

Secondary Publication



de Souza Cardoso, Luis Fernando; Schwandt, Tobias; Schleising, Andy; u. a.

Behavioral and Perceptual Responses to 3D Character and Volumetric Avatars Under Conflicting Guidance in Mixed Reality

Date of secondary publication: 19.06.2026

Version of Record (Published Version), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-115657x

Primary publication

de Souza Cardoso, Luis Fernando; Schwandt, Tobias; Schleising, Andy; u. a. (2026): Behavioral and Perceptual Responses to 3D Character and Volumetric Avatars Under Conflicting Guidance in Mixed Reality, in: IMX '26: Proceedings of the 2026 ACM International Conference on Interactive Media Experiences, New York, NY: ACM, pp. 26–34, doi: 10.1145/3788851.3805007.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

Behavioral and Perceptual Responses to 3D Character and Volumetric Avatars Under Conflicting Guidance in Mixed Reality

Luis Fernando de Souza
Cardoso
University of Bamberg
Bamberg, Germany
Virtual Worlds and Digital Games
Group
Ilmenau University of Technology
Ilmenau, Germany
luis.cardoso@uni-bamberg.de

Andy Schleising
Virtual Worlds and Digital Games
Group
Ilmenau University of Technology
Ilmenau, Germany
andy.schleising@tu-ilmenau.de

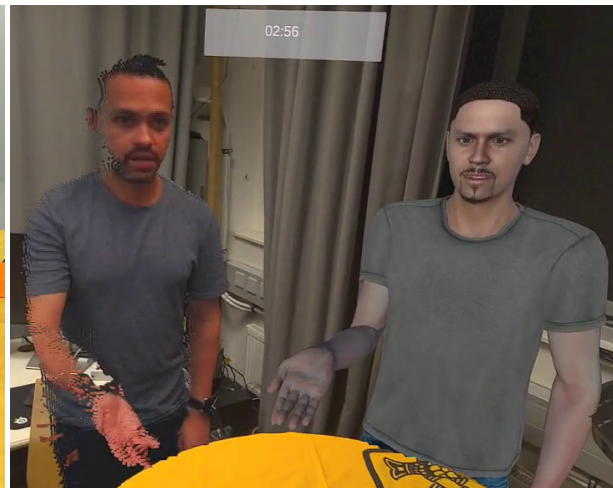
Tobias Schwandt
Virtual Worlds and Digital Games
Group
Ilmenau University of Technology
Ilmenau, Germany
tobias.schwandt@tu-ilmenau.de

Sophie Jörg
University of Bamberg
Bamberg, Germany
sophie.joerg@uni-bamberg.de

Wolfgang Broll
Virtual Worlds and Digital Games
Group
Ilmenau University of Technology
Ilmenau, Germany
wolfgang.broll@tu-ilmenau.de



(a) Participant performing the task while receiving real-time instructions from a virtual avatar.



(b) First-person view of the participant within the MR environment, observing the two avatars providing guidance.

Figure 1: The experimental setup with two avatars (a character avatar and a volumetric avatar) giving divergent instructions during the cable-connection task.

Abstract

In Mixed Reality (MR), in-the-moment guidance can support physical tasks under time and attention constraints. In contexts where such guidance is embodied, users must rapidly evaluate whether to trust and follow an advisor's recommendations. Prior work has

examined how avatar representation shapes trust and social perception, but less is known about user responses when guidance conflicts and multiple embodied agents compete for attention in real time. We study trust-relevant perception and behavior in an MR paradigm where two co-present agents represent the same advisor and deliver step-by-step instructions with predefined disagreements. In a within-subjects study, we compared a volumetric avatar (VA) and a high-fidelity 3D character avatar (CA), holding identity and expressive behavior constant, and measured trust, social presence, and behavioral adherence during contradictory steps. Rendering style produced no detectable differences in trust, social



This work is licensed under a Creative Commons Attribution 4.0 International License.
IMX '26, Athlone, Ireland

© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2448-0/26/06
<https://doi.org/10.1145/3788851.3805007>

presence, or co-presence. Instead, participants managed disagreement through mixed adherence and switching between agents, suggesting that instruction competition and time pressure strongly shape decision-making. These findings position conflicting guidance as a distinct context for studying trust in embodied agents, where compliance may reflect rapid prioritization rather than stable commitment to a representation. We discuss implications for designing and evaluating MR agents when timing and attention constraints can overshadow representation cues.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**.

Keywords

Embodied Interaction, Virtual/Augmented Reality, Avatar Realism, Quantitative Methods

ACM Reference Format:

Luis Fernando de Souza Cardoso, Andy Schleising, Tobias Schwandt, Sophie Jörg, and Wolfgang Broll. 2026. Behavioral and Perceptual Responses to 3D Character and Volumetric Avatars Under Conflicting Guidance in Mixed Reality. In *ACM International Conference on Interactive Media Experiences (IMX '26)*, June 09–11, 2026, Athlone, Ireland. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3788851.3805007>

1 Introduction

Mixed Reality (MR) systems can provide immediate, actionable guidance for tasks such as navigation, procedural assembly, and safety-critical operations. In time-critical situations, users may need to act under time constraints and rapidly assess information reliability before deciding whether to comply. A key challenge arises when guidance is inconsistent or conflicting: users may receive competing instructions from multiple sources that vary in quality, urgency, or clarity.

Such conflicting moments can be specially challenging in MR, where guidance may be conveyed through embodied agents and spatial cues. Users must prioritize information streams while managing cognitive load and maintaining task performance, often relying on perceptual and social cues rather than deliberate comparisons of instruction content. The embodied avatar is one candidate cue, as visual and behavioral design can shape perceived credibility, social presence, and trust.

Prior work shows that avatar characteristics can increase social presence and interpersonal trust, particularly through expressive nonverbal cues [1, 3]. Other work examines how appearance [20], stylization [9], and voice modulation [32] shape perceived trustworthiness. However, trust is typically studied in controlled or sequential settings that present a single avatar without conflict or time pressure. As a result, we know little about real-time trust decisions when multiple co-present avatars provide divergent guidance.

In this preliminary study, we investigate trust-relevant behavior and perception in an MR paradigm modeling real-time conflicting guidance from two co-present agents. We hold agent identity constant while varying rendering style (Volumetric Video Generated Avatar (VA) vs. High-fidelity 3D Character Avatar (CA)) and measure subjective evaluations (trust and social presence) alongside behavioral responses to contradictory instructions. Rather than

treating conflict as noise, we use it to probe how users prioritize guidance under competition and whether rendering cues contribute within these dynamics.

Our main research questions are as follows:

- **RQ1:** Under conflicting guidance in MR, what prioritization strategies emerge in behavioral responses?
- **RQ2:** How do users perceive each agent (trust/social presence), and do these perceptions align with their moment-to-moment compliance?
- **RQ3:** Does rendering style (VA vs CA) modulate prioritization or perception in this paradigm?

Our main contribution is an exploratory MR study design that captures how users resolve conflicting guidance from two co-present virtual agents. We report initial behavioral prioritization patterns and their relationship with perceived trust and social presence, and we outline design recommendations for future experiments on avatar representation under competing, time-sensitive instruction delivery.

2 Related works

2.1 Avatar representation and rendering fidelity

Avatar visual fidelity influences how users interpret virtual agents and their guidance [35]. Prior work has examined a broad spectrum of representations, from simplified and stylized embodiments to detailed character models and volumetric captures, and has linked these choices to social perception and interaction outcomes [16, 35, 37].

A common approach relies on character-based avatars built through digital modeling and animated via facial and body tracking. Such avatars can support self-identification and behavioral alignment, especially when the representation resembles a realistic human figure [9, 39]. They also allow designers to tune expressive and responsive behaviors, which contributes to perceived social presence and competence [3, 20]. Volumetric video offers a different pipeline by capturing a person's appearance and motion directly, producing photorealistic full-body representations that can strengthen authenticity and embodiment in immersive environments [10, 18, 28]. In this context, naturalistic movement and vocal affect have also been shown to shape perceived trust [32].

Still, realism is not a reliable shortcut to better experiences. Highly realistic avatars can become visually demanding, evoke uncanniness, or draw attention away from the task [13]. In Augmented Reality (AR) consultation settings, improvements in scale or detail do not necessarily translate into higher trust, satisfaction, or performance, and overly salient realism can reduce comfort [7, 8]. Users have also been reported to prefer smaller or less intrusive embodiments over full-size humanoids, which may reduce both uncanniness and perceptual load [30, 33, 36]. Overall, the literature points to a context-dependent trade-off between fidelity, salience, and usability rather than a monotonic benefit of realism.

2.2 Social and co-presence in remote assistance

Social presence concerns the sense of interacting with a social entity, including attention, understanding, and the potential for affective and behavioral interdependence [15]. Co-presence refers

more specifically to the mutual awareness of sharing a space with others [11, 23]. Both concepts are central in MR collaboration, shaping engagement, communication quality, and interpersonal trust [4, 31]. Prior work shows that embodiment choices, interaction fidelity, and environmental synchronization can meaningfully affect how collaborators perceive and respond to each other in remote MR settings [34].

A common way to assess these constructs is through self-report instruments such as the Networked Minds Social Presence Inventory (Networked Minds Social Presence Inventory (NMSPI)), which captures dimensions including co-presence, attention allocation, perceived message understanding, and affective interaction [6, 15]. Complementary behavioral measures have also been used, where gaze alignment, gesture coordination, and response timing serve as indicators of shared attention and interaction naturalness [2, 38].

The relationship between representation fidelity and presence is not straightforward. Higher-fidelity and volumetric representations can strengthen social presence and emotional connection [10, 17], but they can also increase cognitive load or introduce occlusion challenges in task-oriented scenarios [36]. Environmental consistency can improve embodiment even with limited graphical detail [27]. Moreover, a higher presence does not necessarily translate into better task performance when interaction complexity and workload constrain attention [37].

2.3 Trust in avatars and virtual agents

Trust is a foundation for effective collaboration and can be defined as the willingness to accept vulnerability under uncertainty based on positive expectations about the intentions or actions of others [21, 29]. In MR, avatars often mediate guidance and collaboration, making perceived trustworthiness relevant for how users interpret advice and decide what to follow. Prior work shows that nonverbal cues such as gaze, posture, and gesture contribute to trust formation [3], and that facial and vocal expressivity further modulate these judgments [32]. Realistic or video-based representations have also been reported to increase perceived reliability relative to simplified embodiments in some contexts [1].

At the same time, appearance-based trust judgments can be misleading. Machneva et al. [20] found strong agreement in trustworthiness ratings based on avatar appearance, yet these perceptions did not reflect the actual behavior of the represented individuals. This motivates complementing self-report measures with behavioral indicators of trust.

Behavioral trust is often operationalized through advice-seeking and compliance decisions. The Virtual Maze paradigm infers trust from whether participants request guidance from a virtual character and subsequently follow it, while controlling task uncertainty so that behavior reflects reliance on the advisor rather than objective performance [14, 19]. Recent adaptations have used this approach to test how variations in avatar appearance, voice, and social behavior influence reliance [19]. Relatedly, Canales et al. [9] showed that stylization effects can be non-linear in a high-stakes decision task, with moderately stylized avatars often preferred over both highly realistic and caricatured designs.

3 Study design

3.1 Experimental Setup

We adopted a within-subjects MR design in which two co-present agents provided step-by-step guidance for the same physical task. Avatar rendering style was manipulated by presenting one agent as a volumetric avatar (VA) and the other as a 3D character avatar (CA). The agents delivered overlapping instruction sequences with predefined points of disagreement to elicit rapid prioritization under time pressure and to capture adherence and switching behavior in real time (Figure 2).

Two of the instructions (“Nothing changed” and “Remove the brown cable”) were identical for both avatars. The mid-task statement ‘Nothing changed’ was intentionally inserted to assess whether participants continued following the same avatar after receiving a neutral, non-actionable update or whether their attention shifted to the other representation. Following the Virtual Maze paradigm, either agent’s instruction sequence could lead to successful task completion, and both sequences resulted in equivalent outcomes. Participants therefore had no intrinsic incentive to switch between agents during the initial steps. This allowed us to observe consistency in early choices before introducing a cue suggesting reconsideration. The mid-task statement “Nothing changed” was intentionally included to examine whether participants would maintain or revise their strategy under unchanged conditions.

Assignment of representation type (VA vs. CA) to Agent A or Agent B was randomized for each participant, and the left-right placement of the avatars was counterbalanced. Regarding instruction timing, the order in which the agents spoke was also randomly assigned at each step. The system ensured that both agents could not start simultaneously and prevented the same agent from speaking twice in a row before the other had interacted with the participant. A minimum temporal offset of 2 seconds between the start of each instruction was enforced, although strict separation between spoken instructions was not required. Both avatars were based on the same actor, and their voices were modified to be distinct but similar using the ElevenLabs voice conversion tool [12], creating the impression of “twin” agents while holding advisor identity constant across representations.

3.2 Avatar generation

The CA was created in Character Creator 4 [24] using the Headshot 2 plugin [25]. Facial animation was captured with an iPhone and recorded in iClone 8 [26]; a state machine controlled transitions between instruction delivery, idle behavior, and blinking.

Facial recordings were synchronized with the volumetric capture to ensure identical expressions across conditions. Body gestures were captured with the Meta Movement SDK, then post-processed to reduce jitter and smooth transitions.

For the VA, we captured raw point clouds with an Azure Kinect using the Azure Kinect SDK 1.4.1 [22]. In Unity, we rendered the volumetric avatar as a particle-based representation by instantiating point samples as mesh particles and mapping the corresponding RGB color values to achieve photorealistic appearance.

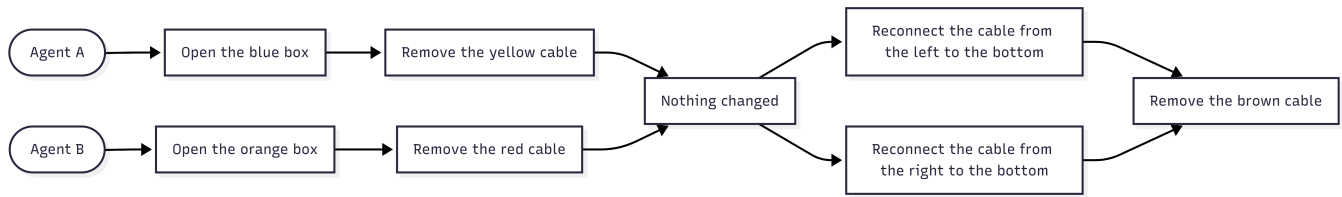
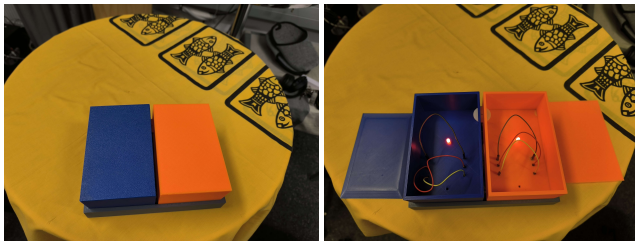


Figure 2: Sequence of instructions from the agents.

3.3 Interactive Device

Our interactive hardware comprised two identical user-facing boxes and a central control unit. The boxes differed only in color (orange and blue) to support consistent presentation across participants. As shown in Figure 3, each box contained three cables: a red and a yellow wire that could be disconnected, and a brown power cable connected to the main circuit.



(a) External view of the two user-facing boxes at the start of the experiment. (b) Internal contents of the boxes during the task, showing the colored cables.

Figure 3: Interactive device used in the experiment.

To induce time pressure, the system generated flashing light, motor vibration, and a beeping sound in a 1 s cycle. A Raspberry Pi Pico controlled the sequence with a 10 s startup delay for supervisor setup and participant orientation. Power was supplied by two batteries (one for the Pico and one for the main circuit), with the brown cable delivering power to the control system.

Each session lasted up to 3 min. During the final phase, the beeping rate increased progressively and ended with a brief high-frequency sequence before automatic shutdown.

3.4 Recording Section

To enable a direct comparison between VA and CA, both representations were based on the same human actor. Our capture pipeline was inspired by Bhattacharya et al. [5], where facial expressions for the character avatar and the volumetric video were recorded simultaneously.

Audio from the actor was recorded concurrently using a unidirectional microphone. At the start of each take, a team member clapped their hands to create a distinct audio-visual cue, enabling synchronization of the video and audio streams during post-processing.

Although facial performances were captured concurrently for both representations, the arm and upper body gestures for the character avatar were produced in a separate session. Using the original

recordings of the actor as reference, we recreated generic arm articulations for the avatar’s body animation. These gestures were deliberately designed to be task-independent, with their duration and amplitude controlled to maintain natural motion while avoiding bias toward either representation. This approach ensured that both avatar types conveyed comparable levels of expressiveness without introducing movements that could influence participant perceptions or task performance.

3.5 Procedure

The study was approved by the Institutional Review Board of the authors’ institution, and all participants provided written informed consent.

Participants first completed a demographic questionnaire, then wore a Meta Quest 3 headset and stood in front of a high table. Before entering the MR passthrough experience, the experimenter activated the interactive device. The first-person view during the task is shown in Figure 1b.

Participants were informed that two agents would provide guidance and that timing would begin once they signaled readiness. Their goal was to deactivate the device within 3 min, and no instructions were given about prioritizing one agent over the other. If the task was not completed in time, the system triggered a blinking-light effect, a short-circuit sound, and a subsequent black screen.

The task flow was controlled by the supervisor, who monitored participant actions via camera and advanced the avatars after each step. If a participant disconnected the brown cable prematurely, the supervisor ended the trial and marked the task as completed. Speaking order was randomized for each task to reduce ordering bias.

After the task (regardless of success), participants completed NASA-TLX and then evaluated each avatar using NMSPI followed by ABI. Avatar evaluation order (CA vs. VA) was randomized, and responses were collected on separate non-revisable screens to reduce direct cross-avatar comparison.

A fixed-position camera recorded the session for post-hoc analysis.

3.6 Measures

Before inferential testing, we assessed normality for each dependent variable using the Shapiro–Wilk test. We used paired-samples *t*-tests when normality held and Wilcoxon signed-rank tests otherwise. Alongside *p*-values, we report Cohen’s *d* for significant and near-significant effects.

Workload was measured with NASA–TLX (mental, physical, temporal demand, performance, effort, frustration), and the overall

score was computed as the unweighted mean across dimensions. Social presence was assessed with the NMSPI; we report co-presence, attentional allocation, perceived message understanding, and perceived behavioral interdependence. We excluded Perceived Emotional Interdependence from confirmatory analyses because the interaction was brief and instrumental, focusing on time-pressured decision-making rather than affect sharing.

Trust was assessed both behaviorally and through self-report. Behavioral trust was coded from video as adherence to each agent’s instructions at divergent steps, and subjective trust was measured with the Ability, Benevolence, and Integrity model of trust (ABI) subscales (ability, benevolence, integrity, and trust).

4 Results

Twenty-one participants (12 women, 9 men; 17–57 years, $M_{\text{age}} = 28.8$, $SD_{\text{age}} = 8.6$) participated in the study. Most (71.4%) reported prior MR experience; 42.9% used it occasionally (once a month or less).

4.1 Task performance and workload

Eighteen participants (85.7%) completed the task within 180 s and three timed out. Including all participants, mean completion time was 88.1 s ($SD = 42.1$; $Med = 71.0$ s).

NASA-TLX indicated moderate to high workload ($M = 58.2$, $SD = 16.1$), with the highest scores in temporal demand ($M = 73.6$, $SD = 19.4$), frustration ($M = 71.0$, $SD = 25.5$), and mental demand ($M = 67.9$, $SD = 24.3$). Effort ($M = 55.7$, $SD = 25.0$) and performance ($M = 52.1$, $SD = 29.4$) were moderate, while physical demand was lower ($M = 28.8$, $SD = 22.6$) (Figure 4).

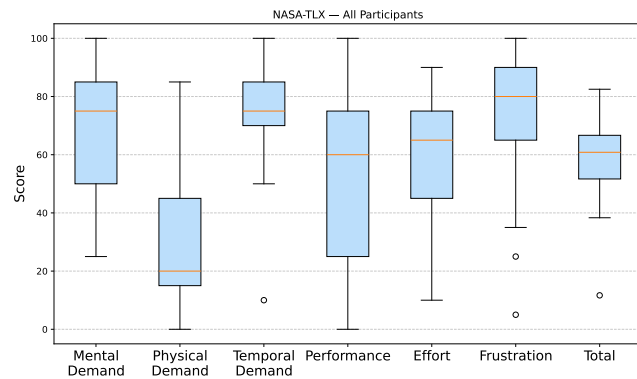


Figure 4: NASA-TLX dimension scores for the experiment.

4.2 Social and co-presence

Across NMSPI dimensions, social and co-presence were comparable between renderings (Table 1).

Paired comparisons showed no differences except for emotional interdependence (Other index), which was higher for VA ($p = 0.003$, $d = 0.7$) (Table 2).

Boxplots confirm similar central tendencies across conditions, with slightly lower dispersion for VA in several indices (Figure 5).

Dimension	Index	CA			VA		
		M	Med	SD	M	Med	SD
Co-presence	Self	5.9	6.0	0.9	5.8	5.7	0.7
	Other	4.7	5.3	1.7	5.0	5.0	1.3
	Global	5.3	5.5	1.1	5.4	5.3	0.9
Attentional Allocation	Self	4.1	4.0	0.8	4.1	4.0	0.6
	Other	4.1	4.0	0.8	4.0	4.0	0.8
	Global	4.1	4.0	0.7	4.1	4.0	0.7
Perceived Message Understanding	Self	4.7	4.7	0.7	5.0	5.0	0.6
	Other	3.9	4.0	0.7	4.1	4.3	0.5
	Global	4.3	4.3	0.6	4.6	4.7	0.3
Perceived Emotional Interdependence	Self	4.8	5.0	1.5	4.7	5.0	1.5
	Other	4.0	4.3	0.9	4.4	4.3	0.8
	Global	4.4	4.5	1.0	4.5	4.7	1.0
Perceived Behavioral Interdependence	Self	4.6	4.7	1.1	4.7	4.7	1.2
	Other	4.2	4.3	1.0	4.2	4.3	1.0
	Global	4.4	4.3	1.0	4.5	4.7	1.0

Table 1: Descriptive statistics (M, Med, SD) for NMSPI dimensions under CA and VA conditions, broken down by Self, Other, and Global indices. All values rounded to one decimal.

Dimension	Index	Test	p	d
Co-presence	Self	t -test	0.493	-0.2
	Other	t -test	0.220	0.3
	Global	t -test	0.655	0.1
Attention Allocation	Self	t -test	1.000	≈ 0
	Other	t -test	0.475	-0.2
	Global	Wilcoxon	0.732	-
Message Understanding	Self	t -test	0.059	0.4
	Other	t -test	0.228	0.3
	Global	t -test	0.069	0.4
Emotional Interdependence	Self	t -test	0.596	-0.1
	Other	t -test	0.003	0.7
	Global	t -test	0.156	0.3
Behavioral Interdependence	Self	t -test	0.574	0.1
	Other	t -test	0.854	≈ 0
	Global	t -test	0.642	0.1

Table 2: Statistical test results (p and d) for all indices under CA and VA.

4.3 Trustworthiness

4.3.1 ABI ratings. ABI ratings were similar between CA and VA across dimensions (Figure 6). Ability was identical ($M = 4.5$), and benevolence ($p = 0.460$, $d = 0.2$), integrity ($p = 0.162$), and overall trust ($p = 0.156$, $d = 0.3$) did not differ significantly between conditions.

4.3.2 Behavior during contradictory steps. Two recordings had technical issues and were excluded, leaving 19 for behavioral analysis. Participants rarely committed to a single guidance source. Instead, many alternated between agents. We therefore focus on the three tasks with contradictory instructions.

In T1 (opening the boxes), three participants opened only one box, always following CA. Only one of these continued to follow CA exclusively in later tasks. In T2 (removing a cable), nine participants worked on both boxes and ten focused on one. Among those who

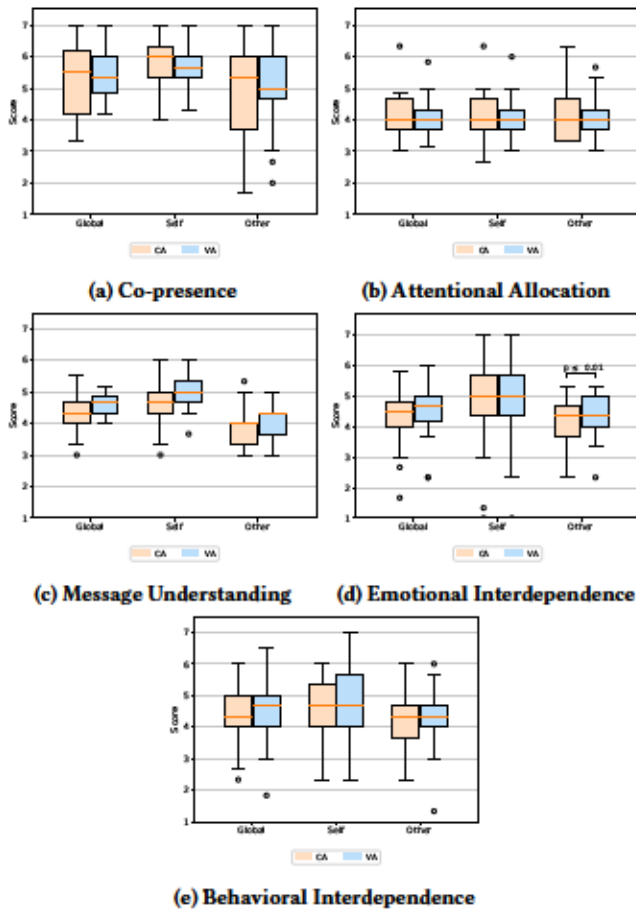


Figure 5: Boxplot visualization of NMSPI dimensions across CA and VA. Each sub-figure represents one of the measured dimensions.

opened both but proceeded with one, all selected the blue (left) box. Eight participants removed both cables, including four who removed the cable from the box each avatar had instructed them to open in T1. Among those who removed only one cable, seven followed CA and four followed VA.

T3 (reconnecting the cable) was completed by 14 participants (two ended early before this step; two recordings were excluded). Within these cases, eight followed VA and six followed CA.

Figure 7 visualizes each participant’s sequence of actions across the three tasks. Each column represents a task (T1–T3), and the horizontal axis indicates whether the action followed CA, VA, both instructions, or neither. Each colored line represents one participant, tracing their choices across steps. Participants are categorized as “both” when they attempted to follow both instruction streams within the same task step, reflecting attempts to combine guidance rather than commit to a single source.

Across 57 contradictory instructions, participants followed the first instruction heard in 65% of cases, with higher rates in T1 (73.7%) and T2 (78.9%).

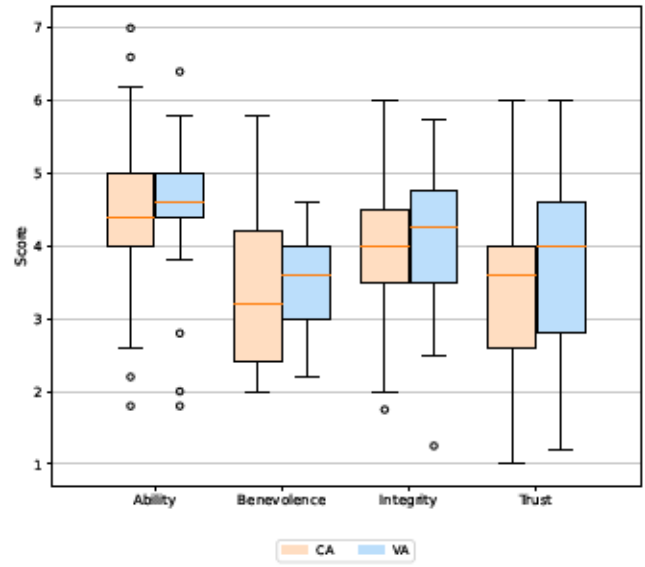


Figure 6: ABI model boxplot for all dimensions.

5 Discussion

Across measures, we did not observe clear differences attributable to rendering style in this conflicting-guidance MR setting. Trust, social presence, and co-presence ratings were comparable for the character avatar (CA) and volumetric avatar (VA), and participants did not develop a stable preference for either agent. Instead, behavior was characterized by mixed adherence, including switching and attempts to incorporate both instruction streams.

This pattern is consistent with prior findings that stylization alone does not reliably shift trust judgments, whereas dynamic expressive cues can be more influential than visual fidelity [9]. Because expressiveness was controlled in our stimuli, the manipulation primarily concerned rendering style, which may partly explain the convergence in subjective ratings. At the same time, the task context likely constrained sensitivity to representation cues. Workload reports indicated substantial temporal demand and frustration, conditions under which users may prioritize task completion over nuanced social evaluation. In this sense, any fidelity-related advantages reported for more naturalistic representations in less demanding interactions [1] may be attenuated by competing contextual pressures.

Behavioral data further emphasize the role of structural features in shaping compliance under conflict. Whereas prior work with a single advisor shows that trust manipulations can influence advice seeking and adherence [19], our setting required participants to arbitrate between two co-present agents delivering contradictory instructions. Participants frequently adopted hedging strategies rather than committing to a single source of guidance. We use this term to describe behaviors in which participants attempted to satisfy both instruction streams, for example by combining actions from each agent within the same task step. These patterns are compatible with prioritization mechanisms such as primacy and attentional competition, and they provide a plausible account

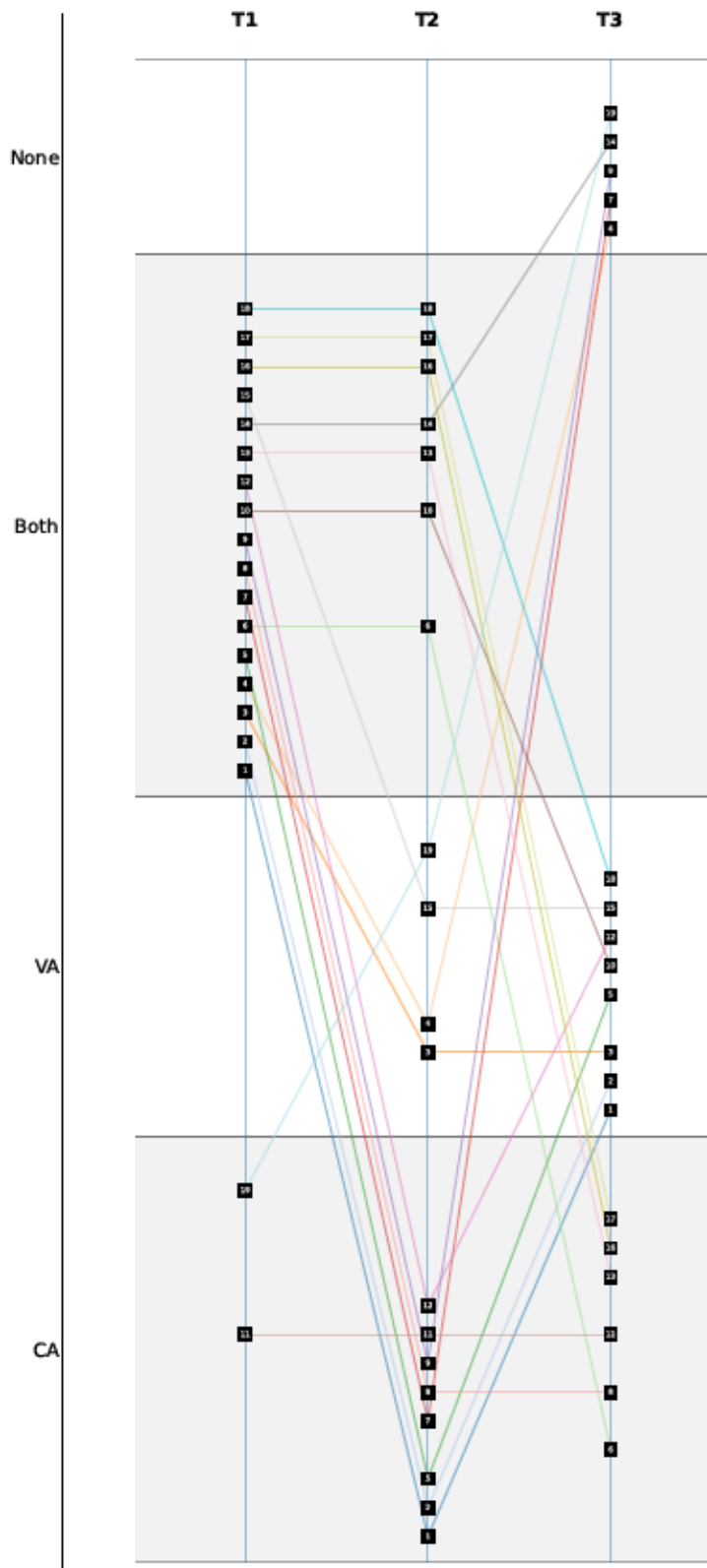


Figure 7: Individual user paths across the three contradictory task steps: box opening (T1), cable handling (T2), and cable connection orientation (T3), represented in the vertical axis. The horizontal axis indicates which source was followed: CA, VA, Both (participant attempted both streams), or None. Each colored line represents one participant, tracing their choices across steps.

for why behavioral differences did not track rendering style in a consistent manner.

Perception measures were similarly stable across representations. Co-presence and engagement scores were largely comparable for CA and VA, consistent with arguments that once an avatar meets a threshold of social bandwidth, additional realism has limited effects on experienced presence [3]. ABI scores also clustered in the mid range for both agents, and neither representation reached values typically associated with strong reliance. In combination with the observed switching and mixed adherence, these ratings point to a setting in which participants maintained a tentative stance toward both agents, rather than committing to one source of guidance.

In relation to our research questions, the behavioral data address **RQ1** by showing that participants frequently managed conflict through switching and hedging, and that instruction timing appeared to influence which guidance was followed when disagreement occurred. For **RQ2**, subjective impressions of trust and social presence were not consistently mirrored in moment-to-moment compliance, indicating that immediate action under conflict was shaped more by situational demands than by stable agent evaluations. Regarding **RQ3**, rendering style (VA vs. CA) did not produce a reliable modulation of either the perception measures or the observed prioritization patterns in this study. Given the limited sample size, these findings should be treated as indicative rather than conclusive, but they reinforce the importance of temporal overlap and competitive information delivery as central design factors in conflicting-guidance MR scenarios.

6 Limitations

This study has several limitations.

First, the sample size was small ($N = 21$), limiting statistical power. Post-hoc power analysis indicated that, given the low observed effect sizes, power remained below 0.2 across metrics. With the current paired design, the study would achieve 80% power only for effects of approximately Cohen's $d = 0.64$ or larger. Accordingly, null results should be interpreted with caution and not as evidence of absence of differences.

Second, although instructions diverged, the agents did not explicitly confront each other (e.g., "do not follow the other"), which may have encouraged participants to treat the situation as parallel guidance streams and remain cautious toward both avatars. Third, compliance was strongly influenced by instruction timing and occasional audio overlap, likely increasing perceptual load and overshadowing subtle effects of rendering style.

Additionally, both avatars were based on pre-recorded performances, ensuring consistency but reducing responsiveness compared to interactive agents commonly expected in MR. We also matched gestures and facial expressions across conditions to isolate rendering fidelity, which improves internal control but reduces ecological validity since real-world systems often couple rendering with expressivity and adaptive behavior. Finally, behavioral trust was operationalized as adherence at divergent steps; future work could incorporate finer-grained measures such as hesitation, correction behavior, and recovery from errors.

7 Conclusion & Future Work

We introduced an exploratory MR paradigm in which users received real-time conflicting guidance from two co-present agents representing the same advisor. Within this setting, rendering style (VA vs. CA) did not yield detectable differences in trust ratings, social presence, or co-presence. Instead, participants frequently managed disagreement through mixed adherence, switching, and attempts to accommodate both instruction streams, suggesting that prioritization under time pressure strongly shapes behavior when guidance diverges.

The results also underscore that behavior in conflicting-guidance scenarios is shaped by how information is delivered, not only by how agents look. Participants rarely committed to one advisor across the task; instead, they often alternated between sources or attempted to satisfy both. This pattern, together with the tendency to follow early instructions at disagreement moments, points to prioritization mechanisms that operate under temporal demand and uncertainty. In such settings, post-task ratings capture overall impressions, but they may not map directly onto the moment-to-moment choices participants make while acting.

Several limitations constrain the conclusions that can be drawn from this work. The sample size limits sensitivity to small effects, and the use of pre-recorded performances removed the responsiveness that often contributes to perceived competence and reliability in interactive systems. In addition, the avatars did not explicitly challenge one another, which may have encouraged participants to treat the situation as competing suggestions rather than a contest between advisors. Future work should build on this paradigm with larger samples, stronger and more controlled conflict manipulations, and adaptive agents that respond to user actions. Reducing audio overlap and introducing decision points that require explicit commitment to one instruction stream would further help separate prioritization effects from representation-driven judgments.

Acknowledgments

This work is supported by the the German Federal Ministry of Education and Research (BMFTR) under the grant 16SV8922 (MULTIPARTIES).

References

- [1] Sahar Aseeri and Victoria Interrante. 2021. The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics* 27, 5 (2021), 2608–2617. doi:10.1109/TVCG.2021.3067783
- [2] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). New York, NY, USA, 1–13. doi:10.1145/3313831.3376550
- [3] 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 (April 2008), 287–318. doi:10.1111/j.1468-2958.2008.00322.x
- [4] Gary Bente, Ralf Schmälzle, Nolan T Jahn, and Andrea Schaaf. 2023. Measuring the effects of co-location on emotion perception in shared virtual environments: An ecological perspective. *Front. Virtual Real.* 4 (April 2023). doi:10.3389/frvir.2023.1032510
- [5] Abhinav Bhattacharya, Luis Fernando de Souza Cardoso, Andy Schleising, Gareth Rendle, Adrian Kreskowski, Felix Immohr, Rakesh Rao Ramachandra Rao, Wolfgang Broll, and Alexander Raake. 2025. AMIS: An Audiovisual Dataset for Multimodal XR Research. In *Proceedings of the 16th ACM Multimedia Systems*

- Conference (Stellenbosch, South Africa) (MMSys '25). New York, NY, USA, 291–297. doi:10.1145/3712676.3718344
- [6] Frank Biocca, Chad Harms, and Jucee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence (Camb.)* 12, 5 (Oct. 2003), 456–480. doi:10.1162/105474603322761270
- [7] Gordon Brown, Julien Hust, Sebastian Büttner, and Michael Prilla. 2022. The Impact of Varying Resolution and Motion Realism of Avatars in Augmented Reality-Supported, Virtually Co-located Sales Consultations. In *Proceedings of Mensch Und Computer 2022* (Darmstadt, Germany) (MuC '22). New York, NY, USA, 255–264. doi:10.1145/3543758.3543772
- [8] Gordon Brown and Michael Prilla. 2020. The effects of consultant avatar size and dynamics on customer trust in online consultations. In *Proceedings of Mensch Und Computer 2020* (Magdeburg, Germany) (MuC '20). New York, NY, USA, 239–249. doi:10.1145/3404983.3405591
- [9] Ryan Canales, Doug Roble, and Michael Neff. 2024. The Impact of Avatar Stylization on Trust. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 418–428. doi:10.1109/VR58804.2024.00063
- [10] SungIk Cho, Seung-wook Kim, JongMin Lee, JeongHyeon Ahn, and JungHyun Han. 2020. Effects of volumetric capture avatars on social presence in immersive virtual environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 26–34. doi:10.1109/VR46266.2020.00020
- [11] James J Cummings and Erin E Wertz. 2022. Capturing social presence: concept explication through an empirical analysis of social presence measures. *Journal of Computer-Mediated Communication* 28, 1 (11 2022), zmac027. arXiv:https://academic.oup.com/jcmc/article-pdf/28/1/zmac027/4843772/zmac027.pdf doi:10.1093/jcmc/zmac027
- [12] ElevenLabs. 2025. *Transform your voice into another while preserving emotion, delivery, and nuance*. https://elevenlabs.io/voice-changer/ Last accessed: 19 Aug. 2025.
- [13] Yue-Ming Guo, Wai-Ling Ng, Fei Hao, Chen Zhang, Shu-Xu Liu, and Adil Masud Aman. 2023. Trust in Virtual Interaction: The Role of Avatars in Sustainable Customer Relationships. *Sustainability* 15, 18 (Sept. 2023), 14026. doi:10.3390/su151814026
- [14] Joanna Hale, Madeleine EM Payne, Kathryn M Taylor, Davide Paoletti, and Antonia F De C Hamilton. 2018. The virtual maze: A behavioural tool for measuring trust. *Quarterly Journal of Experimental Psychology* 71, 4 (2018), 989–1008. doi:10.1080/17470218.2017.1307865 PMID: 28303739.
- [15] Chad Harms and Frank A. Biocca. 2004. Internal consistency and reliability of the networked minds social presence measure. In *Seventh Annual International Workshop: Presence 2004*, Mariano Alcaniz and Beatriz Rey (Eds.). Valencia.
- [16] Andrew Irlitti, Mesut Latifoglu, Qishui Zhou, Martin N Reinoso, Thuong Hoang, Eduardo Velloso, and Frank Vetere. 2023. Volumetric Mixed Reality Telepresence for Real-time Cross Modality Collaboration. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). New York, NY, USA, Article 101, 14 pages. doi:10.1145/3544548.3581277
- [17] Simon Kimmel, Frederike Jung, Andrii Matvienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). New York, NY, USA, Article 429, 16 pages. doi:10.1145/3544548.3580707
- [18] Christos Kyrilitsias and Despina Michael-Grigoriou. 2022. Social Interaction With Agents and Avatars in Immersive Virtual Environments: A Survey. *Frontiers in Virtual Reality* 2 (2022). doi:10.3389/frvir.2021.786665
- [19] Jinghuai Lin, Johrine Cronjé, Ivo Käthner, Paul Pauli, and Marc Erich Latoschik. 2023. Measuring Interpersonal Trust towards Virtual Humans with a Virtual Maze Paradigm. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2401–2411. doi:10.1109/TVCG.2023.3247095
- [20] Maria Machneva, Anthony M. Evans, and Olga Stavrova. 2022. Consensus and (lack of) accuracy in perceptions of avatar trustworthiness. *Computers in Human Behavior* 126 (2022), 107017. doi:10.1016/j.chb.2021.107017
- [21] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (July 1995), 709. doi:10.2307/258792
- [22] Microsoft. 2025. *Microsoft Speech SDK (MSAiface) Download*. https://www.microsoft.com/en-us/download/details.aspx?id=101454 Last accessed: 19 Feb. 2025.
- [23] Daniel Pimentel and Charlotte Vinkers. 2021. Copresence With Virtual Humans in Mixed Reality: The Impact of Contextual Responsiveness on Social Perceptions. *Frontiers in Robotics and AI* 8 (2021). doi:10.3389/frbot.2021.634520
- [24] Reallusion. 2025. *Character Creator*. https://www.reallusion.com/character-creator/ Last accessed: 19 Aug. 2025.
- [25] Reallusion. 2025. *Headshot Plugin for Character Creator*. https://www.reallusion.com/character-creator/headshot/ Last accessed: 19 Aug. 2025.
- [26] Reallusion. 2025. *iClone - 3D Animation Software*. https://www.reallusion.com/iclone/ Last accessed: 19 Feb. 2025.
- [27] Holger Regenbrecht, Katrin Meng, Arne Reepen, Stephan Beck, and Tobias Langlotz. 2017. Mixed Voxel Reality: Presence and Embodiment in Low Fidelity, Visually Coherent, Mixed Reality Environments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 90–99. doi:10.1109/ISMAR.2017.26
- [28] Gareth Rendle, Adrian Kreskowski, and Bernd Froehlich. 2023. Volumetric Avatar Reconstruction with Spatio-Temporally Offset RGBD Cameras. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 72–82. doi:10.1109/VR55154.2023.00023
- [29] Jenna-Lyn R. Roman, Juane Greene, James A. Swaim, Graham H. Lowman, and Michael J. Maloni. 2025. Re-examining the ability, benevolence, and integrity (ABI) model of trust: Do business professors really need all three? *The International Journal of Management Education* 23, 3 (Dec. 2025), 101207. doi:10.1016/j.ijme.2025.101207
- [30] 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 *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 103–108. doi:10.1109/ISMAR-Adjunct.2018.00044
- [31] Ulrike Schultze and Jo Ann M Brooks. 2019. An interactional view of social presence: Making the virtual other “real”. *Inf. Syst. J.* 29, 3 (May 2019), 707–737. doi:10.1111/isj.12230
- [32] Sebastian Siehl, Kornelius Kammler-Sücker, Stella Guldner, Yannick Janvier, Rabia Zohair, and Frauke Nees. 2024. To trust or not to trust? Face and voice modulation of virtual avatars. *Frontiers in Virtual Reality* 5 (Feb. 2024). doi:10.3389/frvir.2024.1301322
- [33] Isaac Wang, Jesse Smith, and Jaime Ruiz. 2019. Exploring Virtual Agents for Augmented Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). New York, NY, USA, 1–12. doi:10.1145/3290605.3300511
- [34] Peng Wang, Yue Wang, Mark Billinghurst, Huizhen Yang, Peng Xu, and Yanhong Li. 2023. BeHere: a VR/SAR remote collaboration system based on virtual replicas sharing gesture and avatar in a procedural task. *Virtual Real.* 27, 2 (Jan. 2023), 1409–1430. doi:10.1007/s10055-023-00748-5
- [35] Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2596–2606. doi:10.1109/TVCG.2023.3247072
- [36] Boram Yoon, Jae-eun Shin, Hyung-il Kim, Seo Young Oh, Dooyoung Kim, and Woontack Woo. 2023. Effects of Avatar Transparency on Social Presence in Task-Centric Mixed Reality Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4578–4588. doi:10.1109/TVCG.2023.3320258
- [37] Kevin Yu, Gleb Gorbachev, Ulrich Eck, Frieder Pankratz, Nassir Navab, and Daniel Roth. 2021. Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4129–4139. doi:10.1109/TVCG.2021.3106480
- [38] Faisal Zaman, Craig Anslow, Andrew Chalmers, and Taehyun Rhee. 2023. MR-MAC: Mixed Reality Multi-user Asymmetric Collaboration. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 591–600. doi:10.1109/ISMAR59233.2023.00074
- [39] Runge Zhu and Cheng Yi. 2024. Avatar design in Metaverse: the effect of avatar-user similarity in procedural and creative tasks. *Internet Res.* 34, 1 (Feb. 2024), 39–57. doi:10.1108/INTR-08-2022-0691