Beyond Replication: Augmenting Social Behaviors in Multi-User Virtual Realities

Daniel Roth[*]
HCI Group
University of Würzburg

Constantin Kleinbeck[†]
HCI Group
University of Würzburg

Tobias Feigl[‡]
Programming Systems Group
Friedrich-Alexander University Erlangen-Nürnberg (FAU)

Christopher Mutschler [§]
Machine Learning and Information Fusion Group
Fraunhofer IIS

Marc Erich Latoschik [¶]
HCI Group
University of Würzburg

ABSTRACT

This paper presents a novel approach for the augmentation of social behaviors in virtual reality (VR). We designed three visual transformations for behavioral phenomena crucial to everyday social interactions: eye contact, joint attention, and grouping. To evaluate the approach, we let users interact socially in a virtual museum using a large-scale multi-user tracking environment. Using a between-subject design ($N = 125$) we formed groups of five participants. Participants were represented as simplified avatars and experienced the virtual museum simultaneously, either with or without the augmentations. Our results indicate that our approach can significantly increase social presence in multi-user environments and that the augmented experience appears more thought-provoking. Furthermore, the augmentations seem also to affect the actual behavior of participants with regard to more eye contact and more focus on avatars/objects in the scene. We interpret these findings as first indicators for the potential of social augmentations to impact social perception and behavior in VR.

Index Terms: Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality

1 INTRODUCTION

Human communication is characterized by a multitude of social behaviors. Participants adapt and coordinate feelings, intentions, and actions with others [18]. They shake hands, establish eye contact, move closer to each other, or mimic their interaction partners to create liking, rapport and affiliation [28] based on a continuous processing of signals on a conscious as well as on a subconscious level. Perception and response of social signals happens "in accordance with an elaborate and secret code that is written nowhere, known by none, and understood by all" [44].

Humans process sensory information, e.g., visual information from social cues and behaviors, based on higher level (top-down) processes such as expectations, pre-acquired knowledge, and the use of contextual information [23], as well as based on bottom-up processes such as stimulation, sensation, and the respective direct information processing [20]. However, computer-mediated communication (CMC) systems often lack capabilities to accurately track and reproduce the important details of social cues. Technical systems will always be subject to potential inaccuracies, e.g., caused by noise [47]. 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" [48].

It is an open question how these shortcomings affect communication in social VR and if potential countermeasures can also be provided by the same technology. For instance, humans can cope with the lack of social cues available in CMC by shifting their attention to, or decode/encode social information into other channels, such as using smilies in text-based communication to display mood or humor [51, 52]. In turn, this indicates that humans have the capabilities to encode/decode social information into/from alternative communication channels and cue presentations, which motivates the general idea of the present paper. To this regard, VR provides communication possibilities which extensively exceed a mere replication of existing channels from the physical world. In VR, representations can in general be decoupled from behavior [6] 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 paper contributes by exploring these possibilities with a novel approach to augment 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). We found several notable effects. First, behavioral augmentations could significantly increase social presence and thought-provocation. Second, we found significant increases in approximated measures of mutual focus (eye contact), focus on active objects in the scene, and a marginally significant increase in the attention towards other participants' avatars. Our approach differs from previous work as we designed augmentations for interactional behavior phenomena. Our results highlight the potential of VR to enhance CMC scenarios beyond a mere replication of social cues from the physical world.

In Sect. 2.1 we review related work and derive our hypotheses. In Sect. 3 we describe our approach and design decisions. Sect. 4 describes the the evaluation and Sect. 5 presents its results. Finally, we critically discuss these results in Sect. 6 and conclude.

2 RELATED WORK

2.1 Virtual Social Interactions

Two visual entities are potentially involved in virtual social interactions: avatars (virtual characters driven by human behavior) and embodied agents (virtual characters driven by algorithms) [5]. Social influence can occur "whether the 'others' present are computer agents or human avatars." [46]. In avatar-mediated communication,

[*]e-mail: daniel.roth@uni-wuerzburg.de
[†]e-mail: constantin.kleinbeck@stud-mail.uni-wuerzburg.de
[‡]Also with Fraunhofer IIS, e-mail: tobias.feigl@fau.de
[§]Also with FAU, e-mail: christopher.mutschler@iis.fraunhofer.de
[¶]e-mail: marc.latoschik@uni-wuerzburg.de

avatars act as the users' virtual representation and provide user embodiment. 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) [...]" [11].

Nowak and Biocca [34] found no significant difference in presence ratings when interacting with agents compared to interacting with avatars. Both representations led to an increased perception of telepresence compared to no representation. Counter intuitively, a low anthropomorphic (simplified) avatar evoked more social presence (i.e. "a mediums ability to connect people" [34] or "the moment-by-moment awareness of the co-presence of another sentient being accompanied by a sense of engagement with the other" [14]) and copresence compared to a high anthropomorphic avatar or no avatar. Bente et al. [13] did not find significant differences between low and high fidelity avatars. Latoschik et al. [29] compared realism of appearance, that users tend to have a higher acceptance of realistic avatars as their own body, as well as a indications that body ownership might be affected by the interactant's appearance. They did not find changes in social presence perception. Recently, Daher et al. [16] found that priming observers to believe that an agent was intelligent increased social presence.

While in [34] no behaviors were replicated, both avatars used in [13] replicated the users gaze and gesture and [29] replicated body motion. The results point at behavior replication to be a dominant factor compared to the form of visual representation. To this regard, Bradler et al. [4] provide an overview on creating, simulating and animating humans, i.e. human figure models. They present a set of desiderata, including that "A human model should move or respond like a human" and "A human model should have a human-like appearance". Whereas [34] and [13] compared representations with either static or replicated social behavior and [4] describe desirable features to replicate physical reality, our approach does not aim to include static expressions or replicate behavior/appearance cues in a direct sense. Instead our goal is to utilize the flexibility of VR to enable message exchange and interpretation by transforming and visually amplifying social phenomena, which was not investigated so far.

2.2 Transforming Social Interaction

Bailenson et al. [10] transformed user behavior to appearance and form changes of a simplified "emotibox" avatar by using facial feature tracking (e.g. width and height of the "emotibox" were controlled by the human mouth movements). The copresence ratings and emotion transmission were lower in the avatar condition compared to video or voice only. However, there was more verbal and nonverbal disclosure of information in the "emotibox" condition [10].

Furthermore, Bailenson et al. [6, 7] introduced the concept of transformed social interaction and argue that nonverbal behavior can be transformed strategically (i.e. rendered nonverdically) as physical behavior and visual representation can be decoupled. For instance, the transformation of gaze or head movements of avatars, the focus of attention of a speaker can be directed to multiple listeners. In a study with three participants gaze augmentation resulted in significantly higher agreement compared to a non-augmented or reduced-gaze. However, participants in the "augmented" condition perceived less social presence. Bailenson et al. also investigated the effect of induced nonverbal mimicry, which was found to faciliate and express social affiliation and likeability. Agents mimicking head movement were perceived more positive than nonmimicking agents [9]. Recently, Oh et al. amplified the smile of an avatar and found [35] higher positive affect and social presence.

Compared to our approach, these 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 using amplifying or substituting transformations increases social presence.*

2.3 Artificial Behavior and Hybrid Models

In order to generate nonverbal behavior for expressive and conversational agents, Cassell et al. [15] presented a system for the rule-based generation of facial expressions, lip motions, eye gaze, head motion and arm gestures based on conversations created by a dialog planner. Vogeley and Bente state 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" [49] including phenomena such as interpersonal synchrony. One approach by Gratch et al. is to create virtual rapport, e.g. by adapting behavioral reactions [21, 22]. These works however clearly distinguish between agents and avatars. In proposing an alternative model, Gerhard et al. define the human as a temporary controller for the virtual representation whereas agents take over control in human absence to foster continuous presence [19]. Specifically focusing on behaviors, Roth et al. [41] introduced the concept of hybrid avatar-agent technologies that actively mediate nonverbal communication using an underlying social artificial intelligence. In turn, these modifications might affect the behaviors of the interlocutors and their impression of the interaction.

While [15, 21, 22] investigated agent behavior, we aim at transforming avatar behavior on a phenomenological level. 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.

3 Design and Implementation

We created a design space for potential augmentations restricted to translational (x,z of the transverse plane) and rotational (x,y,z) data input. Our goal was to find constraints relating the input, the intermediate behavioral phenomena, as well as visual abstractions for the transformation, amplification, and substitution of the 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 (Fig. 1). The three augmentations were chosen to cover multiple dimensions (bidirectional, environment interactivity, multi-person). All augmentations are based on visual feedback and can be described as substitutionary, amplifying and transformational in their characteristics. To prevent third variable bias, we chose a reduced avatar model. In the following we describe the decisions on avatar appearances as well as each transformation in detail.

3.1 Avatar Appearance

According to Watzlawick, one "cannot not communicate" [53], i.e. every present behavior or social cue will have a meaning for interlocutors and will be interpreted. Our study aimed at investigating 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 grey (respectively colored) pillars. This avatar representation specifically avoids any additional artificial social information that may be derived from a more humanoid or realistic representation and influence the participants' perception based on direct social information processing or contextual norms [20, 23, 42]. E.g. Bailenson et al. [7] acknowledge that having avatars with eyes but no replication of eye movement is problematic. We therefore do not render avatar eyes but rather use an approximate (the head direction) to derive visual transformations. Participants immersed in the simulation could derive the forward direction of other participants by their locomotive behavior and noise

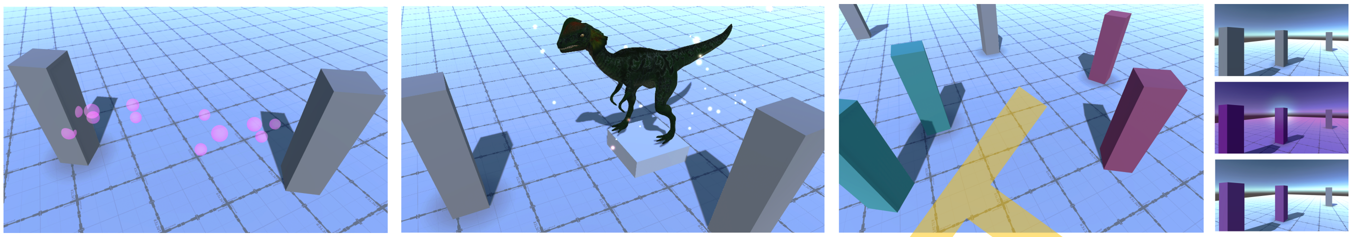

Figure 1: Isolated augmentation effects. Left: The eye contact visual transformation in the form of floating "bubbles". 2nd: The joint attention augmentation using a particle system highlight. 3rd: The grouping augmentation using group colors and visual effect when joining a group.

location as well as voice location during verbal exchange which we perceived sufficient during pretesting. The pillar size was uniform (50x50x180cm), as height can be a strong cue for the perception of dominance [31]. The virtual camera was adapted to fit the height of each participant.

3.2 Eye Contact

Eye contact is important for social interaction [3]. It acts as a form of contact establishment and 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 [30]. Similar to our study, Bailenson et. al [7] used the head direction to describe and render gaze in a virtual environment, as head and eye direction highly correlated.

Early prototype tests identified form, color and frequency as most important aspects. For example, a spike-like particle system was perceived rather negative. We therefore chose a shared-elements visualization that was perceived as soft and related to the idea of exchanging thoughts or gaze by pretesters, see Fig. 1. 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 type of 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 [37]. To identify approximated eye contact between two participants (i.e. two participants focus on each other), we build a ray cast map that gathers all objects in focus (active exhibits as well as other avatars within a predefined distance of 4m) of participants. In a second step, we check all value pairs for whether or not mutual gaze is present and evoke the augmentation effect if eye contact pairs are found.

3.3 Joint Attention

Joint attention is a phenomenon of shared attention or focus toward an object developed in infant stages [32]. Initiating joint attention shows the desire to share a pleasureable experience with others. It therefore inherits "processing of information about the attention of self and others" [33, 269] and signals common interest or a common point of reference [33]. As joint attention includes an interactive process with the environment we designed the transformation as a highlighting particle system on an object that appeared if two or more participants were within the 4 m social distance and focused (head direction) at the same object (Fig. 1). Appearing particles had a small movement radius concentrated in the up axis until disappearing. Other prototypes such as an active color change of the object were perceived as rather irritating by pretesters and furthermore would have changed the appearance character of the exhibit. 16 exhibits (active objects) in the virtual museum were capable of evoking the transformation. We build a ray cast map to collect object hits throughout the simulation. Once duplicates are found, we evaluate whether or not the watchers are within a social distance in order for the highlight to appear.

3.4 Grouping

Grouping is a spatial behavior derived from proxemics [2] that encodes group affiliation, intimacy, or power [1], and is associated with interpersonal attraction. Humans form more positive attitudes towards ingroup members [25] and regarding a distinct communicative aspect, spatial movements often indicate the beginning and ending of interactions [2]. Hall [24] differentiates 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 show similar proxemic behavior to those in the physical world [8]. We considered the social space (<4 m) as the dimensions for groups in our approach.

Our grouping transformation was chosen with regard to group identification (i.e. appearance) and uses color changes to identify group formation and group members, aiming at amplifying the grouping effect. Participants within social distance are grouped together. We chose 4 meters, as the pillar avatars are slightly more extensive in dimensionality than humans, to avoid bumping into each other. To further promote and signal the initiation of a group to the participant, we implemented a fade in/fade out camera effect (visual flare in group color, 2 seconds duration, see Fig. 1 right). Our grouping algorithm was adapted from the k-means algorithm [26] and uses the distances between all participants. Each has an internal predefined HSV color. The neutral gray value was 0,0,0.8 in HSV coordinates. The respective color values are H,0.5,0.8. Only the hue value changes, in order to avoid changes in brightness. With each group formation, the group color is determined from the group member constellation. We do not respect psychological effects on the perception of color in the grouping metaphor as the possible constellations are manifold so a systematic impact was not expected. To avoid disturbing and fast color changes, group constellations need to hold stable for two seconds until the visual coloring was applied. Fig. 2 shows the augmentations as they would have been experienced by the participants in the simulation.

3.5 Virtual Museum

Our main goal in the scenario design was to find a shared social space that inherits both 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 museum for nature in [city, country] for one day, we found that all interactions that we included in our augmentation set (grouping, joint attention, directed gaze) were also performed by people visiting a museum, which is why we found a museum scenario to be a valuable usecase. We choose the topic of dinosaurs as we found that the subject of primeval times is taught in early education and paleontologic exibitions are a topic all participants could relate to in a similar way. 6 exhibits included short audio information for the specific dinosaur that was extracted from Wikipedia. Once audio information was evoked by one of the participants, each user in 4 m distance could hear the

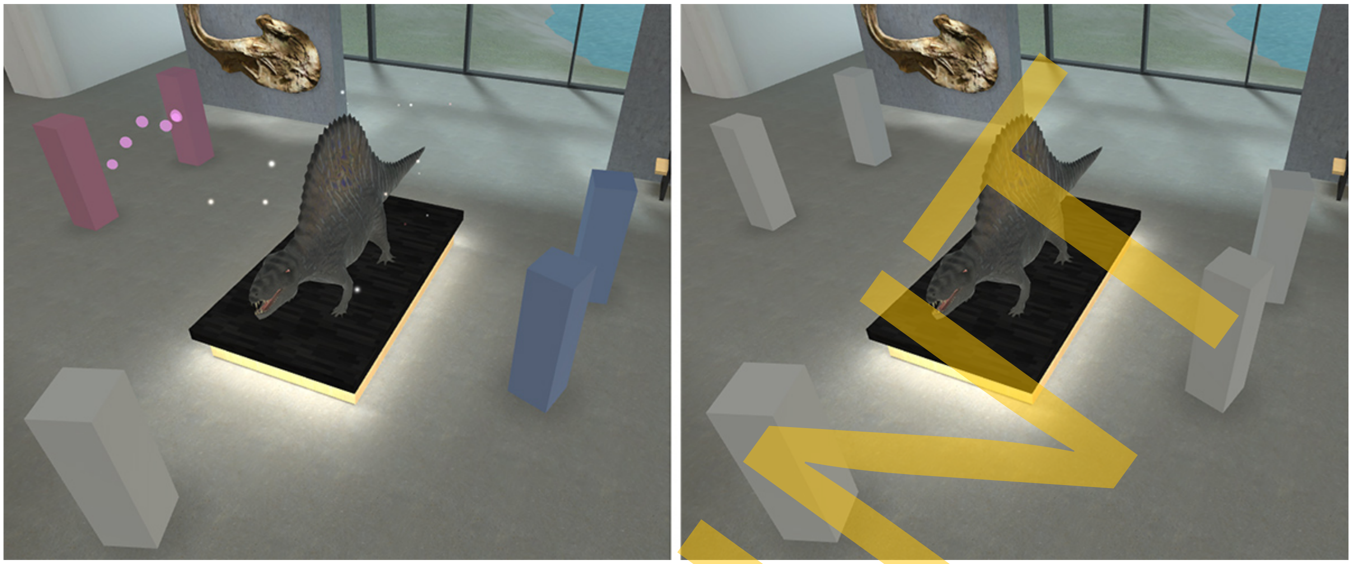

Figure 2: Illustration of the scenario as experienced by the participants in the simulation. Left: Condition with transformations for eye contact (floating bubbles), joint attention (particle highlights on object) and grouping (avatar colors). Right: Condition without transformations.

audio via a one-eared headset, similar to an audio guide in a physical museum.

Fig. 3 depicts the virtual museum. Its virtual dimensions were identical to the measures of the tracking area (20x30 m). Among 16 active objects, six large dinosaur exhibits included prerecorded audio information. The audio information could be triggered by selection via gaze (head direction) when looking at a speaker icon label for 2 seconds, while a loading spinner element appeared. Each audio information is about 1 minute long. An activated audio was shared by all participants close to the activated source and could be heard through the one-ear headphone in moderate volume simultaneously within close range of the exhibit.

4 Methods

4.1 Design

The study was conducted in a between-subjects design comparing the conditions "augmented" (active augmentations for grouping, direct gaze, joint attention, see Fig. 2 left) and "non-augmented" (see Fig. 2 right). User groups experienced the museum either with or without active behavioral augmentations. Participants were blind to the actual goal of the experiment.

4.2 Task

Participants in groups of five participants were told to explore the museum and to learn about the exhibits in a natural way, as they would do in a physical museum. We told the participants that they could freely move and were free to interact with each other. We advised participants not to pass through virtual walls or objects and to not move further than the dimensions of the virtual museum as partly physical walls and objects beyond the virtual dimensions limited the motion space. Participants started from defined start positions depicted in Fig. 3 (bottom).

4.3 Procedure

Participants gave their informed consent and were assigned a random number 1-5 in order to ensure a correct relationship measures. We asked participants to fill out the pre-study questionnaire (demography, personality, media habits, interpersonal relationship, simulator sickness). For the main part of the experiment, we then equipped each subject with an HMD, tracker and audio earphone and instructed them how to recalibrate their HMD in case they experienced perceivable drift, which was done by pushing a button and taking three to five steps in a straightforward line. The participants were informed about their upcoming task using an oral instruction based on a script. We guided the participants to the start positions (Fig. 4, top) and gave them about 30 seconds of acclimatization time in a slighly detached anteroom of the museum (see Fig. 3). We started data logging and the experiment with an oral "go". The participants were exposed to the virtual museum for 15 minutes (see Fig. 4) until the we stopped the exposition. Participants could raise their hand during the exposition if they experienced trouble (e.g. drift). In this case one of the present experimenters assisted them with a recalibration which 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 the participants with sweets or credit points. The overall experiment approximately took about 1-1.25 hours time. The experiment was conducted at the Fraunhofer IIS in Nrnberg. The study was approved by the ethical commission of the institute for Human-Computer Media of the University of Würzburg.

4.4 Apparatus

The simulation was implemented in Unity 3D using a server-client network architecture. Video information was displayed by Samsung S7 and S8 Smartphones in combination with the Gear VR HMDs. Audio information was displayed by a Beyerdynamic DT-1 one-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 (Fig. 4) [50]. To limit the load on the wireless transmission bandwidth and assure a constant stream, the positioning system was used with 20 Hz tracking refresh rate. To smoothen 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 on a straight line, deriving a trajectory vector which was then used to correct the orientation

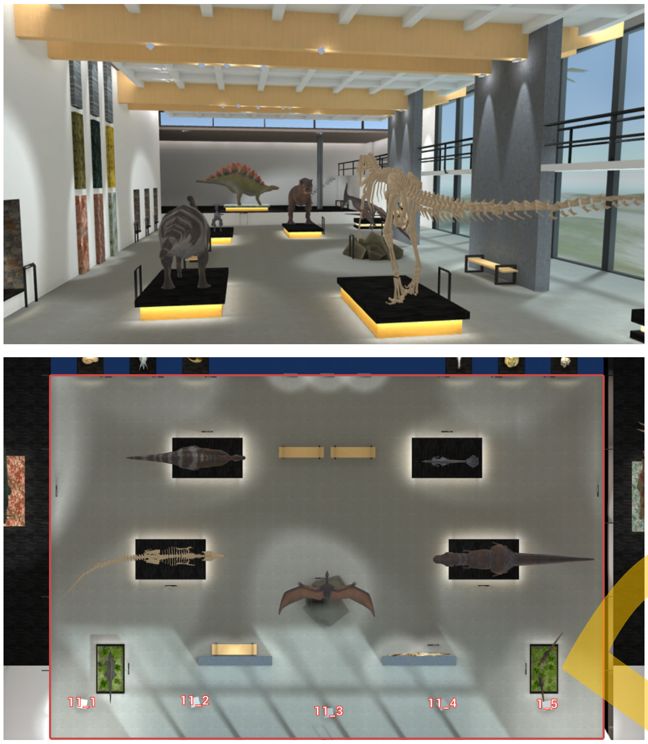

Figure 3: Top: Side view of the final museum environment used for the study. Six main exhibits included audio information. Bottom: Birdseye view on the experimental simulation in the start position. The active walking area is marked in red.

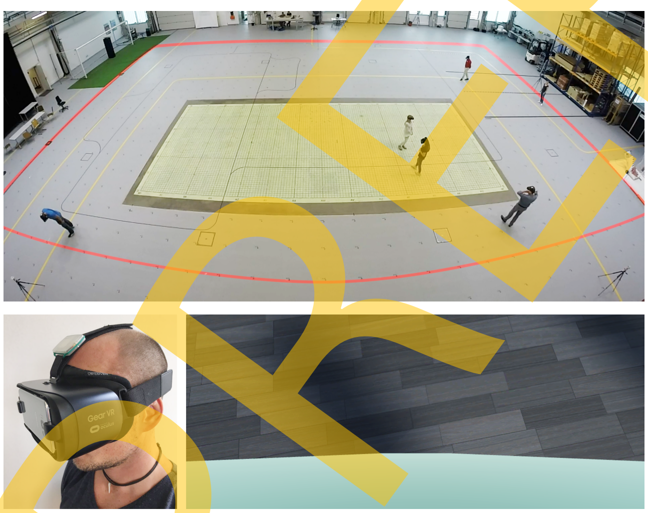

Figure 4: Top: Overview of the tracking space and apparatus. Five participants are immersed simultaneously. Bottom left: User wearing the HMD. A RF transmitter is attached to the HMD for position tracking (transverse plane). Bottom right: User embodiment (first person look at the own avatar).

offset. As especially the S8 mobile phones tended to drift over time, participants had to occasionally recalibrate their simulation (approximately between zero and two times during the exposure).

4.5 Measures

4.5.1 Subjective Measures

We measured *social presence*, *self-reported copresence*, *perceived other's copresence*, and *telepresence* [14] to test the impact of the augmentations. We adapted the scales from [34] and reformulated the scales for a multi-user scenario (i.e. "my interaction partner" = "my interaction partners"). The items were assessed using a 7-point scale (1=strongly disagree, 7=strongly agree; respectively 1=not at all, 7=very much). To measure impacts on the general impression we measured *enjoyment, lasting impression, thought-provocation, suspense, and artistic Value* with the scale from [36] (1=strongly disagree, 7=strongly agree). Furthermore, participants could add qualitative comments concerning the experience. In addition we measured rapport and group accord, which are not in focus of the present paper.

4.5.2 Behavioral Measures

To investigate *H2*, we developed objective behavioral measures. We assessed the time *eye contact* occurred, the time participants were *looking at other avatars*, the time participants were *looking at any objects* (dinosaurs, exhibits, other participants), the time *joint attention* occured, the average *interpersonal distance* to all group members and the time participants were *grouped*. To gain better insights on the behavior over time we sliced the full exposure data (15 minutes) into 6 slices of 2.5 minutes. The data were logged using a data logger running on the server with 20 *Hz*.

4.5.3 Control Measures

We introduced control measures to avoid any third variable bias. We measured the *subjective closeness index* (SCI, $r = .916$) [17]. All participants evaluated their relationship to all other group members on a scale from 1 = "not close at all" to 7 = "very close" with two questions. The SCIs of each subject for the four group members were then averaged. To control for differences in personality, we measured the *big five inventory* short form (all r's $> .203$) from [39]. We assessed the *simulator sickness questionnaire* (SSQ) [27] 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 a lower amount of knowledge acquisition as a hint to cognitive distraction. That is, an augmentation condition runs the risk of demanding extra cognitive capacity that in consequence 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 [38,45] adapted 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 information and stated five multiple choice questions.

4.6 Participants

We tested 37 groups. We removed 9 groups from the analysis because participants did not show up or due to technical issues. 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 125 (41 female) participants in 25 groups, 65 participants in the "augmented" condition, and 60 participants in the "non-augmented" condition. Conditions were randomly assigned. The mean age was 32.34 ($SD = 10.64$). 75 participants were employed, 44 students. 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.

Table 1: Results from the dependent variables assessed by questionnaires. Bold values indicate two-tailed significances at the 5% level.

		augmented			*non augmented*						*90% CI*	
		N	*M*	*SD*	*N*	*M*	*SD*	*t*	*p*	*d*		
Presence	**Social presence**	65	3.58	1.24	60	3.14	1.06	2.17	**.032**	0.38	[0.038	0.859]
	Self-reported copresence	65	4.03	0.97	59	3.97	1.05	0.31	.757	0.06	[-0.302	0.415]
	Perceived other's copresence	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]
Appreciation & Artistic Value	**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]

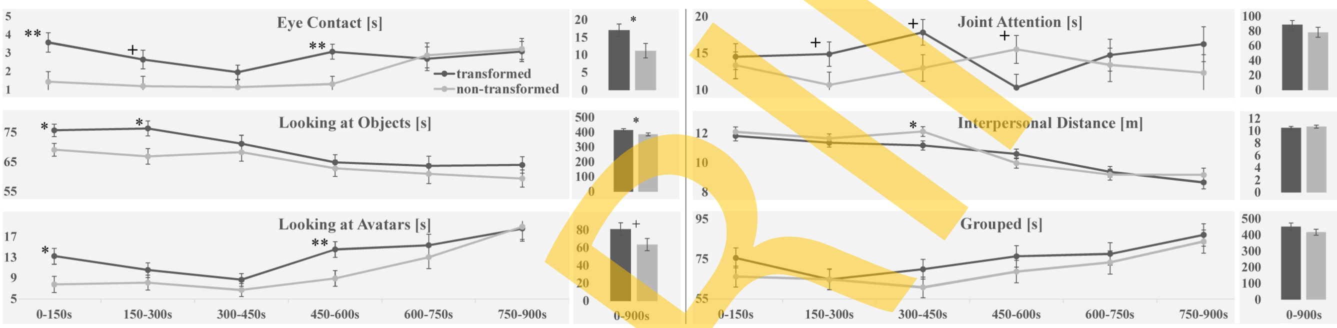

Figure 5: Results from the behavioral measures. Mean values across the 6 different time slices and overall values of 15 minutes exposure. Error bars denote standard errors. ** indicates $p < .01$, * indicates $p < .05$, + indicates $p < .1$ (two-tailed).

5 Results

5.1 Subjective Results

To investigate *H1* and the impact on other presence factors were we conducted independent samples t-tests. Results showed a significant difference in perceived social presence between the "augmented" condition ($M = 3.58$, $SD = 1.24$) and the "non-augmented" condition ($M = 3.14$, $SD = 1.06$; $t(123) = 2.165$, $p = .032$) 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 ($M = 3.47$, $SD = 1.27$; $t(121) = 2.142$, $p = .034$). Results of the subjective measures are presented in Table 1.

5.2 Behavioral Results

The six time slices of the behavioral measures are presented in Fig. 5. We calculated mixed ANOVAs with condition serving as between-subject factor and time serving as repeated-measurement factor. We found significant longer times of eye contact in the augmented condition ($M = 17.06\ s$, $SD = 14.48\ s$) compared to the non-augmented condition ($M = 11.22\ s$, $SD = 15.89\ s$; $F(1,123) = 4.61$; $\eta_p^2 = .036$; $p = .034$). Pairwise comparisons showed significant differences in slice 1 and 4 ($p's < .01$) and a marginal significant difference in slice 2 ($p = .05$). The time difference participants focused on other avatars was marginally significant ($F(1,123) = 3.13$; $\eta_p^2 = .025$; $p = .079$). 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 ($M = 63.65\ s$, $SD = 52.64\ s$). We found differences to be significant in slice 1 ($p = .013$) and slice 4 ($p = .009$). Participants in the augmented condition focused significantly longer on active objects (other avatars, exhibits; $M = 415.72\ s$, $SD = 78.10\ s$) than participants in the non-augmented condition ($M = 387.45\ s$, $SD = 77.73\ s$; $F(1,123) = 4.106$; $\eta_p^2 = .032$; $p = .045$). ANCOVA calculations with previous VR experiences as covariate did not change subjective or behavioral results significantly.

5.3 Control Measure Results

Independent samples t-tests for the SCI and Big Five factors did not reveal significant differences between the two conditions. We analyzed the SSQ using the aggregation procedure described in [27] with a mixed ANOVA. A significant main effect showed that total sickness was significantly lower in pre exposure ($M = 10.36$, $SD = 10.41$) compared to post exposure ($M = 11.93$, $SD = 11.13$; $F(1,122) = 14.47$, $\eta_p^2 = .106$, $p < .001$). The analysis of the SSQ did not show a significant difference between the two conditions. Differences in the subjective or objective (number of correct answers) knowledge measure were not significant.

6 Discussion

Supporting hypothesis *(H1): The augmentation of social behaviors using amplyfing or substituting transformations increases social presence*, we found a significantly higher perception of social presence in the "augmented" condition compared to 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", pointing out that the user actually perceived the substitute with a positive character. One indication that participants tried to interpret the augmentations is that the "augmented" condition was found as significantly more thought-provoking, while there were no differences in subjective or objective learning. We did not find significant differences in the co-presence or telepresence measures, which could indicate that these relations were not affected by the augmentations. It is to be noted that the virtual environment and experimental scenario could be considered as a mixed-reality scenario due to the fact that interactants still heard each other's physical voice and were in the same space with each other. This may have impacted effects and compensation mechanisms of the different augmentations, as grouping was amplified whereas augmentations for eye contact or joint attention might be of substitutionary in their character. Furthermore, participants in the non-augmented condition could also have rely more on prosody as compensation. This does however not 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 i) evoked more approximated eye contact, and ii) focused marginal significantly 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, e.g. 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. Our environment was relatively static which could have made users pay attention to any visual "attractions" and could have biased the social presence and behavior results. Similar ratings of enjoyment and suspense counter-argue this assumption to some extend. We did not find significant differences in the average distance of participants or the amount of 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 to outgroup members [12]. With our data we cannot state any conclusion to 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, thus the experience was not perceived more negative or positive in either condition. This could partly be explained by a ceiling effect (qualitative comments indicate a general 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 increases thought-provocation. Furthermore, we conclude that the proposed augmentations can foster the behavioral interactivity between participants with regard to eye contact and the interactivity between the participants and the environment. We controlled for potential third variable biases such as previous status of relationship or personality. 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 now state multiple limitations that this study faces and our argumentation on the impact.

6.1 Limitations

First, there are indications (user comments, sickness measure) 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 expect 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 the participants. Similar to drift and latency, there are no indicators that there were different amounts of recalibrations between the groups. Third, participants were aware of their interaction partners and saw their partners upfront and were aware that the pillars represent other users, as a blind procedure was not possible due to the extensive setup. Our control measures did not identify a potential bias. Fourth, the fact that we used multiple transformations do not allow interpretation of any finding with relation to a single transformation. For example it seems that grouping and joint attention were not as affective as eye contact. We can however not conclude on this interpretation.

7 Conclusion and Future Work

In this paper, we 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 impact an experience in VR. Our findings extend the results of previous work [40] and suggest that applied augmentations can significantly impact social presence and user behavior. These findings can inform the development of social VR applications or training and therapy of individuals suffering from social disorders. We believe that our approach is an initial step to to explore the potentials of VR as a medium to actively mediate 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. For example, Roth et. al provide a platform to individually investigate behavioral channels [43]. 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.

Acknowledgments

We thank Samantha Straka, Lisa Hartberger, Florian Kern, Diana Rieger, Eric Novotny, Jean-Luc Lugrin, and Gary Bente for support and feedback. This work was supported by the federal state of Bavaria, Germany in the ZD.B special program for innovation labs.

References

[1] P. Andersen, J. Gannon, and J. Kalchik. Proxemic and haptic interaction: the closeness continuum. In *Nonverbal Communication*. De Gruyter, Berlin, Boston, 2013.

[2] M. Argyle. *Bodily Communication*. Methuen, 2nd edition ed., 1988.

[3] M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, pp. 289–304, 1965.

[4] N. I. Badler, C. B. Phillips, and B. L. Webber. *Simulating humans: computer graphics animation and control*. Oxford University Press, 1993.

[5] J. Bailenson and J. Blascovich. Avatars, 2004.

[6] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence*, 13(4):428–441, 2004.

[7] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk. Transformed social interaction, augmented gaze, and social influence in immersive virtual environments. *Human communication research*, 31(4):511, 2005.

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

[9] J. N. Bailenson and N. Yee. Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science*, 16(10):814–819, 2005.

[10] J. N. Bailenson, N. Yee, D. Merget, and R. Schroeder. The effect of behavioral realism and form realism of real-time avatar faces verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372, 2006.

[11] S. Benford, J. Bowers, L. E. Fahln, C. Greenhalgh, and D. Snowdon. User embodiment in collaborative virtual environments. In *Proc. the SIGCHI conf. Human factors in computing systems*, pp. 242–249. ACM Press/Addison-Wesley Publishing Co., 1995.

[12] G. Bente, D. Roth, T. Dratsch, and K. Kaspar. Emotions of my kin: disambiguating expressive body movement in minimal groups. *Culture and Brain*, 4(1):51–71, 2016.

[13] G. Bente, S. Rüggenberg, N. C. Krämer, and F. Eschenburg. Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations. *Human Communication Research*, 34(2):287–318, 2008.

[14] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In *4th annual Intl. Workshop Presence, Philadelphia, PA*, pp. 1–9, 2001.

[15] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. 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*, pp. 413–420. ACM, 1994.

[16] S. Daher, K. Kim, M. Lee, R. Schubert, G. Bruder, J. Bailenson, and G. Welch. Effects of Social Priming Social Presence with Intelligent Virtual Agents. In *Intl. Conf. Intelligent Virtual Agents*, pp. 87–100. Springer, 2017.

[17] S. Gächter, C. Starmer, and F. Tufano. Measuring the closeness of relationships: a comprehensive evaluation of the'Inclusion of the Other in the Self'scale. *PloS one*, 10(6):e0129478, 2015.

[18] M. Gallotti and C. D. Frith. Social cognition in the we-mode. *Trends in cognitive sciences*, 17(4):160–165, 2013.

[19] M. Gerhard, D. J. Moore, and D. J. Hobbs. Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation. In *Intl. Workshop Intelligent Virtual Agents*, pp. 137–155. Springer, 2001.

[20] J. J. Gibson. The senses considered as perceptual systems. 1966.

[21] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. J. van der Werf, and L.-P. Morency. Virtual rapport. In *Intl. Workshop Intelligent Virtual Agents*, pp. 14–27. Springer, 2006.

[22] J. Gratch, N. Wang, J. Gerten, E. Fast, and R. Duffy. Creating rapport with virtual agents. In *Intl. Workshop Intelligent Virtual Agents*, pp. 125–138. Springer, 2007.

[23] R. L. Gregory. The intelligent eye. 1970.

[24] E. T. Hall. The hidden dimension. 1966.

[25] M. A. Hogg and J. C. Turner. Interpersonal attraction, social identification and psychological group formation. *European journal of social psychology*, 15(1):51–66, 1985.

[26] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[27] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The Intl. J. Aviation Psychology*, 3(3):203–220, 1993.

[28] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *J. nonverbal behavior*, 27(3):145–162, 2003.

[29] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *23rd ACM Symposium on Virtual Reality Software and Technology (VRST)*, pp. 39:1–39:10, 2017.

[30] J. M. Loomis, J. W. Kelly, M. Pusch, J. N. Bailenson, and A. C. Beall. Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, 37(9):1443–1457, 2008.

[31] A. Mignault and A. Chaudhuri. The many faces of a neutral face: Head tilt and perception of dominance and emotion. *J. Nonverbal Behavior*, 27(2):111–132, 2003.

[32] C. Moore. *Joint attention: Its origins and role in development*. Psychology Press, 2014.

[33] P. Mundy and L. Newell. Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, 16(5):269–274, 2007.

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

[35] S. Y. Oh, J. Bailenson, N. Krämer, and B. Li. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLoS ONE*, 11(9):e0161794, 2016.

[36] M. B. Oliver and A. Bartsch. Appreciation as audience response: Exploring entertainment gratifications beyond hedonism. *Human Communication Research*, 36(1):53–81, 2010.

[37] L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3):232–240, 2004.

[38] P. S. Raju, S. C. Lonial, and W. G. Mangold. Differential effects of subjective knowledge, objective knowledge, and usage experience decision making: An exploratory investigation. *J. consumer psychology*, 4(2):153–180, 1995.

[39] B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J. research in Personality*, 41(1):203–212, 2007.

[40] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. [poster] social augmentations in multi-user virtual reality: A virtual museum experience. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 42–43, Oct 2017. doi: 10.1109/ISMAR-Adjunct.2017.28

[41] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente. Hybrid Avatar-Agent Technology A Conceptual Step Towards Mediated "Social" Virtual Reality and its Respective Challenges. *i-com*, 14(2), 2015.

[42] D. Roth, J. L. Lugrin, D. Galakhov, A. Hofmann, G. Bente, M. E. Latoschik, and A. Fuhrmann. Avatar realism and social interaction quality in virtual reality. In *2016 IEEE Virtual Reality (VR)*, pp. 277–278, March 2016. doi: 10.1109/VR.2016.7504761

[43] D. Roth, K. Waldow, M. E. Latoschik, A. Fuhrmann, and G. Bente. Socially immersive avatar-based communication. In *Virtual Reality (VR), 2017 IEEE*, pp. 259–260. IEEE, 2017.

[44] E. Sapir. 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), 1927.

[45] F. M. Schneider, C. Weinmann, F. S. Roth, K. Knop, and P. Vorderer. 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, 2016.

[46] R. Schroeder, D. Diaper, and C. Sanger, eds. *The Social Life of Avatars*. Computer Supported Cooperative Work. Springer London, London, 2002.

[47] C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37(1):10–21, 1949.

[48] M. Slater. Grand Challenges in Virtual Environments. *Frontiers in Robotics and AI*, 1(3), 2014.

[49] K. Vogeley and G. Bente. Artificial humans: Psychology and neuroscience perspectives embodiment and nonverbal communication. *Neural Networks*, 23(8):1077–1090, 2010.

[50] T. von der Grün, N. Franke, D. Wolf, N. Witt, and A. Eidloth. A real-time tracking system for football match and training analysis. In *Microelectronic systems*, pp. 199–212. Springer, 2011.

[51] J. B. Walther. Interpersonal Effects in Computer-Mediated Interaction: A Relational Perspective. *Communication Research*, 19(1):52–90, 1992.

[52] J. B. Walther, B. Van Der Heide, A. Ramirez Jr, J. K. Burgoon, and J. Pea. Interpersonal and Hyperpersonal Dimensions of Computer-Mediated Communication. In *Handbook the Psychology Communication Technology*. John Wiley & Sons, Inc, 2015.

[53] P. Watzlawick, J. H. Beavin, and D. D. Jackson. *Some tentative axioms of communication. I Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes*. New York, NY: WW Norton & Company. Ina von Schantz Lundgren & Mats Lundgren, 1967.