we included two control measures assessing the perceived similarity of the virtual character ("To what extent did you have the feeling that the virtual body was similar to your own?") as well as the correct perception of the coherence of clothing ("To what extent did you have the feeling of wearing different clothing from the clothes you were actually wearing?" (0 - not at all, 10 - totally) — these were adapted from [304]). Furthermore we measured simulator sickness with the simulator sickness questionnaire [156], valence, arousal, and dominance through a self-assessment measurement [51] as well as the positive and negative affect. The results of the latter measures are not part of the present discussion (see [348] for further details).

Procedure

We welcomed participants and informed them about the study. After they agreed to participate and provided written consent, we randomly assigned them to one of the clothing conditions and scanned them with either casual or motion capture clothes. After the scan was completed, the avatar was created and participants completed the pre-study questionnaire. Participants were then calibrated for the first trial and instructed to perform body movements, similar to the procedure described in Section 3.3.1 (e.g,"Lift your right arm and wave to your mirror image in a relaxed way", or "Now stretch out both arms to the front and perform circular movements", etc.). In this form, six audio instructions were played back via loudspeaker. After each instruction, the participants were instructed to "Look at the movement in the mirror - at your own body - in the mirror - at your own body" in order to assure a controlled exposure to the stimulus. When the participants completed all instructions but were still immersed, the "in-situation" questions were assessed. Participants were released from the display equipment and filled out the post-experimental questionnaire in digital form. In the same manner, all participants performed three trials (generic avatar created by a character generator, a generic avatar created by photogrammetry, and a personalized avatar created by photogrammetry) in each of the *medium* conditions, leading to six total trials. Depending on the *clothing* condition, they saw avatars in either casual or motion-capture clothing. After finishing all trials, we compensated the participants and wished them farewell.

Participants

The final sample consisted of $N = 29$ participants (15 female, 14 male, $M_{age} = 23.62$, $SD_{age} = 3.53$), of which 27 were students at the time. None of the participants reported severe sensory or motor disorders. Any visual impairment was corrected for during the experiment. All reported having spoken the German language for 10 years or longer, which ensured their correct understanding of the audio instructions and questions. Out of the final sample, 13 were assigned to the motion capture clothing scan condition, and 16 participants were assigned to the casual clothing scan condition. The study was conducted at the University of Bielefeld.

3.4.2 Results

A factorial mixed analysis of variance (ANOVA) was calculated for each dependent variable. Greenhouse–Geisser corrected values are reported, when the sphericity

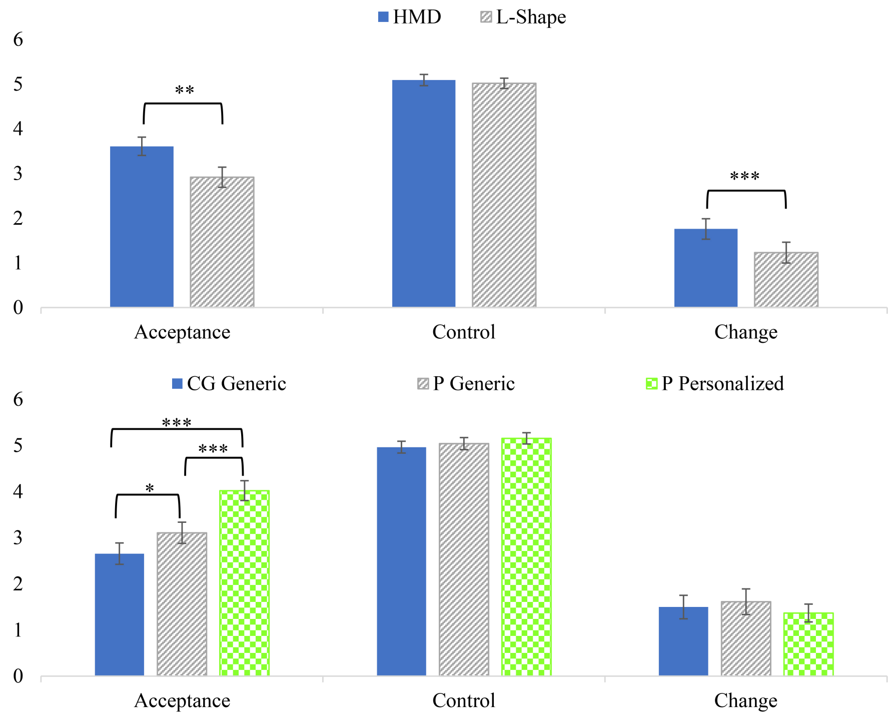

Figure 3.8: Comparisons for the main measures. *Top.* Comparisons split by *medium*. *Bottom.* Comparison split by *personalization*. Character generator avatar (CG) vs. generic photogrammetry avatar (P Generic), vs. personalized photogrammetry avatar (P personalized). *Note.* *** $p < .001$; ** $p < .01$; * $p < .05$; *bars denote the mean value; error bars denote the standard error.*

assumption was violated. When necessary, Huynh–Feldt corrections of DOF were applied.

Medium

Significant main effects for the factor medium (HMD vs. L-shape) were found for the in-situation body ownership ($F_{1,27} = 17.66$, $p < .010$, $\eta_p^2 = .40$), the in-situation agency measure ($F_{1,27} = 7.71$, $p = .010$, $\eta_p^2 = .22$), the presence measure ($F_{1,27} = 32.04$, $p < .001$, $\eta_p^2 = .54$), the post-immersion acceptance measure ($F_{1,27} = 14.57$, $p = .001$, $\eta_p^2 = .35$) and the post-immersion change measure ($F_{1,27} = 18.78$, $p < .001$, $\eta_p^2 = .41$). With regard to the post-immersion alpha-IVBO measure, pairwise comparisons revealed that the acceptance of the virtual body and the perception of change were higher in the HMD condition when compared to the L-shape condition ($p \leq .001$). Table 3.7 depicts marginal means. Figure 3.8 (Left) illustrates the results.

Personalization

We found significant main effects for the in-situation body ownership ($F_{2,54} = 27.43$, $p < .001$, $\eta_p^2 = .50$), the in-situation presence ($F_{2,54} = 32.04$, $p = .001$, $\eta_p^2 = .21$), and for the post-immersion acceptance ($F_{2,54} = 25.16$, $p < .001$, $\eta_p^2 = .48$) measures. Pairwise comparisons revealed that acceptance was higher for the personalized avatar as compared to both generic avatars ($p \leq .001$). The level of body acceptance of the generic photogrammetry avatar was also higher compared to the generic character generator avatar ($p = .033$). Table 3.8 depicts the results of all comparisons.

Clothing and Control Measures

The perception scale for clothing showed a significant main effect for the between-factor *clothing* ($F_{1,27} = 18.83$, $p < .001$, $\eta_p^2 = .41$) and for *personalization* ($F_{1.64,44.25} = 4.45$, $p = .023$, Huynh-Feldt-$\epsilon = .82$, $\eta_p^2 = .14$). The assessment of similarity showed a significant main effect for *personalization* ($F_{2,54} = 55.45$, $p < .001$, $\eta_p^2 = .67$) and *medium* ($F_{1,27} = 5.00$, $p = .034$, $\eta_p^2 = .16$).

Table 3.6: Univariate Main Effects.

Scale	Personalization†	Medium†	Clothing†
(In-situ) Ownership	*** (.50)	*** (.40)	
(In-situ) Agency		.010 (.22)	
(In-situ) Presence	.002 (.21)	*** (.54)	
Acceptance	*** (.48)	.001 (.35)	
Control			
Change		*** (.41)	
Clothing Perception	.023 (.14)		*** (.41)
Similarity	*** (.67)	.034 (.16)	

Note. † p (η_p^2); *** $p < .001$.

Table 3.7: Estimate Marginal Means for the Within-Factor Medium.

Scale	HMD†	L-shape†	p
(In-situ) Ownership ^{a}	5.00 (± .31)	4.66 (± .41)	***
(In-situ) Agency ^{a}	8.13 (± .27)	7.75 (± .27)	.010
(In-situ) Presence ^{a}	6.77 (± .30)	4.56 (± .45)	***
Acceptance ^{c}	3.61 (± .21)	2.92 (± .23)	.001
Change ^{c}	1.76 (± .23)	1.23 (± .23)	***
Manipulation Check: Similarity ^{a}	4.80 (± .30)	4.26 (± .32)	.034

Note. † *Mean [± standard error of the mean (SEM)];* *** $p < .001$;
In situation measures (In-situ) and post-experimental measures.
Likert scale range from low to high: ^{a}*0-10,* ^{c}*0-6;*

Table 3.8: Estimate Marginal Means and Pairwise Comparisons for Personalization.

Scale	(1) CG Gen. †	(2) P Gen.†	(3) P Pers. †	(1 to 2)§	(1 to 3)§	(2 to 3)§
(In-situ) Ownersh.^{a}	4.42 (± .42)	4.75 (± .38)	6.82 (± .35)		***	***
(In-situ) Presence^a	5.28 (± .35)	5.58 (± .36)	6.14 (± .34)		.002	.015
Acceptance^c	2.66 (± .23)	3.11 (± .23)	4.02 (± .22)	.033	***	***
Clothing^a	3.80 (± .49)	4.01 (± .43)	2.80 (± .41)			.016
Similarity^a	2.92 (± .45)	3.11 (± .46)	7.56 (± .30)		***	***

Note. † *Mean (± SEM);* § *pairwise comparison of indicated levels;* *** $p < .001$; *CG Character Generator; P Photogrammetry; Likert scale range from low to high:* ^{a}*0-10,* ^{b}*1-9,* ^{c}*0-6.*

Correlations

We calculated bivariate Pearson correlations between the main dependent variables. Specifically, we were interested in how the ownership components relate to presence, and whether the scale components pinpoint the intended aspects of ownership and control. For all trials, we found significant correlations between in-situation ownership and in-situation presence ($r \geq .377$, $p \leq .044$), between in-situation ownership and similarity ($r \geq .462$, $p \leq .012$), between in-situation ownership and acceptance ($r \geq .438$, $p \leq .017$), and between similarity and acceptance ($r \geq .500$, $p \leq .006$), thus pointing to a interdependency between these measure combinations.

We furthermore found a significant correlation between in-situation presence and *acceptance* that was stable across all trials ($r \geq .371$, $p \leq .048$).

With regard to the control component, there was a significant correlation between in-situation agency and *control* ($r \geq .452$, $p \leq .014$), which one would have expected on the basis of the questioning. No further significant correlations were observed.

3.4.3 Discussion

With regard to personalization, the study confirmed previous findings that suggest an impact of appearance on ownership and presence [148, 199, 305, 323]. More precisely, we found a statistically significantly higher *acceptance* for personalized avatars in comparison to generic avatars. As one would expect from the construction of the study that used identical tracking methods and quite similar rendering latencies, there were no impacts on the perceived *control*. Our findings did *not* show that personalization also *change* the perception of body scheme. Compared to Study 1 however, a *change* in body scheme would not necessarily have been expected, as the photogrammetric personalization that was performed most likely neglected differences in the body scheme. In contrast to the photogrammetry approach, Study 1 utilized avatar bodies that were constructed with 3D authoring software (and thus approximates rather than photogrammetric captures) for the personalized condition.

Our findings confirm a suggested impact with regard to immersion, more precisely an increase of the *acceptance* of the virtual body as the own, and the perception of perceptual *change* of the body scheme. A logical explanation is the respective reference frame, which differed in the L-shape condition (looking down I see the physical body) compared to the HMD condition (looking down, I see the virtual body). This relates to previous findings that link the ownership to the perspective constraint [199], yet, in our study the main changes were the reference frame rather than a variation of first vs. third-person perspective. Therefore, we could clearly pinpoint that not the differences in avatar appearance but the actually higher immersion evoked a difference in perceived body change. This effect is accompanied by a greater perception of presence. With regard to the correlations, we could substantiate our assumptions that *acceptance* relates to the dimension of ownership, and that the *control* subscale assesses agency. Adding to the argument of the personalization results, similarity and acceptance correlated significantly. Furthermore, we found that the in-situation presence assessment correlated with in-situation ownership and acceptance which points at a relation between presence and virtual body ownership. However, based on our analyses, we cannot draw conclusions regarding the causality of this relation.

Limitations

One limiting factor of the study was that the personalization condition did not include a personalized character constructed through a character generator. Despite the fact that we found significant differences between a scanned personalized/-generic character and a generic character from a generator, there was no implication that these results would also hold true for a personalized character created using a character generator. This limitation is addressed in Section 3.6.2. Furthermore, compared to the previous study, we did not employ any facial or gaze tracking, which is why the behavioral realism might be slightly reduced compared to the study described in Section 3.3.

Conclusion

In this study, we could identify that both, the type of medium (i.e., the immersion) and personalization strongly affect the level of ownership (the *acceptance* component) over a virtual body. We further found that the level of immersion affects the perceived *change* of the body schema. The proposed scale performed reliable, yet generalizing the measurement instrument to non-mirror setups requires further research, which is described in Section 3.6. The presented results do rely on single user simulations. One aspect that was widely neglected in previous work is the potential impact of social situations and, consequently, the impact of the perception of others on the self. Therefore, we conducted a third study, reported in the following Section 3.5.

3.5 Study 3: Impacts of Social Situations

A third study aimed to investigate the effect of avatar realism on the VBO in the context of social interactions. Following the model presented in Section 2.4.2, we could assume that there are interactions between the perception of others and the perception of the self in a VE that may affect the self-embodiment in immersive social interactions. In this regard, one influential factor identified is consistency [308], which was shown to also affect the perception of familiarity [193] as one aspect of a suggested Uncanny Valley [210]. However, previous studies found that a reduced consistency in a single avatar affected the perception towards that avatar [193]. In this study, we followed another approach and investigated avatar appearance consistencies across interactions, namely human-agent interactions, and how these affect virtual embodiment.

Figure 3.9: Avatars used in Study 3 (adapted from [174], Copyright © 2017 Latoschik et al.).

3.5.1 Method

Design

In a two-factor (self-representation type, agent representation type) within-subject design, we evaluated how social interaction with an agent affects the self-perception and the perception of others. Regarding the first factor (self-representation type), participants were represented either by an abstract wooden mannequin or by a realistic avatar (see Figure 3.9)created through the photogrammetry approach presented by Achenbach, Waltemate, Latoschik, and Botsch [1] (see also Section 3.4.1). In a similar manner, the animated agent (agent representation type) was represented as either a wooden mannequin, or a realistic human character, leading to consistent and inconsistent conditions (see Figure 3.10). We induced body ownership through controlled movement instructions. To explicitly induce a social interaction, the participants were instructed to return an animated waving gesture of the agent.

Prestudy

In order to find avatar types that result in an inconsistent perception, we pretested eight avatars. Five avatars were created via photogrammetry scans, one was created with a character generator[20], and two abstract avatars, namely the wooden mannequin and the robot avatar also used in the previous study (Section 3.3), were tested regarding the perception of uncanny valley factors [136] and emotion [51]. We chose the wooden mannequin over the robot avatar because in addition to a limited humanness, the robot avatar evoked feelings of eeriness and differences in arousal and dominance that could potentially be confounding factors for the experiment. A male and female avatar were selected from the photogrammetry set on the basis of having the most equally balanced ratings in comparison to the wooden mannequin (see [174] for details). The avatars finally selected are depicted in Figure 3.9.

[20] Autodesk Character Generator, Autodesk, `https://charactergenerator.autodesk.com/`

Figure 3.10: Setup to assess the impact of social situations. *Top:* Congruent and incongruent interaction situations. The participant interacts with a visitor (agent) on the outside of the window. *Middle:* Procedure. In each trial an embodiment induction phase was performed before a social interaction phase. *Bottom:* Induction phase. Movements were performed in front of a virtual mirror. Parts of the figure are reprinted from [174].

Apparatus

A VE in the form of a living room was developed using Unity3D[21]. Users were tracked with an optical infrared passive-marker tracking system (OptiTrack Flex3, 16 cameras, 120 Hz[22]). The tracking system was interfaced with the simulation, and the retrieved skeletal motion was used to retarget the participants' behavior to avatar movement. The simulation was rendered to an Oculus Rift CV1[23] HMD (2160px $\times$ 1200px, 90 Hz). We used the HMD internal inertial measurement unit (IMU) data, which has a higher refresh rate, to control the virtual camera. The relative orientation data of the IMU was fused with the absolute orientation data (up-axis) of the tracking system in order to match coordinate systems via a calibration routine. Audio instructions were displayed via the HMDs earphones. The overall end-to-end motion-to-photon latency (tracking to display) was measured by manual

[21] Unity Technologies, `https://unity3d.com/`
[22] NaturalPoint, Inc. DBA OptiTrack, `https://optitrack.com/products/flex-3/`
[23] Facebook Technologies, LLC. `https://optitrack.com/products/flex-3/`

frame counting similar to the method by He, Liu, Pape, Dawe, and Sandin [134]. Measurements revealed an overall motion-to-photon system latency of $M = 80.8$ ms, $SD = 14.62$ ms.

Measures

We assessed the Alpha-IVBO questionnaire (see Table 3.4). Scale reliabilities were acceptable to good, with $\alpha \geq .733$ with the exception of the *control* factor scale in the self–wooden mannequin/other–wooden mannequin condition. We further assessed how the participants judged the interactant (i.e., the agent) after the interaction with regard to the uncanny valley factors of humanness, eeriness, and attractiveness, by assessing the scale from Ho and MacDorman [136] ($\alpha \geq .762$).

We further assessed communicative aspects of the interaction, the perceived copresence, telepresence, social presence [220], and virtual rapport [121]. Furthermore, we assessed trust in the agent with three questions adapted from Chun and Campbell [66] to fit the context: "I think the virtual character has good intentions.", "I would count on the virtual character", "I would trust the virtual character", (1 - does not apply at all, 7 - applies totally, see [174] for further details). We also assessed simulator sickness [156] and trait empathy [311] along with demographic variables, which are not subject to the present discussion.

Procedure

We welcomed participants and informed them about the study. After providing written consent to agree to participate, the participants dressed in the motion capture clothing and filled out the pre-study questionnaire. From this point, the participants performed all conditions in randomized order. In each trial, the participants were calibrated for the motion tracking and equipped with the HMD. All instructions were pre-recorded and played through the earphones. At the beginning of the exposure, we induced body ownership using motion instructions that lasted for approximately 120 s. Participants were asked to step on a marked spot in front of the mirror and perform four specific arm movements (right/left arm up to the front, right/left arm up in front of the breast) accompanied by instructions where to focus (e.g., "Hold your right arm up straight in front of you, the inner hand faced to the floor; focus on your right hand; focus on the same hand in the mirror; let your arm rest again"). Participants were then asked to turn 90 degrees and step on another marking that appeared in front of the window. Participants were instructed to look out of the window and a fixation cross was displayed. They were instructed to return the wave from the virtual character that would appear in order to induce a social gesture (the waving was mirrored, meaning that for right handed participants, the agent would wave with the left hand, and vice versa). After the agent appeared, the virtual agent put up his hand and waved to the participant, animated by the pre-recorded animation, for approximately 10 s. After the interaction, participants were helped to remove the HMD and guided to the digital questionnaire assessing the dependent measures. After finishing all trials, participants were compensated and released from the study.

Participants

The final sample consisted of 20 participants ($M = 20.25$, $SD = 1.21$, 11 female, 9 male, all students at the time). None of the participants reported severe motor, auditory or visual disabilities/disorders. Two participants had slight visual impairments that were not corrected for during the experiment. Sixteen participants had previous experience with VR, and four of the participants were left-handed.

3.5.2 Results

Data were analyzed using two factor ANOVAs for repeated measures of all measurements taken.

Humanness, Eeriness, and Attractiveness

We analyzed the participants judgment of the virtual agent. Results revealed a significant main effect for agent representation type; $F(1, 19) = 10.91$, $p = .004$, $\eta_p^2 = .365$. As expected by the pretest results, the analyses showed higher levels of humanness for the realistic avatars created using photogrammetry ($M = 2.97$, $SE = 0.16$) than for the wooden mannequin ($M = 2.23, SE = 0.20$).

Virtual Body Ownership

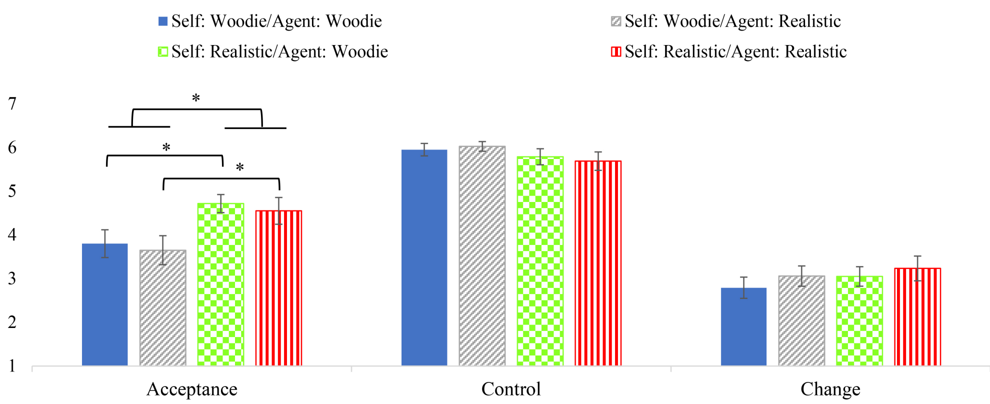

Figure 3.11: Comparisons for acceptance, control, and change. *Note.* * $p < .05$; *bars denote the mean value; error bars denote the standard error.*

Figure 3.11 depicts the results of the IVBO questionnaire. Analyses showed a significant main effect for self representation type; $F(1, 19) = 6.19$, $p = .022$, $\eta_p^2 = .246$. *Acceptance* was rated higher for the realistic photogrammetry avatar ($M = 4.63$, $SE = 0.24$) compared to the wooden mannequin ($M = 3.72$, $SE = 0.31$). Pairwise comparisons revealed significant differences for both agent representation types ($p \leq .034$).

We observed a trend for agent representation type in the factor *change*, which measures the change in self-perception of the own body; $F(1, 19) = 4.21$, $p = .054$, $\eta_p^2 = .181$. Pairwise comparisons showed a higher value for the perceived *change* if the agent representation type was a realistic human ($M = 3.15$, $SE = 0.23$) than if it was a wooden mannequin ($M = 2.91$, $SE = 0.20$). Yet the difference was not backed by statistical significance and, thus, further research is needed.

No further interaction or main effects were observed.

Other Measures

We did not find any significant effects in the analyses for the different aspects of presence, trust, or rapport.

3.5.3 Discussion

The study could confirm the impact of realism with regard to the *acceptance* component of the Alpha-IVBO scale, that is, the realism of the representation seems to have an effect on the VBO. It is safe to assume that the humanness item of the scale impacted these results. While it is no surprise that the manipulation did not strongly impact the perceived *control*, one could have assumed an effect of self representation type on the perceived change. However, the data did not show this effect. We believe this might be due to the fact that the wooden mannequin was still humanoid to some degree, and the body measures applied to many participants. More specifically, we also scaled all avatars according to the participant's size, and thus, height differences would not have had any impact. These results are in line with Study 1, where only a personalized character was found to significantly impact the perceived change in comparison to the wooden mannequin. One interpretation of this finding is that a perceived change in body scheme is moderated by the level of acceptance or ownership perceived of a virtual body. The latter interpretation requires further research. While we observed a trend that may indicate that the perceived ownership of a virtual body is affected by the interactant's virtual character type, or in other words by the consistency of the representations, this effect was not backed by statistical significance and thus further research is of need.

Limitations

A main limitation with regard to the concept in our study was the use of a virtual agent rather than a real human interactant. The reason for this choice was mainly due to the fact that human-agent interactions can be controlled, and thus a controlled exposure to the stimulus was possible. Yet, this prohibited the emerging of dynamics that are typical for human-human interaction. Further research is needed in this area. A second limitation was that the characters in the congruent conditions were identical. This may have been misleading to the participants.

Conclusion

Study 3 substantiated the positive impact of a human-like appearance on the perception of ownership, meaning the acceptance of the virtual body as the own in the case of the more immersive HMD-driven condition is fostered by an avatar that shows human form and texture realism. The study finds that this effect also holds true for social interactions with congruent and incongruent representations of the communicators.

Reflecting on Studies 1-3 more generally, a limitation of the findings is that a manipulation of agency was not yet performed, which is why such a manipulation is addressed in Section 3.7. The presented scale did show good reliability, and the results presented in Study 1-3 are in line with the previous research (see Section 3.1). Yet, with the current phrasing, the scale does not generalize to all scenarios, such as for example simulations without virtual mirrors, which is why an improvement is the topic of the next section.

3.6 Scale Generalization and Validation

While the previously presented measurement for embodiment is usable for many experiment types, three down sides can be identified: 1) the measure is strongly constrained to "virtual mirror" scenarios due to the phrasing, and adapting the scale to more general scenarios may not necessarily result in similar reliability, 2) some of the questions could be improved in rephrasing, and 3) the components are not balanced with regard to the number of items, especially the acceptance component only consists of 3 questions. To further improve the balance of the scale, we added questions to balance the component assessment based on the review of previous work and on the assumption they load on the *acceptance* (14. belongsOtherPerson, 16. belongsToMe) and *control* factors (15. synchronousMovements), respectively. The following scale, shown in Table 3.9, was constructed. A confirmatory factor analysis (CFA) was calculated on the basis of data from three studies that assessed impacts of particular manipulations on embodiment, which will be discussed in more detail in the following Sections 3.6.1, 3.6.3, and 3.6.2.

Confirmatory Factor Analysis

We performed a CFA ($N = 169$) using R[24] with the lavaan package. The reporting of fit indices is based on the recommendation made by Kline [162].

As the assumption of multivariate normality was violated, we conducted a robust maximum likelihood estimation and computed Satorra-Bentler (SB) corrected test statistics (see [53]). A first attempt did not yield an acceptable model fit, and as the modification indices indicated, there were covariations in the error terms of a particular item loading on several factors (1 in factor change: 8. ownOtherBody). Therefore, this item was dropped.

A second attempt did yield an acceptable model fit. However, inspection of the modification indices revealed covariations in the error terms of two items loading on several factors (1 in factor acceptance: 14. belongsOtherPerson, 1 in factor control: 5 controlEnjoyment). Problematic items were thus excluded. Furthermore, items with a factor loading $< .40$ were excluded (which was the case for: 10. myBodyCheck).

[24] The R Foundation, `https://www.r-project.org/`

Table 3.9: Scale Generalization and Extension (Beta-IVBO) and Comparison to the Alpha-IVBO Questionnaire.

Acceptance	**1. myBody**
Alpha	*I felt as if the body I saw in the virtual mirror might be my body.*
Beta	It felt like the virtual body was my body.
	2. myBodyParts
Alpha	*I felt as if the body parts I looked upon were my body parts.*
Beta	It felt like the virtual body parts were my body parts.
	3. humanness
Alpha	*The virtual body I saw was human-like*
Beta	The virtual body felt like a human body.
New	**14. belongsOtherPerson ***
Beta	I felt like the virtual body belonged to someone else. *
New	**16. belongsToMe**
Beta	It felt like the virtual body belonged to me.
Control	**4. myMovements**
Alpha	*The movements I saw in the virtual mirror seemed to be my own movements.*
Beta	The movements of the virtual body felt like they were my movements.
	5. controlEnjoyment
Alpha	*I enjoyed controlling the virtual body I saw in the virtual mirror.*
Beta	I enjoyed controlling the virtual body.
	6. controlMovements
Alpha	*I felt as if I was controlling the movement I saw in the virtual mirror.*
Beta	I felt like I was controlling the movements of the virtual body.
	7. causeMovements
Alpha	*I felt as if I was causing the movement I saw in the virtual mirror.*
Beta	I felt like I was causing the movements of the virtual body.
New	**15. syncMovements**
Beta	The movements of the virtual body were in sync with my own movements.
Change	**8. ownOtherBody**
Alpha	*The illusion of owning a different body than my real one was very strong during the experience.*
Beta	I had the illusion of owning a different body to my own.
	9. myBodyChange
Alpha	*At a time during the experiment I felt as if my real body changed in its shape, and/or texture.*
Beta	I felt like the form or appearance of my own body changed.
	10. myBodyCheck
Alpha	*During or after the task, I felt the need to check that my body really still looked like what I had in mind.*
Beta	I felt like I had to check that my own body still looked like I remembered.
	11. echoHeavyLight
Alpha	*I felt an aftereffect as if my body had become lighter/heavier.*
Beta	I felt like the weight of my own body had changed.
	12. echoTallSmall
Alpha	*I felt an aftereffect as if my body had become taller/smaller.*
Beta	I felt like the size (height) of my own body had changed.
	13. echoLargeThin
Alpha	*I felt an aftereffect as if my body had become larger/thinner.*
Beta	I felt like the width of my own body had changed.

*Note. * needs recoding.*

A third CFA with the remaining 16 items yielded a more parsimonious solution with a good model fit, SB $\chi^2 = 52.50$, $df = 51$, $p = .416$, root mean square error of approximation ("RMSEA") $= .013$, 90% confidence interval of robust root mean square error of approximation $[.000; .052]$, standardized root mean square residual ("SRMR") $= .047$, and a robust comparative fit index ("CFI") $= .998$. Thus, the obtained solution was deemed acceptable to characterize the dimensionality of virtual embodiment. The reliabilities for *acceptance* ($\alpha = .783$), control ($\alpha = .764$), and change ($\alpha = .765$) were acceptable. The final scale is depicted in Table 3.10.

Table 3.10: The finalized Gold-IVBO questionnaire.

Component	Gold-IVBO Question
Acceptance	**AC1. myBody** It felt like the virtual body was my body.
	AC2. myBodyParts It felt like the virtual body parts were my body parts.
	AC3. human-like The virtual body felt like a human body.
	AC4. belongsToMe It felt like the virtual body belonged to me.
Control	**CO1. myMovement** The movements of the virtual body felt like they were my movements.
	CO2. controlMovements I felt like I was controlling the movements of the virtual body.
	CO3. causeMovements I felt like I was causing the movements of the virtual body.
	CO4. synchronousMovements The movements of the virtual body were in sync with my own movements.
Change	**CH1. myBodyChange** I felt like the form or appearance of my own body had changed.
	CH2. echoHeavyLight I felt like the weight of my own body had changed.
	CH3. echo-Tall-Small I felt like the size (height) of my own body had changed.
	CH4. echo-LargeThin I felt like the width of my own body had changed.

Note. Participant instructions: Please read each statement and answer on a 1 to 7 scale indicating how much each statement applied to you during the experiment. There are no right or wrong answers. Please answer spontaneously and intuitively.
Scale: horizontally oriented, 7 point (1–strongly disagree, 4–neither agree nor disagree, 7–strongly agree).
A professional service (tolingo) was consulted for the translation (DE-EN).
A backtranslation (Lengua) was performed to ensure a similar meaning. Scoring:
Acceptance = (AC1+AC2+AC3+AC4)/4
Control = (CO1+CO2+CO3+CO4)/4
Change = (CH1+CH2+CH3+CH4)/4

3.6.1 Study 4: Validating Impacts of Immersion

The first study aimed at confirming the functionality of the scale on the basis of a manipulation of the simulation's immersiveness and at investigating potential correlates to related constructs, namely aspects of presence.

Method

Design In a one-factor between-subjects design modifying the level of *media immersiveness*, participants were either exposed to a projection-based simulation that followed a fish tank VR principle [353] or an immersive simulation presented using an HMD. We induced VBO with similar methods as we used in the previous experiments, by instructing the participants to perform certain movements and focus procedures.

Figure 3.12: Study setup to assess the impcat of immersion. *From left to right:* User in the projection trial (lower immersion, physical reference), user in the HMD trial, mirror vision of the participant in the HMD trial, male and female avatars.

Apparatus The HMD based simulation was developed in a similar manner as the apparatus in Section 3.5.1. Figure 3.12 depicts the setup. The simulation was rendered using an Oculus Rift CV1[25] HMD (2160px × 1200px, 90 Hz). The projection based method was displayed using an Acer H6517ST[26] low-latency, active stereoscopic projector via back projection. The image was cropped and projected in form of a "fake mirror" sized 1.31 (width) × 1.46 (height) meters. The resulting resolution was 480px × 1080px for each eye (transmitted via side-by-side 3D stereo). The projection matrix of the image rendered to the projector was calibrated by measuring the projection's center and edges, utilizing a rigid body and the passive marker tracking system, and matching the according virtual projection preferences. In this way, virtual and physical projection properties are matched. The off-axis stereoscopic projection [49] behaves according to the physical distance of the user. To account for the user's distance from and position in relation to the screen, a head model based on the head-markers of the tracking system was used as reference,

[25] Facebook Technologies, LLC. `https://optitrack.com/products/flex-3/`
[26] Acer Inc. `https://www.acer.com/ac/en/US/content/professional-model/MR.JLA11.009`

placing the center of interest at the approximate left and right eye positions for each virtual camera.

The end-to-end latency of the simulation was approximated via a video-based measurement with frame counting [134] to $M = 77.17$ ms for the projection setup, assuming slightly less latency with the HMD.

Measures We assessed the above items from the Beta-IVBO (respectively Gold-IVBO) scale (αs $\geq$.751). To confirm impacts of immersion on presence and to assess relations between presence and embodiment components, we measured the igroup presence questionnaire (IPQ) [27] [290] in a reduced form in order to fit the study scenario. In addition, we assessed demographic variables.

The IPQ adaptation assessed *general presence* ("In the computer-generated world, I had a sense of 'being there"'), *spatial presence* (" I did not feel present in the virtual space", "I had a sense of acting in the virtual space, rather than operating something from outside", "I felt present in the virtual space"; $\alpha = .786$), *involvement* ("I was not aware of my real environment"," I still paid attention to the real environment","I was completely captivated by the virtual world"; $\alpha = .851$), and *realness* ("How real did the virtual world seem to you?","How much did your experience in the virtual environment seem consistent with your real-world experience?","How real did the virtual world seem to you?"; $\alpha = .672$). The responses were given using a 7-point Likert-type scale (see original source for the anchors).

We further assessed aspects of the MEC spatial presence questionnaire [346], immersive tendency [364], simulator sickness [156], perception of humanness and eeriness [136], positive and negative affect [325, 145], and qualitative feedback. The results of these measures are not subject to the current discussion.

Procedure We welcomed participants and informed them about the study. After the participants provided their written consent to take part in the study, they dressed in the motion-capture clothing and filled out the pre-study questionnaire. Participants were then calibrated for the motion tracking and equipped with an HMD or active stereoscopy glasses. Following the exposure and a short acclimatization, pre-recorded audio instructions were displayed via the earphones. Participants were asked to step on a marked spot in front of the mirror and perform specific movements (e.g., right/left arm up to the front, right/left arm up in front of the breast) accompanied by instructions where to focus (e.g., "Hold your right arm up, straight in front of you, with the inner hand faced to the floor; focus on your right hand; focus on the same hand in the mirror; let your arm rest again"). The instructions lasted for about 150 seconds. After the tasks were finished, participants were helped to remove the HMD or stereoscopy glasses and were guided to the digital questionnaire assessing the dependent measures. Afterward, participants were compensated and released from the study.

Participants Participants were recruited through the recruitment system of the Institute for Human-Computer-Media (University of Würzburg). We excluded

[27] `http://www.igroup.org/pq/ipq/index.php`

participants who had problems during the procedure or when severe tracking errors were noted by the experimenter or the participants. The final sample of $N = 50$ participants included 32 females and 18 males, $M_{age} = 22.18$, $SD_{age} = 2.83$), and 49 were students at the time. There were 46 participants who had previous experience with VR technologies. The sample was equally distributed between conditions (25 per condition).

Results

Comparisons T-tests were conducted for each individual measure. In the case of unequal variances, corrected values are reported. The descriptives results are depicted in Figure 3.13.

Regarding the Gold-IVBO, the perception of change was significantly different between the projection condition ($M = 2.61$, $SD = 1.46$) and the HMD condition ($M = 3.40$, $SD = 1.31$; $t(48) = -2.013$, $p = .0498$, $d = 0.61$). As expected, the immersive HMD condition resulted in a stronger perception of change. Neither acceptance nor control showed a significant difference, the results are depicted in figure 3.13.

With regard to the IPQ presence measures, we found a significant difference with general presence between the projection condition ($M = 3.88$, $SD = 1.45$) and the HMD condition ($M = 5.32$, $SD = 0.99$); $t(42.30) = -4.098$, $p < .001$, $d = 1.16$.

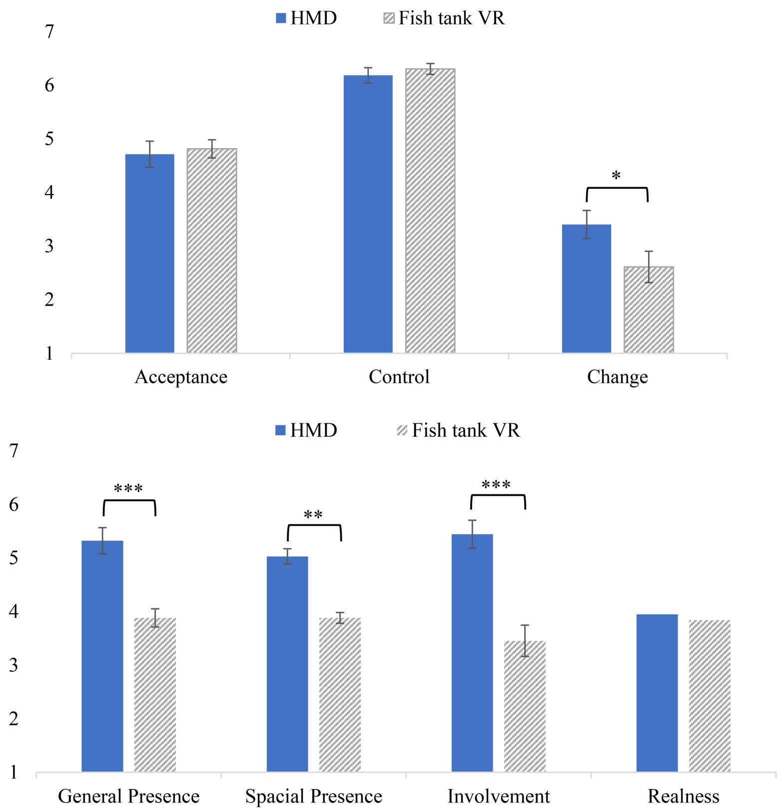

Figure 3.13: Comparisons of the main measures. *Top:* Gold-IVBO factors. *Bottom:* IPQ factors. *Note.* * $p < .05$; *** $p < .001$; *bars denote the mean value; error bars denote the standard error.*

Table 3.11: Significant Bivariate Pearson Correlations.

	Control	Change	General Presence	Spatial Presence	Involvement	Realness
Acceptance	.460**		.360*	.338*		.417**
Change			.301*			
General Presence				.752**	.591**	.528**
Spatial Presence					.630**	.397**

Note. * $p < .05$; ** $p < .01$.

This effect was substantiated by a significant difference in the spatial presence measure between the projection condition ($M = 3.88$, $SD = 1.32$) and the HMD condition ($M = 5.03$, $SD = 1.19$); $t(48) = -3.218$, $p = .002$, $d = 1.19$. as well as for the involvement measure between the projection condition ($M = 3.45$, $SD = 1.28$) and the HMD condition ($M = 5.44$, $SD = 1.18$); $t(48) = -5.712$, $p < .001$, $d = 1.62$. There was no significant difference in the realness measure $t(48) = -0.388$, $p = .70$.

Correlations of Constructs In order to assess how the Gold-IVBO factors correlate with the related constructs of the IPQ, we calculated bivariate Pearson correlations between the Gold-IVBO scores and the IPQ scores. Table 3.11 depicts the results. We found significant correlations between the *acceptance* and *control*, as well as between the acceptance and the general presence, spatial presence, and realness. Furthermore, the *change* factor significantly correlated with the general presence assessment. Further, we found correlations within the presence measures.

Discussion

These results substantiate our previous comparison of media immersiveness (see Section 3.4.2), that showed that a higher level of immersion impacts self-perception in the form of a perceived body *change*. While Study 2 (discussed in Section 3.4) reported that the medium also affected the participants' *acceptance* of the virtual body, this was not confirmed with the applied measure in the current study. This finding indicates a more consistent pinpointing of the effect based on the updated scale items, which was confirmed by the correlation findings. While there was a correlation with medium effect between *control* and *acceptance*, the *change* factor was not found to have strong relationships with either of the other Gold-IVBO factors. Several presence dimensions were significantly affected by the manipulation and the HMD condition resulted in higher presence. The *acceptance* correlated with general presence, spatial presence, and realness, whereas the change factor correlated only to the general presence assessment. The reason for this may lie in the fact that not every participant perceived similar changes with regard to body measures when adapting to the avatar. We scaled the avatars in a universal form according to the height of participants. For some participants, the male or female virtual character body forms and appearance might have been a relative match to their real-world physical appearance. Others may have experienced stronger discrepancies between physical and virtual appearance. The *change* factor's questioning assesses these discrepancies and thus seems to cover the assessment of a perceived change in body schema which. Presence may to some degree be a prerequisite of such a perception, whereas it does not seem strictly related to spatial presence, involvement, or realness in the VE. The non-existent individualizaion or personalization also represents a limitation to our study, as we cannot generalize the findings to different character types.

We conclude that immersion is a driving factor for the change component, and therefore a driver for the perception of a change in body schema. Both, the acceptance (ownership) of a virtual body as well as a perceived change in body schema are related to aspects of presence. In Study 1 and Study 2 we found indications that personalization affects the acceptance of a virtual body. In the following Study, we investigated whether this impact holds true for user-performed avatar personalization.

3.6.2 Study 5: The Impacts of User-Performed Personalization

A second validation study was conducted with the aim of investigating what specific impacts a user-performed personalization has on the perceived VBO. We therefore designed a study that compared the effects of a generic avatar created by a character generator versus a character created by the participant through the same character generator.

Method

Design The study was conducted using a one-factor between-subjects design. We modified the level of *personalization* by either exposing the participants to a virtual embodiment scenario in which they were represented by a generic avatar from a character generator (the same male or female character for all participants), or in which they were represented by a personalized avatar they created themselves using a character generator. These participants were permitted to modify the characters appearance in form of body measures, facial appearance, and hair style; however, the clothing was kept similar to the generic avatars throughout both genders.

Apparatus We used a similar setup as the one used in the HMD condition of the immersion experiment (see Section 3.6.1), except that the participants for this study were immersed using a FOVE 0[28] HMD (2560px $\times$ 1440px, 70 Hz) that allows for a 100 degree field of view (FOV).

Measures We assessed demographic variables and the above items from the Beta-IVBO (respectively Gold-IVBO) scale (αs $\geq .744$). Similar to the work in [348], we assessed positive and negative affect using the PANAS scale in the short form [325, 145]. The positive affect component showed a reliability of $\alpha = .840$ whereas the negative component was of low reliability ($\alpha = .319$) and was therefore dropped from the analyses. To investigate a related construct, we measured self-presence

[28] FOVE, Inc., San Mateo, USA, `https://www.getfove.com/`

with the aspects defined by Ratan and Hasler [250] using the items they included in their factor analyses and adapted them to fit the context of our study (however, two items were removed; "To what extent does your avatar's profile info represent some aspect of your personal identity?", and "To what extent does your avatar's name represent some aspect of your personal identity?"). The proto self-presence measure ($\alpha = .781$) was assessed with items such as "To what extent do you feel like your arm is elongated into the virtual environment through your avatar?" or "How much do you feel like your avatar is an extension of your body within the virtual environment?" Core self-presence ($\alpha = .838$) was assessed with items such as "When arousing events would happen to your avatar, to what extent do you feel aroused?" or "When surprising events happen to your avatar, to what extent do you feel surprised?". Extended self-presence ($\alpha = .661$) was assessed with items like "To what extent is your avatar's gender related to some aspect of your personal identity?" or "To what extent is your avatar's clothing related to some aspect of your personal identity?". To identify a successful manipulation, we measured the perceived similarity of the avatar against the question "Please rate how much the virtual character is similar to you on the following scale, where 1 equals no similarity and 11 equals a digital twin." This measure was assessed both before the exposure, based on an image of the character on a desktop monitor, as well as after the exposure (reflecting the experience in the simulation).

We further assessed the eeriness and humanness of the character [136], similarity and naturalness of movements, simulator sickness, and qualitative comments. The latter measures are not part of the reporting of the present thesis.

Procedure We welcomed participants and informed them about the study. After providing written consent for participation, the participants dressed in the motion-capture clothing and filled out the pre-study questionnaire. Participants were then introduced to the character creator tool (personalized character condition) and were given about 10 minutes to adapt the hair, eyes, facial appearance, and body form of the character. Afterward, the character was presented to them. In the generic character condition, participants were presented with the generic character. After the presentation of the characters, the experimenter imported the character, and the participants were calibrated for the motion tracking system and equipped with an HMD.

Following the exposure and a short acclimatization, audio instructions (pre-recorded) were given via the earphones. Participants were asked to step on a marking in front of the mirror and perform specific movements, accompanied by instructions where to focus (see Section 3.6.1). In addition, the participants were specifically instructed to pay attention to features of the character's appearance (e.g., "Look at the eyes of mirrored self", "Look at the mouth of mirrored-self", "Turn 90 degrees and look at the mirrored self from the side").

The instructions lasted for about 180 seconds. After the tasks were finished, participants were helped to remove the HMD and were guided to the digital questionnaire assessing the dependent measures. When they had completed the questionnaire, participants were compensated and dismissed from the study.

Participants Participants were recruited using the recruitment system of the Institute for Human-Computer-Media of the University of Würzburg. We excluded participants if there were problems during the procedure or if severe tracking errors were noted by the experimenter or the participant. The final sample consisted of $N = 48$ participants (27 female, 21 male, $M_{age} = 21.64$, $SD_{age} = 2.25$), of which all 48 were students at the time. There were 45 participants who had previous experience with VR technologies. The final sample was similarly distributed between conditions (generic character: $N = 25$; personalized character: $N = 23$). All participants reported they had been speaking German for five or more years.

Results

Comparisons T-tests for the dependent measures did not reveal significant differences between the factors on the Gold-IVBO. We found a significant effect for positive affect, with higher positive ratings for the personalized avatar condition ($M = 3.64$, $SD = 0.68$) compared to the generic avatar condition ($M = 3.14$, $SD = 0.64$; $t(46) = 2.655$, $p = .011$, $d = 0.75$).

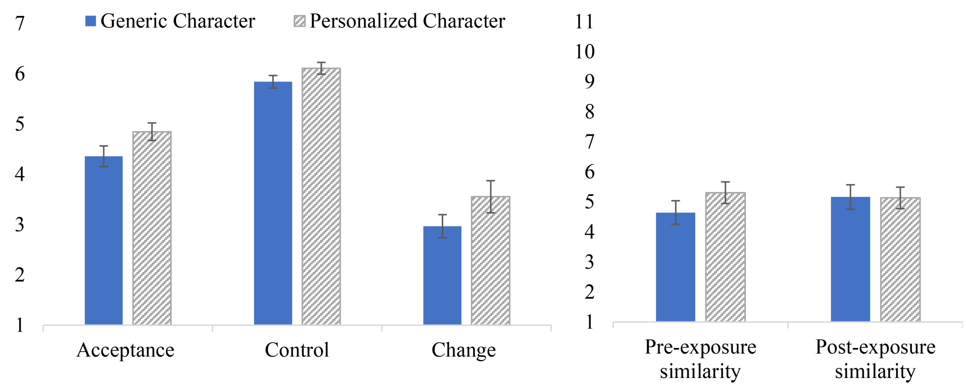

Figure 3.14: Comparison of the embodiment and similarity measures. *Left:* Gold-IVBO factors. *Right:* Similarity. *Note.* *$p < .05$; *bars denote the mean value; error bars denote the standard error.*

With regard to the self-presence measure, we found a significant effect with the proto-self–presence, showing higher ratings for the personalized avatar condition ($M = 3.59$, $SD = 0.67$) as compared to the generic avatar condition ($M = 3.04$, $SD = 0.81$; $t(46) = 2.528$, $p = .015$, $d = 0.734$).

Despite a higher level of similarity being perceived in the pre-exposure similarity measure, this difference was not significant. Descriptive results are depicted in Figure 3.14 and 3.15.

Correlations of Constructs In order to assess how the Gold-IVBO factors correlated with the perception of self-presence, positive affect, and avatar similarity, we calculated bivariate Pearson correlations between the Gold-IVBO scores, the self-presence factors, and the similarity perception. The results are given in Table 3.12. The *acceptance* factor correlated with both, the *control* and the *change* factors of the Gold-IVBO in the present study. Both *control* and *acceptance* correlated with the

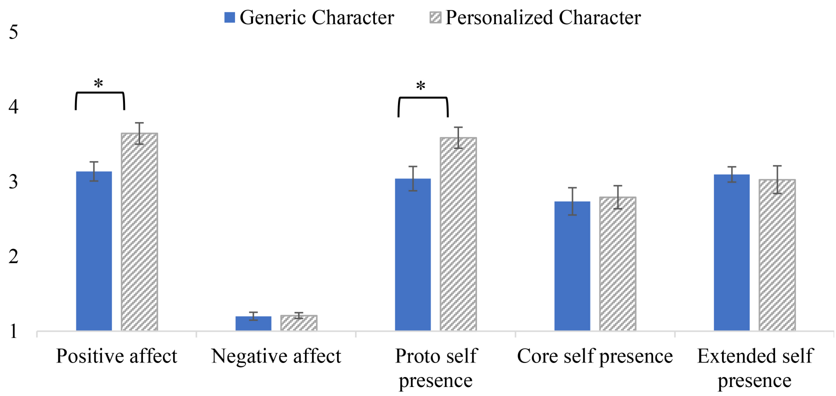

Figure 3.15: Comparison of the affect and presence measures. *Note. * $p < .05$; bars denote the mean value; error bars denote the standard error.*

Table 3.12: Significant Bivariate Pearson Correlations (r) of Related Constructs.

	Control	Change	Proto SP	Ext. SP	Core SP	Pos. Aff.	Simi. Pre	Simi. Post
Accept.	.506**	.374**	.696**					.332*
Control			.652**		.348*	.432**		.311*
Change			.297*					
Proto-SP					.365*	.326*		.427**
Ext.-SP					.402**		.356*	.321*
Core-SP						.311*		.428**
Pos. Aff.							.353*	.310*
Simi.-Pre								.441**

Note. * $p < .05$; ** $p < .01$.

post-exposure similarity measure, whereas the *change* factor did not. All Gold-IVBO factors correlated with the proto-self–presence measure. Interestingly, the *control* factor correlated with a perceived positive affect.

Discussion

Summarizing the results, we found that the personalization procedure did not evoke a lasting perceived similarity with a personalized avatar generated through a character generator, as assessed by the post-exposure similarity measure. Thus, we cannot substantiate the results of the findings in section 3.4 in this regard.

One explanation for this could be the limits placed on the manipulation of the character, which we did due to maintaining comparability. Furthermore, the time the participants had spent with the character generator might not have been suffcent to accurately configure the avatar's form and appearance. One further explanation might be the discrepancy between a desktop visualization and the immersive counterpart [279]. Further, participants might have had a different impression compared to what was presented in the simulation, which is also reflected when comparing the pre-exposure and post-exposure similarity measures.

However, we did find a positive association between the post-exposure similarity measure and both the *acceptance* and *control* factors of the Gold-IVBO, which therefore seemed to stand in relation. The proto-self–presence assessment correlated with the Gold-IVBO factors, to a large effect with *acceptance* and *control*. The *control* factor was further correlated with the core self-presence measure, which points at a relationship between agency and the perception of feelings and impressions that may be evoked by simulation events when controlling an avatar. A limitation of the study was that the possibilities to personalize the character were limited by the character generator's features and the predefined clothing, as was the time for the procedure. In this regard, we found higher values in the perceived change in the personalized condition, accompanied by higher values in the perceived acceptance (non-significant). This is interesting, yet difficult to interpret. One possible interpretation is that with the attribution of the avatar to the self, as participants spent time to configure and create their own avatar, they may have been more cautious with regard to the discrepancies between virtual and physical appearance and form. We conclude that the act of self-personalization did in fact have an impact on self-presence and positive affect, but did not show significant effects with regard to the embodiment components. In the following, we will investigate the impacts of increased degrees of replicated behaviors, namely facial expression and gaze.

3.6.3 Study 6: Exploring the Impacts of Behavioral Realism

While previous studies utilized technology to replicate more behavioral realism [191, 173], a comparison of different degrees of behavioral realism and its impact on the VBO had not been investigated. We therefore conducted a study investigating the addition of gaze replication and facial expression replication in immersive scenarios.

Method

Design In a two-factor between-subjects design, we evaluated the impact of the additional tracking and replication of *facial expressions* and *gaze* on the VBO and related constructs. Participants were represented by generic avatars and were exposed to the simulation while their body movement was replicated to the virtual simulation, either with or without additional facial expression replication, and with or without gaze replication, leading to four conditions: body movement only (BO), body and facial feature replication (BF), body and gaze feature replication (BG), and body, gaze, and face feature replication (BFG).

Apparatus The apparatus described in Section 3.6.2 was extended to utilize the FOVE 0 gaze tracking feature (120 Hz). The FOVE 0's binocular eye-tracking system (120 Hz) captures the reflected image on a hot mirror, showing infrared light emitted toward a user's cornea. Combining the two individual eye vectors, the FOVE integration calculates the intersection point to estimate the convergence point in the virtual scene, which is the approximate focus point the user is gazing at [362]. From this position, we recalculated the eyeball rotation, or more precisely, the eye "muscle value" in Unity's human pose model. Furthermore, we interfaced the BinaryVR Dev Kit V1 [29] to gather information about lower facial deformation. The 3D septh-sensing device Pico Flexx[30] (up to 45 Hz), using infrared illumintation, was affixed to the HMDs. Its 2D and depth images were processed by the BinaryVR Dev Kit to generate facial expression deformation parameters by tracking landmark locations [147, 371]. Tracked expression parameters were sent in the form of blendshapes and included, for example, jaw open, puff, and smile. After preprocessing, the sensor data was fused in the simulation to drive the avatar model, based on Unity's human pose solving model.

Measures We assessed the Gold-IVBO (αs $\geq .732$), a rating of the avatar assessing humanness, eeriness, and attractiveness (αs $\geq .688$) [136], the adapted presence measures previously applied in Study 5 (αs $\geq .645$) [250] (see Section 3.6.2) as well as positive and negative affect (αs $\geq .674$) [325, 145]. We further asked the participants how real, how natural, and how synchronous the movement (motion behavior) of the avatar appeared to them: "The movements were realistic," "The movements were naturalistic," "The movements were in synchrony to my own movements" (1–strongly disagree, 4–neither agree nor disagree, 7–strongly agree).

Procedure The procedure followed the pattern of the previous studies. We welcomed participants and informed them about the study. After providing written consent for participation, the participants dressed in the motion-capture clothing and filled out the pre-study questionnaire.

The participants were randomly assigned to one of the conditions. In each condition, participants were then equipped with an HMD and calibrated for the motion tracking, gaze tracking, and facial expression tracking systems.

Following the exposure and a short acclimatization, pre-recorded audio instructions were given via the earphones. Participants were asked to step on a marking in front of the mirror and perform specific movements, accompanied by instructions where to focus (see Section 3.6.1). In addition, the participants were specifically instructed to move closer, to another marking in front of the mirror, and let their gaze wander to specific focus points, trying to ensure an influence in perception of the manipulation (e.g., "Fixate on the left eye of your mirrored-self", "Fixate the right ear of your mirrored-self" etc.). For the facial expressions, we asked participants to perform certain expressions (e.g., "Open and close your mouth", "try to express happiness by smiling at your mirrored-self," etc.).

The instructions lasted for about 219 seconds. After each trial, participants were helped with removing the HMD, guided to the digital questionnaire assessing the dependent measures, and when all conditions were finished, participants were compensated and dismissed from the study.

Participants Participants were recruited using the recruitment system of the Institute for Human-Computer-Media of the University of Würzburg. We excluded par-

[29] BinaryVR, Inc. `https://www.binaryvr.com/`
[30] pmdtechnologies AG, `https://pmdtec.com/picofamily/`

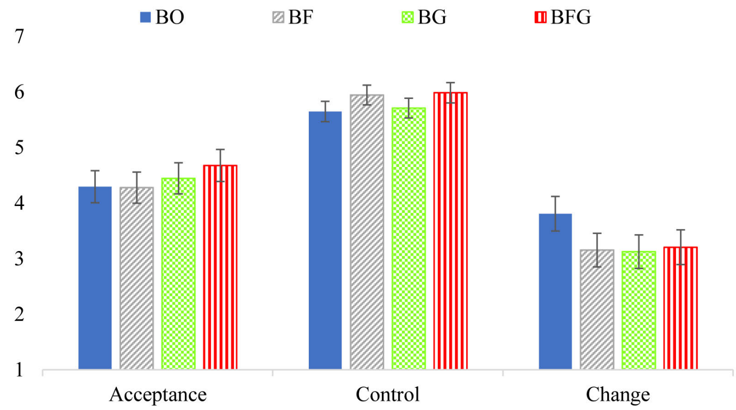

Figure 3.16: Comparison of the Gold-IVBO factors. *Note. bars denote the mean value; error bars denote the standard error.*

ticipants in cases of problems during the procedure or severe tracking errors noted by the experimenter or the participants. One participant had strong uncorrected visual impairment and was excluded from the main analyses. The final sample for the analysis consisted of $N = 70$ participants (46 female, 24 male, $M_{age} = 21.3$, $SD_{age} = 1.82$), of which all were students at the time. 17 participants were assigned to the BO condition, 18 to the BF condition, 18 to the BG condition, and 17 to the BFG condition. There were 65 participants who had previous experience with VR technologies. All participants reported speaking German as their mother tongue. The sample was similarly distributed among the conditions of BO ($N = 17$), BF ($N = 18$), BG ($N = 18$), and BFG ($N = 17$).

Results

Comparisons We analyzed each dependent variable using a two-factor (gaze, facial expression) ANOVA. In short, none of the Gold-IVBO factors revealed a significant effect. Figure 3.16 shows the descriptive results. While the BFG condition was rated highest in acceptance and highest in control, the differences from the other conditions were not backed by statistical significance.

We found a significant main effect of gaze on the perceived humanness; $F(1, 66) = 7.826$, $p = .007$, $\eta_p^2 = .106$, indicating the the perceived humanness was higher in the conditions with enabled gaze tracking and replication ($M = 3.07$,$SD = 0.78$) compared to the conditions with disabled gaze tracking and replication ($M = 2.60$, $SD = 0.60$).

No further significant effects were observed in the analyses with regard to the other dependent measures.

Discussion

The reason for missing effects on embodiment in the presented study could be manifold. First, it could be that facial expression and gaze replication does not have a strong influence in comparison to body movement, which was always present. Body movement arguably resulted in stronger visual stimulation in the simulation.

Future research should, therefore, investigate body movement as an additional manipulation factor. Second, and more likely, the induction phase might have been too short or too vague for the participants to grasp the manipulation, and as a result, the time spent on the gaze interaction and facial expression interaction might not been sufficient for an impact. This is backed by the fact that none of the realism ratings (naturalness, realism, synchronicity) showed a significant effect. Despite higher humanness ratings it is, therefore, most likely that the manipulation was not obvious enough to have an effect. Future studies should take care in the induction phase to make the modification visible to the participant. Nevertheless, we cannot exclude nor confirm impacts of behavioral realism based on the present data. As the presented Studies 1-6 mostly investigated effects of appearance, immersion and personalization, we will investigate the impacts of another aspect that may hinder SMCs to be established in VR simulations, namely latency, in the following.

3.7 Study 7: The Impacts of Latency and Latency Jitter

In order to confirm the impact of sensorimotor coherence found in previous studies [350], we designed a forth study utilizing the Gold-IVBO questionnaire. Due to the within-design structure of the present study, the data were not included in the above CFA (Section 3.6)that led to the questionnaire structure, and thus may provide additional insights. We were specifically interested in the impact of latency jitter, and therefore used the approach by Stauffert, Niebling, and Latoschik [313] to induce jPitter in the simulation. We further investigated associations with related measures.

3.7.1 Method

Design

In a one-factor (latency) within-subjects design, we evaluated the impact of latency and latency jitter, meaning the non-periodic spontaneous peaks of latency (see Section 3.6), on the perception of VBO and the factors of the Gold-IVBO scale. Participants were exposed to a simulation inducing body ownership with four conditions of delayed and jitter-delayed simulation display (minimal latency, medium latency, large latency, latency jitter), which altered the response of the visual feedback to motor actions.

Apparatus

We used a similar apparatus as the one used in Section 3.6.2 with minor differences in the implementation. By cuing (i.e., buffering) the tracking input data from the motion tracking system, artificial delays were introduced into the simulation, similar to the approach used by Waltemate, Senna, Hülsmann, Rohde, Kopp, Ernst, and Botsch [350]. However, because we exposed the participants to an immersive simulation by using an HMD, we tried to prevent biasing sickness effects and *did* *not influence the delay of the virtual camera*, which therefore transformed according to the raw system delay without further modifications. Thus, the body motion was delayed, whereas the visual field-of-view was not. We used an adaptation of the procedure described by Stauffert, Niebling, and Latoschik [313] to introduce latency jitter, which uses a stochastical model for latency spike distribution, thus introducing high latency spikes into the simulation loop. To assess the MTP latency of the resulting simulation, we utilized frame counting based on images from a digital video camera (1000 Hz). The measures resulted in $M = 90.12$ ms ($SD = 16.14$ ms) for the simulation baseline L_B, $M = 206.93$ ms ($SD = 16.47$ ms) for the small delay L_S, $M = 353.07$ ms ($SD = 15.38$ ms) for the larger delay L_L, and $M = 102.58$ ms ($SD = 49.71$ ms) for L_J. L_B, L_S, and L_L were measured with $N = 60$ samples. It is to note that despite measuring with $N = 165$ repetitions in the jitter condition L_J, the resulting mean and SD may not accurately reflect the induced jitter, due to some spikes that could not be captured using the motion-apex measurement applied.

During the simulation, users were embodied with the generic male and female characters used in the previous study (see Section 3.6.2).

Measures

We assessed demographic variables and applied the Gold-IVBO measure (αs $\geq .771$). To assess for relating constructs and previous measures, we used the questionnaire from Kalckert and Ehrsson [151], which was partly adapted from Longo, Schüür, Kammers, Tsakiris, and Haggard [184] and Botvinick and Cohen [47], that measures the constructs ownership (αs $\geq .801$), ownership control (αs $\geq .655$), agency (αs $\geq .636$), and agency control (αs between $.239$ and $.563$). We hypothesized these constructs to stand in direct relation to the proposed Gold-IVBO scale. Questioning was adapted to fit the virtual scenario, as these questionnaires were previously used to assess an RHI scenario (e.g., "The rubber hand moved just like I wanted it to, as if it was obeying my will" = "The virtual body moved just like I wanted it to, as if it was obeying my will"). Questions were assessed with a 7-point Likert type scale (1–strongly disagree, 4–neither agree nor disagree, 7–strongly agree). To control for our manipulation, we asked the participants how real, how natural, and how synchronous the movement (motion behavior) of the avatar appeared to them: "The movements were realistic," "The movements were naturalistic," "The movements were in synchrony to my own movements" (1–strongly disagree, 4–neither agree nor disagree, 7–strongly agree). We further assessed simulator sickness [156] and adapted two questions originating from Longo, Schüür, Kammers, Tsakiris, and Haggard [184] which loaded on the agency component in their analysis. The results of the latter two measures are not part of the present discussion.

Procedure

We welcomed participants and informed them about the study. After providing written consent for participation, the participants dressed in the motion-capture clothing and filled out the pre-study questionnaire.

The conditions were presented to the participants in randomized order. In each trial, the participants were calibrated for the motion tracking system and equipped with an HMD.

Following the exposure and a short acclimatization, pre-recorded audio instructions were given via the earphones. Participants were asked to step on a marking in front of the mirror and perform specific movements, accompanied by instructions where to focus (see Section 3.6.1). In addition, the participants were specifically instructed to perform more rapid and fluid movements (e.g., “Raise your left arm in moderate speed in front of you, and lower it back down next to your hip. Repeat this movement ten times, and focus on your arm while doing so”).

The instructions lasted for about 282 seconds. After each trial, participants were helped with removing the HMD, guided to the digital questionnaire assessing the dependent measures. After the four conditions were completed, participants were released from the study.

Participants

Participants were recruited using the recruitment system of the Institute for Human-Computer-Media. We excluded participants in the case of problems during the procedure or severe tracking errors noted by the experimenter or the participants. The final sample consisted of $N = 22$ participants (17 female, 5 male, $M_{age} = 21.77$, $SD_{age} = 3.62$), of which all 22 were students at the time. There were 21 participants who had previous experience with VR technologies. All participants reported to speaking German as their mother tongue.

3.7.2 Results

Comparisons

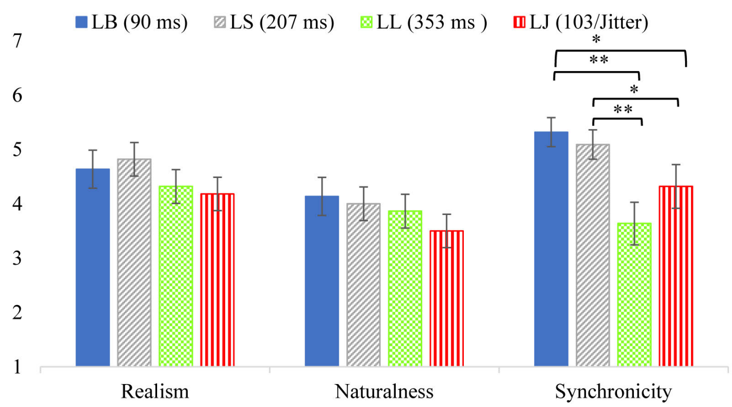

Figure 3.17: Comparison of the motion perception measures. *Note.* $^{*}p < .05$;$^{**}p < .01$ *bars denote the mean value; error bars denote the standard error.*

We compared the conditions by calculating repeated measures ANOVAs for the individual dependent variables. Where the assumption of sphericity was violated,

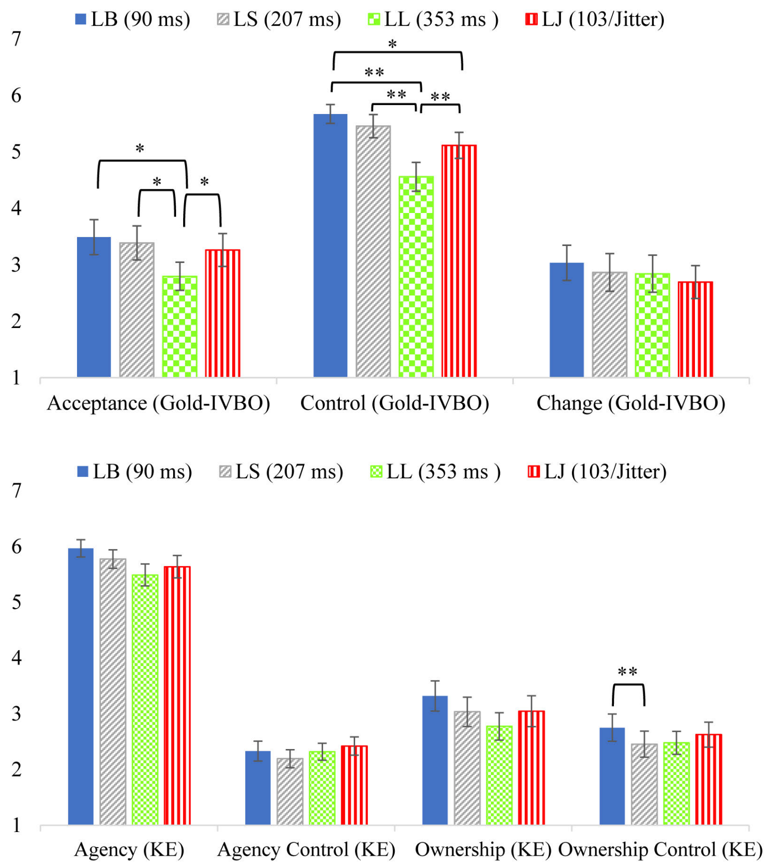

Figure 3.18: Comparison of the embodiment measures. *Note.* $^{*}p < .05$;$^{**}p < .01$ *bars denote the mean value; error bars denote the standard error.*

we report Greenhouse–Geisser corrected values. We found a significant main effect for the *acceptance* score of the Gold-IVBO scale; $F(3, 63) = 3.57$, $p = .024$, $\eta_p^2 = .138$. Similarly, the *control* measure was affected by the manipulation; $F(1.741, 63.553) = 7.50$, $p = .003$, $\eta_p^2 = .263$. No significant impacts were observed for the change factor.

We did not observe any significant effect in the agency and ownership measures of the scale adapted from Kalckert and Ehrsson [151] (KE). The synchronicity assessment showed a significant main effect $F(1.905, 40.005) = 7.077$, $p = .003$, $\eta_p^2 = .252$. Interestingly, neither the realism, nor the naturalness ratings showed a significant main effect. The descriptives and pairwise comparisons are shown in Figure 3.18 and 3.17.

As visible from the *control* and *acceptance* results, the strongest linear latency injection yielded to the lowest scoring of avatar control. The jitter condition was similarly affected, but resulted in significantly better ratings than the condition with a strong latency injection. The results of the condition with the lower level of linear latency injected had comparable results to the jitter-injected condition. In a similar manner, the synchronicity ratings were affected, but the jitter-injected condition was rated significantly lower than the condition with the lower linear latency. In all of the conditions, baseline condition performed best with regard to *acceptance*, *control*, and *synchronicity*.

Correlations of Constructs

To investigate the relations of the constructs, we assessed correlations between the dependent measures. The results are shown in Table 3.13. As expected, the synchronicity assessment showed great correlation with the *control* factor of the Gold-IVBO and the agency factor of the scale from Kalckert and Ehrsson [151] (KE). Furthermore, synchronicity showed a medium to large correlation with the *acceptance* factor, and the ownership (KE) factor.

Table 3.13: Significant Bivariate Pearson Correlations of Constructs for the Latency Manipulation

		Mov. Nat.	Mov. Synch.	Acc. (GI)	Cont. (GI)	Cha. (GI)	Agen. (KE)	A. Cont. (KE)	Owner. (KE)	O. Cont. (KE)
Mov. Real.	L_B	.788**	.675**		.576**		.712**			
	L_S	.690**		.477*	.498*		.674**			
	L_L				.489*					
	L_J	.648**	.523*	.533*	.676**		.678**		.492*	
Mov. Nat.	L_B		.512*	.449*	.508*		.551**			
	L_S			.494*						
	L_L									.446*
	L_J		.535*	.656**	.459*		.432*		.586**	.596**
Mov. Synch.	L_B				.574**		.797**			
	L_S			.609**	.664**		.540**		.438*	.501*
	L_L			.466*	.801**		.663**		.616**	
	L_J			.589**	.774**		.553**		.606**	.454*
Acc. (GI)	L_B				.484*			.519*	.843**	.712**
	L_S				.527*			.503*	.884**	.722**
	L_L				.442*				.901**	.624**
	L_J				.562**				.934**	.665**
Cont. (GI)	L_B						.744**			
	L_S						.706**		.473*	
	L_L						.831**		.581**	
	L_J						.794**		.530*	
Cha. (GI)	L_B									
	L_S							.478*		
Agen. (KE)	L_B							-.462*		
	L_S									
A. Cont. (KE)	L_B								.648**	.669**
	L_S								.681**	.732**
	L_L									.629**
	L_J									.692**
Owner. (KE)	L_B									.745**
	L_S									.793**
	L_L									.674**
	L_J									.653**

Note. * $p < .05$; ** $p < .01$; *GI Gold-IVBO; KE Kalckert & Ehrsson Measure.*

In turn, *acceptance* had medium to large correlations with control, and a large correlation with ownership (KE) and ownership control (KE) across all measures. The change factor did not show any significant correlations with the other measures.

It is interesting to note that we did not find stable correlations between the perceived naturalness, synchronicity, and realism of the movement, which may give room for further discussion on how to distinguish these constructs.

3.7.3 Discussion

In summary, our results present useful insights about the effects of simulation delays, and more specifically, the effects of jitter delays. The descriptive statistics and comparisons reveal that a jitter (in the applied spectrum) can result in worse effects than an increased latency of approximately 100 ms with regard to the perception of virtual body ownership, specifically *acceptance* and *control*. Yet, an additional artificial delay of about 250 ms performed worse than the jitter condition, which quantifies the impact of jitter to some extent. With respect to the correlating constructs, we can confirm that the Gold-IVBO scale picks up modifications in movement synchronicity. The results also confirm that, by nature of the questioning, the perceived realism correlated with the *control* factor. While we found correlations between the *control* factor and the *agency* and *ownership* factors (KE), variance analysis revealed that the Gold-IVBO scale was more sensitive to the present scenario. This may result from the fact that the (KE) questions were originally used in RHI scenarios. Finally, while the results confirm correlations between *acceptance* and *control*, we confirm that the *change* factor was not affected by modifications in the control or agency dimensions. One limitation of the study is that the agency control measure applied from the (KE) questionnaire had overall low reliabilities. In conclusion, we found significant impacts of latency on the control component of the proposed measurement instrument, and could quantify the impact of jitter with regard to linear latency. In the following, we will summarize the results of this chapter.

3.8 Summary

Summarizing the results from Study 1-7, the implemented manipulations revealed that a) the level of immersion, b) the level of humanness/personalization, and c) latency affected virtual embodiment in different dimensions, and thus the results answer *RQ1a: How do simulation properties affect virtual embodiment?*

3.8.1 Impacts of Appearance Realism and Personalization

In Study 1, the modification of avatar's appearance was shown to modify the level of acceptance as measured with the initial Alpha-IVBO scale. This was supported by Study 2, and Study 3, which confirms previous findings on texture and shape realism (e.g., [199], see [158] for a review). Study 5 did not explicitly support or contradict this finding.

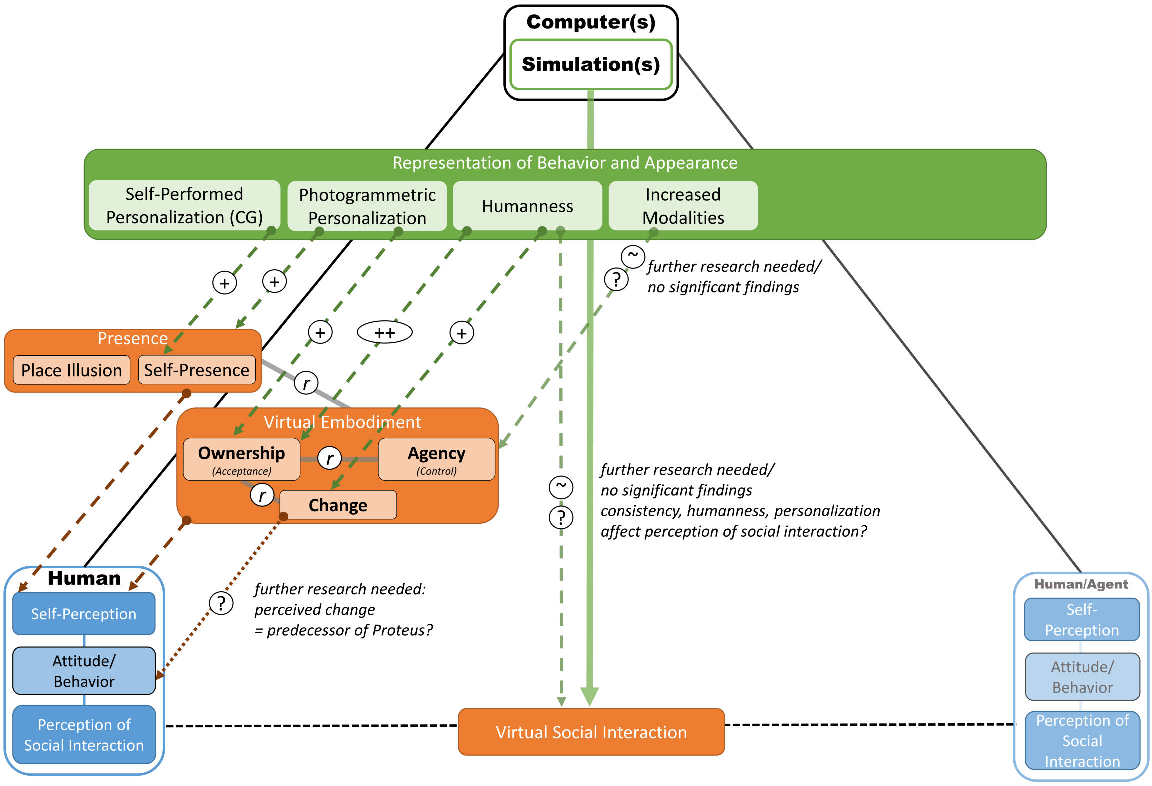

Figure 3.19: Summary of effects from representation of appearance and behavior on the Gold-IVBO factors and presence. Self-performed personalization and photogrammetric personalization increased presence. Photogrammetric personalization as well as humanness were found to positively impact the acceptance (ownership) of a virtual body. (Personalized) humanness did further affect the perceived change. None of the representation manipulations significantly affected agency. Presence and virtual embodiment seem to share some variance. The causality of these correlations cannot be explained by our study designs.

In contrast to Study 1 and 2, Study 3 used a method of constructing a computer-generated personalized avatar rather than a photogrammetry scanned (Study 2), or a facial-model–scan based avatar (Study 1) which may indicate a stronger impact with these procedures, as they were most likely to evoke a greater similarity than was achieved in Study 5 as assessed by the post-exposure manipulation check. Yet the post-exposure similarity measure positively correlated with *acceptance* and perceived *control*. While the construction of an avatar in the personalized condition of Study 5 yielded to a more positive affect and greater self-presence, a significant positive correlation was only observed between the proto-self–presence measure and the *acceptance* factor, and *not* between the level of perceived *acceptance* and positive affect.

The findings of Study 6 do not provide supporting evidence that the vividness of behavior affects the VBO. We attributed this partly to the limited exposure of the scenario which may have biased the study. Figure 3.19 summarizes the findings and still open questions.

3.8.2 Impacts of Immersion

Study 2 showed that the level of media immersiveness affected the perceived *acceptance* toward the virtual body, as well as the perceived self-perception (*change*) as assessed by the Alpha-IVBO measure. Study 4 confirmed the finding of a perceived *change*, but immersion did not have a significant impact on *acceptance*, which might be a result of the better construct separation of the improved Gold-IVBO measure.

The found impact of media immersiveness is in line with previous findings [142], which suggest the need for a spatial 3D representation, but extends the previous work further. All conditions used stereoscopic rendering of the virtual body, whereas the main difference was the visual frame of reference (physical vs. virtual), and thus points at an impact of *context* and *stimulation congruency*. An immersive simulation allows for natural SMCs. Both studies confirmed the impact of the medium on presence (i.e., more immersion resulted in more presence). The presence measures in Study 4 showed correlations to acceptance and change.

We therefore interpret the impact of immersion to specifically address the perceived change of the body schema through virtual embodiment. That is, the *display of a virtual body seen from a first person perspective, as in comparison to the physical body as reference, changes the perception of the own body schema*. The effect is most likely the result of shape discrepancies between the physical and virtual self. Yet, this perceived change is not perceived in low immersive setups with PR references.

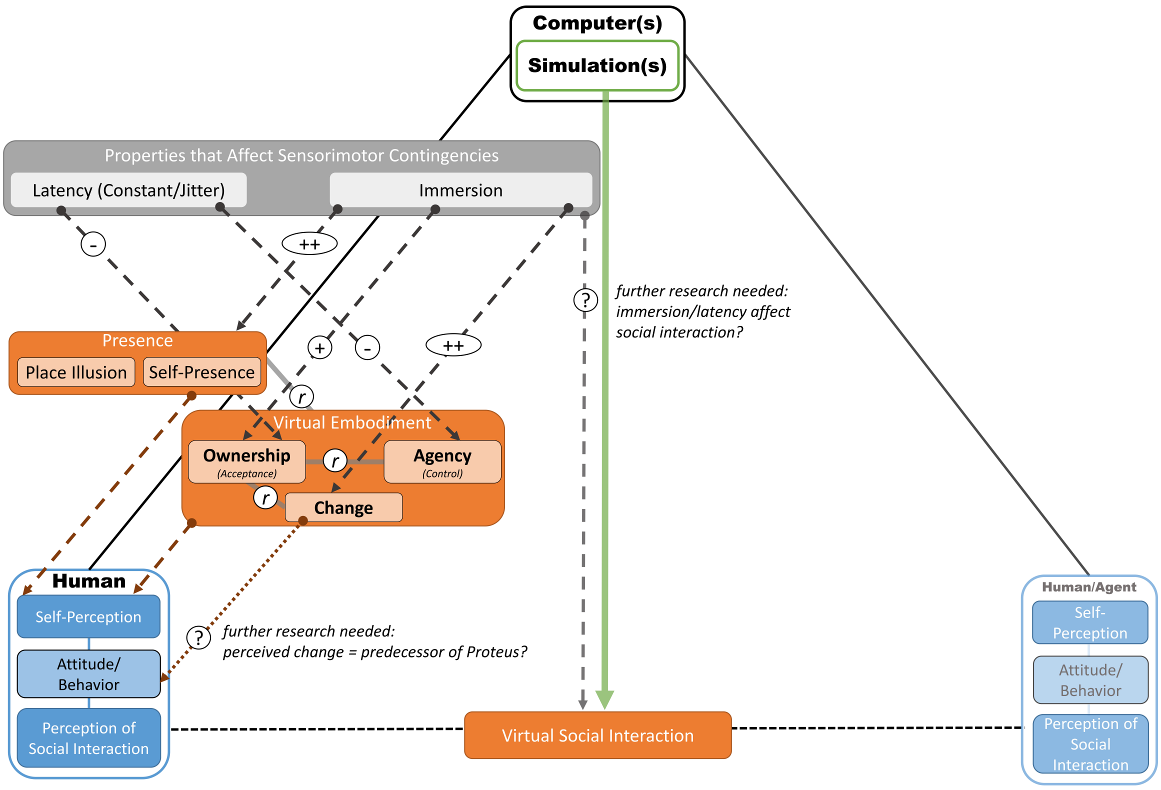

Figure 3.20: Model relating the findings for immersion and latency manipulations. Latency negatively affected agency (control) and ownership (acceptance). Immersion positively affected ownership and the perceived change in body schema. Immersion had a positive effect on presence.

3.8.3 Impacts of Simulation Latency

Study 7 did confirm previous findings on visuomotor coherence [350], in that a longer MTP delays affected the VBO. The results showed impacts on both the *acceptance* and *control* components. In addition to previous works, the study could partly quantify the impact of latency jitter. Jitter had a severe impact on the results, but did not yield a decrease in VBO perception similar to the decrease of the condition with the highest constant MTP latency (353 ms). This is in line with previous findings that suggested 300 ms is an upper limit for visuo-tactile feedback [293, 292].

Figure 3.20 summarizes the effects we found for immersion and latency modifications on virtual embodiment and presence, summarized under the umbrella term "sensorimotor contingencies".

3.8.4 Gold-IVBO Measure

Summarizing the scale-related analyses of Studies 4-7, the constructed improved measure seemed to successfully detect modifications of the simulation that affects its specific sub components. Its construction and validation answers *RQ1b: What latent variables are responsible for the adaptation of a virtual body?* We identified three latent variables of embodiment in the present work. As expected, *acceptance* and *control* showed correlations, whereas the *change* component did partly correlate with *acceptance* but was mostly seperable from *control* in the analyzed data. In our results, we found correlations between presence components and embodiment, for example with regard to the general presence assessment (Study 4) that correlated with acceptance and change, or the proto self-presence measure (Study 5) that correlated with all 3 Gold-IVBO dimensions. While we cannot clearly point out causalities between these concepts with the design of our studies, we interpret that aspects of presence are closely to virtual embodiment. Further, Study 5 showed a correlation between *control* and positive affect.

The *control* component responded strongly to the visiomotor manipulation in Study 7, and did show correlations to the corresponding agency and ownership measures of the scale proposed by [151], similar to *acceptance*.

3.8.5 Implications

From the presented results, we can draw a number of implications.

I1.1: A Higher Visual Immersion Fosters Body Scheme Changes

With our results we could show that by immersing the user into a virtual simulation that does suppress a reference frame to the physical world, components of virtual embodiment, especially the change of the own body scheme are fostered. This is especially important for applications that specifically address body scheme changes, such as therapy applications that aim to assess or treat eating disorders [239, 208] on the basis of body scheme assessments or manipulations. For such applications as well as applications that aim at utilizing the Proteus effect, an HMD-based rendering may be beneficial compared to a projection-based approach.

I1.2: Humanness and Photogrammetric Personalization Foster Ownership

Throughout the presented results we found supporting evidence that the humanness of the virtual character as well as a photogrammetric personalization foster VBO, as assessed by the acceptance measure and further validation measures. This may inform applications such as exposure therapy applications, immersive games, or social media applications. While a certain individualization (such as currently available in many games) yielded to increased aspects of self-presence, it may be necessary to provide a convincing reproduction and rendering of the physical self in the virtual world to achieve very high levels of ownership over a virtual representation.

I1.3: Increased Latency and Latency Jitter Negatively Affects Virtual Embodiment

Adding to previous research [350], fact that both, a higher latency as well as a latency jitter introduced to the simulation was found to especially impact the feeling of control stresses the overall importance of timeliness for applications that allow for virtual embodiment. The design and development of applications should consider this fact, especially in the context of motor performances or motor rehabilitation that support virtual embodiment.

3.8.6 Conclusion

In summary, we could show distinct effects of simulation properties on user embodiment, that could be assessed with the components *acceptance* (ownership), *control* (agency), and *change*. We showed that the proposed measure correlates with related constructs but results in a more sensitive measure and discrimination. Future work should further assess the internal consistency of the measure with regard to alternative manipulations and scenarios. Furthermore, the scale could be extended with an assessment of self-location or the perception of location discrepancies.

Chapter 4

Social Interaction and Interpersonal Effects of User Embodiment

Chapter 3 investigated the effects of virtual embodiment on the self, that is, how perception (ownership, agency, perceived change of the body scheme) is affected when embodied by an avatar. Consequently, regarding social interactions, one could argue that such user embodiment may also affect the perception of social interactions. The idea that immersive VR may become a medium for social communication is now widely investigated under the umbrella term SVR, which broadly describes the use of immersive media for social interactions. Differing from previous shared virtual environments, these applications in general allow for a greater degree of immersion, and in turn a more accurate coherence of the human senses between the physical and virtual worlds [297]. Most importantly, they allow for user embodiment. As described in Sections 1.3 and 2.2, immersion is "a quantifiable description of a technology. It includes the extent to which the computer displays are extensive, surrounding, inclusive, vivid and matching." [300, p.3]. Extending this formulation to the context of SVEs, Kanamgotov, Christopoulos, Conrad, and Prakoonwit [152, p.1] state that on the basis of their experimental results "immersion mostly depends on co-presence and communication of users."

Throughout the application space of immersive environments, one can agree that shared spaces and relationships to computer-supported collaborative work are broad areas that may benefit from improvements of realism aspects in order to achieve copresence. However, the transmission of communicative data takes time, and the appearance of the individual user is typically not matched with her/his virtual surrogate. Therefore, current applications are limited by three major factors: 1) limited *appearance-consistency* and *appearance realism* replicating the visual appearance of the physical user, 2) limited *sensorimotor coherence* because of the replication delay caused by sensor and simulation latencies, and 3) the *limited sensor and simulation capabilities to reproduce the modalities of nonverbal behavior* with high levels of realism and to full extent (for example, facial expression and gaze). In this regard, the latter point is of great importance because it affects the believability of simulation, the believability of the interaction partner (plausibility illusion), and the virtual embodiment of the self.

For example, in current SVR applications users typically navigate through the virtual world via interaction metaphors like teleportation triggered by input devices such as gamepads, motion controllers, or touchpads. Communicative nonverbal behaviors such as gestures, facial expressions, or gaze are either executed by additional commands such as button presses, replicated by forms of inverse kinematic approximations [270], or not replicated at all. With regard to communicational aspects, reduced behavioral realism specifically could affect presence [117], rapport [327], efficiency [35], or social understanding [40]. In consequence, social interactions performed through these avatar-based immersive environments might lack essential capabilities to replicate social behaviors needed for perceiving a social connection or for entraining and rhythmic bidirectional communication processes, such as the building of nonverbal rapport [205].

This chapter will, therefore, tackle this question by first comparing general differences in the perception of physical and virtual interactions, and then by systematically modifying the level of behavioral realism. It is guided by two overarching research questions:

> *RQ2a: How do virtual social interactions with user embodiment compare to physical social interactions?*
>
> *RQ2b: How do technological properties affect user-embodied virtual social interactions?*

Section 4.1 compares two typical social scenarios, a collaborative motor task and a negotiation task in the PR and in an immersive simulation that allows for user embodiment and the replication of body motion. Section 4.2 replicates the negotiation part of the experiment using an alternative design and extending the set of measures. Section 4.3 differentiates different levels of behavioral realism and their resulting impact on the social interaction.

4.1 Study 8: Physical vs. Virtual Worlds: Similarities and Differences

In Study 8 we aimed at investigating the boundaries of immersive user-embodied collaboration and how it affects social interaction in terms of physical collaboration and everyday activities. In repeated measures, participants were exposed to two interaction tasks in PR as well as in VR. In the VR simulation, the behavior replication was reduced to body motion. At the time of the study, there was no apparatus available to us to replicate the photogrammetric appearance of the participants. We therefore specifically used a simplified character (see Figure 4.2), to not provoke bias through artificial social information from appearance or static behavior (i.e., social information gleaned from visual appearance, such as clothes, static facial features, and static gaze). This approach allowed us to assess the importance of the nonverbal channels to be transmitted through subjective questioning. We designed a collaborative motor task (ball game) and a verbal negotiation task that were executed by participants in both the PR and as an immersive embodied VR simulation. We investigated the impacts on performance and presence aspects, as well as the subjective assessment of attentional focus.

4.1.1 Method

Design

This study was conducted in a two-factor mixed design comprising the between-factor *order* of scenario (PR first vs. VR first) and the within-factor scenario type (PR vs. VR). The order of scenario was randomized across the sample and balanced with regard to the gender of participants.

Apparatus

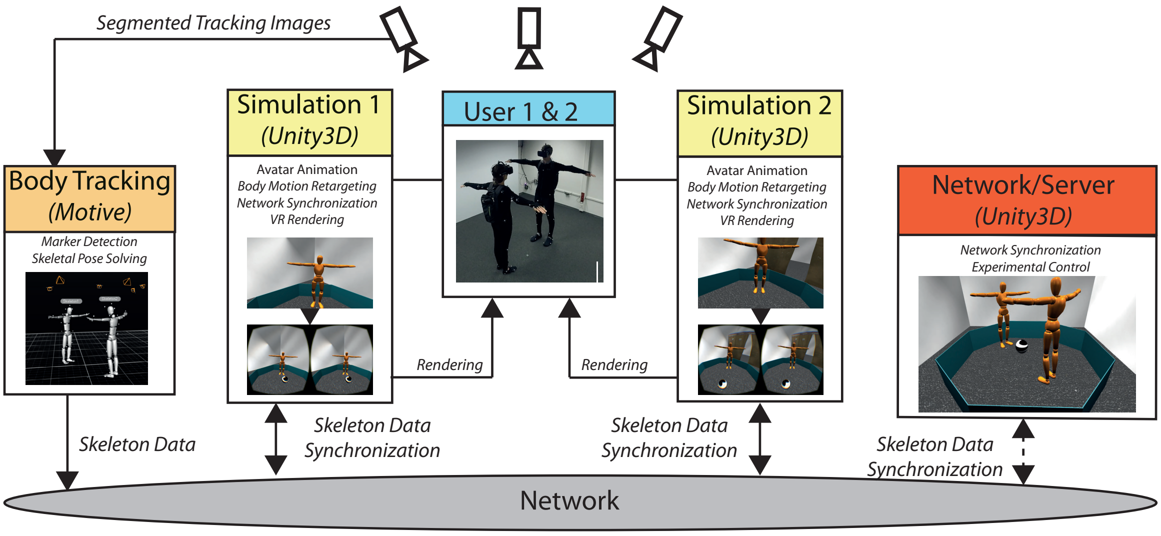

Figure 4.1: System diagram of the VR simulation. The user's body motion was tracked by the tracking system and distributed as skeleton data via network to each client and the server. Each user carried a backpack with a laptop that rendered individual simulations to an Oculus Rift DK2 HMD. The network server served for the experimental control and the synchronization of interactive objects.

Similar to the studies described in the beginning of Section 3.5, we used an optical infrared passive-marker tracking system[31] (OptiTrack Flex3, 16 cameras, 120 Hz max.) with a 37 marker set to perform motion tracking. Figure 4.1 depicts the system principle. VR immersion was achieved with two Oculus Rift DK 2[32] (960px × 1080px per eye, 100 deg FOV). To render the simulation, each participant was equipped with a backpack carrying a 13" Alienware[33] laptop (i75500U@2.40 Ghz, 8 GB memory, GeForce GTX 960M) that was connected to a wireless access point to receive data from the tracking server. The tracking data (acquired in 100 Hz) in the form of the resulting skeleton joint transformations was streamed using NatNet SDK[34] via local area network and wireless network by a separate tracking server machine running Optitrack Motive[35] that served as tracking server. A simulation server running

[31] NaturalPoint, Inc. DBA OptiTrack, `https://optitrack.com/products/flex-3/`
[32] Facebook Technologies, LLC., `https://www.oculus.com/`
[33] Dell GmbH, `https://www.dell.com/de-de/gaming/alienware`
[34] NaturalPoint, Inc. DBA OptiTrack, `https://optitrack.com/support/software/natnet-sdk.html`
[35] NaturalPoint, Inc. DBA OptiTrack, `https://optitrack.com/products/motive/`

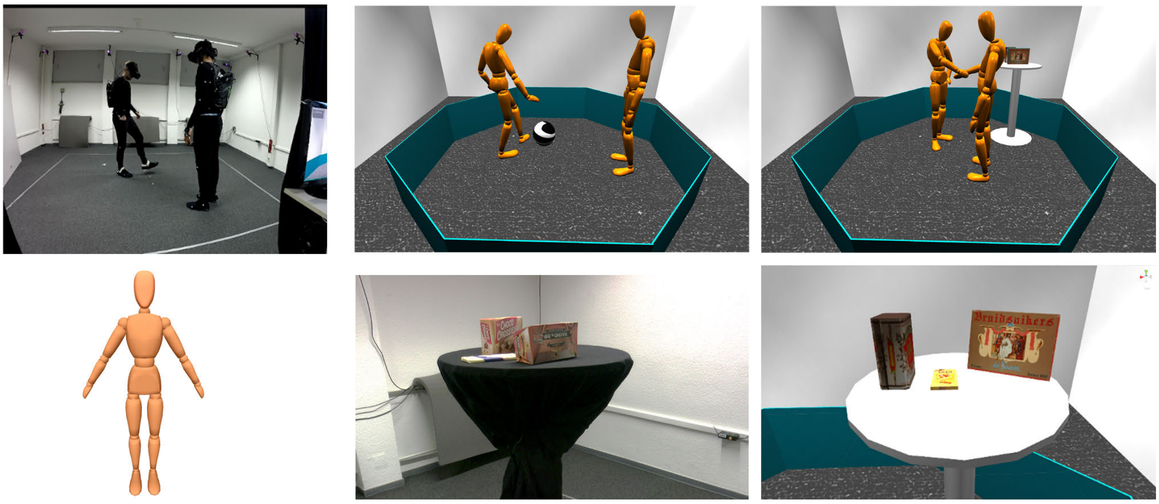

Figure 4.2: Scenario and materials. *Top left:* Physical scenario. Participants immersed in the simulation performing the ball game task. *Top center:* Simulation screenshot of the ball game task, third-person perspective (reprinted from [271] ©2016 IEEE). ***Top right.*** Simulation screenshot of the negotiation task (reprinted from [271] ©2016 IEEE). *Bottom left:* Avatar that was used in the study. *Bottom center/right:* Physical and virtual table with interactive objects.

the Unity3D[36] application was employed to synchronize interactive objects and for experimental control. The simulation scene consisted of approximately 250000 triangles. A soccer ball, along with a table with interactive objects (see Figure 4.2) that were present in the tasks were physical, meaning that they obeyed approximate laws of physics when interacted/collided with. Speech was not mediated. We used a neutral wooden mannequin as the avatar to embody each participant. Each participant could see their own movements as well as their partner's behavior. The mannequin was scaled in a universal fashion to the height of the participant in order to approximately fit her/his measurements.

We conducted a latency measurement of the overall end-to-end latency (user movement to laptop display reproduction, 240 Hz camera), as it has been argued to have an impact on cooperative performance [230]. Using a video frame-counting method similar to the one described by He, Liu, Pape, Dawe, and Sandin [134] to analyze the apices of repetitive user movements resulted in an average latency of 128.75 ms (67.83 ms motion capture system, 2.0 ms LAN, 16.67 ms to experimental server display, 4 ms WLAN, and an additional 38.33 ms to client/laptop display). The real room size was about 25 m^2 (11.5 m^2 active tracking area). For the VR simulation, we introduced a "virtual safety area" to prevent participants from accidently colliding with the room walls. The space for free movement was approximated to 9.3 m^2.

Procedure

Participants were welcomed and informed about the study. After the participants signed the consent form for participation, we guided them to dressing rooms where

[36] Unity Technologies, `https://unity3d.com/`

they could put on the motion capture suits. Participants then filled out the pre-study questionnaire, and we then assigned the participants to one of the condition orders. If participants were assigned to the VR first condition, they filled out the pre-exposure sickness assessment and then were equipped with the backpacks and HMD and calibrated for tracking. Participants started with the ball game. Their task was to pass the virtual soccer ball 10 times in three repetitions, using their non-intuitive foot (i.e., opposite their dominant hand, see Figure 4.2). During the task, video recordings were made and the experimenter counted the number of errors as well as measured the time to finish the second and third trials (the first was a learning trial). Before each trial, participants were recalibrated using a T-pose.

After finishing the ball game task, participants were given a short break and were provided with written instructions about the verbal negotiation task. One topic considered the was the selling/buying of a used TV (buyer vs. seller), and the other topic dealt with the pickup point of a shared ride in a big city (driver vs. fellow passenger). The topics were randomly distributed to the conditions which were balanced throughout the study. The roles were described in detail and some arguments were provided to the participants (e.g., would buy/sell TV for a certain price, would rather have a pickup point close to home). Participants had a maximum of five minutes for negotiation). After the role-playing, participants answered the questions on the digital post questionnaire (including the post-sickness assessment). All dependent questions specifically addressed the role-play. Next, the participants performed the same tasks in PR, without HMDs but wearing motion capture suits and backpacks, and then answered the same post questionnaire without the sickness assessment. Depending on either of the scenario orders, the procedure was vice versa. Finally, we debriefed the participants, answered remaining questions, compensated them, and released them from the study.

Measures

Performance We assessed objective measures for both tasks. In the collaborative motor task, we measured the time to finish the task and the number of errors made during the each ball game trial. Trial 1 was set as learning trial, and trials 2 and 3 were analyzed. With regard to the verbal negotiation task, we assessed whether a consensus was reached and, if so, the time to reach it, in addition to counting how often the participants interacted with the environment (i.e., table, objects).

Presence We assessed copresence using the network minds questionnaire [40, 131] after each scenario, with the subscales attentional allocation (e.g., "I remained focused on [my partner] throughout our interaction", $\alpha s \geq .67$), perceived message understanding (e.g., "It was easy to understand (my partner)," $\alpha s \geq .84$), perceived affective understanding (e.g., "I could tell how (my partner) felt," $\alpha s \geq 73$.), perceived behavioral interdependence (e.g., "The behavior of the other was a direct response to my behavior" $\alpha s \geq .84$). A Likert-type scale was used (1–does not apply at all, 7–fully applies).

Further, we included the questionnaire from Nowak and Biocca [220] that was initially constructed in an appearance and co-existence context and includes the factors self-reported copresence (e.g., “I was interested in talking to my interaction partner,” $\alpha s \geq .67$), perceived other’s copresence (e.g., “My interaction partner created a sense of distance between us” [reversed item], $\alpha s \geq .86$). A Likert-type scale was used (1–do not agree, 5–agree).

Focus of Attention To assess how participants changed their focus, we asked them “To which behaviors did you pay the most attention during the conversation (role-play)?.” In the VR scenario, we further assessed: “Which behaviors of your communication partner that have not been transmitted through VR did you miss the most in order to assess your conversational partner (concerning the role-play)?.” Participants had to answer by marking checkboxes (gesture, facial expression, body movement, speech, gaze, others–yes/no).

Simulator Sickness To control for a potential bias, we assessed simulator sickness [156] before and after the VR sequence ($\alpha s \geq .69$). The pre-study questionnaire further assessed demographic variables and media use.

Participants

We excluded participants in the case of tracking or if a simulation malfunction that significantly disturbed the study procedure occurred. The final sample for the analysis consisted of $N = 36$ participants (20 female, 16 male, $M_{age} = 25.25$, $SD_{age} = 4.64$), of which 30 were students at the time. None of the participants knew each other personally, although one pair of participants reported that they knew their partner from sight. Three participants had previous experience with VR technology, and one participant had used an HMD before. All participants reported speaking German for five or more years. Participants also reported exercising or playing sports regularly, averaging an equivalent of 2 to 2.5 hours per week.

4.1.2 Results

Due to the dyadic structure of the data, we assessed the need for a multi-level model and calculated goodness of fit criteria (Akaike information criterion and Bayesian information criterion, respectively). We tested for changes in the likelihood ratio for the presence variables as proposed by [91, p. 878]. The results did not show significant improvements in the fit of the model ($p > .05$), in which case, Field, Miles, and Field [91] recommend following a regular analysis procedure. We assessed the effects by calculating two-factor mixed ANOVAs with the *scenario order* (VR first vs. PR first) for between-factor and the *scenario type* (VR vs. PR) for within factor.

Simulator Sickness The results of a mixed ANOVA revealed no indication of a significant increase in instance of simulator sickness between the assessment before ($M = 33.45$, $SD = 38.57$) and the assessment after ($M = 38.54$, $SD = 31.59$) the VR exposure, $F(1, 34) = 1.968$, $p = .17$, $\eta_p^2 = .055$. A few participants verbally reported minor sickness during the collaborative motor task, but not to a severe level.

Table 4.1: Estimate Marignal Means and Univariate Main Effects for Scenario Type.

	$PR^{\dagger}$	$VR^{\dagger}$	F(1,34)	p	ηp^2
Attentional allocation	5.85 (0.13)	4.94 (0.21)	17.08	***	.33
Message understanding	5.77 (0.13)	5.08 (0.19)	18.91	***	.36
Affective understanding	5.01 (0.15)	4.29 (0.18)	12.45	.001	.27
Behavioral interdependence	5.26 (0.17)	4.92 (0.19)	4.14	.0496	.11
Self-reported copresence	3.51 (0.12)	3.37 (0.11)	1.40	.246	.04
Perceived other's copresence	3.82 (0.94)	3.62 (0.11)	4.06	.052	.11
Ball game errors	0.91 (0.21)	3.28 (0.35)	53.95	***	.64
Ball game time	24.63 (0.65)	85.88 (4.98)	157.45	***	.84
Negotiation time	161.46 (8.29)	162.88 (9.53)	0.05	.83	.00

Note. † *Mean* ($\pm$ *SEM*); *** $p < .001$.

Performance With regard to the ball game task, a mixed ANOVA revealed a significant interaction effect between the scenario order and the scenario type $F(1, 30) = 6.314$, $p = .018$, $\eta_p^2 = .174$. Overall, the participants made more errors (passing with the wrong foot, touching the ball more than once) in the VR scenario (see Table 4.1), but the difference was greater when the PR scenario was first in order ($M_{PR} = 0.44$, $SD_{PR} = 0.62$; $M_{VR} = 3.63$, $SD_{VR} = 1.54$) compared to when VR was first ($M_{PR} = 1.38$, $SD_{PR} = 1.54$; $M_{VR} = 2.94$, $SD_{VR} = 2.26$). Participants also needed more time to finish in VR ($M = 85.88$ s, $SD = 28.19$ s) compared to PR ($M = 24.63$ s, $SD = 4.00$ s).

Regarding the verbal negotiation task, two couples in PR as well as VR did not find a consensus in time (five minutes), limited to the first scenario in sequence. An interaction showed that when participants started in the PR condition, it took them longer to negotiate in PR ($M = 181.5$ s, $SD = 48.54$ s) compared to VR ($M = 151.63$ s, $SD = 50.44$ s) whereas vice versa in the alternative order comparing PR ($M = 141.23$ s, $SD = 41.22$ s). Overall however, there was no significant difference between the scenario types regarding the negotiation measures. Table 4.1 depicts the univariate main effects for scenario type. We did not find any significant impacts of the order of scenario and did not observe any further significant effects.

Presence Mixed-ANOVAs revealed that all measures for social presence assessed by the networked minds questionnaire were significantly affected by the difference of the scenarios. Table 4.1 depicts the univariate main effects for scenario type. The order of scenario showed a significant main effect for self-reported copresence $F(1, 34) = 7.632$, $p = .009$, $\eta_p^2 = .183$. Pairwise comparisons showed that when participants experienced PR first, the difference between ratings of self-reported copresence in PR ($M = 3.16$, $SD = 0.68$) was significantly lower than when they started with the VR scenario first ($M = 3.86$, $SD = 0.73$; $p = .005$). We did not observe any further significant effects.

Focus of Attention Table 4.2 summarizes the results of the assessment regarding attentional focus. The results show that more focus in VR was set on speech and body motion. However, the difference was much greater in the body motion

Table 4.2: Results of the Cues in Focus Assessment.

	What cues did you focus on?		*What cues have you been missing most?*
Cue	**PR**	**VR**	**VR**
Gesture	17	16	9
Facial expression	24	(0)	31
Body movement	10	23	4
Speech	24	28	4
Gaze behavior	26	(0)	31
Others	0	1	1

Note. Comparisons as indicated by checkbox items.

($PR = 10$, $VR = 23$) compared to the speech category ($PR = 24$, $VR = 28$), indicating a potential shift of attention toward body motion in VR. To quantify the differences, we calculated related-samples McNemar change tests [203] for the measures. The test revealed a significant difference in the frequency that participants reported special attention on body movement (exact $p = 0.07$; two-sided). Out of 36 participants, 10 reported to have paid special attention to body movement in the PR scenario, whereas 23 participants reported having done so in the VR scenario. The comparisons for speech and gesture were not significant.

4.1.3 Discussion

The results of the study suggest that the affected level of visiomotor coherence (that is, less coherence in the VR condition due to the delay between motor action and visual feedback–latency) significantly affected the motor task performance but did not affect performance in the verbal negotiation task. One interpretation is that these types of social interaction (i.e., "functional" negotiations) strongly depend on speech.

The reduced level of behavior realism (as well as appearance realism) based on the reduction of nonverbal channels replicated to the VE was most likely the cause of reduced social presence in VR. The VR negotiation role-play was negatively affected with regard to the judgment of attentional allocation, perceived message understanding, affective understanding, and behavioral interdependence. These differences were visible in the study results regardless of the same audio transmission (local, physical conversation).

The assessment of the focus of the participants on certain behavior cues showed a significant increase in the focus on body movement by the participants in the VR scenario. Two interpretations arise from this result. First, it could be a mere effect of the availability of the cues, as less cues were present in VR due to the reduced avatar animation. However, the lack of difference with regard to the attention on gestures speaks against this interpretation. Therefore, this result might also show a certain compensatory mechanism for missing behaviors, as proposed by the social information processing theory [351].

One limitation of the study may be the within-subjects design we applied. As seen by the effect for self-perceived copresence, the order of scenario affected this measure. However, one initial argument was that social interaction is highly individual, and therefore mixing different individuals to groups may have had an even stronger bias. Nevertheless, the next Study that is presented in Section 4.2 considers a between-subjects design.

In conclusion, the study showed that motor performance was reduced in VR as was the overall quality of communication with regard to social presence. However, the functional negotiation outcome did not suffer from the restrictions of the virtual environment. Further, we found indications for a shift in focus regarding the attention toward nonverbal behaviors, namely, participants focused more on body movements in the VR condition, where facial cues or gaze were not present.

4.2 Study 9: Perceived Efficiency and Impact on Affective Perception

In the second study, we replicated a similar setup and approach to the one described in 4.1, and replicated the study, introducing additional measures and stronger controls to the procedure. As several aspects could have been biased by the previous procedure, such as a cross speak from the collaborative motor task on the role-play, we decided to perform a between-subjects measure and further assess more specifically the potential deficits or benefits arising from such interactions.

4.2.1 Method

Design

The study was conducted in a one-factor (PR vs. VR) between-subjects design. In contrast to the previous study, we exclusively assessed the verbal negotiation task. We decided to assess a group comparison in order to clarify whether differences in presence aspects are stable if no direct comparison is provided to the participants.

Apparatus

The apparatus (see Figure 4.3) used was similar to the one described in Section 4.1.1. We performed minor adjustments to the virtual environment and removed the safety area because no motor collaboration task was used in the procedure.

Procedure

Participants were welcomed and informed about the study, and then they signed the consent forms agreeing to participate. After the participants used separate dressing rooms to change their clothes and put on the motion capture suits (both conditions), they answered the pre-study questionnaire in digital form, and were handed written instructions about the role-play. The role-play topic (buying/selling a used TV) was identical for both conditions. After participants indicated they were ready to start, the experimenter positioned them in the tracking environment and they were

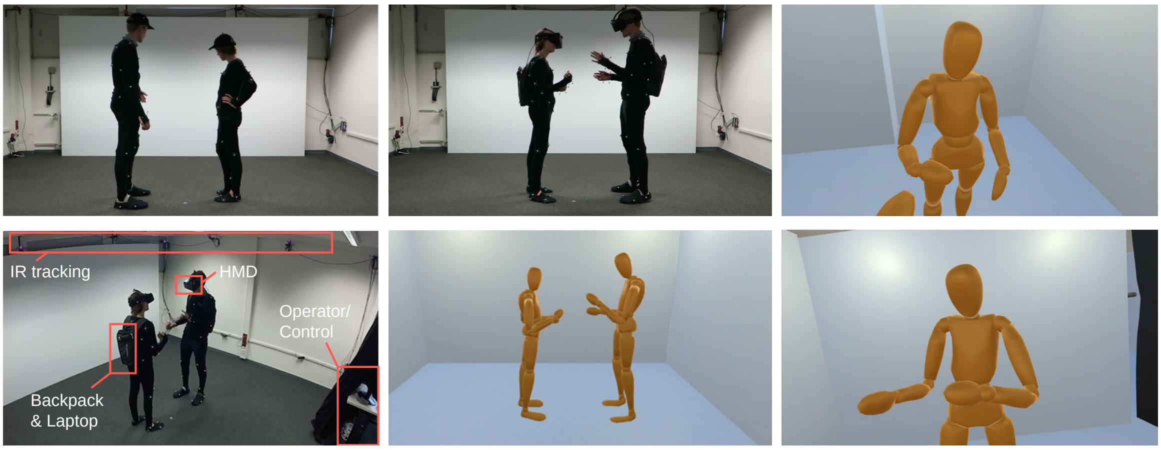

Figure 4.3: Apparatus and adaptations. *Top left.* PR scenario. *Bottom left.* Apparatus. *Top/Bottom center.* Physical scenario in the VR condition and virtual simulation from a third-person perspective. *Top/Bottom right.* First-person perspectives.

calibrated for the motion tracking (both conditions) and for the simulation (VR condition). In the VR condition, the simulation was then started, and participants were given a short acclimatization phase for VR. Following the acclimatization, they began the role-play. If no consensus was found, the experimenter stopped the roleplay after five minutes. Next, the participants were helped to remove the equipment and asked to answer the post-experimental questionnaire in digital form. Finally, we answered any remaining questions and then compensated the participants and dismissed them from the study.

Measures

Presence We assessed *social presence* as a presence factor that was affected in the previous study. Again we assessed the subfactors of the networked minds questionnaire [40, 131]: attentional allocation (Cronbach's $\alpha = .59$), perceived message understanding ($\alpha = .81$), perceived affective understanding ($\alpha = .80$), and perceived behavioral interdependence" ($\alpha = .77$). We investigated *copresence* using the scale from Nowak and Biocca [220] that assesses self-reported copresence ($\alpha = .70$), and perceived other's copresence ($\alpha = .85$).

Affect We further introduced a measure for affect, by using the self-assessment mannequin [51] that assesses arousal, valence, and dominance on the basis of a visual scale. Instead of a self-rating, we asked participants to *judge the other person*, meaning at what level would they ascribe these dimensions to their partner when reflecting the role-play. Each self-assessment mannequin image represented one scale point (1–low, 5–high).

We further asked participants to report what level they would ascribe basic emotions (fear, happiness, anger, sadness, surprise, disgust, contempt; [84]) to their communication partner when reflecting the role-play (-3 – does not apply at all, and 3 – totally applies; for the analyses, the data was recoded to 1-7 values to ease the interpretation).

Behavior Perception and Attentional Focus We asked the participants to assess their *attentional focus*: "On which behaviors did you focus most during the interaction?" using a 7-point scale (-3 – not at all focused on, 0 = neither/nor, +3 – very strongly focused on. For the analyses, the data was recoded to 1-7 values to ease the interpretation). Items to be rated were: gesture, facial expression (mimics), body movement, speech, eye gaze, body posture, body contact, and proximity. The measure was slightly adapted from the previous study to gain more insight into the specific adaptations.

In the VR condition, we also asked *which behaviors participants missed* most: "Which behaviors that did not get transmitted using the VR condition did you miss most in order to assess your communication partner and his/her statements?." The ratings were presented using a 7-point scale (-3 – not missed at all, 0 neither/nor, +3 – missed very much. For the analyses, the data was recoded to 1-7 values to ease the interpretation). Behaviors listed were gestures, facial expression (mimics), and eye gaze.

Furthermore, we added a subjective assessment of the *communication partner's behavior* during the role-play, with the questions to be answered on a 5-point scale (1 – do not agree at all, 3 – neutral, 5 – fully agree): "My communication partner: (a) had continuous eye contact with me, (b) fostered verbal conversation, (c) was restless, (d) had very open body posture, (e) described much by words, (f) had a stiff body posture, (g) moved a lot during the conversation, (h) gesticulated a lot."

Negotiation Performance As objective measures, we assessed whether or not subjects reached a consensus, the time it took them to do so, and the final price negotiated.

Communication Effectiveness We adapted the questionnaire on communication quality/effectiveness from Bente, Rüggenberg, Krämer, and Eschenburg [35] in order to assess the participants' evaluation of the communication. The 21 items were aggregated to the four factors of communication effectiveness as stated in [35]: Satisfaction with outcome ($\alpha = .81$), clarity of partner's contributions ($\alpha = .70$), competent impression ($\alpha = .78$), and relevance of partner's contributions (measure was discarded due to low reliability, $\alpha = .31$).

Simulator Sickness We assessed simulator sickness based on the simulator sickness questionnaire (SSQ) [156] (αs between $\alpha = .39$ and $\alpha = .88$).

In addition, we measured demographics, body ownership, immersive tendency, and a basic emotion rating. The latter three are not subject to the present discussion and were assessed for secondary meta-comparison.

Participants

From the recruited sample, we discarded participants when severe tracking errors occurred that disrupted the interaction or when the task was not properly understood. The final sample for the analysis consisted of $N = 58$ participants (30

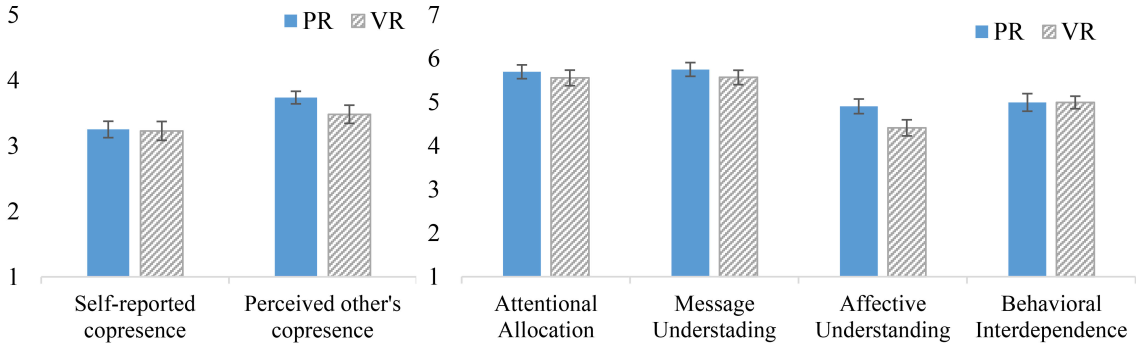

Figure 4.4: Comparisons for the copresence measures and the networked minds measures of social presence. *Note. Bars denote the mean value; error bars denote the standard error.*

female, 23 male, $M_{age} = 24.16$, $SD_{age} = 4.84$), of which 53 were students at the time. All participants had normal or corrected to normal vision. The participants were blind to the experimental goal and did not know each other personally, although four pairs reported that they knew each other from sight. The participants were randomly assigned to either the PR condition ($N = 28$ of, which 14 were female and 14 male), or to the VR condition ($N = 30$ of, which 16 were female and 14 male). All participants were fluent German speakers, although one participant indicated that he/she had been speaking German only for three years. In the VR condition, 10 participants indicated they had previous VR experience.

4.2.2 Results

We assessed the need for a multi-level model and calculated goodness of fit criteria (Akaike information criterion and Bayesian information criterion, respectively). We tested for a change in the likelihood ratio for the social presence, copresence, affect, and communication effectiveness variables following the procedure proposed by Field, Miles, and Field [91, p. 878]. Two factors (perceived other's copresence, and clarity of partner's contributions) showed that the model would significantly benefit from a multi-level analysis (ps$> .05$), but 11 factors did not show significant improvements in the fit. For the sake of consistency, we performed regular analyses. We calculated t-tests where normality could be assumed and non-parametric tests where normality was violated (assessed by Shapiro-Wilk tests). Figure 4.4 depicts the descriptive results.

Presence We found no significant differences in the networked minds measures for social presence or in the co-presence measures. Figure 4.4 depicts the descriptive results. Perceived affective understanding was affected the strongest, comparing the PR condition ($M = 4.90$, $SD = 0.89$), and the VR condition ($M = 4.41$, $SD = 1.01$), $t(56) = 1.97$, $p = .054$.

Affect Analyzing the affect the participants attributed to their communication partner, Mann-Whitney U tests showed significant differences for the valence and

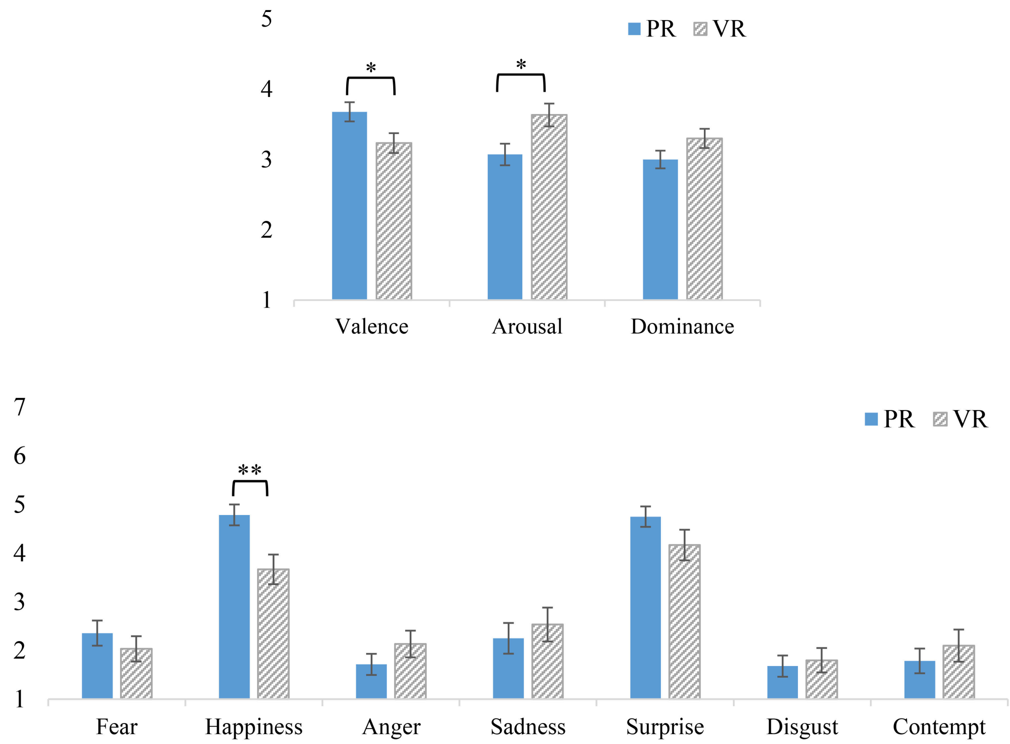

Figure 4.5: Comparisons for the affect attributed to the communication partner. *Top:* Valence, arousal, and dominance perception of the communication partner. *Bottom:* Attribution of the basic emotions to the communication partner. *Note.* * $p < .05$; ** $p < .01$; *bars denote the mean value; error bars denote the standard error.*

arousal dimensions (distributions were similarly shaped). Median scores for valence showed more valence in PR ($Mdn = 4.0$) compared to VR ($Mdn = 3.0$); $U = 294$, $z = -2.153$, $p = .031$, thus they rated the partner more positive. This was confirmed by a significant impact on the level of happiness they attributed to their communication partner in PR ($Mdn = 5.0$) compared to VR ($Mdn = 4.0$); $U = 248$, $z = -2.782$, $p = .005$. The level of arousal was higher in VR ($Mdn = 4.0$) compared to PR ($Mdn = 3.0$); $U = 564$, $z = 2.381$, $p = .017$, indicating that participants had a higher excitation perception of their partner in the VR condition. Figure 4.5 depicts the descriptive results.

Behavior Perception and Attentional Focus The subjective perception of the other person's behavior did not significantly differ. Figure 4.6 depicts the descriptive results. Based on the VR assessment, participants missed facial expression the most ($M = 6.15$), followed by gaze ($M = 5.62$), gestural details ($M = 3.88$), and others ($M = 0.77$).

Objective measures for task performance In each condition, 12 dyads reached consensus. Neither time to reach consensus, nor final price significantly differed.

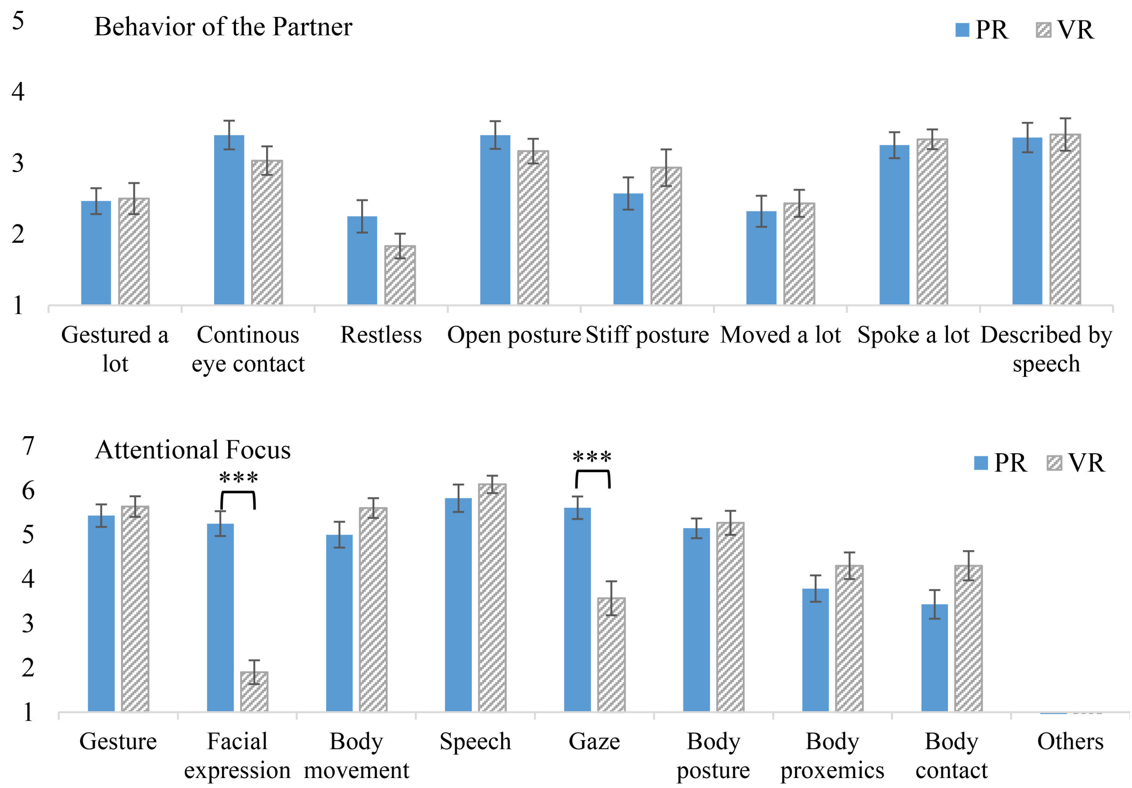

Figure 4.6: Comparisons for the nonverbal behavior dimensions attributed to the communication partner's overall behavior and the reported focus of attention (Study 8). *Top:* Responses to the question which behaviors the communication partner executed. *Bottom:* Responses to the question asking to what degree the participant focused her/his attention to a specific cue. *Note. *** $p < .001$; bars denote the mean value; error bars denote the standard error.*

Communication effectiveness Although task performance was equal for both conditions, a Mann-Whitney U test for the factor "satisfaction with outcome" showed a significant difference between PR ($m_{Rank} = 37.12$) and VR ($m_{Rank} = 22.38$); $U = 206.5$, $z = -3.329$, $p = .001$; insofar, that PR was rated more satisfactory. Figure 4.7 depicts the descriptive results for the communication effectiveness responses.

Simulator sickness We calculated the total sickness score by the sum of the subscores and weighted this sum with the factor proposed by Kennedy, Lane, Berbaum, and Lilienthal [156]. A Wilcoxon signed rank test revealed that people felt significantly sicker after the VR simulation ($M = 40.64$, $SD = 6.55$), than before ($M = 21.44$, $SD = 3.55$) in the total sickness measure ($z = 2.83$, $p = .005$).This may indicate a potential bias by the simulation. Yet, none of the participants reported severe sickness or had to abort the experiment.

We did not observe any further significant effects.

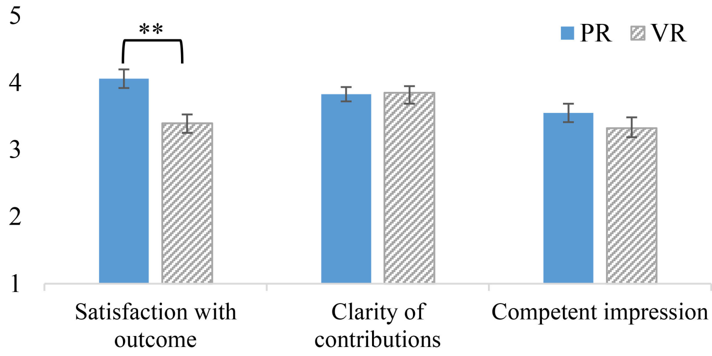

Figure 4.7: Comparisons for the perceived effectiveness. *Note. ** $p < .01$; bars denote the mean value; error bars denote the standard error.*

4.2.3 Discussion

The present study replicated the negotiation task of Study 8 and we slightly adapted the design of the experiment and apparatus. Despite higher ratings in PR, the results indicated no significant differences in presence aspects. Two possible interpretations of the lack of differences in the social presence factors could be proposed: 1) Participants did not have a direct comparison to the PR scenario, and 2) participants interpreted social information by compensating for missing behaviors via other channels, as proposed by the social information processing theory [351, 352]. As in the previous study, the ratings of the focus behavior indicate so, but do not support this assumption based on a level of statistical significance. While the results do not contradict the findings of the previous study and the mean values show a similar trend, a failure to observe significant differences on social presence and attentional focus might have occurred because of the limited sample size. However, they might also be explained by the fact that participants did not have a direct comparison to their communication partner initially (as opposed to the within-subjects design of Study 8). In the study procedure, they were only allowed to quickly introduce themselves (i.e., saying hello) but did not share any communicative behavior upfront. The result is also in line with a recent study by Smith and Neff [309], which found differences in social presence only for non-embodied VR compared to face-to-face conversations, but not for the comparison of face-to-face versus embodied VR. Overall, the affect ratings show a negative impact of the VR simulation on the judgment of the communication partner. We attribute this mainly to the difference appearance with regard to the virtual character representing the partner. For example, important facial features were missing to infer the emotional state from the partner [265], and thus a general negativity bias [280] might explain these negative ratings. However, the finding is important as it stresses the fact that the appearance of a character matters in regard to how the controlling user is perceived. The fact that the participants perceived communication as less efficient in VR because they were less satisfied with the outcome might have been affected by the negative impression. Although in terms of numbers, their performance did not differ, their impression did.

Besides a restricted sample size, a limitation of the study is that the results may have been biased by the increase of perceived sickness. In comparison to Study 8 where a physical task (more dynamic motion) may have reduced these effects [143].

In conclusion, the study shows that the VR condition did, in fact, have a negative impact with regard to affect and perceived, but not objective, efficiency. The results regarding social presence impacts do not contradict the findings of Study 8. With regard to perceived behaviors and social cues, further work should explore different levels of transmitted behavior, which is the target of the next study.

4.3 Study 10: Impacts of Behavioral Realism

In a third study that investigated virtual social interaction, we constructed an alternative apparatus to transmit facial cues and placed the users in separate rooms. They were equipped with sensors that allowed our simulation to replicate body motion, facial expression, and gaze. We decided to have an introductory conversation as rationale, and varied the behavioral DOF to see how additional modalities affected presence and the perception of the character.

The study aimed at investigating more precisely how behavioral realism, specifically the number of modalities transmitted, affect virtual encounters.

4.3.1 Method

Design

In a one-factor between-subject study, we investigated three degrees of *behavior modalities*. Each participant saw either only the body movement (B) of the interaction partner, only the facial expressions (including gaze, F) of the interaction partner, or a combination of both (BF). To accomplish this, we developed an avatar-mediated communication platform described in the following.

Apparatus

The developed platform consisted of two remote optical infrared passive-marker tracking systems[37] (OptiTrack Flex3, eight cameras each, 120 Hz) with a 37 marker set to perform body motion tracking. We further interfaced Carmine 1.09 short range RGB-depth sensors[38] with facial expression and gaze tracking performed by faceshift[39] run on laptops in each room (Alienware, i75500U@2.40 Ghz, 8GB memory, GeForce GTX 960M). To increase the robustness of the facial and gaze tracking, we attached an additional LED light stripe to the steady-shot camera rig that carried the RGB-depth sensor (see Figure 4.9). The rig was worn by the participants in front of their belly, similar to a backpack, to best allow for freedom

[37] NaturalPoint, Inc. DBA OptiTrack, `https://optitrack.com/products/flex-3/`
[38] PrimeSense, Israel, acquired by Apple Inc. in 2013, `https://www.apple.com/`
[39] Faceshift AG, Switzerland, acquired by Apple Inc. in 2015, `https://www.apple.com/`

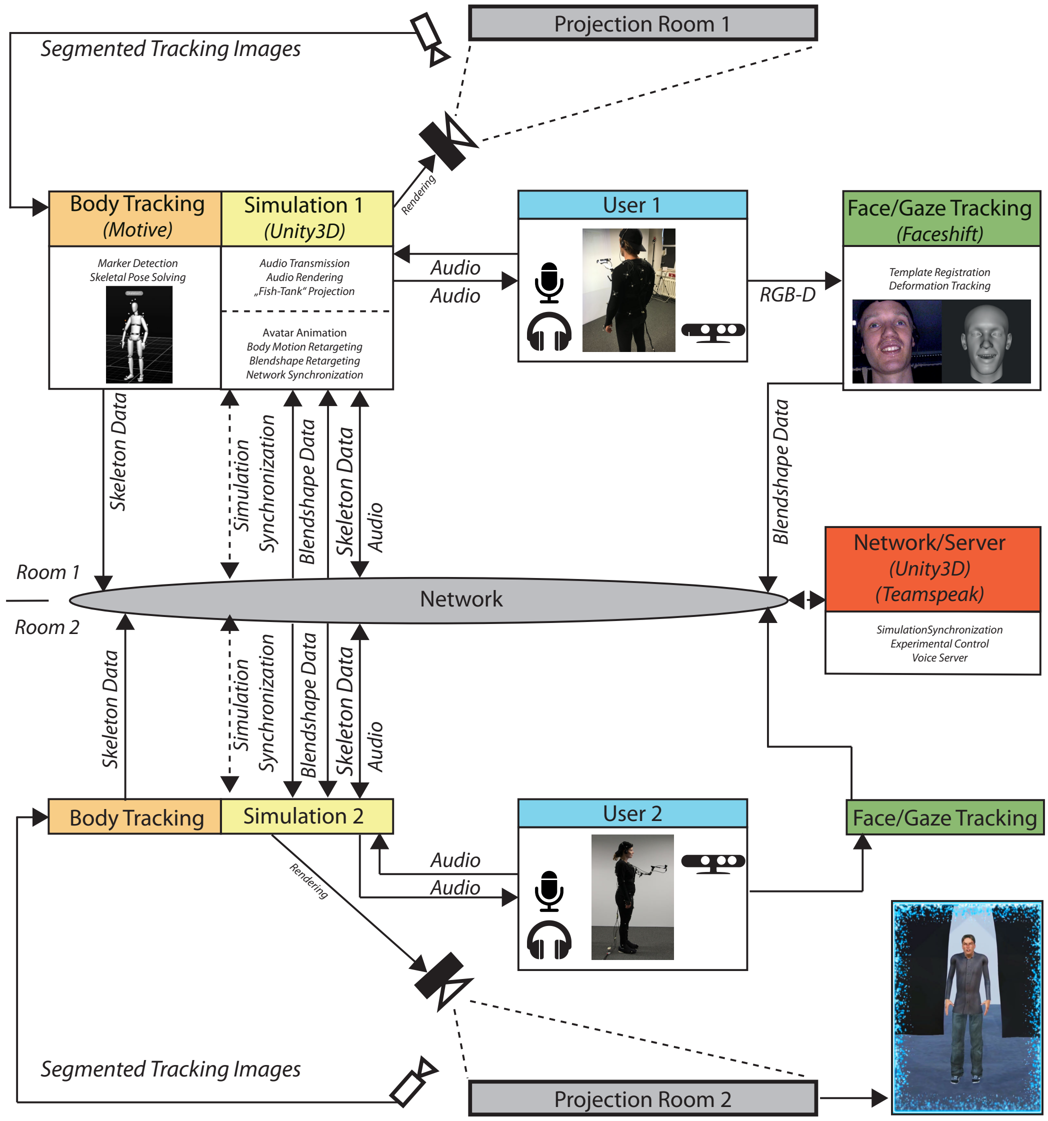

Figure 4.8: System diagram. The Figure depicts the data flow for one projection. The two rooms were identically equipped and were connected via LAN. A server application provided a control for the experimenter to start the experiment and the visualization of the avatars, and to stop the experiment.

of movement within the motion capture tracking space (see Figure 4.10). Two short-throw digital light processing (DLP) projectors (Acer H6517ST[40]) were connected to two rendering clients. The virtual environment was implemented in Unity3D[41] and designed similar to the physical environment in the two remote rooms. In order to minimize latency, the motion tracking and facial expression tracking data were streamed directly to each client and a server for experimental control. Skeleton data that were streamed by the motion tracking system were interfaced with Unity3D's Mecanim animation system (see Figure 4.10).

[40] Acer Inc., `https://www.acer.com/ac/en/US/content/professional-model/MR.JLA11.009`
[41] Unity Technologies, `https://unity3d.com/`

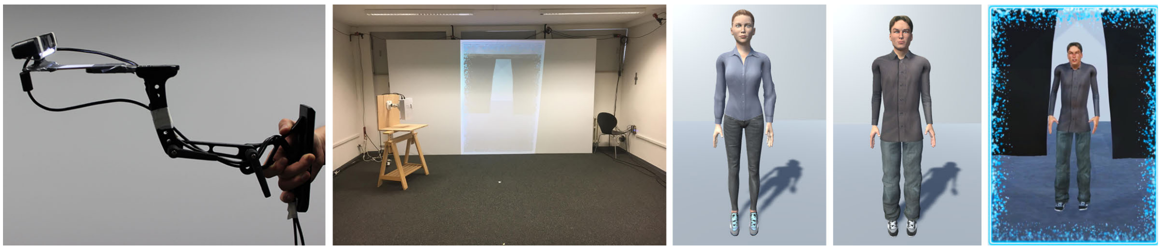

Figure 4.9: Apparatus and materials. *From left to right.* 1. The steady-shot rig with attached sensors that was worn by the user in order to hold the facial expression tracking and audio level stable while allowing free movement. 2. One of the experiment rooms. 3. Male and female avatars used in the study. 4. Rendered stimulus presented on the projection.

Facial expressions in the form of blendshapes were retargeted to the avatars by mapping tables to best suit the human-like characters (see Figure 4.9) that were generated with Autodesk Character Generator[42].

The experiment server was used to configure and control both render clients in the study (i.e., trigger a synchronous start of the simulation and so on).

We used microphones (Sennheiser ME4[43]) attached to the rig to capture audio, which was reproduced by headphones. The TeamSpeak software enabled transmission of the audio data separately from the main simulation. The system is depicted in Figure 4.8. We measured MTP latency for all components in our system using a 240 Hz camera resulting in $M = 210$ ms ($SD = 23$ ms) for body motion and $M = 220$ ms ($SD = 34$ ms) for facial expressions. Audio latency was assessed by recording original and resulting impulse responses ($M = 416$ ms, $SD = 27$ ms). Hence, the audio was delayed slightly longer compared to the visual feed.

Each short-throw projector was placed close to the projection screen in order to minimize occlusion by the user. The projectors were rotated to display in portrait mode to allow life-sized avatar projections, resulting in an image with a height of 2.25 meters when the projector was placed approximately 80 cm away from the screen. The participant stood approximately 1.2 m from the screen, and occlusion started to occur at about 40cm, depending on the participant's distance to the procjector, see Figure 4.10.

The VR projection followed the technique described by Ware, Arthur, and Booth [353]. Using head coupled perspective projection to allow a view of life-sized virtual interaction partners, the perspective projection changed with the position of the user. The image viewport was calibrated based on a tracking measure of the corners of the pyhsical image projection. We assumed the projection screen (target) to be a rectangle. Hence, it sufficed to calibrate three of its corners using tracking markers. The motion capture system defines a unique coordinate system for all tracked markers, and thus, the screen can be registered in this system. The

[42] Autodesk Character Generator, Autodesk, `https://charactergenerator.autodesk.com/`

[43] Sennheiser electronic GmbH & Co. KG, `https://de-de.sennheiser.com/mini-lavalier-mikrofon-clip-on-liveton-me-4`

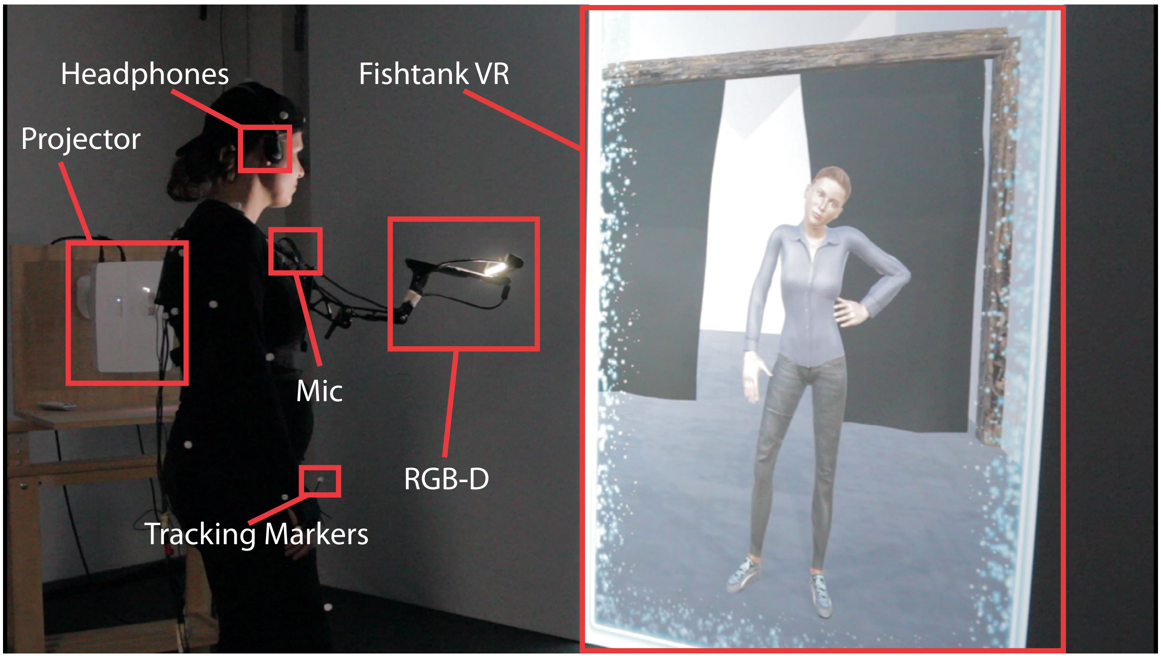

Figure 4.10: Apparatus and a user during the interaction. Using the fish tank VR metaphor, the virtual camera's perspective changed according to the users head transformation. The avatars were rendered in life-size, taking into account the distance the users had to the portal. The image was slightly brightened in the dark areas for the sake of information detail.

projection feed was presented in 1080p 60 Hz. Technically, the systems supports stereoscopic rendering, achieved by off-center projection [49, 278].

However, the facial tracking suffers when users wear active stereo glasses, which is why we used non-stereoscopic rendering to achieve better behavior tracking quality.

Procedure

At first, a short oral introduction was given to each participant in both rooms for the purpose of clarifying the study. This was followed by health and safety information, data acquisition consent, and a participation consent form. In order to minimize the time between the pre-questionnaire and the start of the social interaction the participants first changed into motion-capture suits with reflective markers and a vest with mounts for the facial tracking camera rig. Afterward, the participants were asked to complete the pre-experiment questionnaire, which included written instructions about the free interaction. In the instructions, participants were asked to get to know their partner without any formal conditions being imposed, for which they had five minutes. Next, the steady-shot rig was mounted, and calibrations were performed. Subjects were then positioned at a fixed starting point and the task began. Following the interaction, the subjects answered the post-experimental questionnaires. In all conditions, participants were equally equipped with technical equipment.

Measures

In contrast to Study 9, we changed the interaction rationale to a free interaction and therefore adapted our measures for presence to better fit the context. Furthermore, we did not assess affect because a rating would have been highly dependent on the topic the participants chose to follow.

Presence We assessed social presence ($\alpha = .80$; 7-point scale, see [220] for anchor names), self-reported copresence ($\alpha = .63$; 1–strongly disagree, 7–strongly agree), perceived other's copresence ($\alpha = .89$; 1–strongly disagree, 7–strongly agree) and telepresence ($\alpha = .86$; 1–not at all, 7–totally) using the questionnaire from [220].

Rapport As we were interested in how the different degrees of behavior affect the rapport [327] among the participants, we assessed rapport ($\alpha = .88$; 1–strongly disagree, 7–strongly agree) with the questionnaire used in [120] that extends copresence questions from [220] to a total of 11 statements: ("I felt I had a connection with my interaction partner," "I think that we understood each other," "My interaction partner was warm and caring," "My interaction partner was respectful to me," "I felt I had no connection with my interaction partner" [reverse coded], "My interaction partner created a sense of closeness or camaraderie between us," "My interaction partner created a sense of distance between us" [reverse coded], "My interaction partner communicated coldness rather than warmth" [reverse coded], "I wanted to maintain a sense of distance between us" [reverse coded], "I tried to create a sense of closeness or camaraderie between us," " I tried to communicate coldness rather than warmth" [reverse coded]). By their nature, both measures are therefore correlated.

Avatar Perception In order to investigate the impacts on the perception of the virtual representations, we assessed the uncanny valley factors humanness ($\alpha = .84$) and eerieness ($\alpha = .67$) based on the questionnaire from Ho and MacDorman [136].

Participants

We recruited 98 participants for the study. Due to the recruitment process, we could not set any prerequisites for participants. We therefore excluded participants with strong uncorrected visual impairments (listwise), participants who marked insufficient language skills (pairwise), or dyads where the tracking was unstable for the analyses (pairwise). The final sample for the analyses consisted of 56 participants (40 female, mainly students, $M_{age} = 25.30$, $SD_{age} = 5.97$). Participants were randomly assigned to one condition, and the final sample we analyzed had the following distributions: Body: 19 (9 females); Face: 16 (9 females); Body & Face: 21 (13 females). Eight participants had previous experience with VR systems.

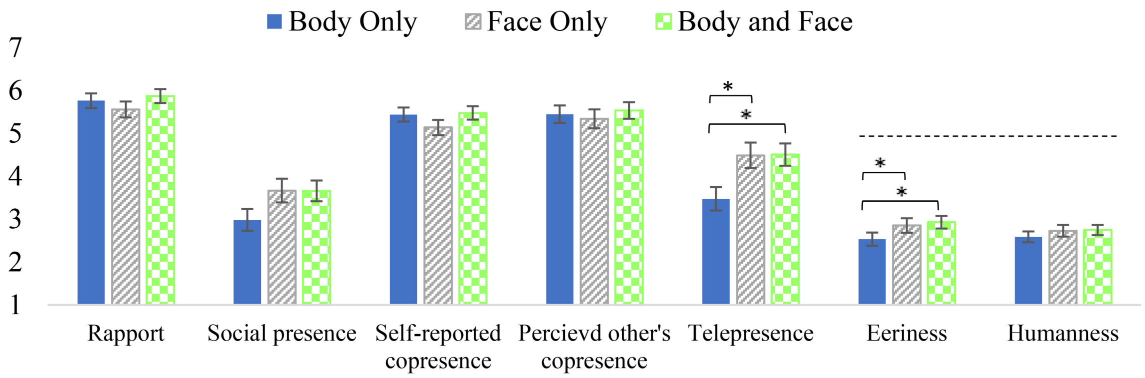

Figure 4.11: Comparisons of the presence and uncanny valley measures. *Note.* *$p < .05$; *bars denote the mean value; error bars denote the standard error.*

4.3.2 Results

Fig 4.11 shows the descriptive results of the second study. We conducted ANOVA for the dependent measures where normality could be assumed based on the results of Shapiro–Wilk tests.

Presence

A Levene's test for social presence showed that equality of variances cannot be assumed. A one-way Welch ANOVA conducted to investigate the impact on social presence showed no significant difference: Welch's $F(2, 30.87) = 3.297$, $p = .050$, $\eta_p^2 = .030$. The BF ($M = 3.66$, $SD = 1.17$) and F ($M = 3.67$, $SD = 1.39$) resulted in greater social presence than B ($M = 2.98$, $SD = 0.70$). Gender did not significantly impact social presence. An ANOVA for self-reported copresence was not significant; $F(2, 53) = 1.184$, $p = .314$. BF ($M = 5.48$, $SD = 0.70$) and B ($M = 5.44$, $SD = 0.71$) were higher than F ($M = 5.14$, $SD = 0.73$). A Shapiro–Wilk test showed that the normality assumption was violated for perceived other's copresence. A Kruskal–Wallis H-test showed no significant difference; $\chi^2(2) = 1.142$, $p = .565$. An ANOVA for telepresence showed significant differences between the groups; $F(2, 53) = 4.635$, $p = .014$, $\eta_p^2 = .149$. BF was highest ($M = 5.51$, $SD = 1.0$), and F ($M = 4.49$, $SD = 1.27$) was still higher than B ($M = 3.47$, $SD = 1.33$). Bonferroni adjusted pairwise comparisons showed that the differences between B and F ($p = .046$), as well as between B and BF ($p = .025$), were statistically significant.

Rapport

An ANOVA conducted for rapport was not significant ($F(2, 53) = .814$; $p = .449$; $\eta_p^2 = .030$). Rapport was highest in the BF condition ($M = 5.87$, $SD = 0.72$), followed by the B condition ($M = 5.76$, $SD = 0.78$) lowest in the F condition ($M = 5.56$, $SD = 0.72$). An ANCOVA including gender as a covariate showed that gender significantly impacted rapport ($F(1, 52) = 4.20$; $p = .045$; $\eta_p^2 = .075$). Over all conditions, female participants reported higher rapport ($M = 5.92$, $SD = 0.59$) compared to male participants ($M = 5.52$, $SD = 0.86$).

Avatar Perception

A Shapiro–Wilk test showed that the normality assumption was violated for the eerieness responses. A Kruskal–Wallis H-test showed a significant difference between groups; $\chi^2(2) = 8.663$, $p = .013$. The data distribution was equal across groups. Following Dunn's procedure for post-hoc comparisons, Bonferroni adjusted pairwise comparisons showed that B and F ($p = .036$) and B and BF ($p = .013$) were significantly different. BF resulted in the highest median ($Mdn = 3.0$) compared to F ($Mdn = 2.81$) and B ($Mdn = 2.63$). An ANOVA for humanness was not significant; $F(2, 53) = 0.208$, $p = .813$.

4.3.3 Discussion

The present study aimed at investigating the effects of different levels of behavior modalities that are replicated to avatars that represent the users in a virtual social interaction. We could not find significant differences in the social presence or copresence factors. However, telepresence was significantly higher in the conditions where facial expression and gaze were available. It seems that the behavioral realism affected this presence factor. Considering the questions asked (e.g., how involving, intense, and immersive the participant found the experience), this is in compliance with the description of immersion as accurate mapping of behaviors and senses [300, 297]. Adding more sophisticated modalities thus seemed to have improved the simulation and, in turn, the perception of telepresence. However, we did find that eerieness increased in the conditions where facial expressions and gaze were included. This points to inaccurate mappings or artifacts with regard to facial expressions and gaze behaviors, which could have also affected other dependent variables. Tracking inaccuracies were partly noticeable and indeed, participants commented on it subjectively: "the turned/not visible eyes were an obstacle for getting to know and assess the other participant", "I could not clearly identify the mimics [German: 'Mimik'–Facial Expression] of the other participant. Whereas I could [identify] her movements." However, the sensor that was used in the simulation can be described as the state-of-the-art for real-time tracking (in years 2015/2016) and, compared to other solutions, was considered best in pretesting by lab members. However, future work should consider this fact with regard to system improvements.

4.4 Summary

This chapter aimed at comparing physical and virtual interactions and investigating the impact of the level of behavioral modalities presented.

4.4.1 Impacts of Embodied Virtual Interactions

Regarding a comparison between PR and VR interactions, Study 8 identified significant impacts on social presence when comparing a PR scenario with a VR scenario.

While Study 9 could not confirm statistical significance, the results show the same trend toward lower social presence values in VR. Study 9 might have suffered from the small sample size collected for the analyses, and the detected sickness could have biased the results. Yet, recent research assessing a similar simulation did show a significant difference of social presence with regard to non-embodied VR in comparison to embodied VR or face-to-face conversations but did not find a significant difference comparing embodied VR and face-to-face [309]. We may therefore conclude that deficiencies can arise from embodied virtual interactions compared to PR interactions, yet, further research is needed to systematically assess the impact on social presence and copresence perception. Specifically, these deficiencies can be attributed to perceptual effects such as presence, connectedness, or efficiency, rather than functional aspects of the communication. While similar shifts in attentional focuses occurred throughout Study 8 and 9, these did not result in significant effects and thus did not find evidence of a conscious shift in decoding social information. Yet, the data also does not contradict the SIPT [351]. In neither of the studies did we find indications for deficits in negotiation performance. While the collaborative performance in Study 8 obviously suffered from the constraints of the simulation (i.e., latency, physical abstraction of the ball behavior, as well as missing haptic feedback), we did not find these deficits in the negotiation outcomes.

The VR condition in Study 9 showed a significant lower valence attributed to the communication partner along with significant higher arousal. This might be due to a negative bias toward the mannequin, or a general more negative experience in the virtual interaction. Furthermore, efficiency was perceived to be lower in VR despite the fact that the objective performance was similar.

4.4.2 Impacts of Transmitted Behavior Modalities

In Study 8 and 9 participants indicated that they missed facial expression the most when interpreting their communication partner. We operationalized a systematic comparison in Study 10. While the presence of facial expressions did result in higher social presence ratings, the difference was not significant and thus we cannot conclude on the basis of our data that the social aspect was specifically addressed by the presence of facial expression. Yet, the presence of facial expression resulted in higher telepresence. Considering the questions stated, one could argue that they address, to part, aspects of immersion and thus the addition of facial expression resulted in the desired effect. Yet, adding facial expressions did also yield to a higher eerieness in the perception of the virtual character driven by the interaction partner. There may be two reasons for this. Either the reproduction of facial expression was unrealistic and therefore uncanny, or the participants were simply very new to the fact that avatars may provide replicate facial expressions in real time. Further research should explore this aspect in more detail. In summary, the presented studies provide an overview of potential impacts that arise from embodied social interactions. Figure 4.12 summarizes the results of the findings presented in the present chapter.

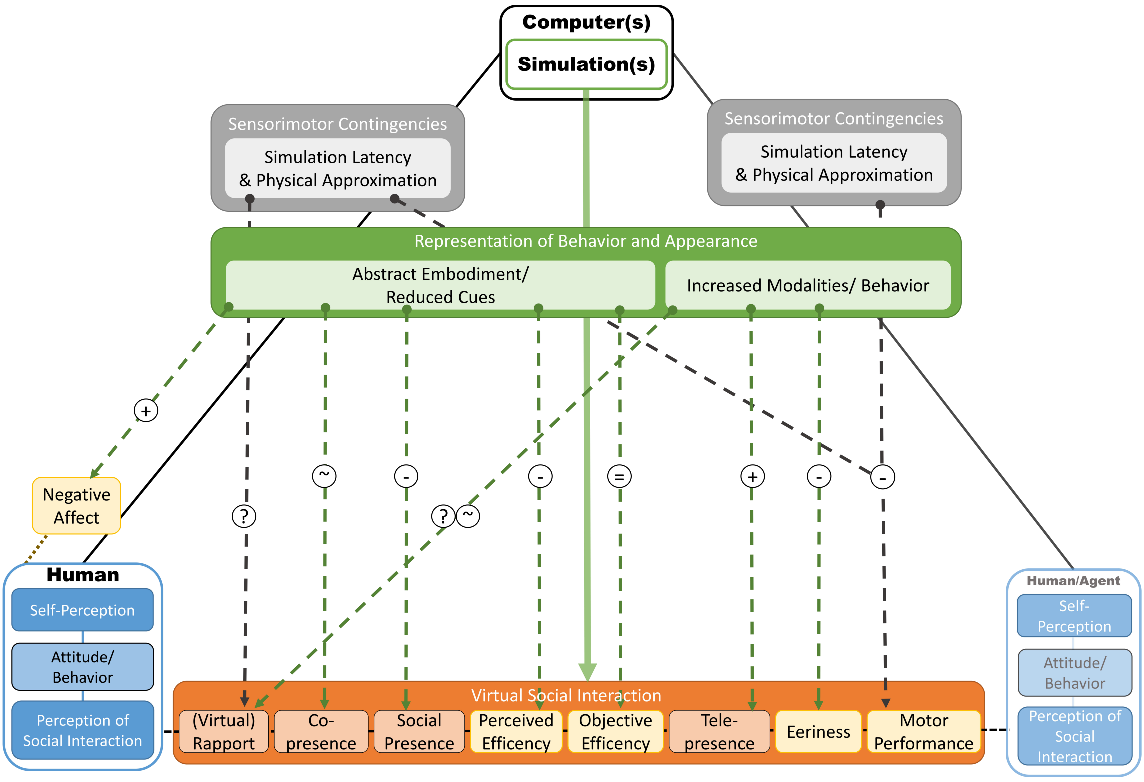

Figure 4.12: A model summarizing the results results of the social interaction studies. Abstraction of appearance (combination with a latency present in the virtual simulation) resulted in a reduced perception of social presence as well as a reduced perceived (but not objective) effiency of the interaction. The simulation latency and thus hindered sensorimotor contingencies negatively impacted collaborative motor performance. Increased degrees of freedom in behavior reproduction (facial expressions) led to a higher telepresence but negatively affected the perceived eerienes of the virtual avatar representing the communication partner (i.e., more eerieness was perceived).

4.4.3 Implications

I2.1: Reduced Affordances Can Reduce Social Presence

Especially in the comparison of physical and virtual interactions it became clear that VR interactions suffer from a reduced realism in user embodiment. Despite potential compensation mechanisms, social presence in applications is likely to be fostered by the presence of social behaviors, as they allow to infer intention and emotion as well as to assess a partner's reactions.

I2.2: Behavioral Affordances Need to Be Present and Convincing

Study 10 showed negative effects of an increased behavioral realism. We thus argue that when such cues are presented, they need to be presented a) visible, and b) with feasible quality. Collaborative and SVR applications may consider this aspect in their design. A mere presence of a cue may therefore not suffice, as a reduced realism could have negative effects.

I2.3: Simulation Latencies Decrease Collaborative Motor Performance

Similar to negative effects in social presence, we found that the abstracted appearance in combination with the latency present in the simulation of Study 8 showed a negative impact on motor performance. While one could argue that, to part, the simulation's physical approximations may have not been sufficent, to large degree we attribute this impact on the latency that inferred with SMCs and thus reduced the performance. Especially applications with collaborative motor interactions should take both into account: to simulate realistic physical behavior and to reduce simulation latency to minimum.

4.4.4 Conclusion

In summary, we could show that embodied social interaction in comparison to physical interactions are inferior with regard to social presence and collaborative motor performance. An increased realism showed indications of a positive impact on aspects of presence. Yet, the realism of the presentation of behaviors may have not been sufficient for a convincing naturalism. Further studies should explore alternative ways of display social behavior and a more engaging role of the simulation, which is the topic of the next Chapter.

Chapter 5

Hybrid Social Interaction

Chapter 3 described the embodiment of the user and related intrapersonal effects, and Chapter 4 described technologies for virtual interactions and how they affect social interactions. The present chapter describes how computers can be used to actively mediate such interactions through technological interventions beyond mere replication of user behaviors and thus an intentional augmentation of user-embodied interactions, thereby tackling the computers as social mediators metaphor outlined in Section 2.4.2 with the development of HAAT, outlined in Section 2.4.5.

The three studies that are presented in this chapter investigate three forms of augmentations: i) real-time body modifications based on the higher-level phenomenon of nonverbal mimicry in a dyadic setting (Section 5.1), ii) direct hybrid augmentation of social gaze behavior in a dyadic setting (Section 5.2), and iii) visual augmentation based on the substitution and transformation of common phenomena in a multi-user setting (Section 5.3).

5.1 Study 11: Hybrid Mimicry

In this study, we investigated how the artificial incorporation of a phenomenon of social interaction, namely nonverbal mimicry (the motor imitation of another's nonverbal behavior [65]) could be achieved and how it was perceived by the participants. For this reason, a hybrid prototype was developed that modified the upper body motion of participants to periodically interject artificial nonverbal mimicry. In a user study, the impacts on the perceptions of participants with regard to presence, rapport, and liking were measured. In this exploratory investigation, we used a naïve approach to a social AI. A simple state machine was developed to mediate the conversation. At distinct time intervals, the upper-body behavior of one interactant was mapped to the visual representation of the other interactant with a slight delay, blending real and prerecorded behavior. Simply speaking, without knowing, subjects periodically experienced their own movements mapped to the interaction partner's avatar in the VR simulation with a small time-delay. An initial study aimed at comparing dyads with transformed behavior to those without transformed behavior, exploring whether the artificial insertion of mimicry impacted the participants' perceptions of the communication, and if so, whether the impact was positive or negative.

5.1.1 Nonverbal Mimicry

Nonverbal mimicry can be described as the the motor imitation of another's physical behaviors in a social interaction [65]. This "chameleon effect" [64], along with other forms of interpersonal adaptation [55], can facilitate the establishment of affiliation and rapport [167] and has distinct impacts on social interactions by serving as "social glue" [168], binding and bonding people together [337]. Mimicking others has been shown to lead to liking, rapport, and prosocial behavior [65, 315, 316, 317].

Previous research also found a positive impact of being mimicked, such as the facilitation of liking and rapport [20], group harmony [166], and empathy [201]. And reciprocally, perceiving rapport and liking leads to more mimicry [202, 315]. So far, there seems to be no clear causality between the correlating constructs, indicating a bidirectional effect.

5.1.2 Virtual Mimicry and Virtual Rapport

A number of previous approaches have investigated aspects of virtual mimicry ("digital chameleons" [16]) and virtual rapport. However, most of these studies investigated this phenomenon based on human-agent interactions [16, 121, 139, 128], and were often reduced to 2D environments without virtual embodiment of the self, leading to a constant physical world reference. For example, Bailenson and Yee [16] investigated the effect of induced nonverbal mimicry. Agents mimicking head movement were perceived more positive than nonmimicking agents.

Gratch and colleagues presented multiple studies regarding their concept of virtual rapport [121, 122, 123, 120]. Rapport is associated with positivity and mutual attentiveness in an interaction, and is strongly related to nonverbal behavior such as directed gaze, head nodding, smiling, mirroring and the like (for an overview see [327]). They built a virtual agent that was trained to react to a speaker's behavior utilizing known rapport facilitating behaviors. In a user study, they found that the rapport agent positively influenced speech quality, motivation and the overall impression compared to an avatar with prerecorded behavior [121]. Furthermore they observed the establishment of a bidirectional relationship that was not present with prerecorded behavior.

In addition to human-agent interactions, these social behavior models could also be used in real time CMC. Applications for avatar-mediated social interaction that utilize such social behavior models could not only serve as passive transmitters of communication, replicating the behavioral data of users, but may also become an adaptive tool, actively modifying and adapting behaviors with HAAT [269]. To this end, a naïve prototype was implemented in the present study to generate initial findings. To the best of our knowledge, our approach is the first truly interactive paradigm [361] that focuses on a method for prompting the naturalness in the social interaction. The following describes the approach to the simulation, the modifications and blending process, as well as the subjective and objective results of the study.

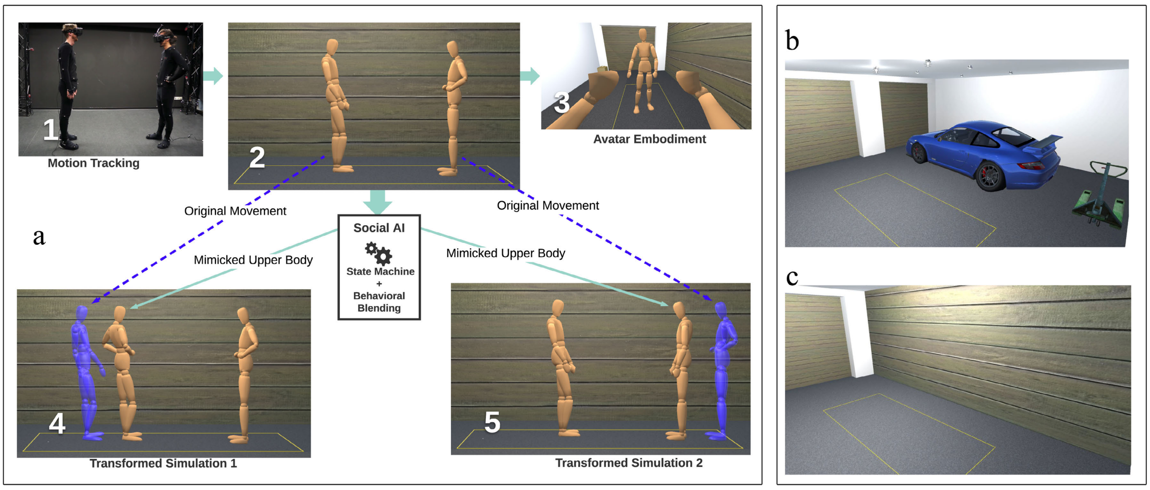

Figure 5.1: Scenario and simulation principle. ***a.*** Apparatus. 1. Users in the tracking environment, immersed in the simulation. 2. The simulation during the task from a third person viewpoint. 3. First person viewpoint during a conversation. 4. Mimicked movement modification for the simulation of the right user. 5. Mimicked movement modification for the left user. ***b.*** Environment during the acclimatization phase. ***c.*** Environment during the negotiation phase. Figure reprinted from [274].

5.1.3 System Design and Implementation

Apparatus

Our apparatus were adapted from previous work (see Chapter 4 Study 8, and 9) and used a 16-camera system (OptiTrack, Flex3, 100 Hz) to perform body tracking (see Figure 5.1). Motion data were synchronized with two client simulations, implemented with Unity3D[44], that were not network dependent, yielding two non-persistent individual simulations. Each user was embodied as a wooden mannequin avatar, and the simulation was rendered to each user's Oculus Rift DK2 HMD[45] (960px × 1080px per eye, 100 ° field of view [FOV]) (see Figures 5.1, 1 & 3). Using video-based measurements, the motion-to-photon latency was approximated to 90ms. To calibrate the absolute rotation of the tracking system (world space) with the HMD IMU rotation data, we recalibrated the y-axis rotation of the virtual camera to the rotation in world space before and during the simulation.

Artificial Injection of Nonverbal Mimicry

The real-time injection of artificial mimicry was based on a simple reflex agent [282]. The motion data (skeleton joint transformations) provided by the tracking system were used to drive the avatars of each user, which were preliminarily scaled according to each participant's height. A state machine controlled the injection of nonverbal behavior, which was timer-based and consisted of four active states and two states to extend the principle into future work (see Figure 5.2). The idle state replicated the original physical movements of each participant. After 20 seconds, the

[44] Unity Technologies, `https://unity3d.com/`
[45] Facebook Technologies, LLC., `https://www.oculus.com/`

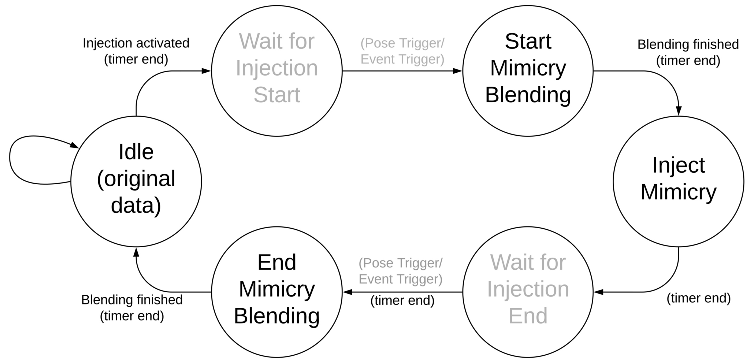

Figure 5.2: Mimicry injection principle. The grayed-out states were introduced for future developments, such as to stop the mimicry injection based on speaker/listener status. Figure reprinted from [274].

start mimicry blending state was activated by the experimenter. This state blended the original upper body motion (excluding the hips) of the partner's skeleton model to one's own buffered (3 second delay) upper body motion using a smooth linear interpolation (2 second window). After the blending, the inject mimicry state (8s) and wait for injection end state (1s) are held to inject mimicry for a total of 9 seconds. Similarly, the buffered own motion presented by the partner's avatar was then blended back to the partner's original motion during a two second frame, after which the system returned to the idle state again for 20 seconds, and so on, until the experimenter stopped the injection. We termed our system "mimicry injector".

5.1.4 Method

Design

We used a one-factor between-subjects design to compare the artificial mimicry condition to a control group. In both conditions, participants saw their original behavior replicated to their own avatar. In the artificial mimicry condition, the mimicry injector was activated to blend the artificial behavior with the real behavior of the interlocutor. In both conditions, participants were participating in role-play that was constructed as a buyer/seller paradigm in which participants had to bargain over the price of a used sports car in a garage environment.

Procedure

Participants were welcomed, briefly introduced to each other, and informed about the study. After agreeing to participate by signing the consent form, participants were asked to fill out a pre-study questionnaire assessing demographics. Participants were then separately guided to dressing areas and asked to put on the motion-tracking suit. After applying the passive optical markers, participants were given written instructions about their role (buyer/seller) that included three arguments to strengthen their position in the negotiation role-play (e.g., that the axle was broken, which was later visible on the simulated car). After carefully reading the instructions, the participants were equipped with the HMD, guided to a predefined starting position, and calibrated for the simulation. The simulation was then started, and participants were given a short, silent acclimatization phase to prevent sickness effects and to control for their impressions. In this phase, the participants could take a look at the object of negotiation (the sportscar) as a simulated model, including the modifications defined by the arguments (e.g., broken axle). After one minute, a garage door blocked the view of the car, and participants were asked to turn to each other and start the negotiation (see Figure 5.1 b, c). In the artificial mimicry condition, the "mimicry injection" was activated after 20 seconds. During the negotation, a yellow square marked the safe tracking area (see Figure 5.1). When participants found a consensus or the 7 minute time window was up, the simulation was stopped. Participants were then asked to complete a post-experimental questionnaire that assessed dependent measures, and that was followed by the debriefing and a compensation (student credits).

Measures

Dependent Measures We asked participants to judge whether the partner's behavior was realistic (1–do not agree at all, 7–fully agree). We further assessed factors for social and copresence using the questionnaire from [220] (1–do not agree at all, 7–totally agree, Cronbach's αs$>.649$) and also measured virtual rapport using the questionnaire from [120] (1–do not agree at all, 7–totally agree, $\alpha = .807$). Furthermore, we included a measure for liking and attraction adapted from [224] (1–do not agree at all, 7–totally agree, $\alpha = .903$). We assessed affect with the positive and negative affect scales [354] (1–little or none, 5–extreme, αs$>.813$) and tested how the manipulation impacts trust with three questions (e.g., "I would rely on my communication partner". 1–does not apply at all, 7–totally applies, $\alpha = .856$). We further assessed if participants had reached a consensus, and if so, how long it took them to negotiate to reach this consensus (stopwatch) and then calculated the difference in interpersonal distance over time. Furthermore, participants were asked to comment on the behavior of the other person and any potential suspicions they may have had.

Control Measures To control for a bias introduced through personality traits or a previous relationship between the participants, we used the social-closeness index [8] (1–not close at all, 7–very close) and a measure of the Big 5 personality traits [247]. Simulator sickness [156] was measured in a pre–post measure.

Participants

From the 70 student participants who took part in the study, we excluded dyads with participants who experienced technical problems, who did not have fluent language skills, or who did not fulfill the task. Participants were blind to the goal of the experiment. One dyad was excluded because a participant had a correct suspicion. Our final sample consisted of 40 German participants (24 females, $M_{age} = 21.87$, $SD_{age} = 2.54$) equally distributed among the conditions.

Table 5.1: Results of the Objective Measures (Adapted from [274]).

	Control†	*Mimicry Injected*†
Interaction time [s]	251 (48)	265 (97)
Negotiated price [e]	35000 (1793)	36217 (1068)
Time to consensus [s]	251 (48)	266 (97)
Distance change (start–end) [mm]	-24 (188)	21 (436)

Note. † *Mean* ($\pm$ *standard deviation*)

5.1.5 Results

System Evaluation

We excluded two dyads from the technical analysis due to corrupt data. Data analysis showed that mimicry was injected 25.44 % of the overall conversation time ($SD = 1.98$%). On average, 8.39 ($SD = 3.13$) mimicry injections took place during a conversation in the mimicry condition. However, only two comments regarding the behavior from participants in the mimicry condition mentioned that they felt their partners were adopting their movements (e.g., "made similar movements as I did"). Thus, we evaluate the technical functionality as rather successful.

Control Measures

T-tests assessing differences for the control measures revealed that differences were non-significant. ANOVA results for pre- and post-simulation measurements of simulator sickness showed that subjects felt sicker ($p < .001$) after the exposure but there was no significant difference between the conditions.

Dependent Measures

T-tests for behavioral realism, social presence, and copresence, as well as the rapport measure, did not yield significant results. Neither the trust rating nor the interpersonal attraction or positive/negative affect yielded to significant differences between the conditions. Figure 5.3 depicts the descriptive results. Overall, six out of ten dyads found a consensus in the control condition, and six out of ten in the mimicry condition. A chi-square test showed that the difference was not significant. Six out of ten male dyads and eight out of twelve female dyads reached a consensus. The difference was not significant. The time participants interacted was slightly longer in the mimicry injected condition ($M = 327.5\ s$, $SD = 108\ s$), compared to the non-mimicry condition ($M = 284.8\ s$, $SD = 82.89\ s$). T-tests showed that neither the total interaction time, the negotiated price, the time to consensus, or the difference in interpersonal distance from the start of the interaction to the end of the interaction differed significantly between the conditions (see Tab. 5.1).

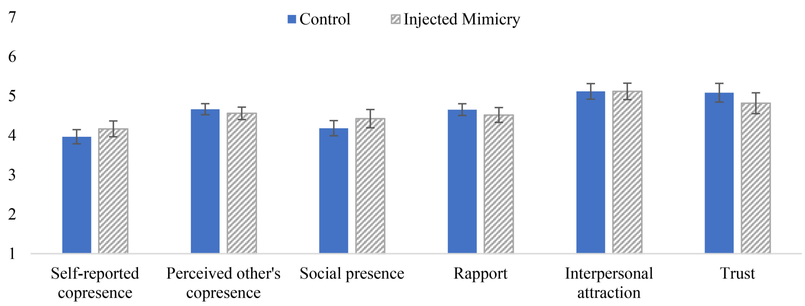

Figure 5.3: Results of the subjective dependent measures. *Note. Bars denote the mean value; error bars denote the standard error.*

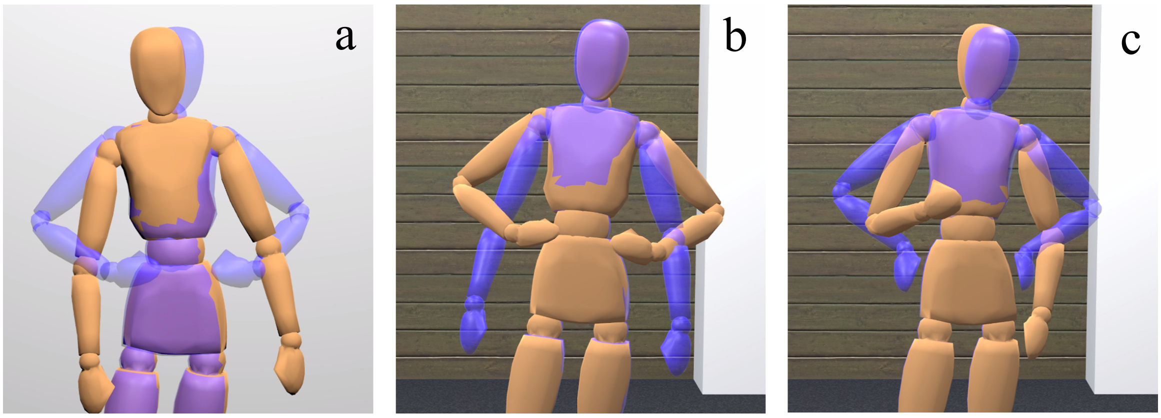

Figure 5.4: Modification examples of the artificial mimicry injection. Blue transparent avatar: original user motion. Wooden mannequin: modified motion presented to the partner. a) First-person view of participant X showing a successful adaptation of the original pose of participant Y, and (b) vice versa. c) Mimicked dialog gesture that might have a negative impact on rapport perception and disrupt nonverbal synchrony. Reprinted from [274].

5.1.6 Discussion

While the results showed no direct impact on the perceptions of the communication, we evaluate the prototype as rather successful. One subject identified the manipulation and had a correct suspicion, and only two subjects consciously noticed the modification. We aim to further investigate blending techniques.

Regarding the impact on the perception as such, our naïve periodic injection of mimicry led to the mimicking of body poses and also of specific dialogue gestures accompanying speech. Thus, while these replications may have had a mirroring appearance it these may have been unfitting a semantic level, see Figure 5.4. Furthermore, the asynchronous simulations led to both partners having overlapping mimicry injections in the simulation. These may have impacted the impressions counterproductively, as entrainment processes of interpersonal synchrony are bidirectional, time-dependent processes, similar to rhythm and instrumental interplay in music. Thus, the injection at random points during the communication may have disrupted processes of coordination and synchronization. Thus, future improvements should include distinct triggers for the injection, such as detecting the speaker and listener of the conversation [268], which implies a networked simulation that includes additional modalities such as voice or gaze. Attentional focus is of importance for a second reason, which is that the injection is not necessary or cannot be detected when the communication partner looks in another direction or focuses on other objects in the scene. In our study, this was prevented as the partners had a narrow scenario and did not have many other objects to focus on. However, in future applications for collaborative and social interaction, this might not necessarily be the case.

There is no doubt that the verbal channel is of tremendous importance for communication in most social interactions. Considering that our study was based on a strong verbal exchange, we can therefore not rule out that the verbal channel had an overruling impact on the outcome. However, we argue that everyday social interaction is mostly accompanied by verbal exchange in addition to nonverbal behavior. Another improvement to the social AI triggering modifications could be a semantic judgment to detect complex gestures and prevent the injection of movements that are not suitable for the context. Furthermore, a status analysis judging the current rapport and the conversational situation in real time could assess the necessity for modifications and prevent inadequate injections that could happen with our naïve system as it that acts periodically.

Limitations

Observing the simulation, we noticed slight "foot-skating" issues due to the kinematic retargeting that was applied, which could have caused a negative bias. Furthermore, we did not test how using more realistic avatars would impact the results. Finally, our results cannot be blindly generalized as the sample size in our study was limited.

Conclusion

In this study we presented a system to artificially inject nonverbal mimicry to embodied virtual social interactions. We tested the performance and how an artificial mimicry injection compares to natural behaviors replicated to immersive embodied social interactions. We did not find specific significant negative impacts of our system in the study. Considering that we compared to natural behavior the prototype already successfully altered the interaction. Future work may consider alternative methods of further controlling the points of injection, as we did, for example, not include a decision process on the basis of a speaker/listener status or a semantic interpretation.

5.2 Study 12: Hybrid Gaze

As outlined in Chapter 2.3.4, in face-to-face encounters, nonverbal behavior such as body movements, facial expressions, and eye gaze are of paramount importance for person perception, impression formation, inference of emotions, and rapport [54, 327]. Unlike spoken language, our behavior is not mutable. In fact, it is analog and "always on." During virtual social interactions, however, we interact with avatars or with agents. Both can vary in the degree to which behavior is replicated and expressed. Consequently, a lack of variation and contingency in displayed behavior, or distortions in the dynamics, will be interpreted as a meaningful signal, whether it is intentional or not.

In this regard, gaze behavior is especially important due to its use for initiating conversations, understanding others, and expressing ourselves, disclosing human comprehension, and turn taking [149, 94, 6, 160]. Following Harper, Wiens, and Matarazzo [132], Kleinke [160] described face-gaze, eye-gaze, mutual gaze, eye contact, looking/gazing, gaze avoidance, gaze omission, staring and glancing as phenomena of gaze behavior that can be present in social interactions. Whereas mutual gaze is associated with two people gazing at each other's faces, eye contact describes two people gazing at each other's eyes. All of these behaviors are parts of our social interactions that we experience in everyday physical life. Humans without visual or social disorders can typically experience, disambiguate, and interpret gaze behaviors and appropriately react to it.

This has major implications regarding the inclusion of gaze in future applications for virtual interactions [278]. First, imperfections in capture, transmission or rendering can lead to subtle variations in observable behaviors and potentially cause misattributions of intentions and emotions, or induce undesirable social impressions. Second, the lack or disturbance of gaze transmission could be either compensated for by artificial social intelligence or completely transformed [12], for example via visual augmentations triggered by social signals and phenomena (see Section 5.3). Third, the technology may be able to serve as a monitoring and repair system that detects disturbances in nonverbal behaviors of humans, for example, a lack of lack of eye contact due to distractions, and compensate for them by synthesizing a more appropriate behavior [269]. Against this background, the potential role of virtual gaze behavior, including opportunities [214] and risks, has to be reassessed.

The present study therefore aimed at a systematic investigation of social gaze in dyadic virtual interactions, and assessing effects on social perception. Based on a social gaze model that took into account speaker–listener coherences that were derived from previous studies, we compared 1) a pure *natural gaze* transmission that was based on eye tracking data, 2) a *hybrid gaze* model based on a natural gaze default transmission augmented with a social gaze model, 3) a *synthesized gaze* model based on a random gaze default transmission augmented with the social gaze model, and 4) a pure *random gaze* transmission that was merely based on statistical distribution without taking into account any social affordances.

Conceiving of natural gaze behavior as the gold standard, we hypothesized that non-contingent random gaze, which merely follows general rules regarding frequency and duration, falls short of being perceived as natural and meaningful. We further hypothesized that a dynamic augmentation of this behavior, meaning making it contingent on the partner's speech activity, using a synthesized gaze model would be evaluated as more similar to a natural gaze. However, applying the same algorithms to augment the natural behavior could result in negative effects as it could destroy the subtle dynamics of natural interactions. To the best of our knowledge, this is the first study to provide a systematic comparison in this manner.

Gaze Behavior Models

Lee, Badler, and Badler [180] presented a gaze model based on empirical models of saccades and statistical models of gaze data retrieved from video analysis. They included an attention monitor module that accounted for statuses such as talking versus listening, head rotation, mutual gaze, or averted gaze. They also found indicators of higher levels of engagement and friendliness compared to random gaze and more perceived liveliness and friendliness compared to static gaze.

With regard to conversational interactions, Garau et al. [109] compared random gaze to inferred gaze based on audio input. In accordance with the literature on gaze behavior [6], they based their inference model on the assumption that people gaze more at their interaction partner while listening [109, 108] and included an audio trigger "while speaking"/"while listening" to infer gaze. They found that inferred gaze affected the evaluation of the interaction, including copresence, the perception of similarity to a face-to-face interaction, and results in a more positive evaluation of the partner.

Bente, Eschenburg, and Aelker [29] investigated different patterns of randomized gaze behaviors in three conditions: short periods of directed gaze (two seconds averted, two seconds directed) versus long periods of directed gaze (two seconds averted, four seconds directed) versus real gaze. They found that longer periods of gaze direction led to more copresence and a better evaluation compared to shorter gaze periods.

Vinayagamoorthy et al. [339] expanded the model from [109] to include inferred gaze, assuming that mean saccade magnitudes are shorter, that people tend to focus on their interaction partner more, and that inter-saccadic intervals between the focus positions are longer when listening [339]. They found that inferring gaze with more realistic characters can lead to increased effectiveness, whereas applying the gaze model on a cartoonish character did not result in any difference.

Ma and Deng [192] synthesized gaze by building a dynamic, coupled, component analysis based statistical model trained with eye-head motion data, which seemed to be superior to the model from Lee and colleagues [180], but the evaluation was limited to the assessment of animation clips. Furthermore, their model did not include speech, which was identified as a modifier of gaze behavior [6, 109, 108], and a trigger of gaze patterns. Le et al. [176] further extended this work with a fully automated framework to generate head motions, eye gaze, and eyelid motion based on live or recorded speech input. Although extending the dimension of behavioral realism, live avatar-based interactions were not investigated in their work, nor in other related work that used a multimodel approach [245], or reactive gaze behavior based on head position tracking [159].

Overall, previous research showed that artificial gaze models can result in positive impacts on the perception of an animation or a communication scenario. However, previous studies did not systematically compare gaze models to natural gaze or hybrid forms, and thus, indications about the level of adequateness and perceived realism of these models are still open questions that will be tackled in the present thesis.

Gaze Transformation and Hybrid Approaches

Previous research investigating social gaze focused on the replication of gaze behavior with avatars, (e.g., [109]), or artificial gaze models for agents solely driven by algorithms (e.g., [281]). An alternative line of research investigated whether transforming social interactions in virtual environments by decoupling the visual representation from the physical behavior can affect the interaction. As already mentioned in Section 2.4.4, Bailenson and colleagues investigated the idea of a non-zero-sum gaze, meaning a user-dependent simulation of another person's gaze [11, 24]. By augmenting gaze behavior, virtual environments can surrogate individual points of attention to each participant of multi-user simulations, and therefore foster a more positive perception of the interaction. For instance, one user can direct his/her gaze on multiple interaction partners simultaneously as each simulation can be rendered individually, which in turn can express increased attentiveness to each interactant. In a study with two participants and one presenting experimenter, Beall, Bailenson, Loomis, Blascovich, and Rex [24] found that participants directed their gaze toward the experimenter more often when non-zero-sum gaze was active (only directed gaze for each participant) compared to reduced gaze (only averted gaze) or natural gaze (approximated via head movements).

Although studies of TSI give indications of the potential impacts of modified behavior, such as increased agreement [11, 24], previous models rely on either linear manipulations or algorithms that do not account for interpersonal adaptations within the virtual interaction. Doing so, however, allows to merge both forms of embodied representations, avatars and agents, into hybrid forms (see Section 2.4.5). Hybrid models have been investigated with regard to their potential to evoke continuous presence [115], and in the context of establishing alternative communication channels on the basis of behavioral phenomena, such as eye contact, joint attention, and grouping (see the next Section 5.3).

The benefits of hybrid gaze approaches could be manifold. First, the introduction of augmented behaviors could compensate for the lack of sensor inputs and transmission errors by compensating for missing or interrupted behaviors based on a underlying social AI. Second, socio-communicative deficits could be studied, further extending "offline" paradigms [111] and allow to study these behaviors in interactions. Third, gaining insights into the effects of hybrid systems and their further development could foster the inclusion and training of individuals with social disorders. In the following, we present our approach end emprical validation.

Approach

Previous research focused on artificial gaze models for the animation of and interaction with virtual characters [180, 192, 176, 245, 159, 281], as well as for the compensation of missing sensory input in shared virtual environments [109, 108, 339, 30, 29]. Two major approaches to constructing artificial gaze can be derived:

1) a randomized gaze model based on statistical distributions of fixations and saccades, and 2) a randomized gaze behavior default transmission which can be augmented using a model respecting for social affordances to a synthesized gaze behavior, which could also compensate for transmission interrupts. These approaches assume a complete lack of gaze information, and could be considered agent approaches. For the compensation of non-adequate social gaze (e.g., in inter-cultural contexts, or with regard to social disorders), and to foster affinity and liking, 3) a third hybrid solution that uses natural gaze as default transmission but also reacts upon social affordances based on an underlying AI, which can be constructed with hybrid avatar-agent technologies (see Section 2.4.5). The present work systematically assesses to what degree adequate social gaze behavior in avatar-mediated communication can be established through such models.

5.2.1 Methods

We adopted a dyadic design with four conditions: "natural gaze, non-augmented" (natural/gold standard), "natural gaze, augmented" (hybrid), "random gaze, augmented" (synthesized), and "random gaze, non-augmented" (random/baseline control). In each dyad, two participants with the same biological sex were assigned to the same condition, and represented by avatars according to their biological sex.

Apparatus

A client-server architecture was developed in Unity3D (v.5.6.0f3). We tracked participants gaze using Tobii 4C eyetrackers[46] (90 Hz), attached to a 28" screen (1920px $\times$ 1200px). We used Sennheiser PC310 headsets[47] for audio transmission. A motion-to-photon latency measurement (240 Hz camera, 50 repeated measures) of the eye movement resulted in $M = 308$ ms ($SD = 33$ ms) latency. Audio latency was measured with source and client end-to-end audio recordings capturing impulse responses ($M = 281$ ms, $SD = 22$ ms). After participants were placed in remote rooms (see Figure 5.5), cartoon-like 3D avatars[48] that matched the biological gender of the participant were displayed against a black background (see Figure 5.5). We chose this type of avatar because it provides a clear gaze indication.

The application was based on five modules: 1) A speech to animation framework [49] was used to process audio input and approximate mouth movements. 2) Voice communication was established via voice over IP framework [50]. 3) Natural gaze/head movement tracking is performed using Tobii AB's Unity SDK. 4) A model-based reflex agent (social AI) to be able to transition between the default gaze state and the social gaze model (Figure 5.6). 5) An experimenter control interface for master control over the project components and the experimental procedure.

[46] Tobii Technology, Danderyd Municipality, Sweden `https://gaming.tobii.com/product/tobii-eye-tracker-4c/`
[47] Sennheiser electronic GmbH & Co. KG, Wedemark, Germany
[48] Faceshift AG, Switzerland, acquired by Apple Inc. in 2015, `https://www.apple.com/`
[49] SALSA, Crazy Minnow Studio, LLC, Cheyenne, USA
[50] Dissonance Voice Chat, Placeholder Software

Figure 5.5: Avatars and apparatus for studying hybrid gaze (reprinted from [268] ©2018 IEEE). *Top left:* Female and male avatars used in the study. *Top right:* The target head area of interest (head AOI) and eyes AOI used to assess the avatar/agent/hybrid gaze behavior as well as the subjects' real gaze behaviors during the interactions. *Bottom:* Study setup with users at maximum distance from the screen. The chairs were fixed to assure tracking quality.

Computed Gaze Models

The simulation transmitted speech directly to the interlocutors and rendered gaze onto their digital representation according to one of the conditions. For the *random gaze model*, fixations lasted between 1 second and 3.3 seconds. To indicate a fixation, a random (normal distribution) screen coordinate was generated by projecting ray from a fixed (user spawn) position to a random coordinate on the near camera frustum. The gaze direction (i.e., the eyeball joint rotation) of the avatar was then modified to be oriented in that direction. To approximate saccades, we used a linear interpolation of 20 frames (i.e., 0.33 seconds at a 60 Hz render refresh rate).

We conceptualized our *social gaze model* similar to the one proposed by [180]. Drawing from previous findings that listeners make more eye contact than speakers, we displayed behaviors depicting an attentive listener (see Figure 5.6). The idle state were the constructed random gaze behavior or the tracked natural behavior, respectively, depending on the condition. We detected listening and speaking behavior through audio input, similar to [109]. Once the user's voice reached an audio threshold, an event triggered the augmentation state. In the augmentation state, the gaze of the listener is directed to the screen center, in order to visualize listening focus (i.e., directed gaze) of the interaction partner on speaker. Starting with this directed gaze, the inter-saccadic interval was randomly selected from a normal distribution ($M = 3.97$ s, $SD = 0.78$ s) in order to prevent a staring effect

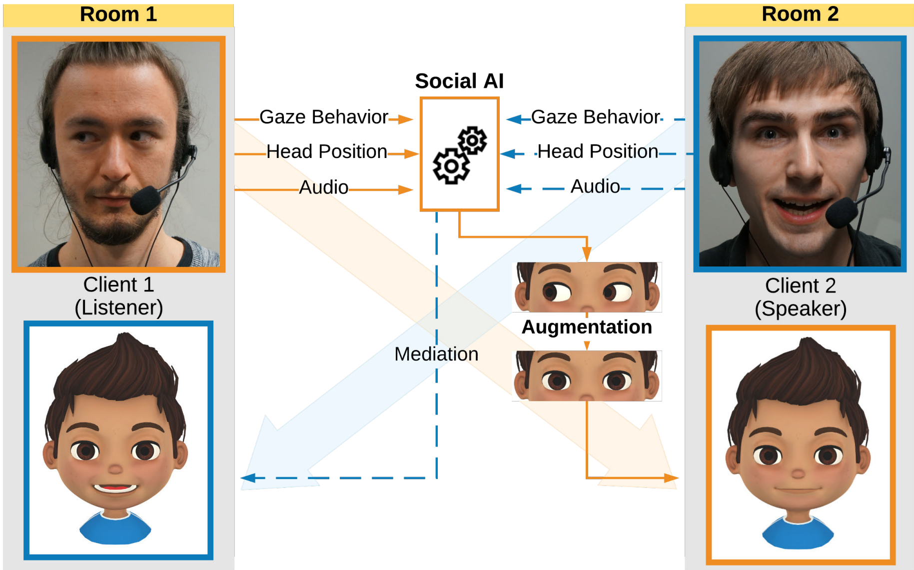

Figure 5.6: Functional principle of the hybrid gaze model (reprinted from [268] ©2018 IEEE). The model transmitted natural gaze and intervened according to appropriate social context. Depicted by the orange pathway, the social AI detected the status of a speaker via audio input. As an appropriate reaction, the gaze behavior of the listener was augmented (directed gaze), altering the natural gaze behavior.

and to simulate a more natural behavior. Upon reaching the directed gaze duration limit, a saccade was introduced, followed by a fixation on a random fixation point on the interlocutor's screen for a shorter interval (normal distribution, $M = 0.22$ s, $SD = 0.12$ s). All fixations were held stable for the time of the inter-saccadic interval. We chose this approach based on findings suggesting a preferred mutual gaze duration for dyadic settings is around three seconds [38].

This model results in virtual characters who face a speaking human to make eye contact, with occasional glances away. In the two augmented conditions, it is thus combined with either a natural or randomized gaze behavior default transmission

Procedure

The study was conducted at the University of Würzburg and took about 45 minutes. Participants were welcomed and quickly introduced to each other. Each participant was guided into a separate room and handed the study information and the consent form. Once participants agreed to participate by completing the consent form, they answered a pre-study questionnaire. Next, participants were instructed about the apparatus, equipped with headphones, and calibrated for eye tracking. The experimenter then started the simulation, which gave written instructions to the participants about their task for approximately 30 seconds ("In the following, you will see your communication partner visualized as a virtual character. Please have a conversation for the next five minutes and get to know each other, as if it is a normal conversation. When getting to know each other, you are free to choose any topics you want to talk about."), followed by the interaction, which lasted for five minutes. Afterward, the audio stream was cut off, and a visual text box popped up asking participants to wait for the experimenter. The participants were then presented with a questionnaire assessing dependent measures and, finally, debriefed in detail. The study was approved by the Ethical Committee of the Human-Computer Media Institute of the University of Würzburg. Student participants were compensated with study credit points.

Measures

Manipulation Check We introduced a *manipulation check*, assessing the amount of time (dwell time) the avatars established directed gaze (head AOI), eye gaze (eyes AOI), and background focus (background AOI). To do so, we calculated gaze vector ray cast hits of dynamic areas of interest (AOI, see Figure 5.5) in the simulation (60 Hz). Dwell times for the analyses were calculated exclusively (i.e., an eyes AOI hit would not change the head AOI statistics). As this measure was introduced during the course of data collection, the sample size was limited to $N = 58$.

Virtual Rapport We asked participants to judge their *(virtual) rapport* with their interlocutors [120] (e.g., "I felt I had a connection with my interaction partner") for each of the 11 items (1–strongly disagree, 7–strongly agree. Cronbach's $\alpha = .85$).

Social Presence We measured *social presence* with six items slightly adapted from [220] to fit the same response format (e.g., "To what extent did you feel able to assess your partner's reactions to what you said?" = "I was able to assess my interaction partner's reactions to my statements") using a the same scale (1–strongly disagree, 7–strongly agree; $\alpha = .83$).

Interpersonal Attraction We also assessed *interpersonal attraction* [76] using six agreement statements (1–strongly disagree, 7–strongly agree), that included items such as "I like my interaction partner" or "My interaction partner is friendly" ($\alpha = .86$).

Trust As trust has been shown to be affected by gaze in multiple studies [218], the *perceived trust* was measured with three items: "I think, my interaction partner has good intentions," "I would rely on my interaction partner," and "I would trust my interaction partner" (1–does not apply at all, 7–totally applies; $\alpha = .745$).

Behavioral Realism and Behavioral Naturalness To investigate the impact on the *movement realism and movement naturalness* of the perceived gaze, we presented participants two statements (1–doesn't apply at all, 7–totally applies): "The eyes of the virtual character moved realistically," and "The eyes of the virtual character moved naturally." To check for any impact on the perception of the avatar, we assessed the avatar with regard to the perceived *humanness* ($\alpha = .816$), and *eeriness* ($\alpha = .736$) using a semantic differential [136].

Resulting Participant Behavior Similar to the manipulation check, we assessed the actual *users' true gaze behavior* by dwell times for the head AOI, eyes AOI, and background AOI (see Figure 5.5). In addition to these measures, we assessed affect and the Big Five inventory. However, the results of the latter are not subject to the current thesis.

Participants

148 participants were recruited at the University of Würzburg. We excluded dyads when tracking failed for longer periods ($N = 28$) and when participants knew each other before the study ($N = 28$). The final sample consisted of $N = 90$ participants ($M_{age} = 22.01$, $SD_{age} = 3.20$). Of these, 48 were female, 89 were German, and 88 were students. Dyads were randomly assigned to one condition, and a chi square test showed that the distribution of gender was not significantly different.

Considerations for the Analysis

We assessed the need for a multi-level model [91] and tested for a change in the likelihood ratio. None of the tests revealed significant decreases, thus rejecting the need for multi-level modeling. To analyze the manipulation check and the participants' gaze behaviors, we calculated one-way ANOVAs with post-hoc comparisons (Tukey's Honest Significant Difference [HSD]). The assumption of equal distributions was assessed by visual inspection of a boxplot. To analyze the subjective outcomes, used the Kruskal–Wallis test. Subsequently, we conducted pairwise comparisons using Dunn's procedure [81] with a Bonferroni correction for multiple comparisons. Furthermore, we conducted a Jonckheere–Terpstra test on ordered alternatives to identify linear trends, ordered from the most natural to the most synthetic condition: "natural-gaze, non augmented" (*natural gaze*, $N = 24$, 12 female) "natural gaze, augmented" (*hybrid gaze*, $N = 22$, 14 female), "random gaze, augmented" (*synthetic gaze*, $N = 24$, 10 female), and "random gaze, non augmented" (*random gaze*, $N = 20$, 12 female).

5.2.2 Results

Manipulation Check

One-way ANOVAs showed significant main effects for the the background AOI $F(3, 57) = 96.63$, $p < .001$, $\eta_p^2 = .843$, the head AOI $F(3, 57) = 9.60$, $p < .001$, $\eta_p^2 = .348$, and the eyes AOI $F(3, 57) = 6.84$, $p = .001$, $\eta_p^2 = .275$. Pairwise comparisons showed that the random gaze evoked significantly more background attention than all other conditions and that with the hybrid gaze there was significantly less attention given to the background compared with all other conditions ($ps \leq .010$). Natural gaze was significantly lower in head AOI dwell time than synthetic gaze ($p = .019$). Random gaze, in turn, was significantly lower in head AOI dwell time than hybrid gaze or synthetic gaze ($ps \leq .001$). Natural gaze was significantly higher in eye AOI dwell time than random gaze ($p = .001$), and random gaze was significantly lower in eye AOI dwell time than natural or hybrid gaze ($ps \leq 0.03$). No further significant effects were found. The social gaze model seems to have successfully altered the gaze behavior behavior by directing the gaze when appropriate (see Figure 5.7), and therefore, the manipulation check was deemed successful.

Virtual Rapport

A Kruskal-Wallis test (equal distributions) showed that rapport scores were statistically significantly different between the conditions, ($\chi^2(3) = 9.13$, $p = .028$). Pairwise comparisons revealed a statistically significant difference in rapport scores between the random gaze condition ($Mdn = 5.14$) and the natural gaze condition ($Mdn = 5.73$, $p = .024$, see Table 5.2). In addition, a Jonckheere–Terpstra test showed that there was a statistically significant trend for rapport scores, $TJT = 1109.00$, $z = -2.941$, $p = .003$.

Social Presence

A Kruskal–Wallis test (equal destributions) did not yield significant results; $\chi^2(3) = 3.77$, $p = .287$. A Jonckheere–Terpstra test did not show a significant trend ($TJT = 1295.00$, $z = -1.593$, $p = .111$).

Interpersonal Attraction

A Kruskal–Wallis test (equal distributions) revealed that median scores were significantly different between the conditions ($\chi^2(3) = 9.13$, $p = .028$). Pairwise comparisons revealed statistically significant differences in interpersonal attraction scores between the random gaze condition ($Mdn = 5.14$) and the natural gaze condition ($Mdn = 5.73$, $p = .024$), but not between any other conditions. A Jonckheere–Terpstra test showed that there was a statistically significant trend, $TJT = 1029.50$, $z = -3.524$, $p < .001$.

Trust

A Kruskal–Wallis test (equal destributions) did not yield significant results ($\chi^2(3) = 5.25$, $p = .137$). A Jonckheere–Terpstra test did however show a significant trend $TJT = 1211.00$, $z = -2.230$, $p = .026$.

Behavioral Realism and Behavioral Naturalness

A Kruskal–Wallis test showed that judgments of behavioral realism were significantly different between the conditions, ($\chi^2(3) = 9.66$), $p = .022$. Pairwise comparisons revealed statistically significant differences in scores between the random gaze condition ($Mdn = 4.50$) and the natural gaze condition ($Mdn = 6.0$, $p = .026$), but not between any other comparison. Similarly, judgments of behavior naturalism were significantly different between conditions ($\chi^2(3) = 11.950$,

Table 5.2: Subjective Results (adapted from [268]).

	Natural gaze		Hybrid gaze		Synthesized gaze		Random gaze	
Measure	*Mdn*	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*	*M* (*SD*)
Virtual Rapport	5.73	5.72 (0.58)	5.64	5.60 (0.58)	5.64	5.38 (0.64)	5.14	5.05 (0.81)
Social Presence	3.83	3.75 (1.10)	3.75	3.65 (1.11)	3.67	3.65 (1.29)	3.17	3.13 (0.99)
Interpersonal Attraction	5.75	5.67 (0.50)	5.50	5.56 (0.65)	5.25	5.24 (0.56)	5.08	4.95 (0.80)
Trust	5.33	5.35 (0.63)	5.33	5.24 (0.64)	5.17	5.13 (0.73)	5.00	4.83 (0.74)
Gaze Behavior Realism	6.00	5.25 (1.29)	5.00	4.55 (1.60)	5.00	5.04 (1.46)	4.50	4.00 (1.49)
Gaze Behavior Naturalism	6.00	5.33 (1.20)	5.00	4.55 (1.68)	5.00	4.88 (1.48)	5.00	4.15 (1.18)
Humanness	3.08	2.97 (0.69)	2.67	2.64 (0.77)	2.83	2.81 (0.79)	2.75	2.61 (0.70)
Eeriness	2.75	2.71 (0.33)	2.88	2.72 (0.62)	2.81	2.70 (0.45)	2.75	2.73 (0.33)

$p = .008$) and pairwise comparisons confirmed a significant difference between the random gaze condition ($Mdn = 5$) and the natural gaze condition ($Mdn = 6.0$, $p = .005$). Jonckheere–Terpstra tests confirmed a significant linear trend for both, behavioral realism ($TJT = 1212.50$, $z = -2.258$, $p < .024$) and behavioral naturalness ($TJT = 1138.00$, $z = -2.849$, $p < .004$). Interestingly, the hybrid condition was slightly inferior to the synthesized condition but not to a significant level.

Humanness and Eeriness

No significant effects were found for humanness or eeriness.

Resulting Participant Behavior

A one-way ANOVA revealed a significant main effect for condition $F(3, 86) = 3.44$, $p = .020$. Pairwise comparisons showed a significant difference in dwell times between the random gaze (longest dwell time) and the natural gaze ($p = .047$), as well as between the random gaze and the random augmented gaze ($p = .044$) for the head AOI dwell time (see Figure 5.7). No further significant effects were found.

5.2.3 Discussion

Our goal was to systematically investigate gaze models in avatar-mediated communication. To do so, we implemented a model based on earlier approaches to augment social gaze [180, 109] and compared four gaze conditions: a) natural gaze b) hybrid gaze, c) synthesized gaze, and d) random gaze. We measured the impact of these models on five minute long social interactions with regard to perceived virtual rapport, social presence, interpersonal attraction, trust, behavioral realism, and naturalness of same-sex dyads. Furthermore, we evaluated the resulting visualized gaze behavior of each condition as well as the resulting gaze behavior of the participants.

Our results are in line with previous findings. Supporting our hypotheses, natural gaze was superior and random gaze was inferior to all other models with regard to the subjective measures. The linear trends found for virtual rapport, interpersonal attraction, and trust indicate that natural gaze suffered from artificial manipulation,

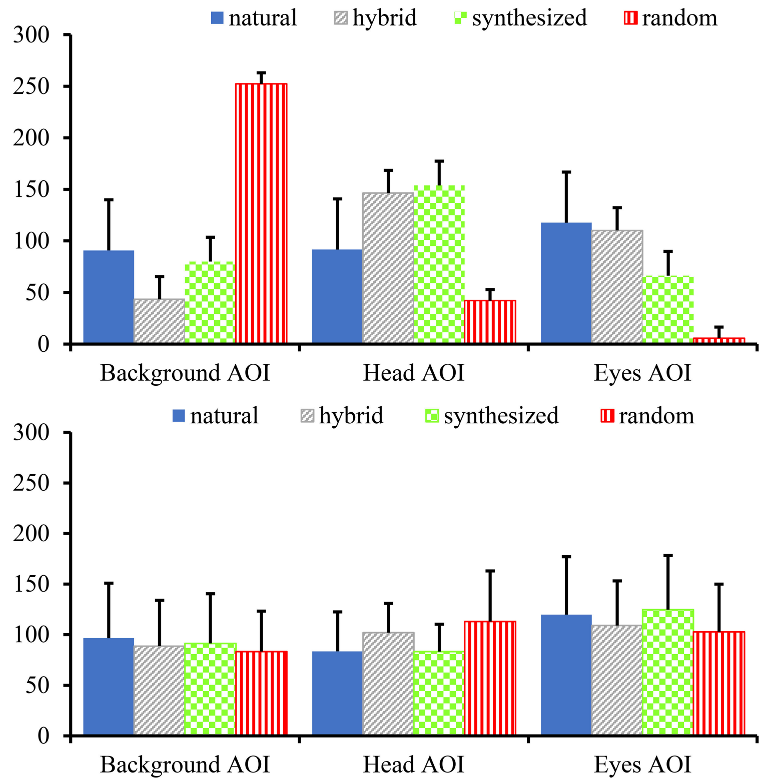

Figure 5.7: Dwell times for each AOI for the avatar behavior and the human user behavior (adapted from [268]). *Top:* Virtual character behavior. Mean dwell times evoked by the behavior ($N = 58$) displayed by the virtual character (respectively, each gaze model) in each condition. *Bottom:* Human behavior. Dwell times of the participants ($N = 90$) for each condition. Dwell times are displayed in seconds. *Note. Bars denote the mean value. Error bars denote standard deviations. Missing samples (human behavior) were discarded.*

whereas the random gaze behavior that was constructed based on statistical distributions of saccades and fixations benefited from the additional social behavior introduced by the model (synthesized gaze). Exploratory ANOVAs conducted for these measures led to similar results. Therefore, we can conclude that both the synthetic and the hybrid gaze models that establish social gaze contingencies based on typical nonverbal patterns are superior to purely random models.

Some interpretations arise from these findings. First, it seems logical that the statistical distribution of behaviors resulting from the default random gaze (i.e., synthesized or random non-augmented) is imperfect in comparison to the natural gaze. However, the results do show that, to some extent, social gaze contingency can be established through a simple rule-based system, which has now been quantified by the results of the study. Thus, supporting evidence is provided that displaying attention and interest (i.e., directing the gaze to the interlocutor while listening) is an important social reaction.

Another interpretation of this finding is that the gaze model we used and the resulting behavioral blending and animation were not sophisticated enough to seamlessly blend with the gold standard of natural gaze behavior. Interestingly, the results for the perceptual judgment of realism and naturalness interestingly give a hint that the hybrid gaze model, meaning the natural gaze transmission that was augmented by the social gaze model, led to a lower rating in realism and naturalness (non-significant) than either the pure natural gaze model or the sophisticated agent (i.e., synthesized) model. This could be a similar effect to the one mentioned in [339], which is a consistency break in realism. However, the effect we found is limited to the behavioral level and might point at a problematic break in behavioral characteristics of gaze, meaning that the model that we applied to augment social gaze in the hybrid condition did not seamlessly blend into the natural gaze behavior. Further development for not only for purely artificial gaze models but also for hybrid models should, therefore, take consistency into account as an important factor. Furthermore, future developments for VR simulations of the model should also use the eye positions of the user as a specific target point for induced eye gaze.

The results of the behavioral analyses indicate that participants in the random gaze condition tried to establish social gaze contacts via directed gaze, as shown by high dwell times in the head AOI. One interpretation of this is that participants initially tried to initiate social gaze contact (i.e., directed gaze) by reflex, but when no adequate reaction resulted, this pattern was followed. Another way of interpreting this finding is that participants tried to make sense of the random movements because there was a to lack of variation, which could then have acted as a distractor and caused additional mental demand. However, we did not assess any measures that could support this interpretation and the behavioral analyses do not provide a clear image due to the slightly reduced focus on the eyes AOI.

Limitations

Some limitations exist in our study. First, the sample was relatively small and populated with typical developed individuals, in our case mainly students, which makes it difficult to generalize the findings. Second, we did not use virtual reality hardware, an immersive setup, or a fish tank VR paradigm in our study. The findings should therefore be interpreted with care in terms of more immersive systems because the Mona Lisa effect (the impression of eye contact irrespective of the participant's perspective) could have had an impact [261]. Third, our augmentation only altered the direction of the gaze, not the direction of the avatar's head, and this was only aimed augmenting a directed gaze and not for establishing eye contact. Fourth, we did not assess interactions with more than two users, such as, for example, Ding, Zhang, Xiao, and Deng [78], which provides room for further research. Fifth, we used cartoon-like characters that could have biased impression formation [339], and this factor may change with more realistic characters.

Conclusion

This study investigated augmented gaze and quantified the resulting impact in comparison to natural, random, and synthesized gaze. The manipulation check indicates that the gaze augmentation model successfully altered the behaviors toward what was intended. Natural gaze (i.e, the gold standard) did not benefit from the augmentation, whereas the social gaze model was an improvement compared to a random model. The presented approach provides a valuable research tool for the future exploration of nonverbal behavior in avatar-mediated communication. We conclude that social gaze models that are based on available modalities could be beneficial for the development of social virtual environments, such as to cope with the lack of sensory inputs to track and replicate gaze behavior in order to substitute behavioral channels or to compensate for dropouts in data transmission. To this regard, future research should explore alternative social messages that do not reflect natural behaviors, which is the target of the following Study.

5.3 Study 13: Visually Augmented Social Interaction

Human communication is characterized by a multitude of social behaviors. Participants adapt and coordinate feelings, intentions, and actions with others [107]. They shake hands, establish eye contact, move closer to each other, or mimic their interaction partners to create liking, rapport, and affiliation [168] based on a continuous processing of signals on a conscious, as well as subconscious, level. The perception of and response to social signals happens "in accordance with an elaborate and secret code that is written nowhere, known by none, and understood by all" [284].

Humans process sensory information, such as visual information, from social cues and behaviors, based on higher level (top-down) processes such as expectations, previously acquired knowledge, and the use of contextual information [124], as well as based on bottom-up processes such as stimulation, sensation, and the respective direct information processing [116] (see Section 2.1.1). However, CMC systems often lack the ability to accurately track and reproduce the important details of social cues. Technical systems will always be subject to potential inaccuracies such as those caused by noise [291]. Despite recent progress, full behavioral realism is currently not available in consumer VR products, and, as Slater stated "The goal of VR to accurately simulate all aspects of reality is physically infeasible" [298].

It is as yet unknown how these shortcomings affect communication in social VR and if potential countermeasures can be provided by the same technology. For instance, humans can compensate for the lack of social cues available in CMC by shifting their attention to, or decoding/encoding social information through other channels, such as using smilies in text-based communication to display mood or humor [351, 352]. This indicates that humans have the capabilities to encode/decode social information into or from alternative communication channels and cue presentations, which prompted the general idea of the present paper. In this regard, VR provides communication possibilities that substantially exceed the mere replication of existing channels from the physical world. In VR, representations can generally be decoupled from behavior [12], and cue representation can be manifold. In conclusion, we argue that VR applications have unlimited potential to *extend* and *transform* the reality of physical communication with regard to the information perceived and displayed. The exploitation of this potential defines the overall research goal of the present work.

This study contributes by exploring these possibilities using a novel approach to augmenting social behavior. We designed three visual transformations for behavioral phenomena: 1) *eye contact*, 2) *joint attention*, and 3) *grouping* and evaluated their impact on social interaction in a shared social space (a virtual museum). Our approach differs from previous work in that designed augmentations for interactional behavior phenomena, and our results highlight the potential of VR to enhance CMC scenarios beyond simple replication of social cues from the physical world.

Appropriate Representations of Appearance and Behavior

In avatar-mediated communication, avatars provide user embodiment and act as the users' virtual representation. User embodiment can be referred to as "the provision of users with appropriate body images so as to represent them to others (and also to themselves)" [28, p. 242]. The question is however, what is an appropriate body image to represent users? For example, is an artifical character that does not render user-specific social appearance cues and does not replicate physical user movement appropriate? These questions become especially important in contexts, where the physical appearance and behaviors cannot be replicated. To this regard, Nowak and Biocca [220] found that, counterintuitively, a low anthropomorphic (simplified) avatar evoked more social presence and copresence than did a high anthropomorphic avatar or no avatar. Bente, Rüggenberg, Krämer, and Eschenburg [35] did not find significant differences between low- and high-fidelity avatars. While in no behaviors were replicated in the study from Nowak and Biocca [220], the avatar used in the study from Bente, Rüggenberg, Krämer, and Eschenburg [35] could replicate the users gaze and gesture. These results point to behavior replication being the dominant factor in comparison to the form of visual representation. In Study 10 (see Section 4.3) we found that a larger degree of behavior replication led to a greater perception of telepresence. Yet, we did not find significant impacts on social presence. With regard to the guidelines to the behavior of humanoid agents, Bradler et al. [10] provided an overview of creating, simulating, and animating humans (i.e., human figure models) and also presented a set of desiderata that included "A human model should move or respond like a human" and "A human model should have a human-like appearance." In order to generate nonverbal behavior for expressive and conversational agents, Cassell, Pelachaud, Badler, Steedman, Achorn, Becket, Douville, Prevost, and Stone [62] presented a system for the rule-based generation of facial expressions, lip motions, eye gaze, head motions, and arm gestures based on conversations created by a dialogue planner. Vogeley and Bente [342] stated that artificial humans of the future should also take into account "the emotional and relational aspects of communication with an emphasis both on understanding and production of nonverbal behavior" [342] including phenomena such as interpersonal synchrony. An approach used by Gratch and colleagues is to create virtual rapport, by, for example, adapting behavioral reactions [121, 122]. These works, however, mainly considered agents. Considering avatar-mediated interactions and simulations that do not allow to replicate the user's appearance or behavior based on appropriate sensing or tracking technology, we presented two approaches to simulate artificial behavior in Studies 11 and 12 that could eventually be used to cope with the lack of sensory channels, namely the construction of artificial social gaze and the injection of artificial mimicry. These were implemented on the basis of natural human behavior. In this study we asked whether we can find alternative social affordances that stimulate virtual interactions through hybrid technologies.

Substituted and Amplified Behavior through Hybrid Technologies

In Sections 2.4.4, 2.4.5, 5.1.2, and 5.2 we presented related works that describe the presentation, augmentation or transformation of human behavior through hybrid technologies VEs, for example to achieve a continuous presence, artificial attention to multiple listeners [24], or the amplification of facial expressions with the case of smile behavior [224]. Without doubt, these artificial modifications of the nonverbal exchange can impact the perceptions of the social interaction. For example, the amplification of smile behavior in the study by Oh, Bailenson, Krämer, and Li [224] led to more social presence, more positive affect, and participants had an overall more positive impression of the interaction. The work of Bailenson, Merget, Yee, and Schroeder [15] are specifically interesting in the context of the present study. They focused on the abstraction of communicative behaviors, namely facial expressions, to a rectangular polygon avatar that was low in form realism. Using facial expression tracking, they changed certain factors of form and color of the avatar. For example, the more a user smiled, the more yellow was introduced to the avatars overall color. Vice versa, the more a person frowned, meaning, making a sad expression, the more blue was introduced to the avatars color. Furthermore, the width end height of the avatar followed the width and height of the users mouth. While these abstractions were inferior in the recognition of emotion compared to voice only or a video conference, and the avatar condition was inferior with regard to the perception of copresence, their results also showed that interaction partners disclosed more information to the abstracted avatar, both verbally and non verbally. Their approach points out an important aspect: Not only the appearance of avatars can be abstracted, but also their behavior.

In this study we specifically address phenomena that are dependent on multiple users. Thus, compared to our approach, the above studies did not respect the contingencies of interactional behaviors but rather focused on general transformations. Similar to Oh et al., our approach for the behavioral augmentations also infers an amplifying character, which is why we hypothesize that *(H1) the augmentation of social behaviors by amplifying or substituting transformations increases social presence.* While [62, 121, 122] investigated agent behavior, we aimed at transforming avatar behavior on a phenomenological level. Our approach did not aim to include static expressions or replicate behavior/appearance cues in a direct sense. Instead, our goal was to utilize the flexibility of VR to enable message exchange and interpretation by transforming and visually amplifying social phenomena that had not yet been investigated. To that regard, the work of Boker, Cohn, Theobald, Matthews, Brick, and Spies [46] and Oh, Bailenson, Krämer, and Li [224] showed that augmented behaviors may also impact the resulting behaviors of the interactants. Considering our approach a hybrid technology that interprets and modifies social interactions we hypothesize that *(H2) the augmentation of social behaviors impacts the respective social behavior of users,* that is, eye contact, joint attention, and grouping behavior.

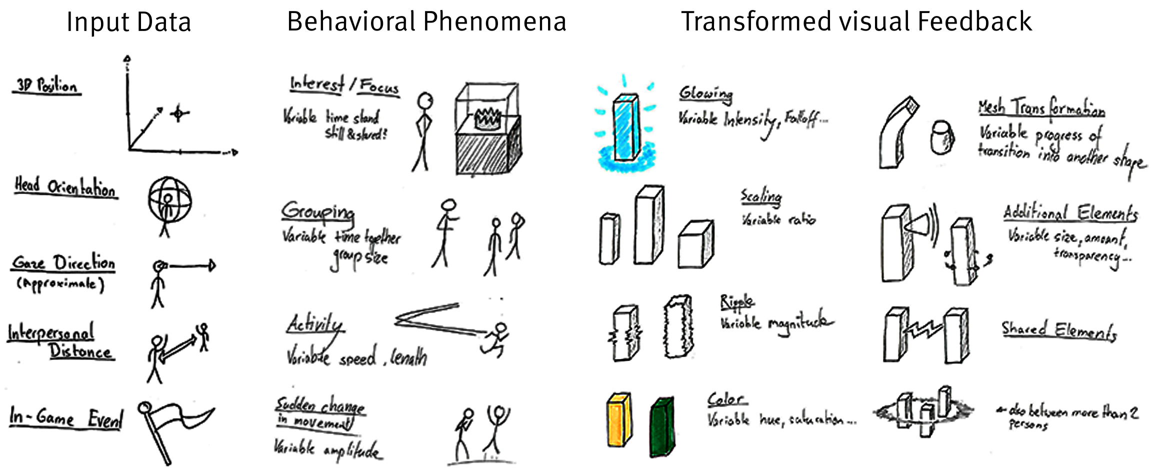

Figure 5.8: The design space considered for the visual augmentations (reprinted from [266] ©2017 IEEE). We explored data that we could access, what intermediate behavioral phenomena we could potentially augment, and what visual effects we could implement.

To a certain degree, we reverse the rational stated by the SIPT [351] by actively implementing compensation mechanisms.

5.3.1 Design and Implementation

We created a design space for potential augmentations that was restricted to translation (x,z of the transverse plane) and orientation (x,y,z) data input (see Figure 5.8). Our goal was to find constraints relating the input, the intermediate behavioral phenomena, and visual abstractions for the transformation, amplification, and substitution of behavioral patterns. We decided on three augmentations for social phenomena: *eye contact* which was augmented with floating bubbles; *joint attention* which was augmented with object highlights; and *grouping* which was augmented with color changes (see Figure 5.9). The three augmentations were chosen to cover multiple dimensions (bidirectional, environmental interactivity, multi-person), and all were based on visual feedback and possessed characteristics that could be described as substitutionary, amplifying, and transformational in their characteristics. To prevent third variable bias, we chose a reduced avatar model. In the following sections, we describe the decisions about avatar appearances and review each transformation in detail.

Avatar Appearance

According to Watzlawick, one "cannot not communicate" [355], meaning that every present behavior or social cue will have and unique specific interpretation for the interlocutors and thus will be interpreted and affect the interaction. Our study aimed to investigate the impact of behavioral augmentations in a controlled way. To avoid any bias from artificial, non-reproducible social or behavioral cues such as appearance, postures, facial displays, or gaze displays, participants were represented as featureless cuboid gray pillars by default. This avatar representation specifically avoids any additional artificial social information that may be derived from a more humanoid or realistic representation and that could influence the participants' perceptions based on direct social information processing or contextual norms [116, 124, 271]. For example, Bailenson, Beall, Loomis, Blascovich, and Turk [11] acknowledged that having avatars with eyes but no replication of eye movement is problematic. We therefore did not render avatar eyes but rather use an approximation (the head direction) to derive visual transformations. Participants immersed in the simulation could determine the forward direction of other participants by their locomotive behavior and the location of noises as well as voice location during verbal exchanges, which we perceived sufficient during pretesting. In addition, pillar size was uniform for all participants (50 cm $\times$ 50 cm $\times$180 cm), because height can be a strong cue for the perception of dominance [204]. Furthermore, the virtual camera was adapted to fit the height of each participant.

Eye Contact

Eye contact is important for social interaction [7]. It acts as a form of contact establishment and it signals that interlocutors pay attention to each other. Although different in their anatomical nature and precision, a user's head direction is typically a good indicator of the attentional focus [185]. Similar to our study, Bailenson, Beall, Loomis, Blascovich, and Turk [11] used head direction to describe and render gaze in a virtual environment. As head and eye direction are highly correlated, it served as an approximate.

Early prototype tests identified form, color, and frequency as the most important aspects for a potential directed gaze augmentation, whereas, for example, a spike-like particle system was perceived rather negative. We chose a shared-elements visualization that was perceived as soft and related to the idea of exchanging thoughts or gazes, according to pretesters (see Figure 5.9). The floating bubbles used to augment eye contact were semi-transparent, and we adapted the frequency and amount based on pretester feedback. We chose a light pink/magenta color (RGBA: 255,159,197,168) as this color is associated with harmony and can be perceptually located as high in activity, low in weight and moderately warm [228]. To identify approximated eye contact between two participants (i.e., two participants focusing on each other), we built a ray-cast map that gathered all of the objects in focal area of the participants (active exhibits as well as other avatars within a predefined distance of 4m). In a second step, we checked all value pairs to determine whether or not mutual gaze was present and evoked the augmentation effect if eye contact pairs were found.

Joint Attention

Joint attention is a phenomenon that develops in infancy and refers to shared attention or focus toward an object [209]. Initiating joint attention shows the desire to share a pleasurable experience with others. It therefore inherits "processing of information about the attention of self and others" [213, page 269] and signals a common interest or a common point of reference [213]. As joint attention includes an interactive process with the environment, we designed the transformation to be a particle system that highlighted an object and that appeared if two or more participants were within the 4 m social distance and were focused (determined by head direction) on the same object (see Figure 5.9). These appearing particles had a small movement radius concentrated along the up axis until disappeared. Other prototypes, such as having the object actively change color, were perceived as rather irritating by pretesters and also would have changed the character of the exhibit. Sixteen exhibits (active objects) in the virtual museum were capable of evoking the transformation. We built a ray-cast map to collect object hits throughout the simulation. When duplicates were found, we evaluated whether or not the watchers were within the social distance and if so, the highlight would appear.

Grouping

Grouping is a spatial behavior derived from proxemics [6] that encodes group affiliation, intimacy, or power [4], and is associated with interpersonal attraction. Humans form more positive attitudes towards ingroup members [137], and with regard to a distinct communicative aspect, spatial movements often indicate the beginning and ending of interactions [6]. Hall [129] differentiated between intimate (0.15 m–0.45 m), personal (0.38 m–1.22 m), social (1.22 m–3.66 m), and public (3.66 m–7.62 m+) distances, and it was shown that participants in VR execute proxemic behaviors similar to those in the physical world [14]. In our approach, we considered the social space (<4 m) to be the dimension for groups, as the pillar avatars are slightly more extensive in dimensionality than the human users (which aimed at avoiding bumping into each other).

Our grouping transformation was chosen with regard to group identification (i.e., appearance) and therefore we used color changes to identify group formation and group members, aiming at amplifying the grouping effect. Participants within a radius of the social distance were grouped together. To further promote and signal the initiation of a group to a participant, we implemented a fade in/fade out color effect rendered to the viewport (the group's color as a visual flare for a duration of 2 seconds; see right side of Figure 5.9). Our grouping algorithm was adapted from the k-means algorithm [144] and used the distances between all participants. Each participant was assigned an internal predefined hue, saturation, value (HSV) color. The neutral gray value was 0,0,0.8 in HSV coordinates, and the respective color values are H,0.5,0.8. In order to avoid changes in brightness, only the hue value changed. With each group formation, the group color was determined by the constellation of group members. We did not take into account any psychological effects from color perception in the grouping metaphor because no systematic impact was expected due to the manifold possible constellations. To avoid disturbing and rapid color changes, group constellations had to hold stable for two seconds until the visual coloring was applied. Figure 5.11 shows the augmentations as they would have been experienced by the participants in the simulation.

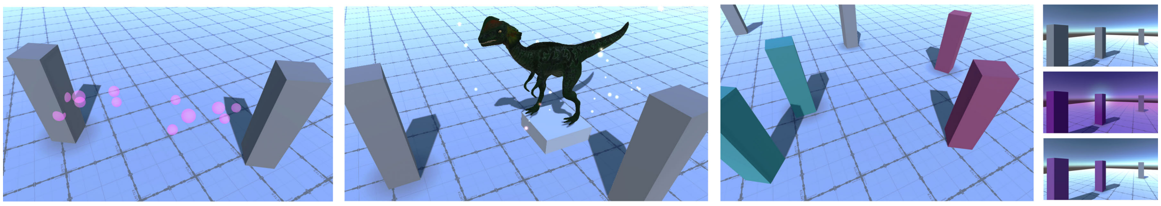

Figure 5.9: Isolated display of the social augmentation effects (reprinted from [267] ©2018 IEEE). *Left:* The eye contact visual transformation in the form of floating "bubbles." *Middle:* The joint attention augmentation using a particle system highlight. *Right:* The grouping augmentation using group colors and visual effects when joining a group.

Virtual Museum

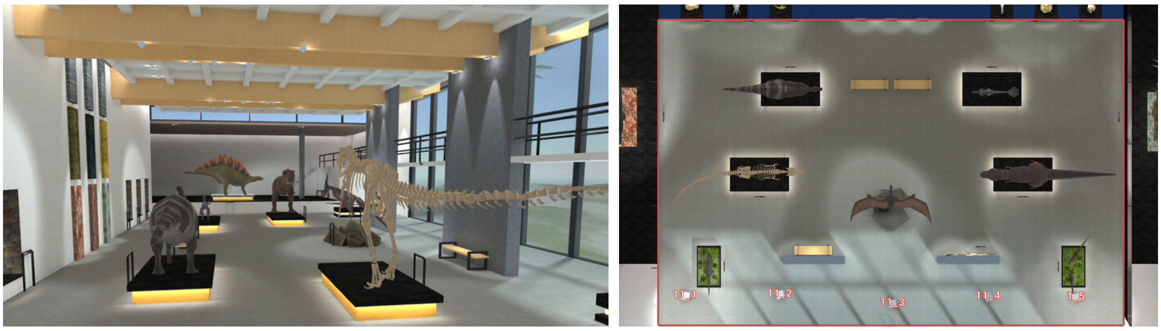

Figure 5.10: The virtual museum environment and the start positioning (reprinted from [267] ©2018 IEEE). *Left:* Side view of the final museum environment used for the study. Six main exhibits included audio information. *Right:* Birdseye view on the experimental simulation in the start position. The active walking area is marked in red.

Our main goal in the scenario design was to find a shared social space that inherited affordances for interactions with the environment as well as possibilities for social interactions and the freedom to explore and interact within a large space. We decided to use a museum setting, as this represents a shared social space with larger dimensions. Observing a medium-sized physical nature museum (Senckenberg museum) in Frankfurt, Germany, for one day, we found that all interactions that we included in our augmentation set (grouping, joint attention, directed gaze) were performed by people visiting a museum, which is why we found a museum scenario to be a valuable use case. We chose the topic of dinosaurs as we found that the subject of primeval eras is taught in early education and paleontologic exibitions are a topic all participants can relate to in a similar way.

Figure 5.10 depicts the virtual museum. Its virtual dimensions were identical to the measurements of the tracking area (20x30 m). Among sixteen active objects, six large dinosaur exhibits included prerecorded audio containing information extracted from Wikipedia. The recording could be triggered by selection via gaze (head direction) by looking at a speaker icon label for 2 seconds, while a loading spinner element appeared. Each audio clip was about 1 minute long. Once the audio information was activated by a participant, all users within a 4 m radius could also hear the information via a one-ear headset, similar to an audio guide in a physical museum.

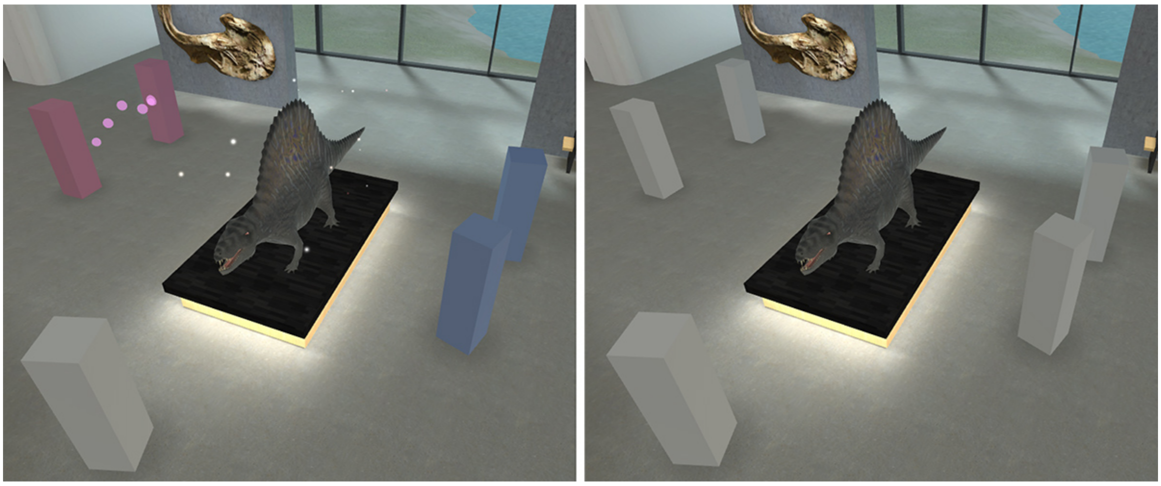

Figure 5.11: Illustration of the experiment conditions (reprinted from [267] ©2018 IEEE). *Left:* Condition with transformations for eye contact (floating bubbles), joint attention (particle highlights on object), and grouping (avatar colors). *Right:* Condition without transformations.

5.3.2 Methods

Design

The study was conducted in a between-subjects design comparing the conditions "augmented" (active augmentations for grouping, direct gaze, or joint attention; see Figure 5.11 left) and "non-augmented" (see Figure 5.11 right). User groups experienced the museum either with or without active behavioral augmentations. The participants were blind to the actual goal of the experiment.

Task

For each assessment, five participants were grouped together and were told to explore the museum and to learn about the exhibits in a natural way, as they would in a physical museum. The participants were advised that they could move freely and were free to interact with each other, but that they should not to pass through virtual walls or objects and or move further than the dimensions of the virtual museum to ensure safety. Participants started from defined start positions, as depicted at the bottom of Figure 5.10.

Procedure

Participants gave their informed consent and were assigned a random number from one to five in order to ensure a correct relationship measures. We then asked the participants to fill out the pre-study questionnaire (demography, personality, media habits, interpersonal relationships, simulator sickness). Next, for the main part of the experiment, we equipped each subject with an HMD, tracker, and audio earphone and instructed them how to recalibrate their HMD (by pushing a button and taking three to five steps in a straight line) in case they experienced perceivable drift. The participants were given oral instructions, based on a script, about their upcoming task, and then we guided them to their start positions (see Figure 5.12, top) and gave them about 30 seconds of acclimatization time in a slighly detached anteroom of the museum (see Figure 5.10). We started data logging and the experiment with an oral "go," and then the participants had 15 minutes to explore the virtual museum (see Figure 5.12). At 15 minutes, we stopped the trial. Participants could raise their hand, if they experienced any problems during the exposition (e.g., drift), and one of the experimenters assisted them. A recalibration took approximately 15 seconds. After the exposure, we asked participants to fill out the post-experimental questionnaires (dependent measures). Finally, we debriefed the participants and compensated them with either sweets or class credit points. The overall experiment approximately took about 1 to 1.25 hours time. The experiment was conducted at the Fraunhofer IIS in Nürnberg. The study was approved by the ethical commission of the institute for Human-Computer Media of the University of Würzburg.

Apparatus

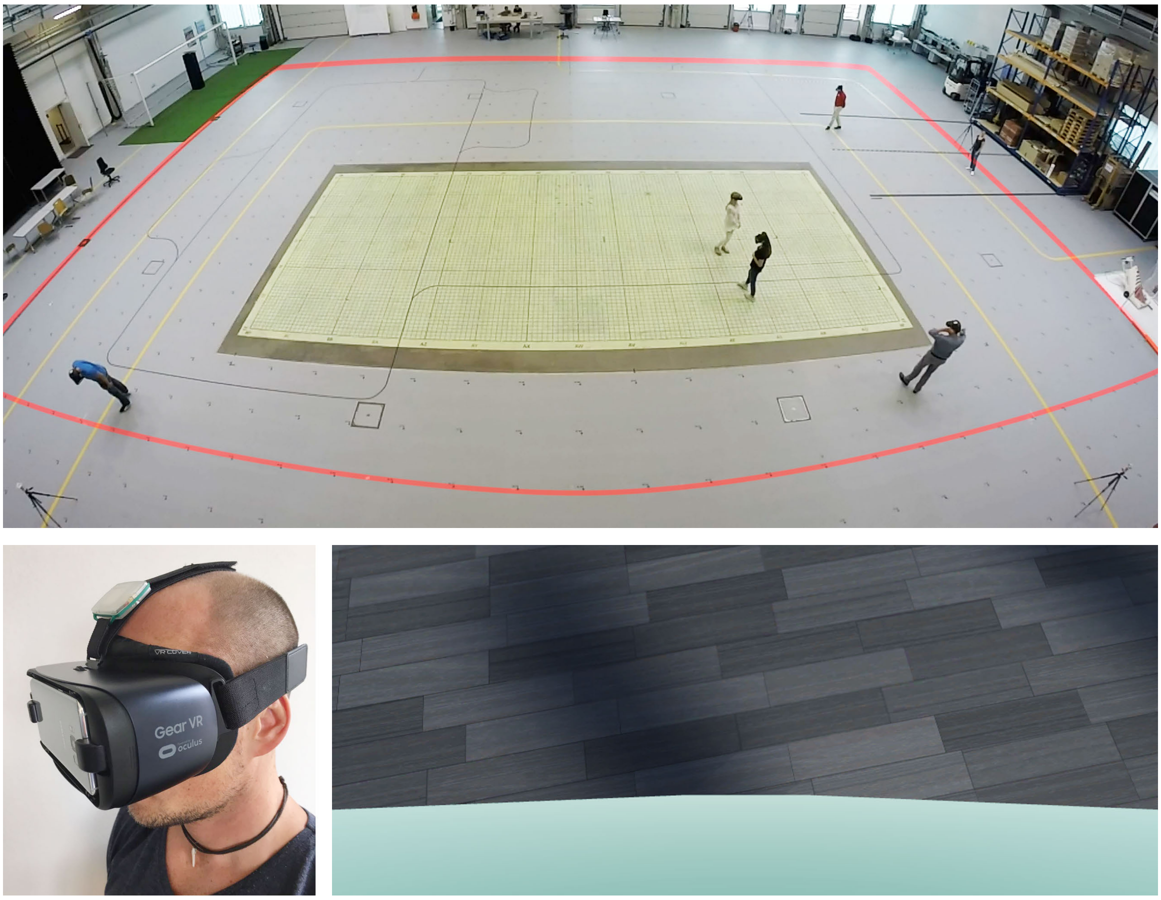

Figure 5.12: Tracking environment, HMD with RF transmitter and first person view on the avatar as seen in the simulation (reprinted from [267] ©2018 IEEE). *Top:* Overview of the tracking space and apparatus. Five participants were immersed simultaneously. *Bottom left:* User wearing the HMD. An RF transmitter was attached to the HMD for position tracking (transverse plane). *Bottom right:* User embodiment (first-person look at the own avatar).

The simulation was implemented in Unity3D[51] using a server-client network architecture. The simulation for each participant was rendered by and displayed on Samsung S7 and S8 smartphones[52] that was attached to a GearVR[53] HMDs. Audio information was delivered by a Beyerdynamic DT 1[54] single-ear headphone. We used a large-scale radio frequency-based real-time location system (RTLS) operating in the gigahertz band to cover a tracked area of approximately 20x30 m (see Figure 5.12) [343]. To limit the load on the wireless transmission bandwidth and assure a constant stream, the positioning system was used with a 20 Hz tracking refresh rate. To smooth the visual simulation, we interpolated these data with a spring-damper like function over three visual frames. The absolute position tracking had a circular error probable in 95% of 22.4 cm. The RTLS position data was combined with the Gear VR rotational tracking. To calibrate and align positional and rotational tracking, we used a short calibration routine recording the user's positions when walking forward in a straight line, deriving a trajectory vector that was then used to correct the orientation offset. As especially the S8 mobile phone orientation data tended to drift over time, participants had to occasionally recalibrate their simulation (between approximately zero and two times during an exposure).

Measures

Subjective Measures We measured *social presence*, *self-reported copresence*, *perceived other's copresence*, and *telepresence* [40] to test the impact of the augmentations. We adapted the scales from [220] and reformulated them for a multi-user scenario (i.e., "my interaction partner" = "my interaction partners"), and the items were assessed using a 7-point scale (1–strongly disagree, 7–strongly agree; respectively 1–not at all, 7–very much). To determine impacts on general impressions, we measured *enjoyment, lasting impression, thought-provocation, suspense,* and *artistic Value* using the scale from [225] (1–strongly disagree, 7–strongly agree). In addition, participants could add qualitative comments concerning the experience. We also assessed *rapport* and *group accord*, which are not the focus of the present reporting.

Behavioral Measures To investigate *H2*, we developed objective behavioral measures. We assessed the amount of time that *eye contact* occurred, the amount of time that participants were *looking at other avatars*, the amount of time that participants were *looking at any objects* (dinosaurs, exhibits, other participants), the amount of time that *joint attention* occured, the average *interpersonal distance* to all other group members, and the length of time that the participants were *grouped* (as assessed by our algorithm). To gain better insights into the behavior over time, we divided the full exposure data (15 minutes) into six time slices of 2.5 minutes each for the

[51] Unity Technologies, `https://unity3d.com/`
[52] Samsung Electronics Co., Ltd., Suwon, South Korea, `https://www.samsung.com/de/smartphones/galaxy-s7/overview/`
[53] Samsung Electronics Co., Ltd., Suwon, South Korea, `https://www.samsung.com/de/wearables/gear-vr-r323/`
[54] beyerdynamic GmbH & Co. KG, Heilbronn, Germany, `https://www.beyerdynamic.de/dt-1-3808.html`

analyses. The data were logged using a data logger running on an experiment server at 20 Hz.

Control Measures We introduced control measures to avoid bias from any third variable. Using the *subjective closeness index* ($r = .916$) [103], we asked all participants to evaluate their relationship to all other group members by answering two questions on a scale from 1–not close at all to 7–very close. The subjective closeness indexes of each subject for the other four group members were then averaged. To control for differences in personality, we measured the *Big Five Inventory-short form* (all rs $> .203$) from Rammstedt and John [247]. We further assessed the *simulator sickness* using the Simulator Sickness Questionnaire [156] before and immediately after the task and asked participants to comment if they experienced visual problems.

To cover our actual interest and to test for cognitive distraction, knowledge variables were assessed. Specifically, one can regard reduced levels of knowledge acquisition as a hint that there was some kind of cognitive distraction. That is, an augmentation condition runs the risk of demanding extra cognitive capacity that therefore can no longer be allocated to the information provided in the virtual environment. We measured *subjective knowledge* ($\alpha = .805$) and *objective knowledge* using the procedures from Raju, Lonial, and Mangold [246] and Schneider, Weinmann, Roth, Knop, and Vorderer [286] respectively, after adapting them to fit our stimulus (i.e., "I felt well-informed by the video" = "I felt well-informed by the museum", 1–strongly disagree, 7–strongly agree). To assess objective knowledge, we extracted facts from the audio recordings to create five multiple choice questions.

Participants

We tested a total of 37 groups. We removed nine groups from the analysis because of technical problems or a reduced number of participants, when one or more participants did not show up. We exluded one group because a participant experienced strong sickness, one group because a participant was aware of the experimental goal, and one group because participants did not fulfill the task. The final sample consisted of $N = 125$ participants (41 female, all others male) in 25 groups. Of these, 65 participants were assigned to the "augmented" condition, and 60 participants were assigned to the "non-augmented" condition. Conditions were randomly assigned. The mean age of the participants was 32.34 ($SD = 10.64$); 75 participants were employed, while 44 were students; and 79 participants had previous experience with VR (17 had 10 or more previous experiences). The number of previous VR experiences did not differ significantly between the conditions.

5.3.3 Results

Subjective Results

To investigate *H1* and the impact on other presence factors, we conducted independent samples t-tests. The results showed a significant difference in perceived social presence between the augmented condition ($M = 3.58$, $SD = 1.24$) and the non-augmented condition ($M = 3.14$, $SD = 1.06$; $t(123) = 2.165$, $p = .032$, $d = .38$) indicating that participants in the augmented condition perceived more social presence. Furthermore, we found that participants in the augmented condition found the experience more thought-provoking ($M = 4.36$, $SD = 2.92$) than participants in the non-augmented condition ($M = 3.47$, $SD = 1.27$; $t(121) = 2.142$, $p = .034$, $d = .39$). Results of the subjective measures are presented in 5.3.

Behavioral Results

The six time slices of the behavioral measures are presented in Figure 5.13. We calculated mixed ANOVAs with the condition serving as the between-subject variable and time serving as the repeated-measurement variable. We found significantly longer eye contact in the augmented condition ($M = 17.06\ s$, $SD = 14.48\ s$) compared to the non-augmented condition ($M = 11.22\ s$, $SD = 15.89\ s$; $F(1, 123) = 4.61$, $p = .034$, $\eta_p^2 = .036$). Pairwise comparisons showed significant differences in slices 1 and 4 ($ps < .01$). Overall, participants in the augmented condition focused on avatars longer ($M = 81.08\ s$, $SD = 57.13\ s$) compared to participants in the non-augmented condition ($M = 63.65\ s$, $SD = 52.64\ s$). While the comparison over the full time of the experiment was not significant ($F(1, 123) = 3.13, p = .079, \eta_p^2 = .025$), we found differences to be significant in slice 1 ($p = .013$) and slice 4 ($p = .009$) assessed by pairwise comparisons. Participants in the augmented condition focused on active objects (other avatars, exhibits; $M = 415.72\ s$, $SD = 78.10\ s$) for a significantly longer time than the than participants in the non-augmented condition ($M = 387.45\ s$, $SD = 77.73\ s$; $F(1, 123) = 4.106$, $p = .045$, $\eta_p^2 = .032$). ANCOVA calculations using previous VR experiences as the covariate did not change the subjective or behavioral results significantly.

Control Measure Results

Independent samples t-tests for the subjective closeness index and the Big Five factors did not reveal significant differences between the two conditions. We analyzed the simulator sickness using the aggregation procedure described in [156] with a mixed ANOVA. A significant main effect showed that the total sickness was sig-

Table 5.3: Comparisons of the Subjective Dependent Measures

	Augmented		*Non augmented*					
	N	*M (SD)*	*N*	*M (SD)*	*t*	*p*	*d*	*90% CI*
Social presence	65	3.58 (1.24)	60	3.14 (1.06)	2.17	**.032**	0.38	[0.038 0.859]
Self-reported CP.	65	4.03 (0.97)	59	3.97 (1.05)	0.31	.757	0.06	[-0.302 0.415]
Perc. other's CP.	64	4.52 (0.84)	59	4.27 (0.94)	1.60	.113	0.28	[-0.062 0.577]
Telepresence	65	5.15 (1.05)	60	5.24 (1.12)	-0.48	.634	- 0.08	[0.476 0.291]
Enjoyment	65	5.72 (1.21)	58	5.77 (1.11)	-0.26	.795	-0.04	[-0.471 0.361]
Lasting impression	65	4.89 (1.42)	58	4.91 (1.36)	-.086	.932	-0.01	[-0.518 0.475]
Thought-provocation	65	4.36 (2.92)	58	3.47 (1.27)	2.14	**.034**	0.39	[0.067 1.708]
Suspense	65	4.03 (1.47)	58	3.84 (1.30)	0.72	.474	0.14	[-0.318 0.679]
Artistic value	65	3.64 (1.56)	58	3.39 (1.26)	0.95	.342	0.18	[-0.264 0.754]

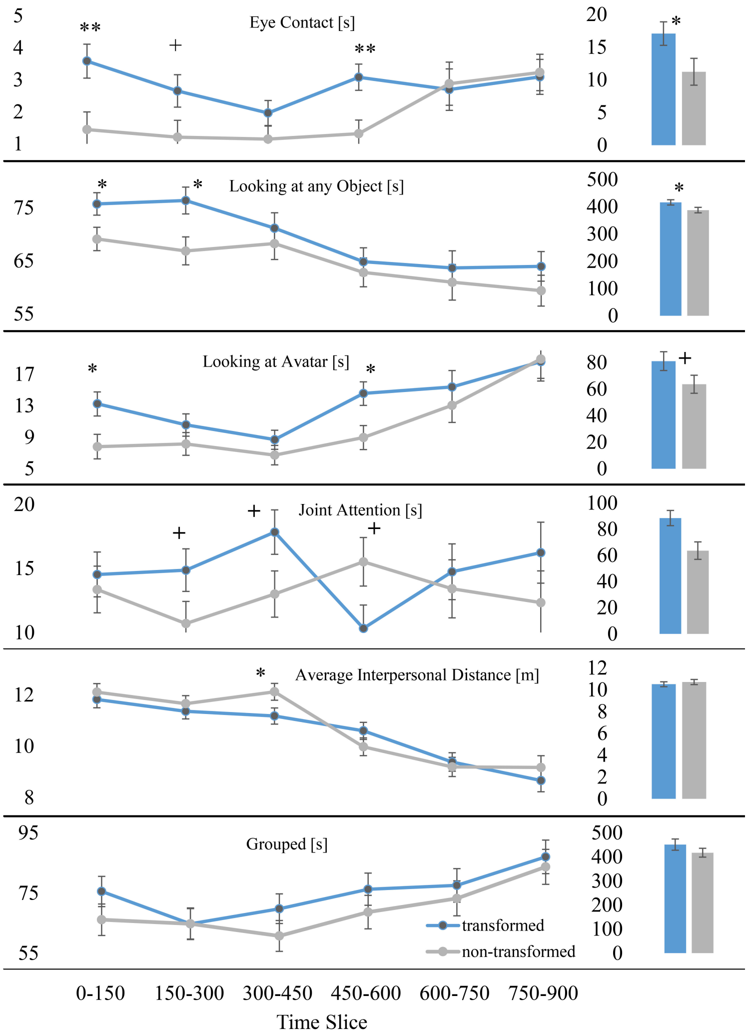

Figure 5.13: Results from the behavioral measures. Mean values across the six different time slices and overall values with 15 minutes exposure. *Note.* * $p < .05$; ** $p < .01$; + $p < .1$ *(two-tailed); bars denote the mean value; error bars denote the standard error.*

nificantly lower in pre exposure measurement ($M = 10.36$, $SD = 10.41$) compared to the post exposure measurement ($M = 11.93$, $SD = 11.13$; $F(1, 122) = 14.47$, $p < .001$, $\eta_p^2 = .106$). The analysis of the sickness measurement did not show a significant difference between the two conditions. In addition, the subjective or objective (number of correct answers) knowledge measure were not significant.

Qualitative Comments

The overall evaluation of the user comments was rather positive. e.g., “Super!”, “Very beautiful and detailed”, “wonderful experience”, or “Very suspenseful, partly asked myself if dinos could become alive”. One user commented that he was “visually very immersed in the scenario. However one cannot blind out the mind”. Some users commented that they missed the y positioning for the perception of height, e.g., “unfortunately no height axis”. Some users experienced minor drift, e.g., minor displacement of the physical and the virtual world: “voices of other participants did not fit exactly to their position”. Some participants experienced minor blurring because of vapor from transpiration. Some participants stated that a face or another indication of the head direction would have helped them when interacting. In accordance with our observations, some participants explored the augmentations more intensively, e.g., “nice effects when looking at a dino together/bubbles when speaking with each other”, “Cool effects with lighting and stars on the skin”, whereas others just noticed them. One participant called the augmentation for directed gaze “love bubbles”.

5.3.4 Discussion

Supporting hypothesis *(H1), The augmentation of social behaviors by amplifying or substituting transformations increases social presence*, we found a significantly higher levels of perception of social presence in the augmented condition than with the non-augmented condition. This implies that visual augmentations for social behaviors can increase the perception of a “shared social virtuality”. The finding is underlined by user comments such as the description of the eye contact transformation as “love bubbles,” signifying that the user actually perceived the substitute as having a positive character. One indication that participants tried to interpret the augmentations is that the “augmented” condition was found to be significantly more thought-provoking, while there were no differences in subjective or objective learning. We also did not find significant differences in the copresence or telepresence measures, which could indicate that these factors were not affected by the augmentations. It should be noted that the virtual environment and experimental scenario could be considered a mixed-reality scenario due to the fact that interactants still heard each other's physical voices and were in the same physical space with each other. This may have impacted the effects and compensatory mechanisms of the different augmentations, because grouping was amplified whereas augmentations for eye contact or joint attention might be considered substitutionary in their character. Furthermore, participants in the non-augmented condition could also have relied more on prosody as compensation. This does not, however, explain the findings for social presence, which we think are robust to the stated interpretation.

We found several indications supporting *(H2), the augmentation of social behaviors impacts the respective social behavior of users*. Participants 1) evoked more approximated eye contact, and 2) focused more on other avatars which suggests that the participants' behaviors changed because of the behavioral augmentations applied to the simulation in the augmented condition. These patterns elucidate that augmenting social behaviors can impact human behavior, for instance in terms of the awareness of others and the awareness of their behavior. Participants in the augmented condition also focused more on active objects (avatars or exhibits), which could indicate that the augmentations made participants more curious and changed their interactivity with the simulation. However, our simulation environment was relatively static, which could have made users pay more attention to any visual affordances and could have biased the social presence and behavior results. Similar ratings for enjoyment and suspense counter this assumption to some extent. We did not find significant differences in the average distance of participants or the amount of time with joint attention, which is why H2 is only partially supported. Although the grouping transformation was meant to be of an amplifying and positive character, the grouping could also induce negative attitudes toward outgroup members [34]. With our data, we cannot state any conclusion in this regard.

We did not find significant differences for enjoyment, suspense, lasting impression, or artistic value, which means that the augmentations did not affect these dimensions, and thus, the experience was not perceived more negatively or positively in either condition. This could be partly explained by a ceiling effect (qualitative comments indicate a generally very positive evaluation), as for many participants, it might have been the first VR experience in a large-scale simulation.

With regard to our overall research goal, we can therefore state that augmenting virtual social interactions can be beneficial for experiences with regard to social presence and an enhanced thought-provoking experience. Furthermore, we conclude that the proposed augmentations can foster behavioral interactivity between participants with regard to eye contact and interactivity between the participants and the environment. We controlled for potential third variable biases such as personality or a previous relationship between participants. Furthermore, the knowledge measures we applied to the experimental procedure do not imply any significant differences in mental distraction. It can thus be concluded that an augmentation does not detrimentally impact cognitive resources. We will next discuss multiple limitations that this study faced and our argumentation on its impact.

Limitations

First, there are indications (user comments, sickness measurements) that participants suffered from minor rotational drift. We did not find any indicators that participants in the augmented condition suffered from more drift, and both conditions used the same hardware. The translational latency of the RTLS was technically evaluated to 206 ms. A motion-to-photon measure of a single client (laptop, frame counting, 1000 Hz camera) resulted in $M = 246.66$ ms latency. We expected slightly higher values in the simulation due to wireless transmission. However, the rotational latency was barely measurable with frame analysis (240 Hz camera). Second, the recalibrations could have distracted some participants. Similar to drift and latency, there are no indicators that there were different amounts of recalibrations between the groups. Third, the participants were aware of their interaction partners, had seen their partners prior to the simulation and were aware that the pillars represented other users; as a blind procedure was not possible due to the extensive setup, and our control measures did not identify a potential bias. Fourth, the fact that we used multiple transformations do not allow for the interpretation of any finding with regard to a single transformation. For example, it seems that grouping and joint attention were not as affective as eye contact, but we cannot make any conclusions regarding this interpretation based on our data.

5.4 Summary

In this Chapter, we presented three studies that conceptualized forms of social augmentations in virtual interactions using HAAT and investigated their impacts. With the presented related work from Chapter 1, the augmentation designs from Studies 11, 12, and 13, and this summarizing discussion we can derive an answer for the overarching research question *RQ3a: What possible modifications can be designed to augment user-embodied social interactions?*.

5.4.1 Categorizing Social Augmentations

We presented three possible approaches to social augmentations, namely augmented mimicry, augmented gaze, and artificial visual affordances of substituting character. The possibilities for the modification of social behavior are, however, manifold. We therefore propose to categorize forms of social augmentations by:

- **The trigger and target modalities**
  The behavior modalities used to trigger or display an augmentation (for example gaze, body motion, voice, proxemics)
- **The phenomenological level**
  The phenomenological level the augmentation is executed on or triggered by (for example, a dyadic phenomenon vs. a group phenomenon, conscious vs. non-conscious behaviors)
- **The ecosystem of origin**
  The origin or the world the augmentation is created by (for example, whether the augmentation aims to present natural behavior such as body motion or facial expressions, or artificial affordances such as visual substitutes or amplifications)
- **The modification persistence**
  Whether or not the augmentation leads to persistent or non-persistent simulations (for example, the gaze augmentation or mimicry augmentation led to non-persistent simulations and thus individual simulation "truths" for each user were rendered, whereas the visual transformations led to persistent simulations shared by all users)

With regard to *RQ3b: What impacts on the perception of interactions arise from augmented social interactions?*, we summarize the insights from the three presented studies.

5.4.2 Impacts of Social Augmentations

Study 11 presented an approach for injecting nonverbal mimicry into user-embodied social interactions in VR. While the results of the preliminary study did not show significant improvements in terms of the ratings of the interaction, they also did not show any significant deterioration of the values. One could argue that the gold standard of natural interaction is to some regard difficult to beat with any artificially constructed or modified behavior. With regard to the latter, the study was a success in indicating a principle functionality of such approaches. Keeping in mind the limited sample size, the study should be replicated, and future studies should investigate different forms of mimicry, for example pose and gesture related mimicry transformations, and facial mimicry.

Study 12 compared a hybrid gaze model and a more sophisticated synthetic gaze model to a less sophisticated gaze model and natural gaze transmission in an avatar-mediated desktop communication setting. While, again, the hybrid and synthetic social gaze models could not attain the level of a natural interaction, the significant trends showed that the social gaze model introduced to the animation of the avatar did cover the social expectations of the behavior to a large degree, especially as compared to a trivial random gaze model. Thus, a model reacting to the conversation partner does show higher levels of perceived rapport, trust, and attraction, as indicated by trend analyses. Therefore, we can conclude that a reactive gaze model has benefits for social interactions and that a hybrid model may be beneficial when sensor data transmission is disrupted. It is left to be investigated whether the approach would also help when inadequate social behavior is performed, such as when people suffer from disorders or do not feel confident. Future studies should also replicate the approach on an immersive level. Current prototypes are works in progress.

Study 13 presented the design, implementation, and evaluation of a concept for the augmentation of social behaviors in multi-user VR. We designed three visual transformations for *eye contact*, *joint attention*, and *grouping* in order to test whether or not these augmentations impacted an experience in VR. Our findings extend the results of previous work [266, 277] and suggest that applied augmentations can significantly impact social presence and user behavior. These findings can inform the development of SVR applications or training and therapy of individuals suffering from social disorders. We believe that our approach is an initial step toward exploring the potentials of VR as a medium for actively mediating human communication. Future work should include the isolated investigation of the presented augmentations. Furthermore, the inclusion of high-fidelity anthropomorphic characters or virtual agents may impact the results and should be investigated along with behavioral degree of freedom. Design considerations could further examine the impact of appearance and form. Pattern-based phenomena such as gaze-cueing and mimicry could extend the presented framework. Figure 5.14 summarizes the results.

5.4.3 Implications

I3.1: Virtual Social Interactions can Benefit from Social Augmentations

Especially reflecting the findings of Study 13, we can conclude that social augmentations can be beneficial for social encounters in SVEs. This can inform the design and development of future applications for entertainment, communication, and research. Study 13 already provides an example, namely the joint exploration of educating simulations. Deriving from our approaches, social augmentations using

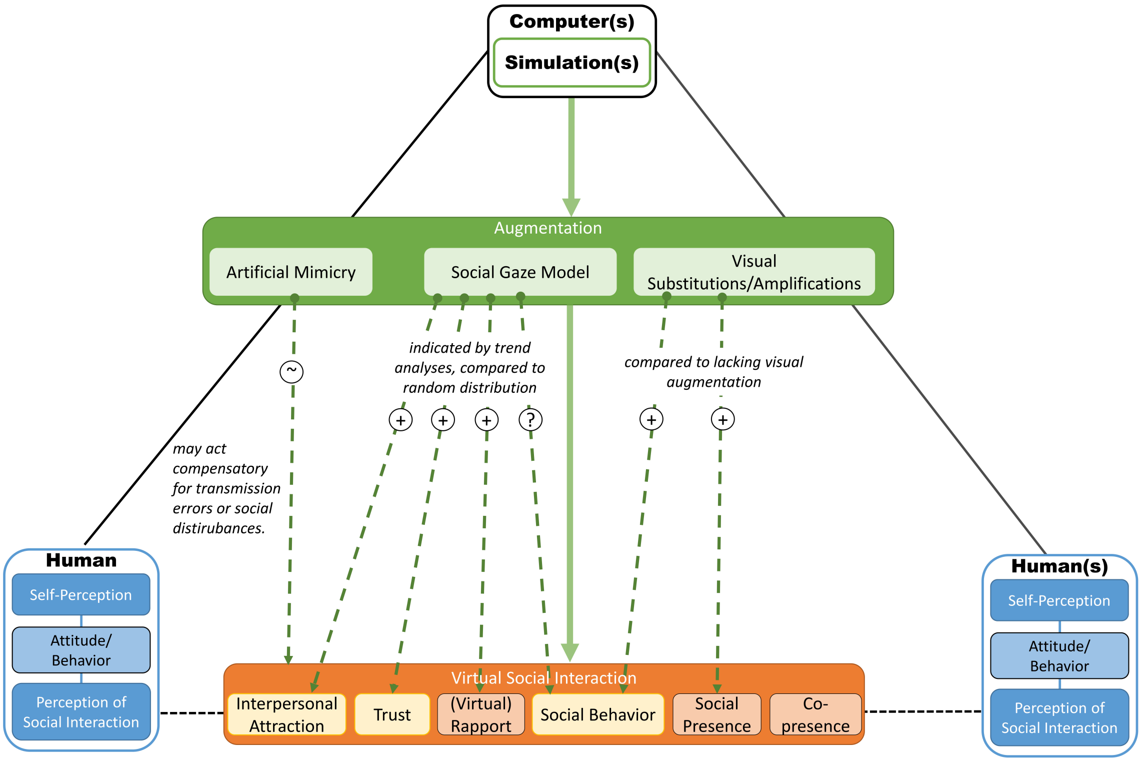

Figure 5.14: Model summary of the findings from the augmentation studies. We did not find significant negative or positive impacts resulting from the injection of artificial mimicry. The social gaze model, compared to a random gaze visualization, improved interpersonal attraction, trust and rapport, whereas it was inferior to a natural gaze of participants without known disorders. Compared to a complete lack of social signals, visual augmentations of social phenomena had a positive impact on social presence, thought provocation, and the social behavior of participants.

HAAT could be beneficial for applications in cases where 1) a lack of sensing and convincing reproduction of natural behavior exists, 2) transmission errors may occur, for example due to a lack of quality of service, and 3) in cases where cultural or distorted social communications occur, such as culture specific nonverbal behavior (e.g., greetings and gestures) or distortions due to non-typical social behavior (e.g., individuals with mental disorders or generally norm-deviances). With regard to the latter, social augmentations could provide assistance to support the interaction, yet further research is of need.

I3.2: Natural, Undistorted Behavior is a Gold Standard

In Studies 11 and 12, we did not find a significant positive impact of the augmentations performed. In contrast to previous work [24, 224], these were not specifically addressing the amplification of social cues or the distribution of social signals to multiple users. For future developments in social augmentations it will therefore be great challenge to compete with natural behavior. While a comparison on the basis of agent behaviors (i.e., mimicking vs. non-mimicking agents) may provide some insights, human-agent interactions are typically less complex than natural social interactions between two people. The advantages of social augmentations in future research may therefore also be stressed when considering the above stated use cases and compare, for example, to situations where the communication is disturbed.

I3.3: Ethical Considerations Need to be Discussed

Throughout the construction of and reflection about the presented studies, it became salient that ethical issues have to be addressed regarding such augmentations. By no means is it justified to blindly apply social augmentations to VR, MR, or AR applications. Considering already identified risks of VR applications [194] social augmentations bring additional challenges regarding ethical and societal aspects. These are, to part, further addressed in Chapter 6.

5.4.4 Conclusion

In summary, we presented three forms of potential augmentations, namely artificial mimicry, augmented gaze, and amplifying or substituting visual transformations in user-embodied virtual interactions using HAAT. We conclude that these could be beneficial in cases of disturbed interactions, or the lack of sensory input. Future work should consider alternative modalities, such as haptics or audio, and improve the triggering and decision making processes of such augmentations. The following discussion will reflect about the present thesis in general, ethical considerations, and future work.

Chapter 6

General Discussion

6.1 Summary and Contributions

Virtual embodiment and embodied social interactions strongly gain in relevance with current technological developments. The present thesis explored intrapersonal and interpersonal effects of user-embodied interactions in VR, and presented a concept and empirical evaluations of the augmentation of social interactions in VEs, respectively VR.

Chapter 1 described the motivation, the presented research themes, and the context. With regard to intrapersonal interactions, Chapter 2 introduced the principles of information processing and the perceptual modeling of user embodiment. Previous work identified the sense of self-localization, the sense of ownership, and the sense of agency as components of (virtual) embodiment. With regard to interpersonal aspects, we reviewed related system developments and empirical findings on computer-mediated social interactions. While the strive for realism may increase social perception within such interactions, the SIPT [351] argues that humans can, to some degree, compensate for missing social cues. On the basis of previous findings and theoretical considerations, we proposed considering *computers as social mediators*. While modifications of parameters for user embodiment have, as found by pervious works, impact on the self-perception (virtual body ownership) and consequences for the social self (Proteus effect), recent works showed that modifications of representation and behavior in social interactions can have impacts on behavior, presence, and perceived affiliation. To design software applications that utilize the latter effect, we contributed the concept of HAAT. HAAT builds upon previous works and extends TSI concepts by allowing to moderate and augment interactions based on bidirectional communication processes and social phenomena (see Section 2.4.5). To explore different aspects of the three research themes, this thesis contributed with multiple prototypes and 13 empirical studies to the knowledge gain on user-embodied interfaces. The main findings and contributions are summarized in the following in relation to the overarching research questions.

6.1.1 Characterization of Impacts of Simulation Properties

Chapter 3 systematically analyzed the impacts of simulation properties and the resulting effects on body ownership, agency, and the perception of body schema change. Multiple empirical results fabricate an answer for *RQ1a: How do simulation properties affect virtual embodiment?*.

Immersion, and thus the support of SMCs provided by the virtual environment was identified to foster the perception of ownership (acceptance) of a virtual body, and a strong facilitator for the perception of a perceived change of the body scheme, confirmed by Study 2 and Study 4. We interpret that the main reason for this is the increased coherence or congruency of the environment, meaning that in contrast to a projection approach where points of reference to physical body and environment are provided to the visual system, HMD-driven simulations do not provide such visual references to PR. In other words, in projection-based approaches two bodies are presented to the user, the physical and the virtual, whereas the HMD-based and more immersive simulation only presents a single body to the visual system. While immersion is a rational factor to influence the perception of a change in the body scheme, the impact on change points to a relation between ownership and body scheme change, meaning that when the acceptance over a virtual body is greater (due to simulation properties fostering top-down and bottom-up factors), the resulting change in body scheme perception is greater, too.

Simulation latency, and thus the hindering of SMCs by a reduced visuomotor synchrony while being exposed to the simulation had a negative effect on virtual embodiment, primarily due to changes in the perceived agency and thus the control over a virtual body. Adding to previous research [350], we found that not only latency but also latency jitter (i.e., reduced visuomotor synchrony) hinder the perception of control over a virtual body and also had a negative impact on the perceived ownership. With our data, we could quantify the impact of jitter on the basis of the distribution we used comparable to an impact of a linear delay between 207 ms and 353 ms.

The *humanness*, and especially a *personalization* of the avatar positively affected the perception of ownership over a virtual body, as well as the perception of a perceived change in the body scheme. More specifically, we found these effects only for procedures where avatars where personalized on the basis of photogrammetry and not when the user performed the personalization based on a character generator. While an increased ownership is rational, the perceived change may require a more complex interpretation and data modeling that is subject to further research. One possible interpretation may be that ownership moderates the perception of change, and thus discrepancies between physical and virtual body become more salient.

To answer RQ1a: In addition to previous findings we found that:

- A higher level of immersion and thus the facilitation of SMCs fosters the ownership of a virtual body as well as the perceived change.
- Simulation latency and latency jitter hinder agency, and thus the perception of control over a virtual body
- Humanness and photogrammetric personalization foster the perceived ownership as well as a perceptual change of the body scheme.

6.1.2 Identification of Latent Variables of Embodiment

By reviewing related work, constructing a measurement instrument through principle component analyses, as well as generalizing and validating this instrument, this thesis can derive an answer to *RQ1b: What latent variables are responsible for the adaptation of a virtual body?*. We identified three latent variables associated with the adaptation of a virtual body:

Acceptance, as the first component derived from the analyses measures the *ownership* over a virtual body. Ratings of this measure will be high if the user accepts the virtual body as the own, that it is his/her body, that the body parts are his/her body parts, that the body is human-like, and that the body belongs to him/her.

Control, the second component we derived from the analyses measures the agency over a virtual body. High ratings in this component result from an accurate movement reproduction and thus if the user feels that the movements he saw are his/her movements, that he/she is controlling and causing the movements of the avatar, and if he/she perceives the movements as synchronous to the physical movements performed.

Change, the third component extends previous approaches to measure the level of embodiment as it explicitly assesses perceived changes of the body scheme that result from being embodied by an avatar. The component will show high ratings when the user perceives that the form or appearance of his/her body changed, and when perceiving his/her body as lighter or heavier, taller or smaller, as well as larger or thinner. As noted above, arguably, this perception is fostered by immersion and may be moderated by the level of acceptance for a virtual body.

To answer RQ1b: With the present approach we could identify and confirm three latent variables:

- Acceptance – Assessing the level of ownership
- Control – Assessing the level of agency
- Change – Assessing the level of a perceived change in body schema

Previous research further identified the sense of self-location as one component of the sense of embodiment [157]. We did not specifically include this component in the scale construction and therefore cannot argue about its contribution to embodiment as a construct. Multiple findings presented in this thesis show that embodiment is related to aspects of presence. Yet, "Self-location and presence are psychological states that refer to different issues" [157, p. 375], as "self-location refers to one's spatial experience of being inside a body and it does not refer to the spatial experience of being inside a world (with or without a body)" [157, p. 375]. For example, out-of-body experiences [83] may affect the sense of self-location. However, these aspects were not in the focus of the present thesis and to include them into the presented scale is subject to future work.

6.1.3 Comparison of Embodied Social Interactions

Studies 8 and 9 compared virtual social interactions that provided user embodiment to physical social interactions and addressed *RQ2a: How do virtual social interactions with user embodiment compare to physical social interactions?* In this regard, Study 8 showed reduced social presence when comparing PR interactions to user-embodied interactions in VR and implications for a compensation of missing social cues. Despite higher social presence in PR in Study 9, the study did not directly support (but also not contradict) this results to a significant level. In both, Study 8 and Study 9 the users showed similar performance in functional communication (i.e., the negotiation). Yet, Study 9 found that they perceive less efficiency in embodied VR, and further a more negative affect attributed to the communication partner. These findings may relate to a reduced presentation of communicative behaviors. However, the assessments of behavioral focus in Study 8 also showed that users shifted their attention towards other cues, and thus are supporting the SIPT [351]. Further research is of need to replicate the findings and support the assumptions derived from the presented empirical work.

To answer RQ2a: We could show that

- Embodied virtual social interactions can result in decreased social presence
- Embodied virtual social interactions can result in a more negative perception of the communication partner
- Embodied virtual social interactions can result in less perceived efficiency

6.1.4 Effects of Technological Properties

Chapter 4 could further provide insights into *RQ2b: How do technological properties affect user-embodied virtual social interactions?*. Study 8 showed a decrease in collaborative motor performance, which is attributed to the introduced latency and the rendering of physics. Study 10 addressed RQ2b more specifically, we investigated the degree of transmitted behaviors. We found a significant increase in telepresence with higher degree of behavioral realism. Yet, this also led to more eerie avatar representations. We interpret this mainly due to a lack in precision (i.e., tracking errors) of the behavior replication.

To answer RQ2b: Our results indicate that

- Latency decreases collaborative motor performance
- Higher degrees of behavior realism increase telepresence
- Higher degrees of behavioral realism can evoke eerieness in the perception avatars

6.1.5 Modifications for Augmented Social Interactions

Chapter 5 designed three possible modes of augmenting social behaviors and in a reflection and relation to previous work answered *RQ3a: What possible modifications can be designed to augment user-embodied social interactions?*. We designed two possible modifications that modify behaviors of natural origin, namely augmented mimicry and augmented social gaze, triggered by either a naïve periodic timing or a specific social contingency, namely the listener status. Further, we designed artificial visual augmentations for behavioral phenomena, namely visual substitutes for the displaying of grouping, joint attention, and eye contact. Whereas augmented mimicry and augmented gaze led to non-persistent simulations, the visualization of artificial cues led to persistent simulations for all users.

To answer RQ3a: We propose that possible modifications can be designed on the basis of the following categories

- The trigger and target modalities
- The phenomenological level
- The ecosystem of origin
- The modification persistence

6.1.6 Impacts of Augmented Social Interactions

Study 11, 12, and 13 further provide answers for *RQ3b: What impacts on the perception of interactions arise from augmented social interactions?*. Modifying body motion on the basis of nonverbal mimicry (Study 11) did not, neither positively nor negatively affect the social perception. On the one side, this shows that a mere "injection" of a phenomenon related to affiliation and a better connection between the communication partner does not per se improve communication. On the other side however, this showed that even a naïve approach such as the one presented in Study 11 did not have a significant negative impact in our study. Study 12 showed that through respecting basic patterns of gaze behavior, an artificial model could improve the perception of the communication. A hybrid approach did not improve the natural (gold standard) communication, but showed better values compared to a pure synthetic construction. Such a model could therefore be beneficial for controlling the modality, for example in the case of transmission errors or social disturbances. Study 13 presented visual substitutions of nonverbal phenomena of everyday interaction, namely social grouping, joint attention, and eye contact. This resulted in an increase of social engagement by the participants and an increased perception of social presence. Further, we could identify that the additional visual affordances impacted the users' (social) behaviors, by higher values of eye contact and more focus on other objects and avatars. In summary, we argue that social augmentations can be beneficial with regard to social presence and social engagement (Study 13), they may enable to provide a fallback in the case of transmission problems (Study 11 and Study 12) and they can, to some degree, reconstruct typical social phenomena (Study 12).

To answer RQ3b: Our results indicate that social augmentations can impact:

- The perceived rapport, trust, attraction positively (hybrid and synthetic gaze compared to a randomized gaze model)
- The perceived social presence and thought provocation (non-augmented interactions compared to visually augmented interactions)
- The resulting social behavior (non-augmented interactions compared to visually augmented interactions)

In Sections 3.8.5, 4.4.3, and 5.4.3 we pointed out some implications for these findings with regard to the design and the development of applications for human-computer interaction and CMC. In the following, we derive some implications and ethical reflections about user-embodied interfaces and augmented social interactions for our society.

6.2 Societal Considerations

The findings of the present studies imply a technological and ethical discussion on how technology shapes our perceptions of the self and the perception of user-embodied interactions. Social VR platforms, such as Facebook VR or Rec Room, are already capable of collocating interaction partners as avatars in navigable multi-user SVEs and transmitting nonverbal cues, for example by using game controllers and HMD tracking. These simulations are inherently different from the capture and replication of static or dynamic images because sensors capture data of human behaviors such as body movement, gestures, facial expressions, and eye gaze, which are then replicated to avatars by the simulation and thus embody each user in real time through virtual representations [271, 278, 266]. In strict consequence, this means that in virtual social interactions, behavioral and neurophysiological data of each user is sensed, transferred, processed, and represented through virtual embodiment by computing machinery. These developments reach beyond perceiving computers as social actors [216] in a pure agent context and instead, allowing them to transform the social behaviors of humans, and thus decouple the virtual from the PR [12]. HAATs aim at utilizing the phenomena of social interaction to support and foster rapport in intercultural and disturbed social interactions. These approaches open up new challenges and possibilities for society. Both ethical and privacy related questions have to be addressed; the transfer to hybrid entities may decrease our autonomy and provide machines with input data machines that is barely perceptible, if at all, by humans, and thus, we have to rethink the rules of engagement in future virtual social interactions. While ethical implications have been discussed for VR in a research context [194], discussions have to start today on how to address the ongoing implementation of human augmentation through embodied virtuality for societal life, which is why we overview current technological trends as well as arising challenges and potentials in the following.

6.2.1 Virtual Augmentation in Everyday Media

Our everyday communication behavior changes at a rapid pace. Most impactfully, mobile phones and social media shaped the way CMC changed over the last two decades. Aside from the fact that we have become permanently online and permanently connected [345] through these technologies, three specific dimensions have changed with these developments: *interactivity*, *realism*, and *autonomy*. In terms of how social media systems for virtual interaction work, the interactivity of applications gained flexibility with regard to the possibilities of communicative exchange of users. Provided with much easier and quicker access to communication platforms, users often freely to decide to join a one-to-one conversation, that may merge into group communications or even mass communication (e.g., second screen media, live streams) and the level of privacy is eventually reduced. Regarding the user autonomy, recent trends in computer graphics imply that data security, as well as ethics, have to be considered with regard to video material and self-representation. The impressive usage of generative adversarial networks has already revealed the capabilities to modify and manipulate video and image content based on two source inputs. With regard to images, using these machine learning approaches can create modified outcomes that match features of two sources [153]. Furthermore, what is referred to as "deepfake," a human appearance template can be matched to dynamic video content to change the identity of the protagonist by training deep neural networks with source and target image data [63]. While these developments mainly consider known algorithmic approaches from computer graphics and, in turn, can be detected using knowledge about the underlying network structures [3, 126], the possibility of modifying behavioral data and identity is impressively demonstrated with these techniques. Yet, video material differs from actual behavioral data, which needs to be sensed and transmitted for immersive, embodied communication.

6.2.2 User-Embodied Interfaces – A Novel Situation

A crucial difference between the current use of non-embodied representations in CMC and avatar-mediated interfaces is the replication and reproduction of behaviors that are tracked through behavior sensing techniques. While the control of avatars is already performed in multi-user online gaming as well as recent social VR applications, the control scheme is often an abstraction of behavior (e.g., waving "hello" by pressing button A rather than physically waving the hand). However, future applications will have higher fidelity, and through sense matching and user control, they will have access to the principles of manipulating bodily self-consciousness, such as by adapting external body parts through synchronous stimulation and visual feedback. Assessing the requirements for embodied interfaces, current developments in ambient sensing and ubiquitous computing let us foresee that, without doubt, it will be reliably possible to sense verbal and nonverbal behavior and display feedback without any restriction of location.

Risks of User-Embodied Interfaces and Hybrid Technologies

As Jaron Lanier puts it, humans are not a gadget [171]. Without a doubt, the modification of behaviors by an artificially driven intelligence decreases human autonomy and information control. The greatest challenges are, therefore, to assess and foresee potential misusage, specifically in the case of a network transmission of behavioral data that tremendously extends the information that can be gathered about human users compared to textual, video, or image materials. In that regard, the most important points of focus are data security, privacy, transparency, and control over personalization instruments.

Because most future SVR systems will most likely use both, reliable and unreliable protocols for data transmission, a first requirement is the encryption of all data on a transport protocol level. Furthermore, local data as well as the trained version of individualized social AIs have to be protected from intrusion. For overall security of individual data, a digital identity, such as a blockchain protected identity, could be used for better protection. It is a matter of importance that the data security councils address the issue of privacy now, including whether and how behavioral data may be used by providers of SVR applications. As noted, SVR applications enable the transmission and gathering of behavioral data. Also, the issue of privacy is not limited to hybrid systems but, rather, applies to all SVR applications. It is the responsibility of lawmakers to foresee and take action to prevent privacy related issues around the control over users' digital identities. For example, the following counter-measures could mitigate the risks evolving from embodied communication:

- Introduction of a digital identity to store and protect social AI data
- Transmission protocol security
- Privacy laws to prevent misuse of photogrammetric, neurophysiological, and behavioral data

Autonomy over Identity and Behavior

As a consequence of the secular challenge that society is facing from data collection in future immersive communications, the difficulties for researches and developers go beyond ethical best-practices [194] for a variety of applications [303], and the protection of vulnerable groups such as children [310]. In fact, we face guiding principles for protecting human autonomy and ownership over the being as such. As to the latter, the question is how we retain autonomy over behavior despite the influence of future social AI systems. It can be argued that brains are "essentially prediction machines" [67], yet, it is the intuition and the divergent that result in individuality, and it is experiences of everyday life that form behavior and responses even though processed as continuous error correction. Conclusively, these descriptions have to be protected from dissolving in a generalization of computation.

While it may be appropriate for experimental research projects and lab studies to retain control over collected data and any modifications thereto, the rules have to be rethought in terms of how much control is given to AIs to retrieve individual data and construct artificial models of behavior. In principle, there is no justification for models to learn how to influence individual behaviors based on societal generalizations. It is therefore critical to discuss how artificial social models will be fed data and to what level a generalized approach can overrule individual choices. While we generally strive for virtual agents that understand the complexity of the rule set for nonverbal behavior, and its operation [61, 164, 342] through the interpretation of social signals, and respective reactions [237], it may not be the goal of hybrid technologies to generalize interpretations of responses. As a result, ownership of the self and reactions that respect for the individual being need to be taken into account in learning and training. Individuality cannot be a "plugin" to a generalized data set of responses and control, and transparency has to be provided to the user with regard to ownership and identity. Through the "sense of agency", humans experience themselves as the agent of their own actions [75]. While to a certain degree we have conscious control over executed behaviors, today's sensor and processing technology is able to detect subtle spontaneous micro expressions and reactions that are difficult for us to control [236]. One could further argue that future human–machine interfaces for virtual simulations will take a shortcut and avoid the detection of actual physical behavior and, rather, foresee behaviors based on neurological signals [177, 298]. At that point, especially when including uncontrollable neurophysiological signals in the loop, the machines will know more about us than we do. It must therefore be considered and made transparent to what degree and on which features social AIs are trained with regard to the neurophysiological correlates of behaviors.

The following countermeasures for mitigating the risks of decreased autonomy could be enforced:

- Constantly present the status of inference of social intelligences to all users
- Constantly allow users to stop any process of interference besides behavior replication
- Constantly inform users about what data is tracked, stored, and utilized by underyling AIs

Chapter 7

Conclusion

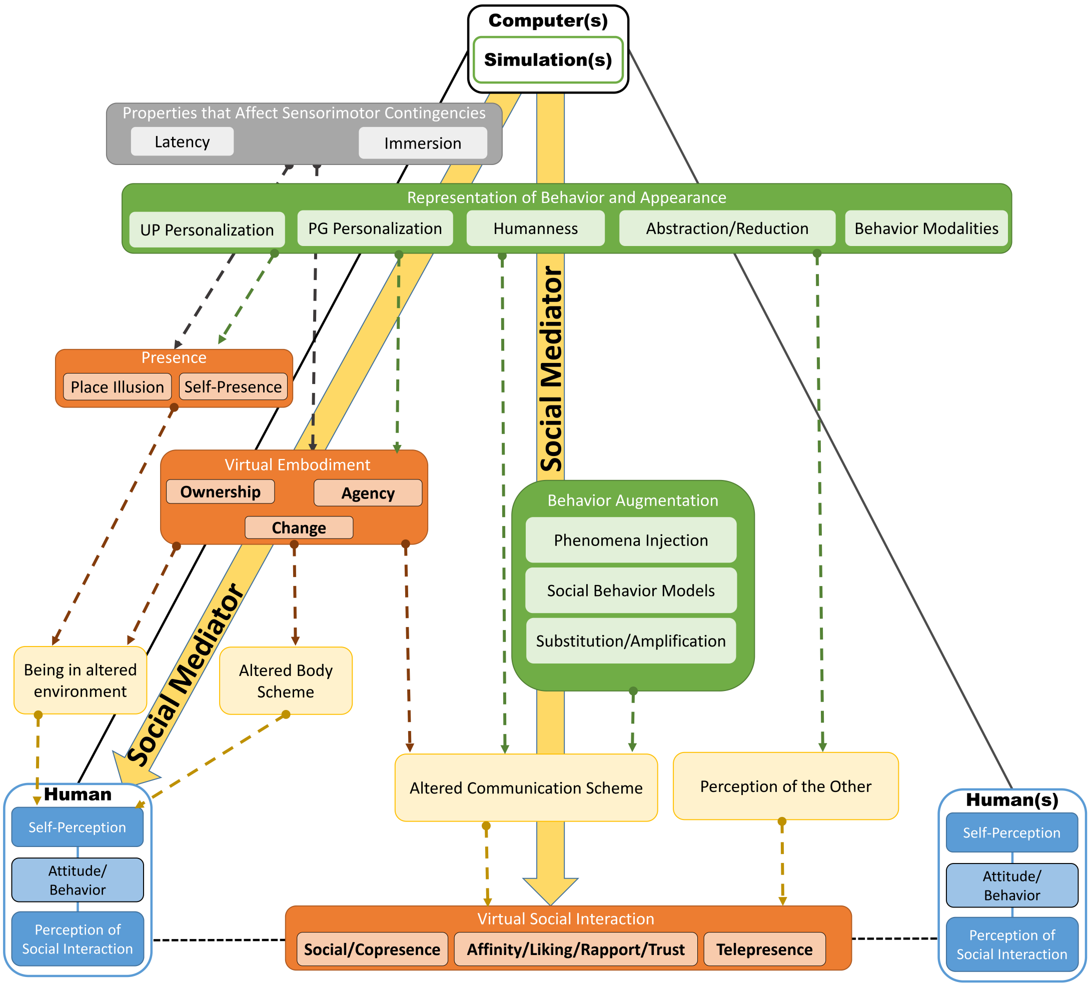

Figure 7.1: Conclusion of the findings. Computers as social mediators impact the self-perception as well as the interaction with others through simulation properties or active engagement. Both of which can be used as design choice and criteria for VEs.

This theses conceptualizes computers as social mediators with respect to user-embodied interactions in VEs. Concluding from the summary of findings and contributions, we can support this conceptualization by showing that properties of the simulation had impact on both, the self-perception (i.e., the social self) as well as the perception of social interactions and resulting behavior in such interactions, and that these can be actively controlled by defining properties and augmentations within the simulation.

Based on related work we presented this model, and the findings of the present research substantiates that computers can have an active part in our perception of the self and the other, as well as the social interaction per se. Through displaying a virtual world not necessarily congruent with our physical world, virtual environments can stimulate our senses and affect long-term and short-term memory, and thus infer with the action-perception cycle. In turn, this results in perceptual impacts that reflect into the physical world, for example by an adaptation of behavior and attitude. Figure 7.1 summarizes the findings based on conceptualizing computers as social mediators.

7.1 Future Work

This thesis contributed with conceptual, methodological, and empirical findings. While we gained further insights into many aspects related to user-embodied interfaces, user behavior and perceptual effects, future work should further explore and progressively build upon the presented findings.

With the construction of the scale to assess components of embodiment, we identified *change* as a latent factor. This is most interesting and as mentioned might be a predecessor that moderates the Proteus effect. Through our results, we cannot gain any insights to that regard. However, further studies should test the hypothesis that a higher acceptance (ownership) fosters the perception of change, which in turn strengthens the Proteus effect for modified characters. However, we found that a higher acceptance is related to personalization, which counter-acts the Proteus concept as it especially aims at using altered forms of representation and not a replica of the self. Future work should consider this aspects and derive further findings from systematic investigation, for example by study designs that allow for a path modeling.

With regard to the technological effects on social interactions, meaning the degree and quality of behaviors transmitted, the present thesis barely scratched the surface. While we could provide relevant insights, tracking technologies will hopefully improve to reduce biasing effects of unwanted artifacts in the reconstruction. The level of immersion (see recently [279]) as well as the form of embodiment seem to have an effect on aspects of social perception. To that regard, we used a very controlled and less affective role-play scenario in two studies, and an interactive scenario in a third study. Further research should therefore find a method to investigate less controlled interactions, or interactions that foster an affective theme. Study 10 may have suffered from the display of less convincing replication of behavior. Ongoing work therefore implements an immersive simulation that supports the replication of facial expressions and gaze with a higher visual resolution (see Figure 7.2.

With regard to the developed model, one could argue that it is not clearly differentiated whether users perceive the interaction as such or the other communication partner(s) differently when performing active augmentations. The architecture and concept should be further refined, especially with regard to solutions that foster

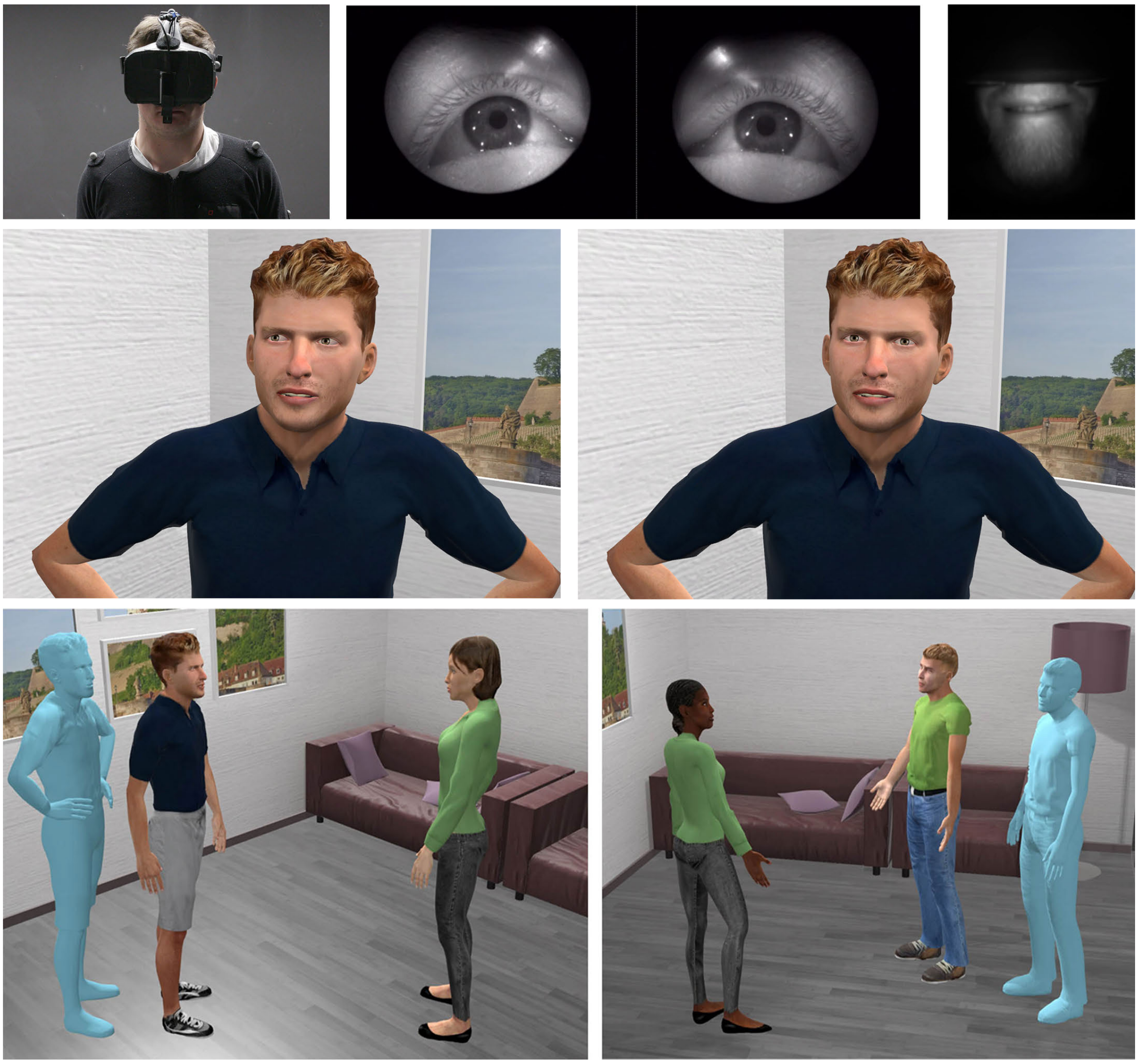

Figure 7.2: Future work: Improving the realism, the triggering and execution of augmentations. Body motion, facial expressions and gaze behaviors are rendered in more realistic simulations to HMDs. Through machine learning and pattern recognition, a more complex model can evolve to trigger and render augmentations.

the persistence of augmented simulations (in multi user approaches) and include further forms of modifications. The present prototype, which we named injectX, additionally merges the previous approaches, namely the augmentation of social gaze and nonverbal mimicry. These are refined on the basis of trigger signals that are detected through more sophisticated machine learning approaches. For example, we differentiate the complexity of gestures, detect speaker and listener states, and analyze the current status of agreement on the basis of head-nod frequencies. This allows a more sophisticated social engine to decide upon and render social augmentations. These may be in form of natural behavior or artificial substitutes. A long term perspective in this regard is to port those algorithms to AR applications.

Finally, neuronal signatures of communicative phenomena could further extent the concept of augmenting social interaction through HAAT (Figure 7.3 shows a current prototype). For example, affective states can be utilized to render additional social information for the user. Further refining and developing underlying social functions could extent the presented approaches. A portation from mere visual to haptic, auditive, and neurophysiological feedback can extent the design space to more ubiquitous methods.

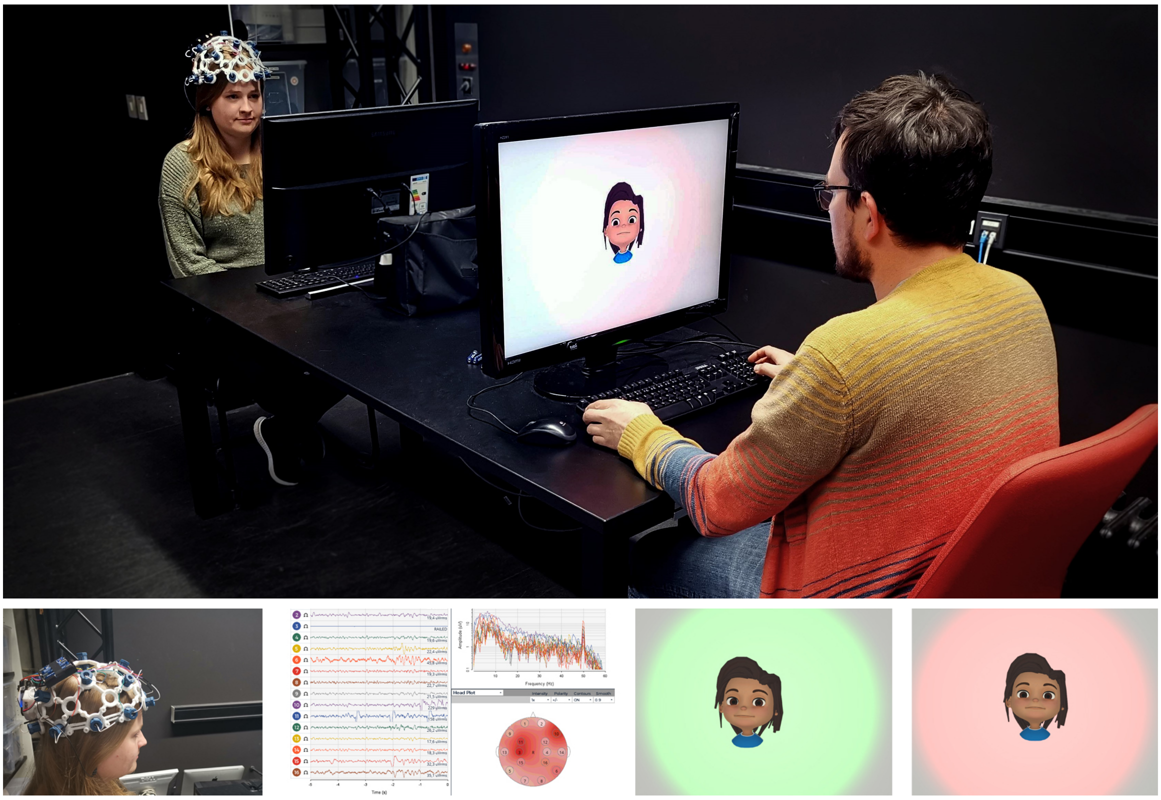

Figure 7.3: Future work: Augmentations based on neuronal signatures. Affective states are derived from eeg signals based on trained neural networks.

7.2 Concluding Remarks

Although user-embodied interactions become more and more salient in everyday life, the investigation of resulting impacts, especially with regard to the augmentation of social behaviors, is still in its early stages. Through continuous research we begin to understand more about the plasticity of the brain to adapt to alternative body representations and social information. The technological approaches described in this thesis can be seen as future and emergent technologies, that enable their application to communication, HCI, diagnosis therapy, research, and entertainment. And to this point, it is very touching to see that not only presented related technological work but also some of our own prototypes described in this thesis are applied in and support other research fields, such as neuroscience and therapy. The field of user embodiment, or virtual embodiment is far from being elaborated. Yet, it is indispensable for future user-interfaces to consider this perspective, as we become hybrid.

Bibliography

[1] Jascha Achenbach, Thomas Waltemate, Marc Latoschik, and Mario Botsch. 2017. Fast Generation of Realistic Virtual Humans. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology*. DOI: 10.1145/3139131.3139154.

[2] Charles Adam and Paul Tannery. 1901. Oeuvres de Descartes.

[3] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. 2018. Mesonet: a compact facial video forgery detection network. *arXiv preprint arXiv:1809.00888*.

[4] Peter Andersen, Jillian Gannon, and Jessica Kalchik. 2013. Proxemic and haptic interaction: the closeness continuum. In *Nonverbal Communication*. De Gruyter, Berlin, Boston.

[5] Page L. Anderson, Matthew Price, Shannan M. Edwards, Mayowa A. Obasaju, Stefan K. Schmertz, Elana Zimand, and Martha R. Calamaras. 2013. Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of consulting and clinical psychology*, 81, 5, 751. DOI: 10.1037/a0033559.

[6] Michael Argyle. 1988. *Bodily Communication*. (2nd edition ed.). Methuen. 390 pages. ISBN: 978-0-416-38150-4.

[7] Michael Argyle and Janet Dean. 1965. Eye-contact, distance and affiliation. *Sociometry*, 289–304. DOI: 10.2307/2786027.

[8] Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63, 4, 596. DOI: 10.1037/0022-3514.63.4.596.

[9] Laura Aymerich-Franch, Damien Petit, Gowrishankar Ganesh, and Abderrahmane Kheddar. 2015. Embodiment of a humanoid robot is preserved during partial and delayed control. In *2015 IEEE International Workshop on Advanced Robotics and Its Social Impacts*. DOI: 10.1109/ARSO.2015.7428218.

[10] Norman I. Badler, Cary B. Phillips, and Bonnie Lynn Webber. 1993. *Simulating Humans: Computer Graphics Animation and Control*. Oxford University Press.

[11] Jeremy N. Bailenson, Andrew C. Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2005. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human communication research*, 31, 4, 511. DOI: 10.1111/j.1468-2958.2005.tb00881.x.

[12] Jeremy N. Bailenson, Andrew C. Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2004. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 13, 4, 428–441. DOI: 10.1162/1054746041944803.

[13] Jeremy N. Bailenson and Jim Blascovich. 2004. Avatars. In *Encyclopedia of Human-Computer Interaction*. Berkshire Publishing Group, 64–68.

[14] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29, 7, 819–833. Retrieved 05/21/2014 from.

[15] Jeremy N. Bailenson, Dan Merget, Nick Yee, and Ralph 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. DOI: 10.1162/pres.15.4.359.

[16] Jeremy N. Bailenson and Nick Yee. 2005. Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science*, 16, 10, 814–819. DOI: 10.1111/j.1467-9280.2005.01619.x.

[17] Jeremy N. Bailenson, Nick Yee, Jim Blascovich, and Rosanna E. Guadagno. 2008. Transformed social interaction in mediated interpersonal communication. *Mediated interpersonal communication*, 6, 77–99.

[18] Domna Banakou, Raphaela Groten, and Mel Slater. 2013. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110, 31, 12846–12851. DOI: 10.1073/pnas.1306779110.

[19] Domna Banakou, Parasuram D. Hanumanthu, and Mel Slater. 2016. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience*, 10, 601. DOI: 10.3389/fnhum.2016.00601.

[20] John A. Bargh and Tanya L. Chartrand. 1999. The unbearable automaticity of being. *American psychologist*, 54, 7, 462. DOI: 10.1037/0003-066X.54.7.462.

[21] Christoph Bartneck, E. Croft, and D. Kulic. 2008. Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. In *Metrics for HRI Workshop, Technical Report*. Volume 471. Citeseer, 37–44.

[22] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2007. Is the uncanny valley an uncanny cliff? In *Robot and Human Interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium On*. IEEE, 368–373. DOI: 10.1109/ROMAN.2007.4415111.

[23] Christoph Bartneck, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. My robotic doppelgänger-A critical look at the uncanny valley. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium On*. IEEE, 269–276.

[24] Andrew C. Beall, Jeremy N. Bailenson, Jack Loomis, Jim Blascovich, and Christopher S. Rex. 2003. Non-zero-sum gaze in immersive virtual environments. In *Proceedings of HCI International*.

[25] Stephan Beck, Andre Kunert, Alexander Kulik, and Bernd Froehlich. 2013. Immersive group-to-group telepresence. *IEEE Transactions on Visualization and Computer Graphics*, 19, 4, 616–625. DOI: 10.1109/TVCG.2013.33.

[26] Mark W. Bell. 2008. Toward a definition of "virtual worlds". *Journal For Virtual Worlds Research*, 1, 1. DOI: 10.4101/jvwr.v1i1.283.

[27] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1997. Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds. *Multimedia Systems*, 5, 2, 93–104. DOI: 10.1007/s005300050045.

[28] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press/Addison-Wesley Publishing Co., 242–249. DOI: 10.1145/223904.223935.

[29] Gary Bente, Felix Eschenburg, and Lisa Aelker. 2007. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain*. Citeseer, 207–14.

[30] Gary Bente, Felix Eschenburg, and Nicole C. Krämer. 2007. Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations. In *Virtual Reality*. Springer, 185–194. DOI: 10.1007/978-3-540-73335-5_21.

[31] Gary Bente, Nicole C. Krämer, and Felix Eschenburg. 2008. Is there anybody out there? *Mediated interpersonal communication*, 131–157. Retrieved 04/28/2014 from.

[32] Gary Bente, Nicole C. Krämer, Anita Petersen, and Jan Peter de Ruiter. 2001. Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior*, 25, 3, 151–166. DOI: 10.1023/A:1010690525717.

[33] Gary Bente, Anita Petersen, Nicole C. Krämer, and Jan Peter De Ruiter. 2001. Transcript-based computer animation of movement: Evaluating a new tool for nonverbal behavior research. *Behavior Research Methods, Instruments, & Computers*, 33, 3, 303–310. DOI: 10.3758/BF03195383.

[34] Gary Bente, Daniel Roth, Thomas Dratsch, and Kai Kaspar. 2016. Emotions of my Kin: Disambiguating Expressive Body Movement in Minimal Groups. *Culture and Brain*, 4, 1, 51–71. DOI: 10.1007/s40167-016-0038-3.

[35] Gary Bente, Sabine Rüggenberg, Nicole C. Krämer, and Felix Eschenburg. 2008. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research*, 34, 2, 287–318. DOI: 10.1111/j.1468-2958.2008.00322.x.

[36] Frank J. Bernieri and Robert Rosenthal. 1991. Interpersonal coordination: Behavior matching and interactional synchrony.

[37] Mark Billinghurst and Hirokazu Kato. 1999. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, 261–284. DOI: 10.1007/978-3-642-87512-0_15.

[38] Nicola Binetti, Charlotte Harrison, Antoine Coutrot, Alan Johnston, and Isabelle Mareschal. 2016. Pupil dilation as an index of preferred mutual gaze duration. *Royal Society Open Science*, 3, 7, 160086. DOI: 10.1098/rsos.160086.

[39] Frank Biocca. 1997. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments. *Journal of Computer-Mediated Communication*, 3, 2. DOI: 10.1111/j.1083-6101.1997.tb00070.x.

[40] Frank Biocca, Chad Harms, and Jenn 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*, 1–9. Retrieved 08/21/2015 from.

[41] Chuck Blanchard, Scott Burgess, Young Harvill, Jaron Lanier, Ann Lasko, Mark Oberman, and Mike Teitel. 1990. *Reality Built for Two: A Virtual Reality Tool*. Number 2-24. ACM.

[42] Olaf Blanke. 2012. Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, 13, 8, 556. DOI: 10/gddnh3.

[43] Olaf Blanke and Thomas Metzinger. 2009. Full-body illusions and minimal phenomenal selfhood. *Trends in cognitive sciences*, 13, 1, 7–13. DOI: 10.1016/j.tics.2008.10.003.

[44] Jim Blascovich. 2002. Social influence within immersive virtual environments. In *The Social Life of Avatars*. Springer London, 127–145.

[45] Jim Blascovich and Jeremy Bailenson. 2011. *Infinite Reality: Avatars, Eternal Life, New Worlds, and the Dawn of the Virtual Revolution*. William Morrow & Co.

[46] Steven M. Boker, Jeffrey F. Cohn, Barry-John Theobald, Iain Matthews, Timothy R. Brick, and Jeffrey R. Spies. 2009. Effects of damping head movement and facial expression in dyadic conversation using real–time facial expression tracking and synthesized avatars. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1535, 3485–3495. DOI: 10.1098/rstb.2009.0152.

[47] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands' feel'touch that eyes see. *Nature*, 391, 6669, 756–756. DOI: 10.1038/35784.

[48] Stéphane Bouchard, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. 2008. Anxiety increases the feeling of presence in virtual reality. *Presence: Teleoperators and Virtual Environments*, 17, 4, 376–391. DOI: 10.1162/pres.17.4.376.

[49] Paul Bourke. 1999. Calculating Stereo Pairs. Retrieved 09/19/2016 from http://paulbourke.net/stereographics/stereorender/.

[50] Doug A. Bowman, Sabine Coquillart, Bernd Froehlich, Michitaka Hirose, Yoshifumi Kitamura, Kiyoshi Kiyokawa, and Wolfgang Stuerzlinger. 2008. 3D user interfaces: new directions and perspectives. *IEEE computer graphics and applications*, 28, 6, 20–36. DOI: 10.1109/MCG.2008.109.

[51] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25, 1, 49–59. DOI: 10.1016/0005-7916(94)90063-9.

[52] Niclas Braun, Stefan Debener, Nadine Spychala, Edith Bongartz, Peter Sorös, Helge Heinrich Oskar Müller, and Alexandra Philipsen. 2018. The senses of agency and ownership: a review. *Frontiers in psychology*, 9, 535. DOI: 10.3389/fpsyg.2018.00535.

[53] Timothy A. Brown. 2014. *Confirmatory Factor Analysis for Applied Research*. Guilford Publications.

[54] Judee K. Burgoon, Laura K. Guerrero, and Valerie Manusov. 2011. Nonverbal signals. *The SAGE handbook of interpersonal communication*, 239–280.

[55] Judee K. Burgoon, Lesa A. Stern, and Leesa Dillman. 2007. *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge University Press.

[56] Joseph N. Cappella. 1981. Mutual influence in expressive behavior: Adult–adult and infant–adult dyadic interaction. *Psychological bulletin*, 89, 1, 101. DOI: 10.1037/0033-2909.89.1.101.

[57] James Carifio and Rocco J. Perla. 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, 3, 3, 106–116.

[58] Christer Carlsson and Olof Hagsand. 1993. DIVE A multi-user virtual reality system. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*. IEEE, 394–400. DOI: 10.1109/VRAIS.1993.380753.

[59] Christer Carlsson and Olof Hagsand. 1993. DIVE—A platform for multi-user virtual environments. *Computers & graphics*, 17, 6, 663–669. DOI: 10.1016/0097-8493(93)90115-P.

[60] Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine*, 22, 4, 67–67.

[61] Justine Cassell. 2000. Embodied conversational interface agents. *Communications of the ACM*, 43, 4, 70–70. DOI: 10.1145/332051.332075.

[62] Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 413–420. DOI: 10.1145/192161.192272.

[63] Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A. Efros. 2018. Everybody dance now. *arXiv preprint arXiv:1808.07371*.

[64] Tanya L. Chartrand and John A. Bargh. 1999. The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, 76, 6, 893. DOI: 10.1037/0022-3514.76.6.893.

[65] Tanya L. Chartrand and Rick Van Baaren. 2009. Human mimicry. *Advances in experimental social psychology*, 41, 219–274. Retrieved 05/10/2017 from.

[66] Ki-Taek Chun and John B. Campbell. 1974. Dimensionality of the Rotter interpersonal trust scale. *Psychological Reports*, 35, 3, 1059–1070. DOI: 10.2466/pr0.1974.35.3.1059.

[67] Andy Clark. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, 36, 3, 181–204. DOI: 10.1017/S0140525X12000477.

[68] W. S. Condon and W. D. Ogston. 1966. Sound and Film Analysis of Normal and Pathological Behavior Patterns. *The Journal of Nervous and Mental Disease*, 143, 4, (October 1966), 338–347. DOI: 10.1097/00005053-196610000-00005.

[69] Mary J. Culnan and M. Lynne Markus. 1987. Information technologies.

[70] Salam Daher, Kangsoo Kim, Myungho Lee, Ryan Schubert, Gerd Bruder, Jeremy Bailenson, and Greg Welch. 2017. Effects of Social Priming on Social Presence with Intelligent Virtual Agents. In *International Conference on Intelligent Virtual Agents*. Springer, 87–100. DOI: 10.1007/978-3-319-67401-8_10.

[71] Mikael Dallaire-Côté, Philippe Charbonneau, Sara St-Pierre Côté, Rachid Aissaoui, and David R. Labbe. 2016. Animated self-avatars for motor rehabilitation applications that are biomechanically accurate, low-latency and easy to use. In *Virtual Reality (VR), 2016 IEEE*. IEEE, 167–168. DOI: 10.1109/VR.2016.7504706.

[72] Charles Darwin. 1872. *The Origin of Species by Means of Natural Selection: Or, the Preservation of Favoured Races in the Struggle for Life and the Descent of Man and Selection in Relation to Sex*. Modern library.

[73] Charles Robert Darwin. 1872. *The Expression of the Emotions in Man and Animals*. John Murray, London.

[74] Tapas K. Das, Gurminder Singh, Alex Mitchell, P. Senthil Kumar, and Kevin McGee. 1997. NetEffect: a network architecture for large-scale multi-user virtual worlds. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. ACM, 157–163. DOI: `10.1145/261135.261164`.

[75] Nicole David, Albert Newen, and Kai Vogeley. 2008. The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and cognition*, 17, 2, 523–534. DOI: `10.1016/j.concog.2008.03.004`.

[76] Deborah Davis and William T. Perkowitz. 1979. Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction. *Journal of Personality and Social Psychology*, 37, 4, 534. DOI: `10.1037/0022-3514.37.4.534`.

[77] Hanne De Jaegher, Ezequiel Di Paolo, and Shaun Gallagher. 2010. Can social interaction constitute social cognition? *Trends in cognitive sciences*, 14, 10, 441–447. DOI: `10.1016/j.tics.2010.06.009`.

[78] Yu Ding, Yuting Zhang, Meihua Xiao, and Zhigang Deng. 2017. A multi-faceted study on eye contact based speaker identification in three-party conversations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3011–3021. DOI: `10.1145/3025453.3025644`.

[79] Richard C. Dorf and Robert H. Bishop. 1998. Modern control systems.

[80] Hubert L. Dreyfus. 1972. What computers can't do.

[81] Olive Jean Dunn. 1964. Multiple comparisons using rank sums. *Technometrics*, 6, 3, 241–252. DOI: `10.1080/00401706.1964.10490181`.

[82] Nat Durlach and Mel Slater. 2000. Presence in shared virtual environments and virtual togetherness. *Presence: Teleoperators & Virtual Environments*, 9, 2, 214–217.

[83] H. Henrik Ehrsson. 2007. The experimental induction of out-of-body experiences. *Science*, 317, 5841, 1048–1048. DOI: `10/ch9m6b`.

[84] Paul Ekman. 1999. Basic emotions. *Handbook of cognition and emotion*, 98, 45–60.

[85] Paul Ekman. 1993. Facial expression and emotion. *American Psychologist*, 48, 4, 384. DOI: `10.1037/0003-066X.48.4.384`.

[86] Paul Ekman. 1970. Universal facial expressions in emotion. *Studia Psychologica*.

[87] Mohammed S. Elbamby, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. 2018. Toward low-latency and ultra-reliable virtual reality. *IEEE Network*, 32, 2, 78–84. DOI: `10.1109/MNET.2018.1700268`.

[88] Irv Englander and Arv Englander. 2003. *The Architecture of Computer Hardware and Systems Software: An Information Technology Approach*. Wiley.

[90] Andy P. Field. 2009. *Discovering Statistics Using SPSS: (And Sex, Drugs and Rock 'n' Roll)*. (3rd ed ed.). SAGE Publications, Los Angeles. 821 pages. ISBN: 978-1-84787-906-6 978-1-84787-907-3.

[91] Andy Field, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. Sage publications.

[92] Lawrence H. Frank, John G. Casali, and Walter W. Wierwille. 1988. Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors*, 30, 2, 201–217. DOI: 10.1177/001872088803000207.

[93] Emmanuel Frecon and Marten Stenius. 1998. DIVE: A scaleable network architecture for distributed virtual environments. *Distributed Systems Engineering*, 5, 3, 91. DOI: 10.1088/0967-1846/5/3/002.

[94] Alexandra Frischen, Andrew P. Bayliss, and Steven P. Tipper. 2007. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133, 4, (July 2007), 694–724. DOI: 10.1037/0033-2909.133.4.694.

[95] S. Friston, A. Steed, S. Tilbury, and G. Gaydadjiev. 2016. Construction and Evaluation of an Ultra Low Latency Frameless Renderer for VR. *IEEE Transactions on Visualization and Computer Graphics*, 22, 4, (April 2016), 1377–1386. DOI: 10.1109/TVCG.2016.2518079.

[96] Sebastian Friston and Anthony Steed. 2014. Measuring latency in virtual environments. *IEEE transactions on visualization and computer graphics*, 20, 4, 616–625. DOI: 10.1109/TVCG.2014.30.

[97] Bernd Froehlich and Doug A. Bowman. 2009. 3D user interfaces. *IEEE computer graphics and applications*, 29, 6, 20–21. DOI: 10.1109/MCG.2009.113.

[98] Henry Fuchs, Gary Bishop, Kevin Arthur, Leonard McMillan, Ruzena Bajcsy, Sang Lee, Hany Farid, and Takeo Kanade. 1994. Virtual space teleconferencing using a sea of cameras. In *Proc. First International Conference on Medical Robotics and Computer Assisted Surgery*. Volume 26.

[99] Thomas A. Funkhouser. 1995. RING: a client-server system for multi-user virtual environments. In *Proceedings of the 1995 Symposium on Interactive 3D Graphics*. ACM, 85–ff. DOI: 10.1145/199404.199418.

[100] Thomas A. Funkhouser et al. 1996. Network Topologies for Scalable Multi-User Virtual Environments. In *Vrais*. Volume 96. Citeseer, 222. DOI: 10.1109/VRAIS.1996.490531.

[101] Joaquín M. Fuster. 2006. The cognit: A network model of cortical representation. *International Journal of Psychophysiology*, 60, 2, 125–132. DOI: 10.1016/j.ijpsycho.2005.12.015.

[102] Joaquın M. Fuster. 2001. The prefrontal cortex—an update: time is of the essence. *Neuron*, 30, 2, 319–333. DOI: 10.1016/S0896-6273(01)00285-9.

[103] Simon Gächter, Chris Starmer, and Fabio Tufano. 2015. Measuring the closeness of relationships: a comprehensive evaluation of the'Inclusion of the Other in the Self'scale. *PloS one*, 10, 6, e0129478. DOI: 10.1371/journal.pone.0129478.

[104] Shaun Gallagher. 2006. *How the Body Shapes the Mind*. Clarendon Press.

[105] Shaun Gallagher. 2000. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4, 1, 14–21. DOI: 10.1016/S1364-6613(99)01417-5.

[106] Shaun Gallagher. 2001. The practice of mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies*, 8, 5-6, 83–108.

[107] Mattia Gallotti and Chris D. Frith. 2013. Social cognition in the we-mode. *Trends in cognitive sciences*, 17, 4, 160–165. DOI: 10.1016/j.tics.2013.02.002.

[108] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. 2001. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 309–316. DOI: 10.1145/365024.365121.

[109] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 529–536. DOI: 10.1145/642611.642703.

[111] Alexandra Livia Georgescu, Bojana Kuzmanovic, Daniel Roth, Gary Bente, and Kai Vogeley. 2014. The Use of Virtual Characters to Assess and Train Non-Verbal Communication in High-Functioning Autism. *Frontiers in Human Neuroscience*, 8. DOI: 10.3389/fnhum.2014.00807.

[112] Michael Gerhard. 2003. A hybrid avatar/agent model for educational CVEs. *PhD diss., PhD Thesis, Leeds Metropolitan University*.

[113] Michael Gerhard. 1997. Avatars and agents in collaborative virtual environments for education. In *Proceedings of the Conference on Human Computer Interaction (HCI), Bristol, UK.*

[114] Michael Gerhard, David J. Moore, and Dave J. Hobbs. 2001. Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation. In *International Workshop on Intelligent Virtual Agents*. Springer, 137–155. DOI: 10.1007/3-540-44812-8_12.

[115] Michael Gerhard, David Moore, and Dave Hobbs. 2004. Embodiment and copresence in collaborative interfaces. *International Journal of Human-Computer Studies*, 61, 4, 453–480. DOI: 10.1016/j.ijhcs.2003.12.014.

[116] James Jerome Gibson. 1966. The senses considered as perceptual systems.

[117] Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. Volume 5. New York: Doubleday.

[118] Mar Gonzalez-Franco Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI*, 5, 74. DOI: 10.3389/frobt.2018.00074.

[119] Mar González-Franco, Daniel Pérez-Marcos, Bernhard Spanlang, and Mel Slater. 2010. The contribution of real-time mirror reflections of motor actions on virtual body ownership in an immersive virtual environment. In *Virtual Reality Conference (VR), 2010 IEEE*. IEEE, 111–114. DOI: 10.1109/VR.2010.5444805.

[120] Jonathan Gratch, David DeVault, Gale M. Lucas, and Stacy Marsella. 2015. Negotiation as a challenge problem for virtual humans. In *International Conference on Intelligent Virtual Agents*. Springer, 201–215. DOI: 10.1007/978-3-319-21996-7_21.

[121] Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, Rick J. van der Werf, and Louis-Philippe Morency. 2006. Virtual rapport. In *International Workshop on Intelligent Virtual Agents*. Springer, 14–27. DOI: 10.1007/11821830_2.

[122] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast, and Robin Duffy. 2007. Creating rapport with virtual agents. In *International Workshop on Intelligent Virtual Agents*. Springer, 125–138. DOI: 10.1007/978-3-540-74997-4_12.

[123] Jonathan Gratch, Ning Wang, Anna Okhmatovskaia, Francois Lamothe, Mathieu Morales, Rick J. van der Werf, and Louis-Philippe Morency. 2007. Can virtual humans be more engaging than real ones? In *International Conference on Human-Computer Interaction*. Springer, 286–297. DOI: 10.1007/978-3-540-73110-8_30.

[124] Richard Langton Gregory. 1970. The intelligent eye.

[125] Victoria Groom, Jeremy N. Bailenson, and Clifford Nass. 2009. The influence of racial embodiment on racial bias in immersive virtual environments. DOI: 10.1080/15534510802643750.

[126] David Güera and Edward J. Delp. 2018. Deepfake Video Detection Using Recurrent Neural Networks. In *IEEE International Conference on Advanced Video and Signal-Based Surveillance (to Appear)*. DOI: 10.1109/AVSS.2018.8639163.

[127] Antal Haans, Wijnand A. IJsselsteijn, and Yvonne AW de Kort. 2008. The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb. *Body Image*, 5, 4, 389–394. DOI: 10.1016/j.bodyim.2008.04.003.

[128] Joanna Hale and Antonia Hamilton. 2016. Testing the relationship between mimicry, trust and rapport in virtual reality conversations. *Scientific reports*, 6, 35295. DOI: 10.1038/srep35295.

[129] Edward Twitchell Hall. 1966. The hidden dimension.

[131] Chad Harms and Frank Biocca. 2004. Internal consistency and reliability of the networked minds measure of social presence.

[132] Robert G. Harper, Arthur N. Wiens, and Joseph D. Matarazzo. 1978. *Nonverbal Communication: The State of the Art.* John Wiley & Sons.

[133] Jonathon D. Hart, Thammathip Piumsomboon, Louise Lawrence, Gun A. Lee, Ross T. Smith, and Mark Billinghurst. 2018. Emotion Sharing and Augmentation in Cooperative Virtual Reality Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, 453–460.

[134] Ding He, Fuhu Liu, Dave Pape, Greg Dawe, and Dan Sandin. 2000. Video-based measurement of system latency. In *International Immersive Projection Technology Workshop*.

[135] Carrie Heeter. 1992. Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1, 2, 262–271.

[136] Chin-Chang Ho and Karl F. MacDorman. 2010. Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26, 6, (November 2010), 1508–1518. DOI: 10.1016/j.chb.2010.05.015.

[137] Michael A. Hogg and John C. Turner. 1985. Interpersonal attraction, social identification and psychological group formation. *European journal of social psychology*, 15, 1, 51–66. DOI: 10.1002/ejsp.2420150105.

[138] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '92*. The SIGCHI Conference. ACM Press, Monterey, California, United States, 119–125. DOI: 10.1145/142750.142769.

[139] Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. 2011. Virtual Rapport 2.0. In *International Workshop on Intelligent Virtual Agents*. Springer, 68–79. DOI: 10.1007/978-3-642-23974-8_8.

[140] Masakazu Ide. 2013. The effect of "anatomical plausibility" of hand angle on the rubber-hand illusion. *Perception*, 42, 1, 103–111. DOI: 10.1068/p7322.

[141] Wijnand A. IJsselsteijn. 2005. History of telepresence. *3D Communication: Algorithms, concepts and real-time systems in human centred communication, John Wiley & Sons: Chichester*.

[142] Wijnand A. IJsselsteijn, Yvonne A. W. de Kort, and Antal Haans. 2006. Is this my hand I see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments*, 15, 4, 455–464. DOI: 10.1162/pres.15.4.455.

[143] Beverly K. Jaeger and Ronald R. Mourant. 2001. Comparison of simulator sickness using static and dynamic walking simulators. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Volume 45. SAGE Publications Sage CA: Los Angeles, CA, 1896–1900. DOI: 10/fx2jqf.

[144] Anil K. Jain. 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*. Award Winning Papers from the 19th International Conference on Pattern Recognition (ICPR) 31, 8, (June 1, 2010), 651–666. DOI: 10.1016/j.patrec.2009.09.011.

[145] S. Janke and A. Glöckner-Rist. 2014. Deutsche Version der Positive and Negative Affect Schedule (PANAS). In *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen. Doi*. Volume 10, 6102.

[146] Jason Jerald. 2015. *The VR Book: Human-Centered Design for Virtual Reality*. Morgan & Claypool.

[147] Y. U. Jihun and PARK Jungwoon. 2017. Head-mounted display with facial expression detecting capability. (March 2017). Patent.

[148] Sungchul Jung and Charles E. Hughes. 2016. The effects of indirect real body cues of irrelevant parts on virtual body ownership and presence. In *Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments*. Eurographics Association, 107–114.

[149] Marcel A. Just and Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87, 4, 329. DOI: 10.1037/0033-295X.87.4.329.

[150] Henry F. Kaiser. 1960. The application of electronic computers to factor analysis. *Educational and psychological measurement*.

[151] Andreas Kalckert and H. Henrik Ehrsson. 2012. Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Frontiers in human neuroscience*, 6. DOI: 10.3389/fnhum.2012.00040.

[152] Aslan Kanamgotov, Athanasios Christopoulos, Marc Conrad, and Simant Prakoonwit. 2012. Immersion in Virtual Worlds - But not Second Life! In IEEE, (September 2012), 107–113. DOI: 10.1109/CW.2012.22.

[153] Tero Karras, Samuli Laine, and Timo Aila. 2018. A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv preprint arXiv:1812.04948*.

[154] Kai Kaspar, Sabine König, Jessika Schwandt, and Peter König. 2014. The experience of new sensorimotor contingencies by sensory augmentation. *Consciousness and cognition*, 28, 47–63. DOI: 10/f6dwvj.

[155] Osamu Kawakami, Yoshiki Kaneoke, Koichi Maruyama, Ryusuke Kakigi, Tomohisa Okada, Norihiro Sadato, and Yoshiharu Yonekura. 2002. Visual detection of motion speed in humans: spatiotemporal analysis by fMRI and MEG. *Human brain mapping*, 16, 2, 104–118. DOI: 10.1002/hbm.10033.

[156] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3, 3, 203–220. DOI: 10.1207/s15327108ijap0303_3.

[157] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21, 4, 373–387. DOI: 10.1162/PRES_a_00124.

[158] Konstantina Kilteni, Antonella Maselli, Konrad P. Kording, and Mel Slater. 2015. Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9. DOI: 10.3389/fnhum.2015.00141.

[159] Michael Kipp and Patrick Gebhard. 2008. Igaze: Studying reactive gaze behavior in semi-immersive human-avatar interactions. In *International Workshop on Intelligent Virtual Agents*. Springer, 191–199. DOI: 10.1007/978-3-540-85483-8_19.

[160] Chris L. Kleinke. 1986. Gaze and eye contact: a research review. *Psychological bulletin*, 100, 1, 78. DOI: 10.1037/0033-2909.100.1.78.

[161] Rune Klevjer. 2012. Enter the avatar: The phenomenology of prosthetic telepresence in computer games. In *The Philosophy of Computer Games*. Springer, 17–38.

[162] Rex B. Kline. 2010. *Principles and Practice of Structural Equation Modeling 3 Rd Ed*. New York, NY, The Guilford Press.

[163] Evelyne Kohler, Christian Keysers, M. Alessandra Umilta, Leonardo Fogassi, Vittorio Gallese, and Giacomo Rizzolatti. 2002. Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297, 5582, 846–848. DOI: 10.1126/science.1070311.

[164] Stefan Kopp. 2010. Social resonance and embodied coordination in face-to-face conversation with artificial interlocutors. *Speech Communication*, 52, 6, 587–597. DOI: 10.1016/j.specom.2010.02.007.

[165] Alexander Kulik, André Kunert, Stephan Beck, Carl-Feofan Matthes, Andre Schollmeyer, Adrian Kreskowski, Bernd Fröhlich, Sue Cobb, and Mirabelle D'Cruz. 2018. Virtual Valcamonica: collaborative exploration of prehistoric petroglyphs and their surrounding environment in multi-user virtual reality. *PRESENCE: Teleoperators and Virtual Environments*, 26, 03, 297–321. DOI: 10.1162/pres_a_00297.

[166] Marianne LaFrance and Maida Broadbent. 1976. Group rapport: Posture sharing as a nonverbal indicator. *Group & Organization Studies*, 1, 3, 328–333. DOI: 10.1177/105960117600100307.

[167] Jessica L. Lakin and Tanya L. Chartrand. 2003. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological science*, 14, 4, 334–339. DOI: 10.1111/1467-9280.14481.

[168] Jessica L. Lakin, Valerie E. Jefferis, Clara Michelle Cheng, and Tanya L. Chartrand. 2003. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of nonverbal behavior*, 27, 3, 145–162. DOI: 10.1023/A:1025389814290.

[170] Ellen J. Langer. 1992. Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and cognition*, 1, 3, 289–305. DOI: 10.1016/1053-8100(92)90066-J.

[171] Jaron Lanier. 2010. *You Are Not a Gadget: A Manifesto*. Vintage.

[173] Marc Erich Latoschik, Jean-Luc Lugrin, and Daniel Roth. 2016. FakeMi: a fake mirror system for avatar embodiment studies. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. ACM, Munich, Germany, 73–76. DOI: 10.1145/2993369.2993399.

[174] Marc Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The Effect of Avatar Realism in Immersive Social Virtual Realities. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology*. VRST '17. Gothenburg, Sweden, 39:1–39:10. DOI: 10.1145/3139131.3139156.

[175] Joseph J. LaViola Jr, Ernst Kruijff, Ryan P. McMahan, Doug Bowman, and Ivan P. Poupyrev. 2017. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional.

[176] Binh H. Le, Xiaohan Ma, and Zhigang Deng. 2012. Live speech driven head-and-eye motion generators. *IEEE transactions on visualization and computer graphics*, 18, 11, 1902–1914. DOI: 10.1109/TVCG.2012.74.

[177] Anatole Lécuyer, Fabien Lotte, Richard B. Reilly, Robert Leeb, Michitaka Hirose, and Mel Slater. 2008. Brain-computer interfaces, virtual reality, and videogames. *Computer*, 41, 10, 66–72. DOI: 10.1109/MC.2008.410.

[178] Bongshin Lee, Petra Isenberg, Nathalie Henry Riche, and Sheelagh Carpendale. 2012. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics*, 18, 12, 2689–2698. DOI: 10.1109/TVCG.2012.204.

[179] Kwan Min Lee. 2004. Presence, explicated. *Communication theory*, 14, 1, 27–50. DOI: 10.1111/j.1468-2885.2004.tb00302.x.

[180] Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler. 2002. Eyes alive. In *ACM Transactions on Graphics (TOG)*. Volume 21. ACM, 637–644.

[181] Bigna Lenggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke. 2007. Video ergo sum: manipulating bodily self-consciousness. *Science*, 317, 5841, 1096–1099. DOI: 10.1126/science.1143439.

[182] Benny Liebold, Michael Brill, Daniel Pietschmann, Frank Schwab, and Peter Ohler. 2017. Continuous measurement of breaks in presence: Psychophysiology and orienting responses. *Media Psychology*, 20, 3, 477–501. DOI: 10.1080/15213269.2016.1206829.

[183] Matthew Lombard and Theresa Ditton. 1997. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3, 2. DOI: 10.1111/j.1083-6101.1997.tb00072.x.

[184] Matthew R. Longo, Friederike Schüür, Marjolein P.M. Kammers, Manos Tsakiris, and Patrick Haggard. 2008. What is embodiment? A psychometric approach. *Cognition*, 107, 3, (June 2008), 978–998. DOI: 10.1016/j.cognition.2007.12.004.

[185] Jack M. Loomis, Jonathan W. Kelly, Matthias Pusch, Jeremy N. Bailenson, and Andrew C. Beall. 2008. Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, 37, 9, 1443–1457. DOI: 10.1068/p5896.

[186] Christopher Lopez, Pär Halje, and Olaf Blanke. 2008. Body ownership and embodiment: vestibular and multisensory mechanisms. *Neurophysiologie Clinique/Clinical Neurophysiology*, 38, 3, 149–161.

[189] Jean-Luc Lugrin, Johanna Latt, and Marc Erich Latoschik. 2015. Anthropomorphism and Illusion of Virtual Body Ownership. In *International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments*.

[191] Jean-Luc Lugrin, David Zilch, Daniel Roth, Gary Bente, and Marc Erich Latoschik. 2016. Facebo: Real-time face and body tracking for faithful avatar synthesis. In *2016 IEEE Virtual Reality*. IEEE VR 2016. IEEE, Greenville, SC, USA, 225–226. DOI: 10.1109/VR.2016.7504735.

[192] Xiaohan Ma and Zhigang Deng. 2009. Natural eye motion synthesis by modeling gaze-head coupling. In *2009 IEEE Virtual Reality Conference*. IEEE, 143–150.

[193] Karl F. MacDorman and Debaleena Chattopadhyay. 2016. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, 146, (January 2016), 190–205. DOI: 10.1016/j.cognition.2015.09.019.

[194] Michael Madary and Thomas K. Metzinger. 2016. Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI*, 3, 3. DOI: 10.3389/frobt.2016.00003.

[195] John F. Maguire and Piers DL Howe. 2016. Failure to detect meaning in RSVP at 27 ms per picture. *Attention, Perception, & Psychophysics*, 78, 5, 1405–1413. DOI: 10.3758/s13414-016-1096-5.

[196] Andrew Maimone and Henry Fuchs. 2011. A first look at a telepresence system with room-sized real-time 3d capture and life-sized tracked display wall. *Proceedings of ICAT 2011*, 4–9.

[197] Massimo Marraffa. 2019. Theory of Mind — Internet Encyclopedia of Philosophy. Retrieved 06/03/2019 from https://www.iep.utm.edu/theomind/.

[198] Antonella Maselli and Mel Slater. 2014. Sliding perspectives: dissociating ownership from self-location during full body illusions in virtual reality. *Frontiers in human neuroscience*, 8, 693. DOI: 10.3389/fnhum.2014.00693.

[199] Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Frontiers in human neuroscience*, 7. DOI: 10.3389/fnhum.2013.00083.

[200] David Matsumoto and Mark G. Frank. 2012. *Nonverbal Communication: Science and Applications: Science and Applications*. Sage.

[201] Richard E. Maurer and Jeffrey H. Tindall. 1983. Effect of postural congruence on client's perception of counselor empathy. *Journal of Counseling Psychology*, 30, 2, 158. DOI: 10.1037/0022-0167.30.2.158.

[202] Daniel N. McIntosh. 2006. Spontaneous facial mimicry, liking and emotional contagion. *Polish Psychological Bulletin*, 37, 1, 31.

[203] Quinn McNemar. 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12, 2, 153–157. DOI: 10.1007/BF02295996.

[204] Alain Mignault and Avi Chaudhuri. 2003. The many faces of a neutral face: Head tilt and perception of dominance and emotion. *Journal of Nonverbal Behavior*, 27, 2, 111–132. DOI: 10.1023/A:1023914509763.

[205] Lynden K. Miles, Louise K. Nind, and C. Neil Macrae. 2009. The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, 45, 3, (May 2009), 585–589. DOI: 10.1016/j.jesp.2009.02.002.

[206] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Photonics for Industrial Applications*. International Society for Optics and Photonics, 282–292.

[207] Marvin Minsky. 1980. Telepresence.

[208] Simone C. Mölbert, Anne Thaler, Betty J. Mohler, Stephan Streuber, Javier Romero, Michael J. Black, Stephan Zipfel, H.-O. Karnath, and Katrin E. Giel. 2018. Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: Attitudinal components rather than visual body size estimation are distorted. *Psychological medicine*, 48, 4, 642–653. DOI: 10/gdcw82.

[209] Chris Moore. 2014. *Joint Attention: Its Origins and Role in Development*. Psychology Press.

[210] Masahiro Mori. 1970. The uncanny valley. *Energy*, 7, 4, 33–35.

[211] Adam Morton. 1980. *Frames of Mind: Constraints on the Common-Sense Conception of the Mental*. Oxford:Clarendon Press.

[212] Andreas Mühlberger, Martin J. Herrmann, Georg Wiedemann, Heiner Ellgring, and Paul Pauli. 2001. Repeated exposure of flight phobics to flights in virtual reality. *Behaviour research and therapy*, 39, 9, 1033–1050. DOI: 10.1016/S0005-7967(00)00076-0.

[213] Peter Mundy and Lisa Newell. 2007. Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, 16, 5, (October 1, 2007), 269–274. DOI: 10.1111/j.1467-8721.2007.00518.x.

[214] Norman Murray, Dave Roberts, Anthony Steed, Paul Sharkey, Paul Dickerson, and John Rae. 2007. An assessment of eye-gaze potential within immersive virtual environments. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 3, 4, 8.

[215] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56, 1, 81–103. DOI: 10.1111/0022-4537.00153.

[216] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 72–78.

[217] Alva Noë. 2009. *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Macmillan.

[218] Aline Normoyle, Jeremy B. Badler, Teresa Fan, Norman I. Badler, Vinicius J. Cassol, and Soraia R. Musse. 2013. Evaluating perceived trust from procedurally animated gaze. In *Proceedings of Motion on Games*. ACM, 141–148. DOI: 10.1145/2522628.2522630.

[219] Tor Nørretranders. 1991. *The User Illusion: Cutting Consciousness down to Size.* Viking.

[220] Kristine L. Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence*, 12, 5, 481–494. DOI: 10.1162/105474603322761289.

[221] J. Kevin O'Regan. 2011. *Why Red Doesn't Sound like a Bell: Understanding the Feel of Consciousness*. Oxford University Press.

[222] J. Kevin O'Regan and Alva Noë. 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24, 5, 939–973. DOI: 10/b3dbnv.

[223] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Front. Robot. AI 5: 114. doi: 10.3389/frobt*. DOI: 10.3389/frobt.2018.00114.

[224] S. Y. Oh, Jeremy N. Bailenson, N. Krämer, and B. Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLoS ONE*, 11, 9, e0161794. DOI: 10.1371/journal.pone.0161794.

[225] Mary Beth Oliver and Anne Bartsch. 2010. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research*, 36, 1, 53–81. DOI: 10.1111/j.1468-2958.2009.01368.x.

[226] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, and Mingsong Dou. 2016. Holoportation: Virtual 3d teleportation in real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 741–754. DOI: 10.1145/2984511.2984517.

[227] Kazuhiro Otsuka. 2016. MMSpace: Kinetically-augmented telepresence for small group-to-group conversations. In *Virtual Reality (VR), 2016 IEEE*. IEEE, 19–28. DOI: 10.1109/VR.2016.7504684.

[228] Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright. 2004. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29, 3, 232–240. DOI: 10.1002/col.20010.

[229] Giorgos Papadakis, Katerina Mania, and Eftichios Koutroulis. 2011. A system to measure, control and minimize end-to-end head tracking latency in immersive simulations. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*. ACM, 581–584. DOI: 10.1145/2087756.2087869.

[230] Kyoung Shin Park and Robert V. Kenyon. 1999. Effects of network characteristics on human performance in a collaborative virtual environment. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE, 104–111.

[231] Sarah Parsons and Peter Mitchell. 2002. The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of intellectual disability research*, 46, 5, 430–443. DOI: 10.1046/j.1365-2788.2002.00425.x.

[232] Tabitha C. Peck, Sofia Seinfeld, Salvatore M. Aglioti, and Mel Slater. 2013. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition*, 22, 3, 779–787. DOI: 10.1016/j.concog.2013.04.016.

[233] Wilder Penfield and Theodore Rasmussen. 1950. The cerebral cortex of man; a clinical study of localization of function.

[234] Patrick Pérez, Michel Gangnet, and Andrew Blake. 2003. Poisson image editing. *ACM Transactions on graphics (TOG)*, 22, 3, 313–318. DOI: 10.1145/882262.882269.

[235] Valeria I. Petkova and H. Henrik Ehrsson. 2008. If I were you: perceptual illusion of body swapping. *PloS one*, 3, 12, e3832. DOI: 10.1371/journal.pone.0003832.

[236] Tomas Pfister, Xiaobai Li, Guoying Zhao, and Matti Pietikäinen. 2011. Recognising spontaneous facial micro-expressions. In *2011 International Conference on Computer Vision*. IEEE, 1449–1456. DOI: 10.1109/ICCV.2011.6126401.

[237] Rosalind W. Picard. 2000. *Affective Computing*. MIT press.

[238] Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27, 2, 169–190. DOI: 10.1017/S0140525X04000056.

[239] Ivelina V. Piryankova, Hong Yu Wong, Sally A. Linkenauger, Catherine Stinson, Matthew R. Longo, Heinrich H. Bülthoff, and Betty J. Mohler. 2014. Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. *PloS one*, 9, 8, e103428. DOI: 10/gf36rz.

[240] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 46. DOI: 10.1145/3173574.3173620.

[241] Iana Podkosova. 2018. *Walkable Multi-User VRThe Effects of Physical and Virtual Colocation*. PhD Thesis. Technical University of Vienna.

[242] Mary C. Potter, Brad Wyble, Carl Erick Hagmann, and Emily S. McCourt. 2014. Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 76, 2, 270–279. DOI: 10.3758/s13414-013-0605-z.

[243] Aaron Powers and Sara Kiesler. 2006. The advisor robot: tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*. ACM, 218–225. DOI: 10.1145/1121241.1121280.

[244] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1, 04, 515–526. DOI: 10.1017/S0140525X00076512.

[245] Rossana B. Queiroz, Leandro M. Barros, and Soraia R. Musse. 2008. Providing expressive gaze to virtual animated characters in interactive applications. *Computers in Entertainment (CIE)*, 6, 3, 41.

[246] Puthankurissi S. Raju, Subhash C. Lonial, and W. Glynn Mangold. 1995. Differential effects of subjective knowledge, objective knowledge, and usage experience on decision making: An exploratory investigation. *Journal of consumer psychology*, 4, 2, 153–180. DOI: 10.1207/s15327663jcp0402_04.

[247] Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality*, 41, 1, 203–212. DOI: 10.1016/j.jrp.2006.02.001.

[248] Fabian Ramseyer. 2011. Nonverbal synchrony in psychotherapy: embodiment at the level of the dyad. *The Implications of Embodiment: Cognition and Communication*, 193–207.

[249] Rabindra Ratan. 2012. Self-Presence, Explicated: Body, Emotion, and Identity. *Handbook of Research on Technoself: Identity in a Technological Society*, 322.

[250] Rabindra Ratan and Béatrice S. Hasler. 2010. Exploring Self-Presence in Collaborative Virtual Teams. *PsychNology Journal*, 8, 1, 11–31.

[251] Rabindra Ratan and Young June Sah. 2015. Leveling up on stereotype threat: The role of avatar customization and avatar embodiment. *Computers in Human Behavior*, 50, 367–374. DOI: 10.1016/j.chb.2015.04.010.

[252] Rabindra Ratan, Michael Santa Cruz, and Peter Vorderer. 2007. Multitasking, presence, and self-presence on the Wii. In *Proceedings of the 10th Annual International Workshop on Presence*. Starlab Barcelona SL, 167–190.

[253] Byron Reeves and Clifford Ivar Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places.* Cambridge university press.

[255] Martin Riemer, Xaver Fuchs, Florian Bublatzky, Dieter Kleinböhl, Rupert Hölzl, and Jörg Trojan. 2014. The rubber hand illusion depends on a congruent mapping between real and artificial fingers. *Acta psychologica*, 152, 34–41. DOI: 10.1016/j.actpsy.2014.07.012.

[256] Giuseppe Riva, Luca Melis, and Milco Bolzoni. 1997. Treating body-image disturbances. *Communications of the ACM*, 40, 8, 69–72. DOI: 10.1145/257874.257890.

[257] Albert Rizzo, Greg Reger, Greg Gahm, JoAnn Difede, and Barbara O. Rothbaum. 2009. Virtual reality exposure therapy for combat-related PTSD. In *Post-Traumatic Stress Disorder*. Springer, 375–399.

[258] Giacomo Rizzolatti and Laila Craighero. 2004. The mirror-neuron system. *Annu. Rev. Neurosci.*, 27, 169–192. DOI: 10.1146/annurev.neuro.27.070203.144230.

[259] Giacomo Rizzolatti and Giuseppe Luppino. 2001. The cortical motor system. *Neuron*, 31, 6, 889–901. DOI: 10.1016/S0896-6273(01)00423-8.

[260] David Roberts, Robin Wolff, John Rae, Anthony Steed, Rob Aspin, Moira McIntyre, Adriana Pena, Oyewole Oyekoya, and Will Steptoe. 2009. Communicating eye-gaze across a distance: Comparing an eye-gaze enabled immersive collaborative virtual environment, aligned video conferencing, and being together. In *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 135–142. DOI: 10.1109/VR.2009.4811013.

[261] Sheena Rogers, Melanie Lunsford, Lars Strother, and Michael Kubovy. 2003. The Mona Lisa effect: Perception of gaze direction in real and pictured faces. *Studies in perception and action VII*, 19–24.

[265] Daniel Roth, Carola Bloch, Anne-Kathrin Wilbers, Marc Erich Latoschik, Kai Kaspar, and Gary Bente. 2016. What You See is What You Get: Channel Dominance in the Decoding of Affective Nonverbal Behavior Displayed by Avatars. In *66th Annual Conference of the International Communication Association*. Fukuoka, Japan.

[266] Daniel Roth, Constantin Kleinbeck, Tobias Feigl, and Christopher Mutschler. 2017. Social Augmentations in Multi-User Virtual Reality: A Virtual Museum Experience. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. ISMAR '17. Nantes, France, 42–43. DOI: 10.1109/ismar-adjunct.2017.28.

[267] Daniel Roth, Constantin Kleinbeck, Tobias Feigl, Christopher Mutschler, and Marc Erich Latoschik. 2018. Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE VR 2018. IEEE, Reutlingen, Germany, 215–222. DOI: 10.1109/VR.2018.8447550.

[268] Daniel Roth, Peter Kullmann, Gary Bente, Dominik Gall, and Marc Erich Latoschik. 2018. Effects of Hybrid and Synthetic Social Gaze in Avatar-Mediated Interactions. In *Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality 2018*. ISMAR '18. IEEE, Munich, Germany. DOI: 10.1109/ISMAR-Adjunct.2018.00044.

[269] Daniel Roth, Marc Erich Latoschik, Kai Vogeley, and Gary Bente. 2015. Hybrid Avatar-Agent Technology – A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com*, 14, 2, 107–114. DOI: 10.1515/icom-2015-0030.

[270] Daniel Roth, Jean-Luc Lugrin, Julia Büser, Gary Bente, Arnulph Fuhrmann, and Marc Erich Latoschik. 2016. A Simplified Inverse Kinematic Approach for Embodied VR Applications. In *Proceedings of the IEEE Virtual Reality (IEEE VR) Conference 2016*. IEEE VR 2016. Greenville, SC, USA. DOI: 10.1109/VR.2016.7504760.

[271] Daniel Roth, Jean-Luc Lugrin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. Avatar Realism and Social Interaction Quality in Virtual Reality. In *2016 IEEE Virtual Reality*. IEEE VR 2016. IEEE, Greenville, SC, USA, 277–278. DOI: 10.1109/VR.2016.7504761.

[272] Daniel Roth, Jean-Luc Lugrin, Marc Erich Latoschik, and Stephan Huber. 2017. Alpha IVBO - Construction of a Scale to Measure the Illusion of Virtual Body Ownership. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. ACM, Denver, CO, USA, 2875–2883. DOI: 10.1145/3027063.3053272.

[273] Daniel Roth, Jean-Luc Lugrin, Sebastian von Mammen, and Marc Erich Latoschik. 2017. Controllers & inputs: Masters of puppets. In *Avatar, Assembled: The Social and Technical Anatomy of Digital Bodies, Jaime Banks (Ed.)* 281–290.

[274] Daniel Roth, David Mal, Christian Felix Purps, Peter Kullmann, and Marc Erich Latoschik. 2018. Injecting Nonverbal Mimicry with Hybrid Avatar-Agent Technologies: A Naive Approach. In *Proceedings of the Symposium on Spatial User Interaction*. SUI '18. ACM, Berlin, Germany, 69–73. DOI: 10.1145/3267782.3267791.

[275] Daniel Roth, Jan-Philipp Stauffert, and Marc Erich Latoschik. 2019. Avatar Embodiment, Behavior Replication, and Kinematics in Virtual Reality (in press). In *VR Developer Gems, William R Sherman (Ed.)* VR Developer Gems. Volume 1. Taylor & Francis.

[277] Daniel Roth, Kristoffer Waldow, Marc Erich Latoschik, Arnulph Fuhrmann, and Gary Bente. 2017. Socially Immersive Avatar-based Communication. In *2017 IEEE Virtual Reality (VR)*. IEEE VR 2017. IEEE, Los Angeles, USA, 259–260. DOI: 10.1109/VR.2017.7892275.

[278] Daniel Roth, Kristoffer Waldow, Felix Stetter, Gary Bente, Marc Erich Latoschik, and Arnulph Fuhrmann. 2016. SIAMC: A Socially Immersive Avatar Mediated Communication Platform. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. VRST '16. ACM, Munich, Germany, 357–358. DOI: 10.1145/2993369.2996302.

[279] Daniel Roth and Carolin Wienrich. 2018. Effects of Media Immersiveness on the Perception of Virtual Characters. In *Ralf Dörner, Paul Grimm, Christian Geiger (Eds.), Proceedings of the 15th Workshop on Virtual and Augmented Reality of the GI Special Interest Group VR/AR. 2018.* Düsseldorf, Germany, 133–144.

[280] Paul Rozin and Edward B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and social psychology review*, 5, 4, 296–320. DOI: 10.1207/S15327957PSPR0504_2.

[281] Kerstin Ruhland, Sean Andrist, Jeremy Badler, Christopher Peters, Norman Badler, Michael Gleicher, Bilge Mutlu, and Rachel Mcdonnell. 2014. Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Eurographics State-of-the-Art Report*, 69–91.

[282] Stuart J. Russell and Peter Norvig. 2016. *Artificial Intelligence: A Modern Approach*. Malaysia; Pearson Education Limited,

[283] Maria V. Sanchez-Vives, Bernhard Spanlang, Antonio Frisoli, Massimo Bergamasco, and Mel Slater. 2010. Virtual hand illusion induced by visuomotor correlations. *PloS one*, 5, 4, e10381. DOI: 10.1371/journal.pone.0010381.

[284] Edward Sapir. 1927. The unconscious patterning of behavior in society. *Selected Writings of Edward Sapir in Language, Culture and Personality, ed. David E. Mandelbaum (Berkeley: U of California P, 1949)*, 556, 11.

[285] Leonhard Schilbach, Bert Timmermans, Vasudevi Reddy, Alan Costall, Gary Bente, Tobias Schlicht, and Kai Vogeley. 2013. Toward a second-person neuroscience. *Behavioral and Brain Sciences*, 36, 04, 393–414.

[286] Frank M. Schneider, Carina Weinmann, Franziska S. Roth, Katharina Knop, and Peter Vorderer. 2016. Learning from entertaining online video clips? Enjoyment and appreciation and their differential relationships with knowledge and behavioral intentions. *Computers in Human Behavior*, 54, 475–482. DOI: 10.1016/j.chb.2015.08.028.

[287] Ralph Schroeder. 2010. *Being There Together: Social Interaction in Shared Virtual Environments*. Oxford University Press, New York. ISBN: 978-0-19-537128-4 0-19-537128-3.

[288] Ralph Schroeder. 2002. Social interaction in virtual environments: Key issues, common themes, and a framework for research. In *The Social Life of Avatars*. Springer, 1–18.

[289] Ralph Schroeder and Ann-Sofie Axelsson. 2006. *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*. Volume 34. Springer.

[290] Thomas W. Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie*, 15, 2, 69–71. DOI: 10.1026//1617-6383.15.2.69.

[291] Claude Elwood Shannon. 1949. Communication in the presence of noise. *Proceedings of the IRE*, 37, 1, 10–21. DOI: 10.1109/JRPROC.1949.232969.

[292] Sotaro Shimada, Kensuke Fukuda, and Kazuo Hiraki. 2009. Rubber hand illusion under delayed visual feedback. *PloS one*, 4, 7, e6185. DOI: 10.1371/journal.pone.0006185.

[293] Sotaro Shimada, Tatsuya Suzuki, Naohiko Yoda, and Tomoya Hayashi. 2014. Relationship between sensitivity to visuotactile temporal discrepancy and the rubber hand illusion. *Neuroscience Research*, 85, 33–38. DOI: 10.1016/j.neures.2014.04.009.

[294] Kang G. Shin and Parameswaran Ramanathan. 1994. Real-time computing: A new discipline of computer science and engineering. *Proceedings of the IEEE*, 82, 1, 6–24. DOI: 10.1109/5.259423.

[295] John Short, Ederyn Williams, and Bruce Christie. 1976. The social psychology of telecommunications.

[296] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. 2013. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56, 1, 116–124. DOI: 10.1145/2398356.2398381.

[297] Mel Slater. 2003. A note on presence terminology. *Presence connect*, 3, 3, 1–5.

[298] Mel Slater. 2014. Grand Challenges in Virtual Environments. *Frontiers in Robotics and AI*, 1, 3. DOI: 10.3389/frobt.2014.00003.

[299] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364, 1535, 3549–3557. DOI: 10.1098/rstb.2009.0138.

[300] Mel Slater, Vasilis Linakis, Martin Usoh, Rob Kooper, and G. Street. 1996. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In *ACM Virtual Reality Software and Technology (VRST)*, 163–172. DOI: 10.1145/3304181.3304216.

[301] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V. Sanchez-Vives. 2009. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience*, 3, 29. DOI: 10.3389/neuro.01.029.2009.

[302] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V. Sanchez-Vives. 2008. Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience*, 2, 6. DOI: 10.3389/neuro.09.006.2008.

[303] Mel Slater and Maria V. Sanchez-Vives. 2016. Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI*, 3. DOI: 10.3389/frobt.2016.00074.

[304] Mel Slater, Bernhard Spanlang, Maria V. Sanchez-Vives, and Olaf Blanke. 2010. First person experience of body transfer in virtual reality. *PloS one*, 5, 5, e10564. DOI: 10.1371/journal.pone.0010564.

[305] Mel Slater and Anthony Steed. 2000. A virtual presence counter. *Presence: Teleoperators & Virtual Environments*, 9, 5, 413–434.

[306] Mel Slater and Anthony Steed. 2002. Meeting people virtually: Experiments in shared virtual environments. In *The Social Life of Avatars*. Springer, 146–171.

[307] Mel Slater, Anthony Steed, and Martin Usoh. 1995. The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In *Virtual Environments' 95*. Springer, 135–148.

[308] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3, 2, 130–144.

[309] Harrison Jesse Smith and Michael Neff. 2018. Communication Behavior in Embodied Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 289. DOI: 10.1145/3173574.3173863.

[310] Erica Southgate, Shamus P. Smith, and Jill Scevak. 2017. Asking ethical questions in research using immersive virtual and augmented reality technologies with children and youth. In *2017 IEEE Virtual Reality (VR)*. 2017 IEEE Virtual Reality (VR). IEEE, Los Angeles, CA, USA, 12–18. DOI: 10.1109/VR.2017.7892226.

[311] R. Nathan Spreng, Margaret C. McKinnon*, Raymond A. Mar, and Brian Levine. 2009. The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of personality assessment*, 91, 1, 62–71. DOI: 10.1080/00223890802484381.

[312] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. 1997. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Volume 41. SAGE Publications Sage CA: Los Angeles, CA, 1138–1142. DOI: 10.1177/107118139704100292.

[313] Jan-Philipp Stauffert, Florian Niebling, and Marc Erich Latoschik. 2018. Effects of Latency Jitter on Simulator Sickness in a Search Task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 121–127. DOI: 10.1109/VR.2018.8446195.

[314] Anthony Steed, William Steptoe, Wole Oyekoya, Fabrizio Pece, Tim Weyrich, Jan Kautz, Doron Friedman, Angelika Peer, Massimiliano Solazzi, Franco Tecchia, et al. 2012. Beaming: an asymmetric telepresence system. *IEEE computer graphics and applications*, 32, 6, 10. DOI: 10.1109/MCG.2012.110.

[315] Mariëlle Stel, Jim Blascovich, Cade McCall, Jessanne Mastop, Rick B. Van Baaren, and Roos Vonk. 2010. Mimicking disliked others: Effects of a priori liking on the mimicry-liking link. *European Journal of Social Psychology*, 40, 5, 867–880.

[316] Mariëlle Stel, Rick B. Van Baaren, and Roos Vonk. 2008. Effects of mimicking: Acting prosocially by being emotionally moved. *European Journal of Social Psychology*, 38, 6, 965–976. DOI: 10.1002/ejsp.472.

[317] Mariëlle Stel and Roos Vonk. 2010. Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction. *British Journal of Psychology*, 101, 2, 311–323. DOI: 10.1348/000712609X465424.

[318] Neal Stephenson. 1993. *Snow Crash*. Bantam Books., New York.

[319] William Steptoe, Oyewole Oyekoya, Alessio Murgia, Robin Wolff, John Rae, Estefania Guimaraes, David Roberts, and Anthony Steed. 2009. Eye tracking for avatar eye gaze control during object-focused multiparty interaction in immersive collaborative virtual environments. In *Virtual Reality Conference, 2009. VR 2009. IEEE*. IEEE, 83–90. DOI: 10.1109/VR.2009.4811003.

[320] William Steptoe, Anthony Steed, and Mel Slater. 2013. Human tails: ownership and control of extended humanoid avatars. *IEEE transactions on visualization and computer graphics*, 19, 4, 583–590. DOI: 10.1109/TVCG.2013.32.

[321] Jonathan Steuer. 1992. Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42, 4, 73–93. DOI: 10.1111/j.1460-2466.1992.tb00812.x.

[322] Evan A. Suma, Belinda Lange, A. S. Rizzo, David M. Krum, and Mark Bolas. 2011. Faast: The flexible action and articulated skeleton toolkit. In *Virtual Reality Conference (VR), 2011 IEEE*. IEEE, 247–248. DOI: 10.1109/VR.2011.5759491.

[323] Kazuaki Tanaka, Hideyuki Nakanishi, and Hiroshi Ishiguro. 2015. Physical embodiment can produce robot operator's pseudo presence. *Frontiers in ICT*, 2, 8. DOI: 10.3389/fict.2015.00008.

[324] The Editors of Encyclopaedia Britannica. 2019. Avatar. In *Encyclopaedia Britannica*. Encyclopædia Britannica, inc. Retrieved 05/14/2019 from https://www.britannica.com/topic/avatar-Hinduism.

[325] Edmund R. Thompson. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of cross-cultural psychology*, 38, 2, 227–242. DOI: 10.1177/0022022106297301.

[326] Evan Thompson and Francisco J. Varela. 2001. Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, 5, 10, 418–425. DOI: 10.1016/S1364-6613(00)01750-2.

[327] Linda Tickle-Degnen and Robert Rosenthal. 1990. The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1, 4, 285–293. DOI: 10.1207/s15327965pli0104_1.

[328] Naftali Tishby and Daniel Polani. 2011. Information theory of decisions and actions. In *Perception-Action Cycle*. Springer, 601–636.

[329] Michael Tomasello. 2009. *Why We Cooperate*. MIT press.

[330] Jolanda Tromp, Adrian Bullock, Anthony Steed, Amela Sadagic, Mel Slater, and Emmanuel Frécon. 1998. Small group behaviour experiments in the coven project. *IEEE Computer Graphics and Applications*, 18, 6, 53–63. DOI: 10.1109/38.734980.

[332] Manos Tsakiris. 2010. My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia*, 48, 3, 703–712. DOI: 10.1016/j.neuropsychologia.2009.09.034.

[333] Manos Tsakiris and Patrick Haggard. 2005. The rubber hand illusion revisited: visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1, 80. DOI: 10.1037/0096-1523.31.1.80.

[334] Manos Tsakiris, Maike D. Hesse, Christian Boy, Patrick Haggard, and Gereon R. Fink. 2006. Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cerebral cortex*, 17, 10, 2235–2244. DOI: 10.1093/cercor/bhl131.

[335] Manos Tsakiris, Gita Prabhu, and Patrick Haggard. 2006. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and cognition*, 15, 2, 423–432. DOI: 10.1016/j.concog.2005.09.004.

[336] Manos Tsakiris, Simone Schütz-Bosbach, and Shaun Gallagher. 2007. On agency and body-ownership: Phenomenological and neurocognitive reflections. *Consciousness and cognition*, 16, 3, 645–660. DOI: 10.1016/j.concog.2007.05.012.

[337] Rick Van Baaren, Loes Janssen, Tanya L. Chartrand, and Ap Dijksterhuis. 2009. Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1528, 2381–2389. DOI: 10.1098/rstb.2009.0057.

[338] Francisco J. Varela, Eleanor Rosch, and Evan Thompson. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge.

[339] Vinoba Vinayagamoorthy, Maia Garau, Anthony Steed, and Mel Slater. 2004. An eye gaze model for dyadic interaction in an immersive virtual environment: Practice and experience. In *Computer Graphics Forum*. Volume 23. Wiley Online Library, 1–11. DOI: 10.1111/j.1467-8659.2004.00001.x.

[340] A. Vinciarelli, H. Salamin, and M. Pantic. 2009. Social signal processing: Understanding social interactions through nonverbal behavior analysis. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference On*. IEEE, 42–49. DOI: 10.1109/CVPRW.2009.5204290.

[341] Kai Vogeley. 2017. Two social brains: neural mechanisms of intersubjectivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, 1727, 20160245. DOI: 10.1098/rstb.2016.0245.

[342] Kai Vogeley and Gary Bente. 2010. ‘‘Artificial humans’’: Psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Networks*, 23, 8, 1077–1090.

[343] Thomas von der Grün, Norbert Franke, Daniel Wolf, Nicolas Witt, and Andreas Eidloth. 2011. A real-time tracking system for football match and training analysis. In *Microelectronic Systems*. Springer, 199–212.

[345] Peter Vorderer, Dorothée Hefner, Leonard Reinecke, and Christoph Klimmt. 2017. *Permanently Online, Permanently Connected: Living and Communicating in a POPC World*. Routledge.

[346] Peter Vorderer, Werner Wirth, Feliz Ribeiro Gouveia, Frank Biocca, Timo Saari, Lutz Jäncke, Saskia Böcking, Holger Schramm, Andre Gysbers, Tilo Hartmann, et al. 2004. MEC Spatial Presence Questionnaire. *Report to the European Community, Project Presence: MEC (IST-2001-37661). Online. Available from http://www.ijk.hmt-hannover.de/presence.*

[347] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM International Conference on Multimedia*. ACM, 831–834. DOI: 10.1145/2502081.2502223.

[348] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24, 4, 1643–1652. DOI: 10.1109/TVCG.2018.2794629.

[349] Thomas Waltemate, Felix Hülsmann, Thies Pfeiffer, Stefan Kopp, and Mario Botsch. 2015. Realizing a low-latency virtual reality environment for motor learning. In ACM Press, 139–147. DOI: 10.1145/2821592.2821607.

[350] Thomas Waltemate, Irene Senna, Felix Hülsmann, Marieke Rohde, Stefan Kopp, Marc Ernst, and Mario Botsch. 2016. The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 27–35. DOI: 10.1145/2993369.2993381.

[351] J. B. Walther. 1992. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research*, 19, 1, (February 1, 1992), 52–90. DOI: 10.1177/009365092019001003.

[352] 2015. *Interpersonal and Hyperpersonal Dimensions of Computer-Mediated Communication*. S. Shyam Sundar. *Handbook of the Psychology of Communication Technology*. John Wiley & Sons, Inc.

[353] Colin Ware, Kevin Arthur, and Kellogg S. Booth. 1993. Fish tank virtual reality. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*. ACM, 37–42. DOI: 10.1145/169059.169066.

[354] David Watson, Lee A. Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, 54, 6, 1063. DOI: 10.1037/0022-3514.54.6.1063.

[355] P. Watzlawick, J. H. Beavin, and D. D. Jackson. 1967. *Some Tentative Axioms of Communication. I Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies, and Paradoxes (s. 48-71)*. New York, NY: WW Norton & Company. Ina von Schantz Lundgren & Mats Lundgren.

[356] Paul Watzlawick, Janet Helmick Beavin, and Don D. Jackson. 1967. *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes*. WW Norton & Company. ISBN: 0-393-01009-0.

[357] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. 2013. Method for facial animation. (2013). Patent.

[358] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. 2011. Realtime Performance-based Facial Animation. In *ACM SIGGRAPH 2011 Papers* (SIGGRAPH '11). ACM, New York, NY, USA, 77:1–77:10. DOI: 10.1145/1964921.1964972.

[359] Daniel A. Weiskopf. 2019. Concepts, Theory-Theory of — Internet Encyclopedia of Philosophy. Retrieved 03/06/2019 from https://www.iep.utm.edu/th-th-co/.

[360] Christopher D. Wickens and C. Melody Carswell. 2012. Information Processing. In *Handbook of Human Factors and Ergonomics*. (4th ed.), 117–151.

[361] Marcus Wilms, Leonhard Schilbach, Ulrich Pfeiffer, Gary Bente, Gereon R. Fink, and Kai Vogeley. 2010. It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social cognitive and affective neuroscience*, 5, 1, 98–107. DOI: 10.1093/scan/nsq024.

[362] Lochlainn Wilson, Bakui Chou, and Keiichi Seko. 2017. Head mounted display. (April 2017). Patent.

[363] Bob G. Witmer, Christian J. Jerome, and Michael J. Singer. 2005. The factor structure of the presence questionnaire. *Presence: Teleoperators & Virtual Environments*, 14, 3, 298–312.

[364] Bob G. Witmer and Michael J. Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7, 3, 225–240. DOI: 10.1162/105474698565686.

[365] Andrea Stevenson Won, Jeremy N. Bailenson, and Jaron Lanier. 2015. Homuncular flexibility: the human ability to inhabit nonhuman avatars. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–16. DOI: 10.1002/9781118900772.etrds0165.

[366] Andrea Stevenson Won, Jeremy N. Bailenson, Jimmy Lee, and Jaron Lanier. 2015. Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20, 3, 241–259. DOI: 10.1111/jcc4.12107.

[367] Nick Yee. 2007. *The Proteus Effect: Behavioral Modification via Transformations of Digital Self-Representtation*.

[368] Nick Yee and Jeremy N. Bailenson. 2006. Walk a mile in digital shoes: The impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments. *Proceedings of PRESENCE*, 24–26.

[369] Nick Yee, Jeremy N. Bailenson, and Nicolas Ducheneaut. 2009. The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*. DOI: 10.1177/0093650208330254.

[370] Sangseok You and S. Shyam Sundar. 2013. I Feel For My Avatar: Embodied Perception in VEs. DOI: 10.1145/2470654.2466428.

[371] Jihun Yu and Jungwoon Park. 2016. Real-time facial tracking in virtual reality. In *SIGGRAPH ASIA 2016 VR Showcase*. ACM, 4. DOI: 10.1145/2996376.2996390.

[372] Vladimir M. Zatsiorsky and Boris I. Prilutsky. 2018. *Biomechanics of Skeletal Muscles*. Human Kinetics.

[374] Peter Ziegler, Daniel Roth, Andreas Knots, Michael Kreuzer, and Sebastian von Mammen. 2018. Simulator Sick but still Immersed: A Comparison of Head-Object Collision Handling and their Impact on Fun, Immersion, and Simulator Sickness. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 743–744. DOI: 10.1109/VR.2018.8446221.

[375] Manfred Zimmermann. 1985. Somato-viscerale Sensibilität: Die Verarbeitung im Zentralnervensystem. In *Physiologie Des Menschen*. Volume 22. Springer, 209–228.

[376] Manfred Zimmermann. 1989. The nervous system in the context of information theory. In *Human Physiology*. Volume 23rd edition. Springer, 166–173.

List of Figures

List of Tables