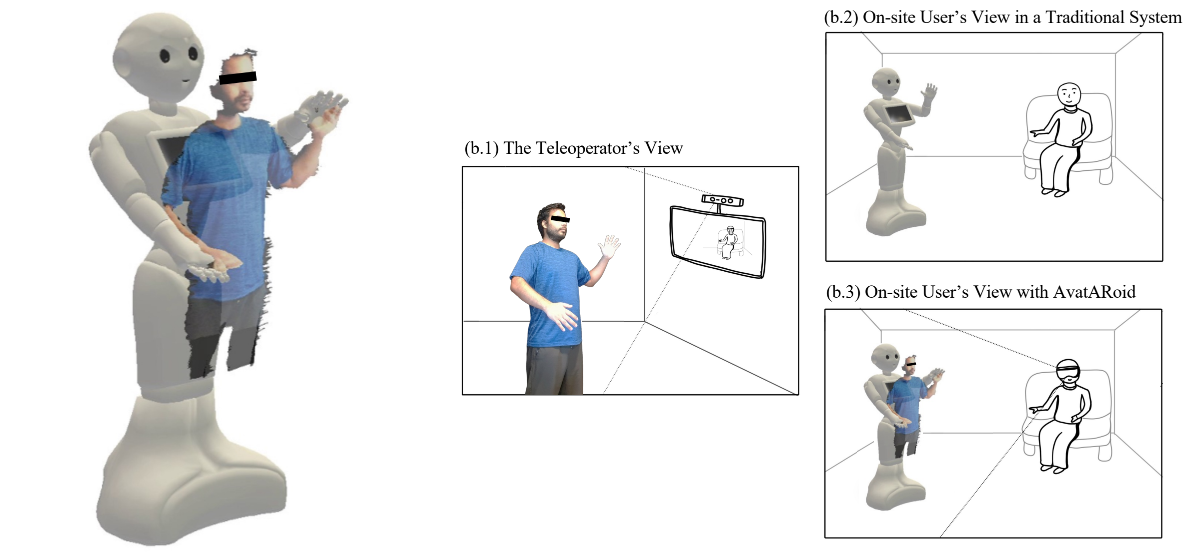
AvatARoid: A Motion-Mapped AR Overlay to Bridge the Embodiment Gap Between Robots and Teleoperators in Robot-Mediated Telepresence

Amit Ghimire
Computer Science
University of British Columbia
Vancouver, British Columbia, Canada
amitghi@cs.ubc.ca

Anova Hou
Computer Science
University of British Columbia
Vancouver, British Columbia, Canada
anova78@student.ubc.ca

Ig-Jae Kim
AI&Robotics Institute
Korea Institute of Science and Technology
Seoul, Republic of Korea
drjay@kist.re.kr

Dongwook Yoon
Computer Science
University of British Columbia
Vancouver, British Columbia, Canada
yoon@cs.ubc.ca



(a) Representation of the Teleoperator in AvatARoid

(b) A Comprehensive Overview of the AvatARoid Telepresence System

Figure 1: The AvatARoid Design: (a) The teleoperator's motion-mapped avatar is overlaid on a humanoid robot; (b.1) The teleoperator interacting from their site; (b.2) The on-site user interacting with teleoperator in traditional system; (b.3) The on-site user interacting with teleoperator in AvatARoid

Abstract

Robot-mediated telepresence promises to facilitate effective social interaction between remote teleoperators and on-site users. However, disparities between the robot's form and the teleoperator's representation cause perceptual conflict in on-site users, degrading interaction quality. We introduce AvatARoid, a novel design that bridges this embodiment gap by superimposing the teleoperator's motion-mapped AR avatar overlay on a humanoid. We evaluated our design in a mixed-method study (n=48) using an immersive simulation where participants interacted with a confederate teleoperator, presented in either (a) a humanoid robot, (b) a humanoid robot with video, or (c) AvatARoid. Results suggest AvatARoid

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713812

significantly improved teleoperator embodiment for on-site users, particularly enhancing co-location, and control perceptions, and providing richer non-verbal gestures. In contrast, video and baseline conditions often resulted in a pronounced disconnect between the teleoperator and the robot for on-site users. Our study offers new insights into designing novel teleoperator representations to promote social interaction in robot-mediated telepresence.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; **Mixed / augmented reality**; **Collaborative and social computing systems and tools**.

Keywords

AvatARoid, telepresence, robot, humanoid, embodiment, avatar, AR overlay,

ACM Reference Format:
Amit Ghimire, Anova Hou, Ig-Jae Kim, and Dongwook Yoon. 2025. AvatARoid: A Motion-Mapped AR Overlay to Bridge the Embodiment Gap Between Robots and Teleoperators in Robot-Mediated Telepresence. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 27 pages. https://doi.org/10.1145/3706598.3713812

1 Introduction

Robotic Telepresence Systems enable a teleoperator to control a physical robot in the space of an on-site user, facilitating interactions as if they were physically together. These systems are gaining traction in healthcare, education, and workplaces [68], fostering engagement between clinicians, long-term patients, and family members [49, 67, 113, 126, 127], boosting student collaboration and motivation in distant learning [48, 112, 124], and enabling in-person like interactions in hybrid work scenarios [73, 130]. The consistent theme across domains is the potential to enhance social interactions in telepresence [68], which relies on on-site users perceiving the robot as the teleoperator's embodiment and experiencing their social presence, which is vital for fostering rapport and creating positive impressions [37, 68, 110].

The visual representation of the teleoperator in the robotic telepresence system plays a crucial role in shaping on-site users' perception of the teleoperator. Embodied social presence theory posits that people interpret interactions through embodiment, which conveys both verbal and non-verbal social cues (e.g., identity, body language, gestures) that enhance emotional connection and the sense of social presence [81, 88]. Effective representation in telepresence systems enhances the perceived embodiment of the teleoperator, helping the on-site users interpret social cues more easily and view the teleoperator as a genuine social entity. This encourages emotionally resonant interactions and positive impressions of the teleoperators [71].

Two distinct embodiments exist for the teleoperator in robotic telepresence: their own body and the robot they control. Current user representation methods do not effectively bridge the gap between these two embodiments. Tablet displays showing the teleoperator's video [68] fail to establish a strong connection, leading on-site users to perceive the robot as an object, causing issues like bullying [97] or invasion of personal space [122]. AR-based 3D avatars overlaid on non-humanoid robots create unrealistic expectations, contributing to "uncanny valley of telepresence" when robots fail to meet these expectations [55]. While enhancing functionality, humanoid robots distort perceptions by attributing robotic traits to teleoperators due to the robot's highly anthropomorphic design [26, 69]. Addressing these challenges in current user representations is crucial for bridging the gap between the robotic and teleoperator embodiments.

To bridge the embodiment gap in robot-mediated telepresence, we introduce *AvatARoid*, a novel telepresence design that superimposes a motion-mapped AR overlay of the teleoperator's avatar onto a humanoid robot (See Figure 1). The humanoid robot provides physical agency and manipulation capabilities to the teleoperator, allowing them to "be in on the action" [22]. Synchronizing motion between the avatar and the robot reinforces the teleoperator's embodiment by establishing a one-to-one relationship between the robot's and the teleoperator's actions which enhances social interaction. To evaluate the efficacy of AvatARoid, we pose a subsequent research question: How does AvatARoid influence the on-site user's sense of the teleoperator's embodiment, social presence, and impression compared to existing user representations in telepresence robots?

To address this question, we ran a mixed-method study, nesting qualitative behavior observation, questionnaires, and interviews within a controlled lab setting. We evaluated how AvatARoid's design—compared to robot-only and video representations—affects the teleoperator's embodiment, presence, and impression formed by on-site users. To mitigate the technical challenges and potential confounds of implementing a physical robot, we utilized a visually realistic virtual robot in an immersive simulation. The results show that AvatARoid significantly improved the on-site users' sense of the teleoperator's embodiment compared to the alternatives, particularly by enhancing the sense of colocation between the robot and the teleoperator, while reinforcing the teleoperator's control over the robot's actions. Interview responses emphasized AvatARoid's potential for fostering rich social and emotional engagement. However, on-site users' impressions of the teleoperator (e.g., trustworthiness, likeability, similarity, and infrahumanization) remained consistent across different representation conditions, indicating the need to consider additional factors in future evaluations.

This study makes the following contributions to the field of human-computer interaction, with a particular emphasis on integrating telepresence and augmented reality:

- *AvatARoid Design*: We introduce the AvatARoid system, a novel telepresence solution combining motion-mapped, photorealistic AR avatars with a humanoid robot to enhance remote interaction by bridging the gap between the teleoperator's embodiment and the robot's physical presence.
- *Empirical Evaluation*: We provide empirical evidence that the AvatARoid design enhances on-site users' sense of spatial presence with the teleoperator, improves perceptions of the teleoperator's control over the robot, and fosters stronger emotional engagement and clearer communication of non-verbal cues in robot-mediated telepresence.

These contributions establish a foundation for the development and evaluation of innovative user representation strategies in augmented telepresence systems, ultimately advancing the potential for fostering meaningful social interactions in remote environments.

2 Related Work

2.1 Challenges in Existing Teleoperator Representations

Currently, widely deployed Robotic Telepresence Systems are predominantly designed as Mobile Robotic Telepresence [68], often described as a “tablet on wheels.” This approach provides mobility for teleoperators but limits their ability to express themselves and manipulate the objects in the environment [22, 105]. With teleoperators represented through a video, the lack of effective identity and gestural cues often causes a "mixing-metaphor" problem [122], where the robotic embodiment of the teleoperators is iatedassoc to many non-human (for instance furniture) as well as some human-like (for example, person with disability) metaphors. When on-site users form mental models of the teleoperator as an impersonal object or machine [55, 73, 122] while the teleoperator feels the robot as an extension of their embodiment, this compromises interaction quality and leads to violations of social and proxemic norms [73, 122].

AR technologies offer solutions by creating 3D avatars or holographic representations of the remote collaborators [12, 90, 94, 96, 102] that embody the teleoperator's gestures and movements. While they enable expressive gesturing, they lack physical presence, mobility, or abilities to manipulate objects in the on-site user's physical setting. Recent studies [55, 61] have combined the mobility of robots with the expressiveness of AR avatars to represent the teleoperator. However, they present a dilemma: the robot's limited physical abilities do not align with expectations set by detailed visual representation of the avatars, leading to uncanny telepresence experiences [55].

Humanoid robots with their mobility, expressiveness, and ability to manipulate objects [5, 7, 18] enable tangible physical interaction like handshakes and fist bumps, which improves social connection and cooperation between teleoperators and on-site users [5, 18]. They are also effective in inducing socially enriching interactions, like reciprocal body gestures such as nodding [56]. However, they risk distorting the on-site user's perception of the teleoperator's personality [25, 26, 69], attributing robotic qualities to the teleoperator referred to as ”robo-morphism” [109], which negatively impacts the teleoperator's social presence due to the strong identity of the robot itself [29].

The AvatARoid design approach addresses operational and representational discrepancies by integrating humanoid robots with direct augmented reality (AR) avatar overlays.

2.2 Telepresence Robots and AR

In the field of AR and robotics, social interaction and telepresence collaboration have emerged as significant application domains, as highlighted by a comprehensive survey [121]. Integrating AR content with robotics has been explored to enhance remote interactions, demonstrated by systems like RobotAR [131]. This toolkit facilitates remote consulting, allowing instructors to overlay AR content into a student's workspace for instructional support during problem-solving tasks.

Exploring AR-based user representation in telepresence robots has provided various insights. For instance, Zhang et al. [138] demonstrated that an AR-rendered generic avatar head on a robot helps on-site users accurately determine the teleoperator's gaze, unlike a conventional tablet display. Hwang et al. [52] introduced a tele-meeting system projecting a 3D video of a teleoperator onto a mobile robot, enhancing co-presence between remote and local participants. More recent innovations combine VR interfaces for teleoperators with HMDs, and AR HMDs for on-site users. These systems integrate mobile robots with HMD-based AR avatars [55] or projection-based AR displays [61], enhancing the teleoperator's avatar on wheeled non-humanoid robots.

AvatARoid builds upon the foundation of the HMD-based AR overlay system used in VROOM [55]. VROOM was significant for enhancing remote collaboration by improving the teleoperator's ability to communicate non-verbally, such as through gestures. Although VROOM's approach showed promise, its implementation using non-humanoid robots limited the ability to fully realize the potential of expressive visual representation. AvatARoid's design aims to address this issue by incorporating a humanoid robot to better match the teleoperator's gestures and expressions while synchronizing the robot's motion with the AR avatar. Furthermore, in contrast to using a pre-rendered 3D avatar, created from a static teleoperator image, which constrained the fidelity of facial expressions, AvatARoid's design employs a real-time point cloud AR avatar that more accurately captures and conveys the teleoperator's dynamic facial expressions, thereby potentially enhancing the overall communication experience.

2.3 Perceived Embodiment and Social Presence in Robot-mediated Social Interaction

Our evaluation of interaction quality in robotic telepresence systems is based on the Embodied Social Presence Theory, which identifies embodiment as a central factor in shaping users' cognitive engagement and perception during social interactions [86]. According to this theory, embodied actions—including verbal and non-verbal cues, such as gestures, movements, and body language—serve as key social signals. Users interpret these signals relative to the embodiment of their interaction partner, constructing meaning through a mutual exchange of actions, thoughts, and feelings. This process fosters a sense of social presence, defined as the feeling of being genuinely connected to another person during an interaction [20, 86, 114]. Social presence influences how on-site users perceive the teleoperator, with studies highlighting its importance in maintaining rapport and ensuring the quality of interaction in robotic telepresence [37, 110].

The physical and representational embodiment of the teleoperator within the robotic system directly influences users' sense of presence.Enhanced physical embodiment, achieved through lifelike robotic movements and gesture coordination, has been shown to heighten social presence and deepen cognitive and emotional engagement with the teleoperator [124]. Additionally, user representation, such as AR avatars or on-screen visuals, offers supplemental cues that amplify social presence. For instance, synchronized non-verbal gestures and gaze behaviors between the robotic system and the teleoperator's representation improve understanding and strengthen the impression of the teleoperator as an authentic social partner [36, 115]. Research also highlights the role of identity cues, such as culturally relevant or personalized avatars, in enhancing users' perception of presence and trustworthiness [41, 60, 76].

Beyond visual representation, the alignment of robotic actions with user representation cues significantly shapes interaction dynamics. For example, synchronized mutual gaze during hallway encounters has been shown to elevate the quality of social interactions and observer impressions of the teleoperator [36]. Similarly, research on robot-human communication reveals that achieving gesture synchrony with avatars markedly improves observers' understanding of the teleoperator's actions and their overall impression of them [115]. These findings suggest that social presence is a multifaceted construct, influenced by the interplay between robotic embodiment, user representation, and interaction design.

To further enhance social presence and impression formation, robotic systems can incorporate detailed AR avatars and ensure synchronization between robotic motion and the avatar's actions. These design principles, as exemplified by the AvatARoid platform, provide a foundation for creating immersive telepresence systems that prioritize authentic and engaging interactions.

2.4 Impression Formation in Robot-Mediated Telepresence

In robot-mediated telepresence, experiential factors such as trust, likeability, homophily, and infrahumanization play a pivotal role in shaping interaction quality and user experience. These impression measures, alongside social presence, capture the social and psychological responses of on-site users to teleoperators, mediated by robotic systems and augmented representations. Together, they define perceptions of teleoperators and significantly influence relationship building, collaboration, and task performance in telepresence systems [35, 107].

Trust, defined as confidence in the reliability and intentions of another party, develops through consistent robotic system performance and alignment between capabilities and teleoperator representations, enhancing cooperation and reducing anxiety [46]. Research shows that physical embodiment and system control by the teleoperator further increase trust between partners, as local users perceive teleoperators with greater autonomy as more reliable [99]. Likability, the degree of positive affect towards a teleoperator, often hinges on aesthetic and behavioral attributes. Features such as personalized and human-like representations with clothes not only enhance rapport and creativity but also improve perceptions of uniqueness and humanness, further strengthening interpersonal dynamics [42, 52, 76].

Homophily, the perception of shared characteristics or values, strengthens interpersonal bonds and facilitates effective communication. When AR avatars incorporate culturally appropriate or context-specific cues, they foster greater alignment and social connection [60]. Intimacy, marked by emotional closeness and mutual synchrony, is facilitated by coordinated gestures and gaze behaviors. Systems that mimic natural human interactions and enable fluid movement of telepresence robots contribute to these perceptions, although individual differences among users shape the impact of these interactions [21, 30, 123]. Finally, infrahumanization, the bias of attributing lesser humanity to out-group members, can hinder inclusivity and empathy. However, detailed and human-like robotic representations reduce this effect, promoting positive social dynamics even in diverse settings [109].

These constructs collectively define the psychological dimensions of telepresence, offering a foundation for designing systems that enable meaningful interactions. For instance, findings suggest that personalization and appropriate system height affect perceptions of dominance and persuasiveness, influencing conversational dynamics and leadership roles during interactions [100]. By leveraging culturally tailored designs, expressive avatars, and features like mimicry, telepresence platforms can foster trust, likeability, and empathy while mitigating biases. Such insights are vital for advancing applications in remote collaboration, education, and healthcare [15, 87].

3 Designing AvatARoid: Bridging Embodiment in Robotic Telepresence

This section details the design goals of AvatARoid and the corresponding design elements developed to achieve them.

3.1 Design Goals of AvatARoid

AvatARoid aims to enhance social interaction by fostering a stronger sense of teleoperator embodiment and social presence for on-site users, guided by three specific goals.

Goal 1. Enhance presence through direct AR overlay of the upper body. AvatARoid enhances presence by providing a dynamic 3D representation of the teleoperator. This includes a realistic and expressive AR avatar that emphasizes detailed non-verbal communication through facial expressions, gaze, gestures, and body language, effectively conveying emotional nuances and improving social presence [89, 91, 132]. By directly overlaying the teleoperator's upper body onto the humanoid robot, the AR avatar enhances spatial awareness, fostering social presence and creating a shared physical-space experience, which is crucial for interaction [19, 90, 91]. In contrast, existing 'tablets on wheels' designs, which limit the teleoperator's presence to a screen, fail to convey these detailed non-verbal cues, thereby reducing human presence and embodiment [73, 122].

Goal 2. Establish teleoperator's agency through real-time motion synchrony. It is necessary to ensure the teleoperator's motions are accurately and intuitively reflected in both the robot and their representation, providing an integrated and coherent representation to help on-site users establish the teleoperator's agency. AvatARoid aims to align the robot's motion with the teleoperator's avatar, which helps on-site users interpret the robot's actions as those of the teleoperator. Previous studies show that synchronizing on-screen gestures of the teleoperator, such as nodding and head-turning, by moving the robot in physical space, improves action interpretation [115]. AvatARoid extends this approach by synchronizing gestures not on-screen but in 3D space, enhancing the realism of the teleoperator's actions.

Goal 3. Maintain the focus on the teleoperator's identity through detailed visual representation. A critical aspect of AvatARoid's design is preserving the teleoperator's identity by emphasizing key visual details, such as facial features and expressive body movements. Although humanoid robots, known for their mobility and expressiveness [5, 7, 18], can facilitate human-like interactions that enhance cooperation [5, 18] and reciprocate non-verbal gestures [56], they also risk overshadowing the teleoperator's individuality and influencing user judgments [25, 26, 69, 109]. By conceptualizing the robot as an avatar that centers on the teleoperator's personality, on-site users can more readily focus on the person behind the technology rather than the robot itself, ultimately reinforcing a genuine sense of human presence.

3.2 Design Elements

AvatARoid's design aims to achieve the aforementioned goals through two primary design elements: motion synchronization between the robot and the AR avatar and detailed rendering of the avatar overlay. Here, we outline the design challenges, present the design elements we opted for in our problem context, and rationalize our choices against the alternatives. Figure 2 shows the integration of our design elements into a cohesive user experience.

3.2.1 Motion Mapping and Synchronization. To achieve Goals 1 and 2, AvatARoid employs a realistic and responsive synchronization between the humanoid robot and the AR avatar, particularly in matching the body articulation, i.e., joint position and motion. This requires leveraging the well-researched correspondence problem [43] in robot teleoperation, which entails mapping the teleoperator's human-like motion onto the mechanical constraints of a robot.

Our design employs upper-body retargeting and Inverse Kinematics (IK) to synchronize the humanoid robot's arm and head movements with the teleoperator avatar's hand and head motions, ensuring accurate reflection of the teleoperator's real-time actions in the on-site user's physical space [33]. IK computes the robot's joint configurations that enable the robot arm's end-effector to reach a desired position and orientation. Despite occasional misalignments due to differing kinematic structures, the robot's hand targeting the avatar's hand ensures spatial congruency, reinforcing the perceived connection between the robot and the avatar. This synchronization enhances the on-site user's intuitive perception of shared actions and strengthens their sense of teleoperator embodiment.

Alternative approaches to motion mapping in robotic telepresence include using pre-rendered, gesture-triggered motions that classify a teleoperator's gestures into predefined animations [8]. However, this reduces expressiveness and may fail to reflect real-time actions. Another method is forward kinematics, which predicts an end effector's position and orientation from joint parameters derived from the teleoperator's pose [33, 58]. In AvatARoid, forward kinematics is complicated by mismatched kinematic structures and reduced degrees of freedom between human and robot arms, causing spatial discrepancies during tasks like handshakes and object handovers and disrupting the sense of embodiment [8].

3.2.2 Overlay Avatar Rendering. To achieve Goal 3, our design leverages a front-view upper body point-cloud avatar generated in real-time through a single RGBD (Red, Green, Blue, and Depth) sensor. This enables us to construct a volumetric avatar, a three-dimensional, color-enriched point-cloud representation of the teleoperator, which we can overlay on the robot. Findings of previous studies also inform our design choice, suggesting point cloud avatars to be a better visualization method for improving social presence in AR as well as full or upper body realistic avatars being the most preferred in collaborative AR settings [132, 137].

We explored the possibility of using a pre-rendered rigged 3D avatar generated from a static image as suggested in prior work [96, 102], but ultimately decided against it due to its limitations in accurately capturing the dynamism and subtleties of facial expressions. While static facial representations avoid rendering artifacts, they risk invoking the 'uncanny valley' effect when striving for photorealism, appearing lifelike failing to be convincingly human [77]. High-fidelity real-time avatar techniques exist [90], but they require complex capture setups beyond the scope of AvatARoid's initial evaluation. Thus, our design enables an authentic, interactive avatar using only a single RGB-D sensor.

4 Evaluating AvatARoid

Our study was designed to evaluate the impact of AvatARoid's motion-mapped avatars on perceptions of teleoperator's embodiment and agency, as compared to existing representations in prevalent telepresence robotic system configurations [5, 18, 68], directly addressing our research question. To this end, we conducted a between-subject, mixed-method study combining qualitative behavior observation and post-task interviews within a controlled experiment. The study utilized a one-way factorial experimental design comprising three conditions (Figure 3):

- Baseline (BL): Only the humanoid robot was presented.
- Video (VD): A tablet attached to the robot's torso displayed a live video feed of the teleoperator's face.
- AvatARoid (AR)

Our participants played the role of the on-site user. A research team member acted as the confederate teleoperator, using a deception protocol to ensure experimental control and standardization, a common approach in HRI studies [98]. We paid careful attention to detail to avoid potential confounds. In all conditions, the physical robot performed gestures and movements driven by the teleoperator's hand motion, and we used the same equipment for voice communication. Additionally, the confederate teleoperator was unaware of participants' conditions to avoid bias towards any of the three conditions. After the experiment, participants were debriefed on the role of the confederate and the deception protocol. The study protocol received approval from the University's institutional review board.

4.1 Participants

We recruited 48 participants (16 per condition, randomly assigned) through the university's online study advertisement platform. The sample consisted of 33 women, 14 men, and one non-binary individual. The majority were aged 19–24, followed by 25–34, 45–54, and

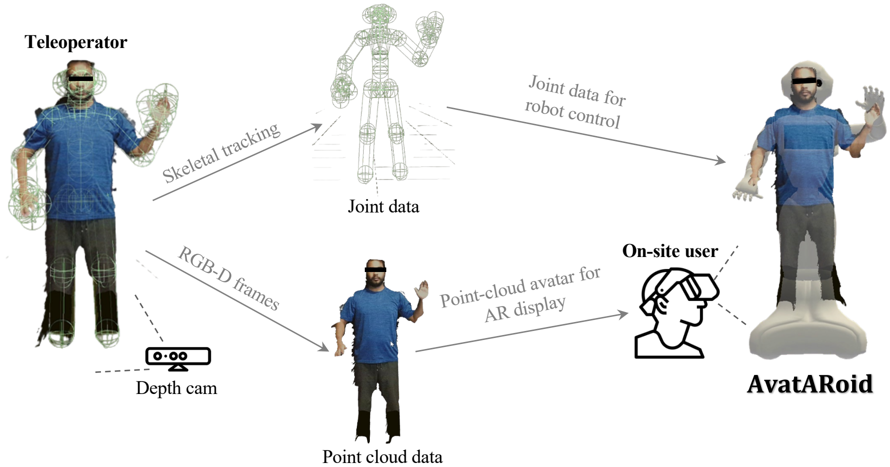
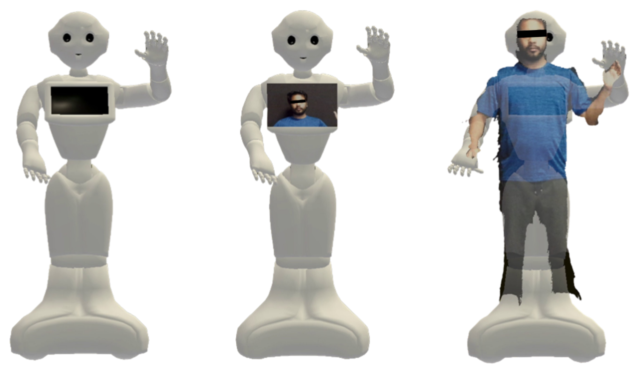


Figure 2: Anatomy of the AvatARoid system design: The upper half illustrates the Motion Mapping and Synchronization Design, where teleoperator joint data, captured by a depth sensor, is transmitted to the on-site system to control the humanoid robot. The lower half depicts the Overlay Avatar Rendering Design, where RGBD images create a point cloud avatar of the teleoperator. This avatar is streamed to the on-site user system which uses AR headset to display the avatar overlaid on the robot.



(a) Baseline (b) Video (c) AvatARoid

Figure 3: Three user representations used in our evaluation: (a) Baseline: a humanoid robot, (b) Video: a humanoid robot with a teleoperator video, and (c) AvatARoid

35–44. In terms of prior experience, 32 participants had no experience with robots, 10 had minimal interaction, and 6 had extensive experience. Regarding AR/VR devices, 28 had no prior usage, 15 had limited familiarity, and 5 had substantial experience. For detailed demographics, see Table 1. Each participant received a $50 honorarium.

4.2 Simulation, Environmental Setup and System Implementation

4.2.1 Simulating AvatARoid. To evaluate AvatARoid, we employed a simulation-based approach in an immersive AR environment representing a typical robot-mediated telepresence scenario. We integrated techniques such as blending the virtual robot into the real environment, enhancing spatial awareness, and incorporating realistic audio cues to achieve co-presence and realism—key elements in simulation-based studies [38, 59, 62, 75].

Our decision to use a simulated environment, rather than a physical robot setup, stemmed from a desire to isolate the effects of different user representations while minimizing confounds introduced by physical implementations. Although virtual robots can have their own confounds (e.g., perceived identity, lack of physicality, user expectations and biases), the simulation enabled precise control and replication of robot behavior across conditions, ensuring that observed differences in user responses resulted solely from visual representation. In contrast, physical robots could introduce variability from mechanical limitations, sensor inaccuracies, and environmental factors, complicating interpretation. The simulation also provided greater flexibility in manipulating robot appearance

Category	BL	VD	AD
Gender			
Women	11	11	11
Men	5	5	4
Non-binary	0	0	1
Age Group			
19–24	11	10	10
25–34	5	2	4
35–44	0	2	1
45–54	0	2	1
Prior Experience (Robots)			
None	10	12	10
Limited	6	1	3
Substantial	0	3	3
Prior Experience (AR/VR)			
None	10	9	9
Limited	4	6	5
Substantial	2	1	2
Total Participants	16	16	16

Table 1: Participant Demographics and Experience

and behavior, ensured experimental consistency and reproducibility, and circumvented the technical and financial challenges of implementing a full AvatARoid system at this stage [129].

Simulation-based approaches are widely used in HRI research for their cost-effectiveness, flexibility, and controlled environments, making them ideal for early-stage assessments [106, 111, 133]. Previous studies show that simulating verbal interactions with virtual humanoid robots in immersive simulations can approximate the experience of interacting with an actual robot [82]. As our primary focus is on the effects of differing visual representations of the teleoperator—and not on physical interactions, which did not vary across conditions—this approach is appropriate. This strategy aligns with recent HRI trends demonstrating that simulations can provide social experiences comparable to those of physical robot interactions [39, 82, 93].

We acknowledge that simulated environments cannot fully capture tangible presence or physical manipulation. However, given our emphasis on verbal interactions and visual representations, simulation proves especially suitable for our research objectives. For a more detailed discussion of our rationale, implementation, and the validity of simulation in HRI studies, please refer to Appendix A.1.

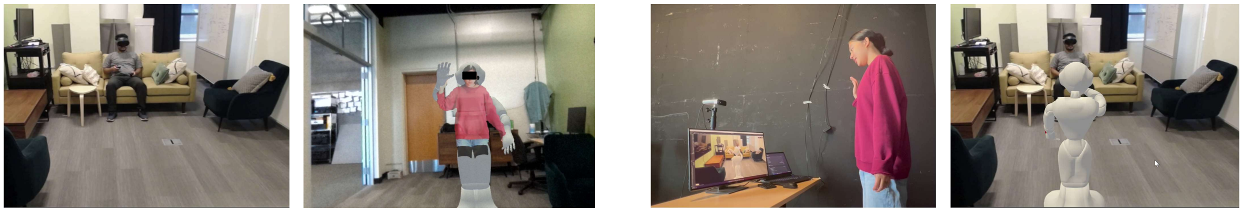
4.2.2 Environment Setup. The teleoperator and on-site user setup in our AR simulation are shown in Figure 4. The on-site user was seated on a sofa, wearing an AR HMD, and interacting with the virtual humanoid robot in the simulation while the teleoperator controlled the simulated robot from a different room using their system.

4.2.3 System Implementation Details. We provide the implementation details of the two systems: the Teleoperator System and the On-Site User (participant) System. We illustrated the architectural design of our simulated AvatARoid system in Figure 5.

Teleoperator System The teleoperator system utilizes an Azure Kinect Sensor to stream RGBD and joint data of the teleoperator for real-time point-cloud avatar creation and motion synchronization. We used 3D model of the Pepper robot [104] simulated in Unity3D (ver. 2021.3f) as the robotic system given its status as a widely used social humanoid robot in HRI research and its compatibility with motion-mapped interaction studies. The teleoperator system ran on Windows 11 and integrated with the Azure Kinect sensor with Azure Kinect SDK ver.1.4.1 and AzureKinectExamplesForUnity package [4] for unity. A camera provided a third-person view of the on-site environment on the teleoperator's screen, and we used Vuforia AR plugin [2] for calibration to align the virtual robot with the real room. The teleoperator triggered the robot's navigation animations via keyboard inputs. Photon Unity Networking facilitated voice communication between the two systems, and Kinect data was streamed over LAN using the Kinect Unity package.

On-site User System The on-site system receives the RGBD and joint data, using an IK algorithm to convert the teleoperator's hand movements into robot arm motions. We used Oculus Integration plugin for Unity to develop the on-site user system for the seated participants to interact with the robotic simulation which ran on a Meta Quest Pro AR headset (106 degrees horizontal and 96 degrees vertical FOV). We used the Scene Understanding API to capture and create the experiment room geometry for the simulation and the eye-tracking SDK to capture the gaze data. We drive the robot's arm movements using the teleoperator's hand pose data using Cyclic Coordinate Descent algorithm [57], using the FinalIK package's [1] implementation and constrained by the robot's kinematic limits [3] simulated using the Rotation Limit component provided by the package. Simultaneously, the AR avatar display superimposes the teleoperator's point-cloud avatar onto the robot in real-time employing a simplistic pipeline explained in a previous study [103] consisting of the acquisition of RGBD data as a point cloud view of the sensor's field of view, segmentation to filter out the teleoperator and remapping to the headset's coordinate system. We ensured visual consistency through pre-calibration by timestamping and buffering the RGBD and skeletal frames for precise motion alignment. In the 3D game engine simulation, the avatar is matched to the robot's pose using the engine's transform hierarchy. In real-world scenarios, external trackers or image markers track the robot's pose to align the avatar. Kinect's RGBD data was used to display the teleoperator's facial video on the robot's torso tablet in the Video condition. The headset's speaker and microphone facilitated communication with the teleoperator.

While precise quantitative analysis of algorithm parameterization on system performance (e.g., avatar arm speed, positioning accuracy, or synchronization) is beyond the scope of this study, we implemented strategies to minimize errors in the system. The IK algorithm, mapping the teleoperator's movements to the robot's arm, was carefully parameterized to ensure smooth and accurate motion within the robot's kinematic constraints. We applied higher weights to bones near the end of the IK chain, which improved



(a) The participant setup in our lab

(b) The participant's point of view as seen on their headset

(c) The Teleoperator Setup

(d) The Teleoperator's Point of View as seen on their screen

Figure 4: Our study setup for the on-site user a) participant setup to be seated in our lab wearing headset, b) the participant's point of view of the teleoperator in AvatARoid condition as seen from the headset, c) the teleoperator setup in a different room, and d) the teleoperator's point of view of the participant as seen in their screen.

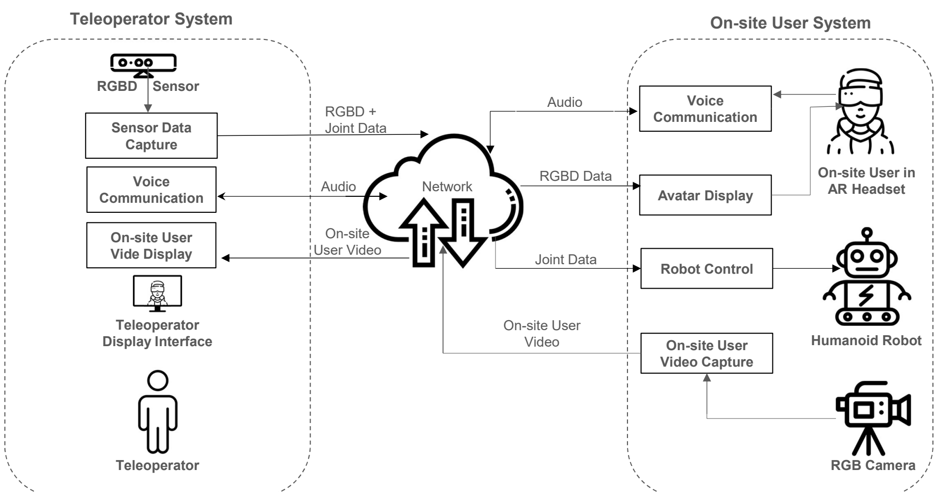


Figure 5: System architecture of the experimental AvatARoid system, software modules shown in a rectangle: On the left, the teleoperator system consists of modules to capture and stream sensor data, enable voice communication, and display the on-site user video. In the middle, the network enables data communication between the two users. On the right, the on-site user system consists of modules to create and display the point cloud avatar, a robot control system to send motion commands to the robot, and a voice and video capture module for audio-visual communication from the on-site user.

hand position accuracy. Additionally, we configured the IK algorithm to complete four iterations with a tolerance threshold near zero for both the robot and avatar hand positions. Input errors from the Azure Kinect depth camera, such as joint tracking noise or occlusions that caused jitter and jumps in robotic hand movements, were mitigated using filtering the joint positions. While this introduced slight delays, they ensured smooth motion alignment between the teleoperator and the avatar. These delays, along with those from data acquisition and transmission, remained imperceptible due to the use of a virtual robot in a low-latency LAN setup. Rotation limits, although slightly affecting accuracy, were essential for maintaining physically plausible movements, for instance, to avoid collision of the robot's hand and torso when the teleoperator's hand approached their torso.

4.3 Procedure

The overview of the procedure is shown in Figure 6. Each experiment session took about one hour. Participants signed the consent form and completed a demographic survey before the study (see Appendix A.4). Our procedure comprised the following five steps:

Step 1 - Greeting and Orientation. Upon arriving at the venue, the experimenter escorted the participants to the study room and oriented them about the study's objectives and procedure.

Step 2 - Playing the 1st Investment Game and Entering the Simulation. Participants began with the investment game, a variant of the trust game [10], designed to measure interpersonal trust. The experimenter explained the rules of the investment game to the participants. Participants started with 5 dollars and could invest between 0 and 5 in increments of 1 dollar. They were told the invested amount would be doubled and passed to the teleoperators, who then chose how much to return. The participant was then handed the money and two envelopes, each corresponding to a decision

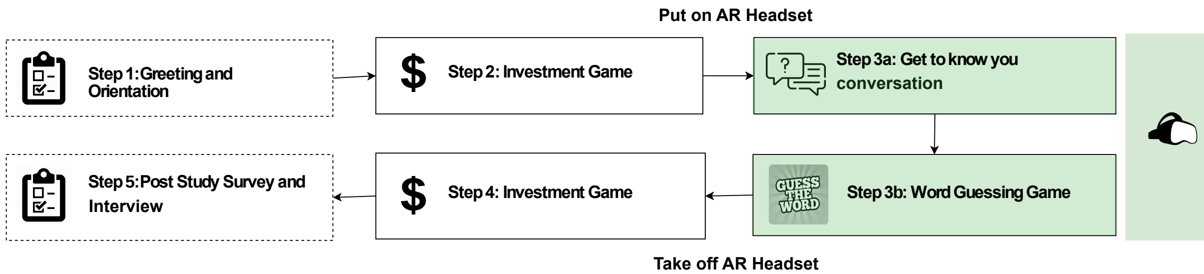


Figure 6: Overview of the study procedure session

to either keep or invest. To incentivize the participants, they were initially told that their participation would yield a 30-dollar compensation, to which earnings in two rounds of this game would be added. However, at the end of the study, everyone received an additional 20 dollars regardless of their game decisions. After they handed their envelopes to the experimenter, the participant was briefed about the headset, its adjustment features, and necessary safety precautions. The experimenter remotely launched the simulation application once the participants were ready and wearing the headset. The teleoperator confederate initiated the robot's entry animation once their system signaled that the participant's system was ready.

Step 3 - Interaction Tasks. The participants then performed the conversational tasks modeling a robot-mediated telepresence interaction in the simulated experimental system. Throughout these tasks, the lead investigator monitored the task's progress on a mobile device connected to the headset showing the participant's POV, and observed participants' behaviors and interactions with the teleoperators.

- *Step 3a: "Get to Know You" Conversation Task* Next, the experimenter provided instructions for the "get to know you" task. This was a casual structured conversation in which participants answered the questions from a predetermined set and asked them back to the confederate. The confederate's responses were loosely scripted to not bias the participants' impression of them. We followed a turn-taking approach for an entire set to avoid participants mirroring the confederate's scripted disclosure of the questions, as disclosure in this conversation was one of our behavioral measures. The questions were displayed on a monitor screen for the teleoperator confederate and in a virtual text panel on the room wall for the participants. Participants used the headset controller to navigate the questions. The experimenter then left the room, and the confederate initiated the task. After completion of the task, the experimenter entered the room to explain the second task.
- *Step 3b: Collaborative Word Guessing Game Task* Participants teamed up with the confederate for a word-guessing game. The confederate described words shown to them on their screen, and participants attempted to guess as many words as possible within 3 minutes. This task modeled collaborative activities common in telepresence settings. After their final greetings, the confederate activated the robot's exit animation sequence.

Step 4. Playing the 2nd Investment Game. Participants were then advised to remove the AR headset and were invited to play the investment game once more. This task aimed to measure the behavioral measure of trust the participants built in the teleoperator using the difference in investment in two rounds of the game.

Step 5. Post-task Survey and Interview. Then, the participants completed a survey with questionnaires for our measures. We also conducted in-depth interviews to gather detailed qualitative data on participants' experiences and perceptions.

At the end of the session, participants were debriefed on the study's deception protocol, including the role of the confederate and the fictitious nature of the trust game's decision on payment implications. The session concluded with participants receiving their due compensation.

4.4 Measures

We identified the following measures to evaluate the effectiveness of user representation in AvatARoid. This evaluation focused on the on-site users' perceptions of the teleoperator's embodiment, social presence, and overall impression.

4.4.1 Subjective Measures. We used validated questionnaires to assess participants' perceptions of teleoperator embodiment, social presence, overall impression, and co-presence with the virtual robot, as follows:

- **Perceived Embodiment**: Adapted from the *Self-Avatar Embodiment Questionnaire* [45], this six-item Likert-scale measure assessed perceptions of embodiment, such as colocation, control, and agency.
- **Social Presence**: Measured using a six-item semantic differential scale adapted from the *Social Presence Questionnaire* [88], evaluating the teleoperator's perceived presence as a social entity (Appendix A.6.2).
- **Overall Impression**:
  - *Trustworthiness*: The 14-item bipolar adjective scale from the *Interpersonal Trust Scale* [134].
  - *Likeability*: The 11-item Likert-scale *Reysen Likeability Scale* [101].
  - *Perceived Similarity*: A nine-item semantic differential adapted from *Homophily Scale* [85].

- *Infrahumanization*: An emotion-selection task from Rae et al. [98] measured human emotion attributions.
- **Co-presence with the Robot**: A seven-item measure adapted from the *Co-presence Questionnaire* [64] evaluated spatial and perceptual co-presence with the robot.

Each questionnaire used a 7-point scale, with scores averaged to produce indices for each construct. Full item details are in Appendix A.6.

4.4.2 Behavorial Measures. In addition to subjective measures, our evaluation included behavioral metrics:

Gaze Duration: Utilizing the headset's eye-tracking hardware and game engine's collision system, we recorded gaze duration on different parts of the robot and teleoperator representation as a measure for visual attention [6, 135].

Word Count: Drawing inspiration from previous studies [98], we assessed participants' response length in the "get-to-know-you" task as a behavioral measure reflecting trust [134] and likeability [31].

Gain from the Investment Game: We used participants' investments in the first and second rounds of the game to quantify trust, with the difference indicating trust gained during interaction with the partner. This approach aligns with established methods in prior study [99].

4.5 Analysis

For a holistic view of our measures of interest, we integrate the quantitative data from questionnaire responses, behavioral measures, and qualitative data from interview responses. We present our findings based on analysis and triangulation of the multimodal data collected during our study, including participants' behaviors and language, where the patterns were consistently observed.

Questionnaire Response. We used Cronbach's alpha for reliability measurement and ANOVA to compare the effects of user representations on our measures, followed by TukeyHSD post-hoc pairwise comparisons. When significant, we applied ordinal logistic regression to analyze individual questionnaire items, identifying the effect of different representations on specific items within the measure. Bar plots illustrate ANOVA comparisons of means, while box plots highlight item-level variations, emphasizing response frequency or likelihood. For co-presence, we conducted descriptive analysis and visualized the frequency distribution of co-presence scores across participants.

Behavioral Measures. We used ANOVA followed by Tukey HSD post-hoc comparisons for our behavioral measures of word count and gain in investment. We used descriptive analysis of gaze data to supplement our interview and survey data results.

Interview Response. In our mixed-methods study, qualitative analysis was used to explicate, broaden, and complement the quantitative findings. We conducted a deductive thematic analysis following Braun and Clarke's framework [23], focusing on embodiment, social presence, and teleoperator impressions. Transcribed interviews were systematically coded, with the lead investigator mapping participant's statements to these constructs. Through iterative team discussions and the use of digital mind-mapping tools such as the Mira board, codes were refined into higher-order themes. These themes were further contextualized within the broader results, providing a cohesive interpretation of both significant and non-significant quantitative findings. This approach allowed us to triangulate data, deepen our analysis, and reveal nuanced insights into how different user representations influenced participants' perceptions and behaviors.

5 Findings

Our evaluation of AvatARoid using the proposed simulation approach reveals its effectiveness in assessing robot-mediated telepresence system designs. Most participants (39 out of 48) rated the robot's co-presence highly, with an average score of 4.87 on a seven-point Likert scale (see A.3 for details on participant feedback regarding the simulation). In post-study interviews, participants highlighted the act of a "robot coming in the room," and the robot being "visually embedded in the room" made them feel that the "interaction was happening in this room." These evaluation results pave the way for assessing the other measures discussed below.

5.1 AvatARoid Improved the Embodiment of the Teleoperator for On-site Users.

A one-way ANOVA revealed a significant difference in perceived embodiment levels among the three conditions, $F(2, 45) = 4.03$, $p = .025$ (see Figure 7). The effect size, η_p^2, was 0.15, indicating a large effect. Post hoc comparisons using Tukey's HSD indicated that the mean score for the AvatARoid condition ($M = 5.07$, $SD = 0.68$) was significantly higher than that of the Baseline condition ($M = 4.12$, $SD = 1.00$), $p = .020$. However, the Video condition ($M = 4.45$, $SD = 1.16$) did not significantly differ from the Baseline condition, $p = .174$.

An item-wise questionnaire analysis determined specific aspects of embodiment where the AvatARoid condition excelled. Figure 8 presents a box plot of responses for each item in the embodiment questionnaire across the three conditions. We used interview response to further deepen our understanding of the results, presented in Table 2.

5.1.1 The Colocation, Control, and Agency Perceptions in AvatARoid Contribute to a Stronger Sense of Embodiment. AvatARoid significantly improved participants' perceptions of colocation and control over the robot compared to the Baseline condition. These aspects contributed to a stronger sense of embodiment relative to the Baseline, although the Video condition showed mixed results.

Co-location. Participants were more likely to feel that the teleoperator and robot were co-located in the AvatARoid condition, significantly contributing to their increased sense of the teleoperator's embodiment. Ordinal logistic regression revealed a significant enhancement in co-location between the teleoperator and the robot for the AvatARoid condition compared to the Baseline, $b = 1.62$, $SE = 0.64$, $z = 2.52$, $p = .012$. The Video condition showed a non-significant increase, $b = 1.06$, $SE = 0.66$, $z = 1.60$, $p = .111$, compared to Baseline.

10 participants indicated that the AR avatar enhanced their sense of co-location between the robot and the teleoperator. This heightened spatial presence was evident in statements of the participants, like P22^{AR} and P29^{AR}, who described the interaction as "conversing

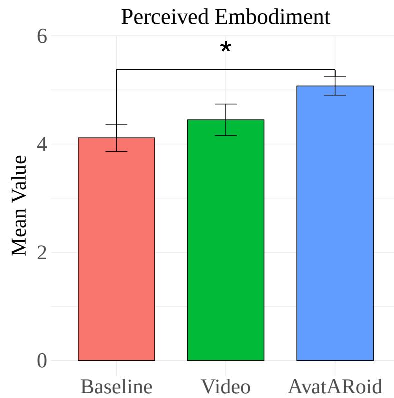


Figure 7: Mean embodiment score with standard deviation error bars across three user representations; (*) indicates a significant difference (p <.05).

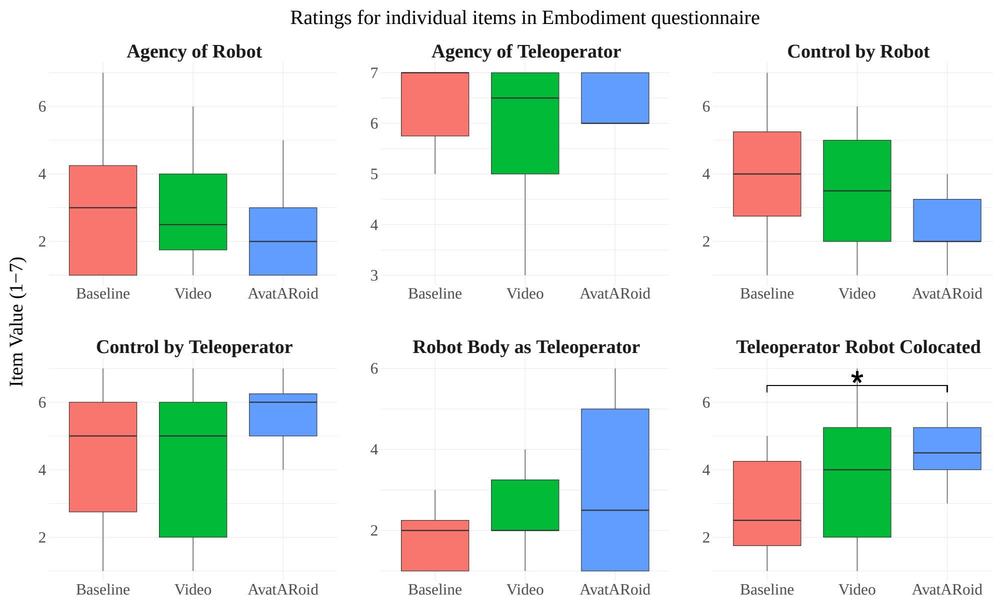


Figure 8: Item-wise rating scores of the embodiment questionnaire; An asterisk (*) indicates a significant difference based on ordinal logistic regression (p <.05).

face-to-face in the same room," a sentiment echoed by eight others in the AvatARoid condition.

AvatARoid's AR overlay and synchronized motion mapping integrated the teleoperator avatar with the humanoid robot, creating a sense of co-location. Four participants described the avatar and robot as a unified entity, with P27[AR] noting how the robot's hands "complemented" rendering artifacts like a "clipping arm" in the AR avatar. This sense of harmony was echoed by P14[AR], who stated, "The robot just felt like a part of Andy—whenever Andy moved, the robot did the same; they felt like one entity."

This sense of spatial presence was notably stronger in the AvatARoid condition compared to the Baseline, with the Video condition showing slight improvements but still lagging. During the interaction, participants in baseline condition expressed wondering, "Where is the actual person?". While acknowledging a slightly enhanced presence compared to standard video conferencing, participants in the Video condition still expressed feelings of disconnection. Specifically, nine participants (P3, P16, P21, P24, P28, P31, P39, P45, P47) in the Video condition noted that the teleoperator felt confined to the screen, leading them to feel they were "speaking to a person on the screen" and perceiving "the robot being the actual thing that was in the room, not the person."

Control. AvatARoid may improve the participants' perception of the teleoperator's control over robot movement and reduce perceived robot autonomy, although these findings are not significant

Theme	Code	Description	Participants (%)
Enhanced Sense of Teleoperator Embodiment	Robot-Teleoperator Co-location Perception	Participants reported the AR avatar made them feel as though the teleoperator was present in the same room, creating a heightened sense of spatial presence.	10 (62.5%)
	Robot Control Perception	AvatARoid's synchronized motion mapping seamlessly integrated the teleoperator's avatar motion with the robot.	14 (87.5%)
	Single Entity	Participants perceived the teleoperator avatar and robot as a unified entity.	4 (25%)
	Teleoperator-Centered View	Participants in the AvatARoid condition largely disregarded the robot, focusing instead on the teleoperator's presence, motion, and expression.	4 (25%)
Challenges in Teleoperator Perception in Video and Baseline	Ambiguous Control Source	Participants in the Baseline and Video condition experienced difficulty discerning whether the robot was controlled by the teleoperator or pre-programmed.	8 in Video (50%), 9 in Baseline (56.25%)
	Agency Confusion	Participants in the Video and Baseline conditions reported confusion regarding whether they were interacting with a human or a robot.	7 in Baseline (43.75%), 5 in Video (31.25%)
	Screen Confinement	Participants in the Video condition felt the teleoperator was confined to the screen, creating a diminished sense of spatial presence.	9 (56.25%)
Robot and Teleoperator as Separate Entity	Robot as Furniture/Icon	Many participants in the Video and Baseline conditions perceived the robot as an inanimate object not representing the teleoperator, rather likening it to "furniture where the screen is", "icons to look at", or "webcams in video calls".	11 in Video (68.75%), 8 in Baseline (50%)
	Robot face vs Teleoperator face	Participants in the video conditions expressed confusion about where to look at because of the "presence of two faces".	6 (37.5%)

Table 2: Themes from thematic analysis on AvatARoid's impact on embodiment

and warrant further investigation. Ordinal logistic regression analysis suggested that, compared to Baseline, the AvatARoid condition approached significance in predicting both a heightened perception of control by the teleoperator over the robot's movements, $b = 1.15$, $SE = 1.81$, $z = 0.36$, $p = .071$, and reduced perceived robot autonomy, $b = -1.16$, $SE = 0.64$, $z = -1.83$, $p = .067$. For the Video condition, no trend toward significant effects was found on either perceived robot autonomy, $b = -0.50$, $SE = 0.65$, $z = -0.77$, $p = .441$, or teleoperator control, $b = 0.23$, $SE = 0.64$, $z = 0.37$, $p = .716$.

Participants in the AvatARoid condition overwhelmingly perceived the robot as "directly controlled by" or "mimicking" the teleoperator, with 14 participants expressing this sentiment. P4^{AR} noted, "Watching the robot move its arm to follow the teleoperator's arm made the two feel connected," while P22^{AR} expressed, that seeing the "the robot arm just following (the teleoperator's arm)", made them think that "the robot was not the one controlling."

All 14 Participants attributed their perception of agency and control to the synchronized motion mapping between the overlay and the robot. The high level of coordination provided users with a strong sense that the robot was "replicating what the teleoperator was doing." As P27^{AR} explained:

> "I felt like it was a good one-to-one match. I don't think the robot was doing something that Andy wasn't. So it was just giving me a reinforcement of what Andy was gesturing to." (P27^{AR})

By reinforcing the teleoperator's gestures and consolidating the robot's movements with the avatar's, AvatARoid appeared to reduce ambiguity in agency perception compared to other conditions, as suggested by qualitative feedback. Eight participants in the Video condition and nine in the Baseline Condition were uncertain whether the robot was being controlled by the teleoperator or pre-programmed. P16^{VD} speculated that "the robot could have been following the teleoperator or programmed to move its arm automatically," while P46^{BL} even suggested that "someone else (the experimenter) could have controlled it".

Human Agency. Ordinal logistic regression revealed that the type of representation did not significantly influence participants' perceptions of human agency, neither in terms of feeling they interacted with a robot nor with a real person. For the perception of human agency, both the video condition, $b = -0.46$, $SE = 0.71$, $z = -0.65$, $p = .515$, and the AvatARoid condition, $b = -0.13$, $SE = 0.68$, $z = -0.20$, $p = .845$, showed no significant difference.

The test for robot agency also did not show any significant difference for AvatARoid, $b = -0.70$, $SE = 0.64$, $z = -1.10$, $p = .270$, or Video, $b = -0.10$, $SE = 0.65$, $z = -0.15$, $p = .878$, when compared to Baseline.

Although not supported by their survey response, interview responses of participants in the Video and Baseline conditions revealed their challenges in discerning the human agency, adding a layer of confusion that was notably absent in the AvatARoid condition. P14^{AR} mentioned they were suspicious initially that it might have been just a robot; AvatARoid's overlay appeared to help the participant resolve doubts about the presence of the teleoperator. Notably, 7 participants in the Baseline and 5 participants in the Video condition reported initial confusion about the agency of their communication partner and either employed strategies by "throwing a bait to confirm its human" or remained confused throughout the study as P47 in video condition said "70 percent of the time I was thinking it was an AI agent and 30 that maybe she's a real person". As this tendency was more common in Baseline and Video conditions, this confusion in perceived agency could have potentially influenced the sense of the teleoperator's embodiment.

> "I did (think I was talking to a robot). And what's interesting is the voice didn't really match the appearance.... if I had just heard the voice, I would probably think it's a human. But it was hard for me to get past the fact that she looked like a robot. So I felt like I was still talking to a robot." (P1^{BL})

5.1.2 The Dissonance of Perceiving the Robot and Teleoperator as Distinct Agents is Pronounced in the Baseline and Video Conditions, Potentially Diminishing the Sense of Embodiment. In examining participants' visual attention allocation towards the teleoperator embodiment, our descriptive analysis of gaze duration on the robot head—expressed as a percentage of the total interaction time—revealed the following pattern: participants spent the highest proportion of gaze duration on the robot head in the AvatARoid condition ($M = 37.76$, $SD = 14.90$), followed by Video ($M = 26.13$, $SD = 21.46$), and least in Baseline ($M = 17.83$, $SD = 13.50$). This was interesting as the participant's face was not displayed on the robot's head but on the tablet attached to its torso in the Video condition. Further examination of the gaze data for the Video condition captures this split: participants spent 15.80% ($SD = 16.18$) of their gaze duration on the teleoperator video and 26.13% ($SD = 21.40$) on the robot head. This showed that the participants perceived the robot and teleoperator as two distinct entities.

In post-study interviews, 11 participants in the Video condition reported viewing the robot as merely an inanimate object in the room, describing it as "just another piece of furniture" or "a wall", serving only as "a support for the screen displaying the teleoperator." This highlighted their dual perception of the robot and teleoperator as separate entities. Similarly, eight Baseline participants likened the robot to "icons" or "webcams" during video calls. P11^{BL} stated the inexpressiveness of the robot making the disconnect between the teleoperator apparent when "Sometimes Andy (the teleoperator) will be very lively", and "the robot will be just like (poker face)" which made them feel like "two different things going on here". Most participants in the AvatARoid condition did not report confusion

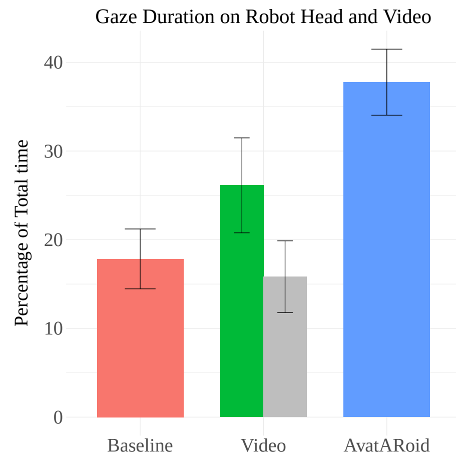
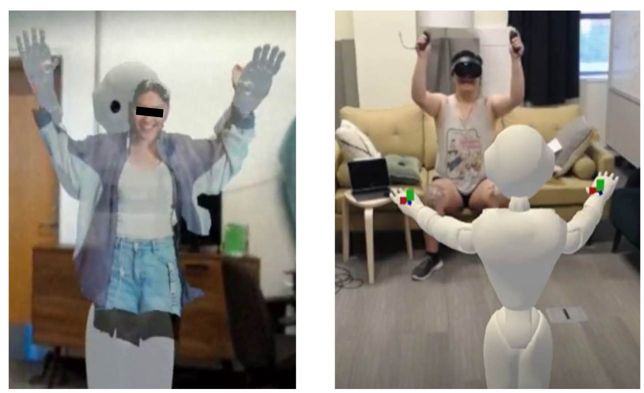


Figure 9: Mean gaze duration with standard deviation error bars, expressed as a percentage of the total time, reflects the time participants spent looking at the robot's head across all three conditions or at the teleoperator video in the Video condition.



(a) The On-site User's View of the Teleoperator

(b) The Teleoperator's View of the On-site User

Figure 10: A snapshot from our study session: Participant and Teleoperator mirror celebratory gesture after completing word guessing game a) The teleoperator as seen in participant headset point of view, b) The participant as seen in teleoperator screen

due to the robot's face, with only one participant expressing such uncertainty.

5.2 AvatARoid Shows Potential to Increase Feelings of Social Presence

The ANOVA test comparing social presence for AvatARoid ($M = 5.06$, $SD = 0.91$), Video ($M = 4.94$, $SD = 1.14$), and Baseline ($M = 4.44$, $SD = 1.22$) did not show any significant effect, $F(2, 45) = 1.45$, $p = 0.246$. The effect size η_p^2 was 0.06, indicating a small effect. We delved deeper into the interview reports to investigate the impact of different visual representations on social presence. We uncovered insights into AvatARoid's potential to enhance social presence, presented in table 3.

5.2.1 Participants Found AvatARoid to be Personable and Intimate. Participants' responses highlighted the potential of AvatARoid to foster emotional connections and offer an engaging, personable experience. 10 participants recognized AvatARoid as an interactive and emotionally resonant platform that enhanced presence and connection. Its ability to convey real-time facial expressions and upper-body gestures was crucial in fostering emotional connections. Many highlighted how reciprocating actions and mirroring gestures facilitated shared experiences, as $P4^{AR}$ described the teleoperator reciprocating their gestures of "raising hands, as hurray when we finished the game" made them feel like "we are celebrating together."

The integration of the avatar and the robot allowed participants to interpret the teleoperator's reactions more effectively. Facial expressions and arm movements acted as grounding mechanisms, enabling participants to discern emotions and gauge responses in real-time, which many felt made interactions "nearly akin to in-person conversations." For example, $P29^{AR}$ described how "mimicked eye-to-eye connections" reinforced their sense of intimacy. AvatARoid's gestural communication added depth and weight to interactions, making participants feel their conversations were impactful and authentic. $P27^{AR}$ summarized this sentiment, saying that the combination of the avatar and the robot enhanced the "sense of presence" and the "impact of responses," making it the "closest thing you can get other than being actually in front of them."

Three participants envisioned AvatARoid's potential in personal and emotionally charged scenarios, such as family gatherings, long-distance relationships, and intimate conversations across borders. $P27^{AR}$ and $P35^{AR}$ highlighted its role in enhancing close family interactions, while $P48^{AR}$ saw it as particularly valuable for situations requiring a deeper sense of presence. $P4^{AR}$ noted that they could envision physical interactions, such as hugs, as part of AvatARoid's immersive potential.

> "I felt it was much more interactive. For example, I felt like I could stand up and maybe hug her; I may have been able to interact with the robot." ($P4^{AR}$)

5.2.2 Video and Baseline Conditions Lack the Emotional and Interactive Depth Achieved with AvatARoid. While AvatARoid was perceived as providing a richer interactive experience, some participants in the Video and Baseline conditions also described added aspects of presence, though without the same depth of non-verbal engagement. Some even reported feeling notably detached and limited in their interactive capabilities. Eight participants in the Baseline condition often felt the lack of emotional or facial expression indicators made the robot appear mechanical rather than an extension of the teleoperator as $P11^{BL}$ mentioned the robot having no "visual indication of the emotions the other participant was expressing" playing a big part in making them feel like "it was just a robot standing than a good representation of the other participant."

Seven Participants in the Video condition found the experience "more fun" and "more present" though they did not always feel it was more interactive compared to traditional video calls. Four participants mentioned that the teleoperator seemed more present and expressive, with the robot's arm movements adding non-verbal gestures to a typical Zoom call. Still, this sentiment was mixed as three participants mentioned that the robot's limited motion did not represent the teleoperator's actual gestures.

5.3 No significant difference was found in subjective and behavioral measures of the teleoperator's impression across conditions.

ANOVA tests revealed that the user representations did not significantly affect participants' affective perceptions of the teleoperator. Specifically, the test results for different impression measures were for: Trustworthiness, $F(2, 45) = 0.22$, $p = 0.806$, $\eta_p^2 = 0.01$; Likeability, $F(2, 45) = 0.19$, $p = 0.828$, $\eta_p^2 = 0.01$; Homophily, $F(2, 45) = 0.39$, $p = 0.680$, $\eta_p^2 = 0.017$; Infrahumanization, $F(2, 45) = 0.70$, $p = 0.505$, $\eta_p^2 = 0.03$; Gain in Investment Game, $F(2, 45) = 0.76$, $p = 0.476$, $\eta_p^2 = 0.03$; and Word Count in Disclosure Interview, $F(2, 45) = 0.15$, $p = 0.865$, $\eta_p^2 = 0.01$.

The ratings for Trustworthiness (for AvatARoid $M = 6.01$, $SD = 0.47$, for Video $M = 5.87$, $SD = 0.76$, for Baseline $M = 5.92$, $SD = 0.68$) and Likeability (for AvatARoid $M = 5.69$, $SD = 0.38$, for Video $M = 5.61$, $SD = 0.89$, for Baseline $M = 5.55$, $SD = 0.68$) were high for all three visual representations. Most participants indicated that their ratings were based on the verbal interaction with the teleoperator and the teleoperator's demeanor and stated they "seemed helpful in the word-guessing game." For perceived similarity (for AvatARoid $M = 4.29$, $SD = 0.84$, for Video $M = 4.08$, $SD = 0.82$, for Baseline $M = 4.32$, $SD = 0.82$), participants reported mostly rating based on the conversation and only weakly guided by appearance when visual cues were present. For infrahumanization (for AvatARoid $M = 3.5$, $SD = 0.73$, for Video $M = 3.44$, $SD = 0.63$, and Baseline $M = 3.18$, $SD = 0.99$), participants generally assigned positively valenced emotions to the teleoperator, regardless of whether the emotion was uniquely human.

5.4 AvatARoid's room of improvement

We identified several areas for improvement in AvatARoid, summarized in Table 4, encompassing refining the avatar for full-body representation, enhancing body language cues through improved motion mapping, and addressing technical limitations such as latency and device resolution.

Five Participants expressed a desire for a full-body avatar, suggesting that the upper-torso-only design limited the sense of the teleoperator's presence. The absence of a complete body created a sense of disjointedness for some participants, impacting their perception of a realistic, face-to-face interaction. Finer details were also called for, particularly articulating smaller body parts like fingers. Four participants mentioned awareness of missing body parts and visual details, such as "clipping" or "transparency" of the avatar, were reported as hindrances to engaging in conversation.

Theme	Code	Description	Participants (%)
Enhanced Emotional Connection and Physical Interaction	Interactive, Intimate Experience	Participants widely acknowledged AvatARoid as fostering emotional resonance and deeper connections through real-time facial expressions, upper-body gestures, and mirroring actions.	5 (31.25%)
	Intimate Use Cases	Participants envisioned AvatARoid's potential in emotionally charged personal scenarios, such as family gatherings and long-distance relationships, where the feeling of presence was crucial.	3 (18.75%)
	Tactile Presence	Participants discussed the potential for physical interactions, such as hugs, and how these could enhance the emotional engagement and interactivity of the experience.	2 (12.5%)
Shared Gestural Feedback and Grounding Through Non-Verbal Cues	Reciprocated Actions	AvatARoid allowed participants to share gestures (e.g., celebratory movements), fostering a sense of connection and togetherness.	3 (18.75%)
	Emotional Interpretation	The integration of the avatar and robot enabled participants to interpret emotions through non-verbal cues, such as facial expressions and arm movements.	3 (18.75%)
Comparison with Existing Video Conferencing Solutions	"Zoom with robot"	While participants in the Video condition considered the experience more "fun" and "present" than zoom, it did not make it more interactive.	4 (25%)
	Limited Non-Verbal Engagement	The robot's limited movements did not serve as a reflective or emotionally responsive medium for the teleoperator's expressions.	3 (18.75%)
Disconnection in Baseline Condition	Mechanical Perception	Baseline participants felt the robot lacked visual indicators for emotions, reducing it to a mechanical proxy that failed to represent the teleoperator's emotional state or presence effectively.	5 (31.25%)
	Lack of Emotional Connection	Participants in the Baseline condition reported a lack of emotional connection, as the robot's limited expressiveness made it feel detached and mechanical.	3 (18.75%)

Table 3: Themes from thematic analysis on AvatARoid's impact on emotional connection and social presence

Eight Participants identified motion mapping as an important area for improvement, particularly for achieving natural and dynamic gestures that align with speech. Four Participants noted concerns about the synchrony of the teleoperator's spoken words and the robot's bodily movements, disrupting the natural flow of communication. Four participants noted stiffness and restricted mobility in the robot's physical articulations, which affected the perception of the experience as less dynamic or immersive. Natural and dynamic movements in both the robot and avatar proved essential for users to effectively interpret body language and feel a higher social presence.

Six Participants identified technical limitations such as latency and device resolution as barriers to achieving a seamless, natural interaction. Three participants cited latency as a barrier to achieving the natural flow of interaction and reported that latency made real-time reaction assessment challenging, requiring more effort to interpret cues. Technical limitations of the AR device, issues like the grainy passthrough AR effect, and the headset's form factor constantly reminded three participants of the "virtuality of interaction," preventing it from feeling face-to-face. Addressing these technical challenges could significantly enhance the sense of immediacy and presence in AvatARoid.

6 Discussion

6.1 Bridging the Embodiment Gap: Unmixing Metaphors?

AvatARoid addresses the *mixing metaphor* problem in robotic telepresence by amplifying the teleoperator's spatial presence and reinforcing the perception that they co-occupy the robot's location as its controlling agent [122]. This heightened presence may help mitigate proxemic violations and support the establishment of social norms [22, 122]. By overlaying AR avatars onto humanoid robots, AvatARoid makes teleoperator gestures more tangible and

Theme	Code	Description	Participants (%)
Improved Avatar Rendering	Full-Body Representation	Participants expressed dissatisfaction with the avatar's limited upper-torso representation, stating it hindered a complete sense of presence. Missing body parts like legs and fingers created an incomplete and less realistic experience.	5 (31.25%)
	Visual Details and Articulation	The avatar's inability to articulate smaller body parts (e.g., fingers) and issues like "clipping" or "transparency" distracted participants from engaging in natural interactions.	4 (25%)
Improved Motion Alignment	Delay in Speech and Gestures	Participants identified a lack of synchrony between the teleoperator's spoken words and the robot's bodily movements, which disrupted the flow of communication. Stiff, unnatural gestures impacted the interpretation of body language and presence.	4 (25%)
	Natural Movements	The robotic system's restricted mobility and unnatural movements were described as contributing to a "virtual" and "immobile" experience, reducing social presence.	4 (25%)
Technical Issues	Speech Latency Issues	Slight Latency was noted as a barrier to achieving natural flow in interactions as if face to face, making it difficult for participants to assess and react to cues in real time.	3 (18.75%)
	Device Resolution and Form Factor	Issues with device resolution (e.g., grainy passthrough AR effect) and the bulky form factor of the headset consistently reminded participants of the virtual nature of the interaction, disrupting immersion.	3 (18.75%)

Table 4: Themes from thematic analysis on AvatARoid's room for improvement

impactful. Synchronizing the avatar's movements with the robot's actions clarifies the teleoperator's influence, reflecting prior findings on consistency in signaling [115]. Building on evidence that AR enhances expressiveness in telepresence [55] and that humanoid robots foster more physical interaction [18], AvatARoid aligns the teleoperator's intentions with the robot's behavior. In doing so, it reinforces the notion that the teleoperator is physically 'right here,' actively controlling the robot, clarifying their role and fostering more harmonious interaction [55, 99, 117]. Moving beyond simply synchronizing screen-based and physical domains, AvatARoid fully integrates the avatar into the local environment, addressing a longstanding gap in teleoperator embodiment.

6.2 From Simulated to Physical Robot Implementations: Applicability, Opportunities, and Challenges

Although our study was conducted within an immersive AR simulation using the AvatARoid system, the core mechanism of superimposing a motion-mapped AR avatar onto a humanoid robot holds significant potential for physical robot implementations. This approach can be translated to real-world settings by employing AR headsets or displays to overlay the teleoperator's avatar onto a physical robot, enhancing the perception of teleoperator embodiment and control. Real robots could further introduce additional sensory cues, such as haptic feedback and tangible spatial presence, which may amplify social presence and enrich telepresence interactions. While the direct application of our findings to physical robots necessitates careful consideration of technical and practical challenges, the insights gained from our simulation suggest opportunities for improving telepresence systems and addressing the embodiment gap identified in prior research [9, 89, 90].

Further research is required to validate the effectiveness of AvatARoid in real-world settings that face challenges, such as hardware limitations, environmental constraints, and safety considerations. Technical challenges can affect the integration of digital avatars with physical robots. For instance, latency may cause discrepancies between the avatar's motion and the robot's physical motion, while limitations in speed and motion dexterity could exacerbate the "uncanny valley" effect, especially when digital and physical elements fail to merge seamlessly [33, 55, 128]. Although our simulation accounted for factors like the robot's limited speed and range, it did not capture all real-world dynamics, and detailed analysis of algorithm parameterization on these issues. Transitioning to physical robots might also introduce challenges related to the perception of agency. As physicality increases, participants may become more aware of the robot's autonomy, complicating the reconciliation of the teleoperator's control with the robot's independent actions. This issue aligns with previous research, which shows that heightened physical presence often complicates perceptions of agency and control [24].

The impact of social presence in AR, when there is a real robot in the actual scene, is a critical consideration for the physical implementations of AvatARoid. In our study, the AR overlay of the teleoperator's avatar onto the robot significantly enhanced participants' sense of co-location, suggesting that AR can effectively augment physical robots to create richer social interactions. In real-world scenarios, the combination of a tangible robot and an AR avatar could potentially amplify these effects, as the physical presence of the robot adds tangible and haptic cues absent in simulations [34, 116]. However, the integration of AR with physical robots may also present other challenges, such as dealing with variable lighting conditions and addressing user discomfort with wearing AR devices for extended periods providing crucial areas for further investigation.

While our simulated robot demonstrated promising results applicable to use with a real robot, the transition to real robots presents both opportunities for improvement and the need for careful design considerations to address these challenges. Future work should include more detailed performance evaluations to optimize the user experience and ensure the successful deployment of AvatARoid in physical robot systems.

6.3 Role of Visual Representations in Social Interaction in Robot-mediated Telepresence

Our findings indicate that while co-location and control perception of embodiment differed significantly between the Baseline and AvatARoid conditions, no significant differences were found between the Video and AvatARoid conditions. However, the absence of significant differences between the Video and Baseline conditions suggests that there was some, albeit not statistically significant, improvement in user experience in the AvatARoid condition over the Video condition. This points to the possibility that, while the AvatARoid representation provided a more dynamic form of interaction, its visual representation did not significantly enhance the other dimensions of social interaction, such as agency, social presence, and emotions, compared to the Video condition.

Interestingly, no significant differences were found across these other variables, despite interview responses offering some support for the notion that the AvatARoid system improved emotional connection and engagement during interactions. This discrepancy between subjective reports and quantitative or behavioral measures suggests that the visual representation of the teleoperator while enhancing co-location and embodiment, did not yield distinct effects in other aspects of social interaction. This highlights the complex and multifaceted nature of social presence and emotional engagement [9, 89], which may not be easily captured through the metrics used in this study.

We identified user representation as a potential factor influencing embodiment, but it is important to recognize that interpersonal interaction and collaboration outcomes also depend on several other factors. These include the nature of the task at hand [6, 51, 118], the relational dynamics between participants [17], and information asymmetries that may affect how users perceive and interact with each other in telepresence settings [117]. In our study, we focused primarily on the visual representation of the teleoperator, which may have limited the impact of these other factors on users' social presence and perceptions of personality.

Future work should take a more holistic, multi-dimensional approach to exploring how these factors interact in social interaction and impression formation within the AvatARoid system. Additionally, it would be valuable to investigate how different avatar configurations, as well as multimodal sensory inputs, could influence these variables, potentially incorporating more dynamic elements or varying levels of interactivity. By expanding the scope of our investigations, we can better understand how avatar representations interact with task characteristics and interpersonal dynamics to shape user experiences in telepresence settings.

6.4 AvatARoid: Design Implications and Alternative Approaches

Our AvatARoid implementation addresses spatial disconnection and agency ambiguity but introduces challenges that influence social engagement and embodiment. Full-body avatars can significantly enhance telepresence by facilitating better interpretation of body language and supporting nuanced non-verbal communication [12, 89, 92]. High-fidelity avatars and technologies like Holoportation [78, 90] further enable embodiment and social presence through the integration of articulated fingers and fine gesture replication, enriching telepresence experiences.

Motion mapping also plays a crucial role in achieving teleoperator embodiment and presence. Our current inverse kinematics (IK) approach tracks the joint positions of the teleoperator's hands, which limits motion fidelity. Advanced teleoperation methods, such as those proposed by Girard et al. [44], Arduengo et al. [11], and Darvish et al. [33], could enhance synchronization, improving both the teleoperator's control and the perception of their actions during telepresence interactions.

The robotic platform's design is another critical factor shaping telepresence systems. Anthropomorphic designs, like Pepper's facial features, enhance relatability but may create a perceptual divide between the robot and the AR avatar. A *Minimal Functional Human Design*, which incorporates a humanoid robotic structure with minimal identity features and a dynamic AR "skin," could unify the telepresence experience while reducing this divide [69, 119]. Additionally, mismatches in action replication can lead to an "uncanny valley" effect [55], emphasizing the need for advanced humanoid teleoperation technologies to improve motion fidelity and synchronization, thereby enhancing embodiment and social presence.

Exploring alternative implementations addressing these improvements in the design elements can further refine telepresence systems. Evaluating intermediary avatars, such as stationary or minimally dynamic representations, could isolate the impact of motion on embodiment. Incorporating bipedal humanoid robots or full-body AR avatars with motion-mapped lower limbs offers the potential to improve embodiment and presence. Hybrid approaches that combine direct teleoperation with predictive algorithms could address latency issues, while testing varying levels of anthropomorphism—from minimal to fully human-like designs—would help identify an optimal balance between familiarity and performance. These refinements underscore the importance of iterative design and comprehensive evaluation in diverse real-world contexts to enhance embodiment, social presence, and usability.

6.5 Limitations and Future Works

Our AR simulation framework provided a high sense of co-presence, but participants noted disruptions from the AR headset's form factor and visual quality, including limited Field of View, resolution issues, a "grainy" passthrough display, and awareness of the headset. Addressing these through improved AR hardware design and fidelity is essential for a more seamless sense of presence. Future simulations could also enhance realism by incorporating technologies like thermochromic ink for dynamic color changes [108], mechanical actuators for real-world movements [72, 120], and human actuators to bridge the virtual-physical gap [28], optimizing AR efficacy for telepresence and HRI applications. Immersive simulation tools are crucial for testing concepts in impractical or hazardous real-world settings but have inherent limitations such as perceived robot identity of the virtual robot, lack of physicality, and user expectation and bias creating differences while interacting with virtual and real robots [82]. Future studies with real robots are necessary to validate and expand these findings. Our study also focused only on the on-site user's experience, excluding the teleoperator's perspective; future research should integrate both sides of the interaction for comprehensive evaluation. Moreover, we did not include the teleoperator interface, overall evaluation with emerging telepresence systems [55] could provide valuable insights.

7 Conclusion

In this paper, we introduced AvatARoid, a novel telepresence system that overlays the teleoperator's avatar on a robotic embodiment and synchronizes their movements. We evaluated the system using a lab study focused on AvatARoid's core mechanisms, including AR avatar superimposition, motion alignment, and their impact on on-site users' perception of the teleoperator's embodiment, social presence, and overall impression— factors essential for high-quality social interaction. For our evaluation, we implemented AvatARoid with a virtual robot using a simulation framework we developed, employing mechanisms to make the robot appear realistic and present in the user's actual environment. AvatARoid shows promise in effectively enhancing the on-site users' perception of unity between the teleoperator's representation and the robotic embodiment, making them feel as if the teleoperator was indeed "located where the robot was" and "in control of the robot". Participants also felt that AvatARoid could serve as an emotionally rich and engaging medium for communication. However, they recommended refinements like displaying the teleoperator's entire body and nuanced gestures for a more authentic sense of presence and body language. Future studies can examine AvatARoid's performance across diverse tasks and application scenarios, consider its impact when combined with the constraints of an actual robot, and expand the evaluation to include both teleoperators and on-site users. As technological advancements continue in the design of humanoid robots, their teleoperation, and mixed reality, AvatARoid offers a foundational step to explore how, through a synergy of these technologies, we could enable telepresence systems with possibilities beyond being there [50], with an ultimate goal of enriching human connections.

Acknowledgments

This research was supported by the Korea Institute of Science and Technology (KIST) institutional program (2E33613, 2E33612). We sincerely appreciate their support, which made this work possible.

References

[1] [n. d.]. Final IK | Animation Tools | Unity Asset Store. https://assetstore.unity.com/packages/tools/animation/final-ik-14290
[2] [n. d.]. Home | Engine Developer Portal. https://developer.vuforia.com/
[3] [n. d.]. Joints Aldebaran 2.5.11.14a documentation. Retrieved 2023-10-13 from http://doc.aldebaran.com/2-5/family/pepper_technical/joints_pep.html
[4] 2019. Azure Kinect Examples for Unity. https://rfilkov.com/2019/07/24/azure-kinect-examples-for-unity/
[5] Safurah Binti Abdul Jalil, Brabbyn Osburn, Jingwen Huang, Mike Barley, Marin Markovich, and Robert Amor. 2012. Avatars at a meeting. In *Proceedings of the 13th International Conference of the NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction (CHINZ '12)*. Association for Computing Machinery, New York, NY, USA, 84–87. https://doi.org/10.1145/2379256.2379270
[6] Ahsan Abdullah, Jan Kolkmeier, Vivian Lo, and Michael Neff. 2021. Videoconference and Embodied VR: Communication Patterns Across Task and Medium. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–29. https://doi.org/10.1145/3479597
[7] Sigurdur Orn Adalgeirsson and Cynthia Breazeal. 2010. MeBot: A robotic platform for socially embodied telepresence. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 15–22. https://doi.org/10.1109/HRI.2010.5453272 ISSN: 2167-2148.
[8] Insaf Ajili, Malik Mallem, and Jean-Yves Didier. 2017. Gesture recognition for humanoid robot teleoperation. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 1115–1120. https://doi.org/10.1109/ROMAN.2017.8172443 ISSN: 1944-9437.
[9] Luis Almeida, Paulo Menezes, and Jorge Dias. 2022. Telepresence Social Robotics towards Co-Presence: A Review. *Applied Sciences* 12, 11 (Jan. 2022), 5557. https://doi.org/10.3390/app12115557 Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
[10] Carlos Alós-Ferrer and Federica Farolfi. 2019. Trust Games and Beyond. *Frontiers in Neuroscience* 13 (2019). https://www.frontiersin.org/articles/10.3389/fnins.2019.00887
[11] Miguel Arduengo, Ana Arduengo, Adrià Colomé, Joan Lobo-Prat, and Carme Torras. 2021. Human to Robot Whole-Body Motion Transfer. In *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*. 299–305. https://doi.org/10.1109/HUMANOIDS47582.2021.9555769 ISSN: 2164-0580.
[12] 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 (May 2021), 2608–2617. https://doi.org/10.1109/TVCG.2021.3067783 Conference Name: IEEE Transactions on Visualization and Computer Graphics.
[13] David J. Atkinson and Micah H. Clark. 2014. Methodology for study of human-robot social interaction in dangerous situations. In *Proceedings of the second international conference on Human-agent interaction*. ACM, Tsukuba Japan, 371–376. https://doi.org/10.1145/2658861.2658871
[14] Franziska Babel, Andrea Vogt, Philipp Hock, Johannes Kraus, Florian Angerer, Tina Seufert, and Martin Baumann. 2022. Step Aside! VR-Based Evaluation of Adaptive Robot Conflict Resolution Strategies for Domestic Service Robots. *International Journal of Social Robotics* 14, 5 (July 2022), 1239–1260. https://doi.org/10.1007/s12369-021-00858-7
[15] Wilma A. Bainbridge, Justin Hart, Elizabeth S. Kim, and Brian Scassellati. 2008. The effect of presence on human-robot interaction. In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. 701–706. https://doi.org/10.1109/ROMAN.2008.4600749 ISSN: 1944-9437.
[16] Federica Bazzano, Federico Gentilini, Fabrizio Lamberti, Andrea Sanna, Gianluca Paravati, Valentina Gatteschi, and Marco Gaspardone. 2016. Immersive Virtual Reality-Based Simulation to Support the Design of Natural Human-Robot Interfaces for Service Robotic Applications. In *Augmented Reality, Virtual Reality, and Computer Graphics (Lecture Notes in Computer Science)*, Lucio Tommaso De Paolis and Antonio Mongelli (Eds.). Springer International Publishing, Cham, 33–51. https://doi.org/10.1007/978-3-319-40621-3_3
[17] Austin J. Beattie, Chad Edwards, and Mallory J. Williams. 2018. Telepresence Group Leaders Receive Higher Ratings of Social Attractiveness and Leadership Quality. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*. Association for Computing Machinery, New York,

NY, USA, 61–62. https://doi.org/10.1145/3173386.3176952

[18] William N. Benson, Zachary Anderson, Evan Capelle, Maya F. Dunlap, Blake Dorris, Jenna L. Gorlewicz, Mitsuru Shimizu, and Jerry B. Weinberg. 2023. Embodied Expressive Gestures in Telerobots: A Tale of Two Users. *ACM Transactions on Human-Robot Interaction* 12, 2 (March 2023), 24:1–24:20. https://doi.org/10.1145/3570908

[19] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM* 45, 7 (July 2002), 64–70. https://doi.org/10.1145/514236.514265

[20] Frank Biocca. 1997. The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication* 3, 2 (Sept. 1997), JCMC324. https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

[21] Frank Biocca, Chad Harms, and Judee K. Burgoon. 2003. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments* 12, 5 (Oct. 2003), 456–480. https://doi.org/10.1162/105474603322761270

[22] Andriana Boudouraki, Joel E. Fischer, Stuart Reeves, and Sean Rintel. 2023. "Being in on the Action" in Mobile Robotic Telepresence: Rethinking Presence in Hybrid Participation. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 63–71. https://doi.org/10.1145/3568162.3576961

[23] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. Publisher: Taylor & Francis.

[24] Cynthia Breazeal, Atsuo Takanishi, and Tetsunori Kobayashi. 2008. Social Robots that Interact with People. In *Springer Handbook of Robotics*, Bruno Siciliano and Oussama Khatib (Eds.). Springer, Berlin, Heidelberg, 1349–1369. https://doi.org/10.1007/978-3-540-30301-5_59

[25] Paul Bremner, Oya Celiktutan, and Hatice Gunes. 2016. Personality Perception of Robot Avatar Tele-operators. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Christchurch, New Zealand, 141–148.

[26] Paul Adam Bremner, Oya Celiktutan, and Hatice Gunes. 2017. Personality Perception of Robot Avatar Teleoperators in Solo and Dyadic Tasks. *Frontiers in Robotics and AI* 4 (2017). https://www.frontiersin.org/articles/10.3389/frobt.2017.00016

[27] Samuel Bustamante, Jan Peters, Bernhard Schölkopf, Moritz Grosse-Wentrup, and Vinay Jayaram. 2021. ArmSym: A Virtual Human–Robot Interaction Laboratory for Assistive Robotics. *IEEE Transactions on Human-Machine Systems* 51, 6 (Dec. 2021), 568–577. https://doi.org/10.1109/THMS.2021.3106865 Conference Name: IEEE Transactions on Human-Machine Systems.

[28] Lung-Pan Cheng, Thijs Roumen, Hannes Rantzsch, Sven Köhler, Patrick Schmidt, Robert Kovacs, Johannes Jasper, Jonas Kemper, and Patrick Baudisch. 2015. TurkDeck: Physical Virtual Reality Based on People. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. Association for Computing Machinery, New York, NY, USA, 417–426. https://doi.org/10.1145/2807442.2807463

[29] Jung Ju Choi and Sonya S. Kwak. 2017. Who is this?: Identity and presence in robot-mediated communication. *Cognitive Systems Research* 43 (June 2017), 174–189. https://doi.org/10.1016/j.cogsys.2016.07.006

[30] Mina Choi, Rachel Kornfield, Leila Takayama, and Bilge Mutlu. 2017. Movement Matters: Effects of Motion and Mimicry on Perception of Similarity and Closeness in Robot-Mediated Communication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 325–335. https://doi.org/10.1145/3025453.3025734

[31] N. L. Collins and L. C. Miller. 1994. Self-disclosure and liking: a meta-analytic review. *Psychological Bulletin* 116, 3 (Nov. 1994), 457–475. https://doi.org/10.1037/0033-2909.116.3.457

[32] Salam Daher, Kangsoo Kim, Myungho Lee, Andrew Raij, Ryan Schubert, Jeremy Bailenson, and Greg Welch. 2016. Exploring social presence transfer in real-virtual human interaction. In *2016 IEEE Virtual Reality (VR)*. IEEE, Greenville, SC, USA, 165–166. https://doi.org/10.1109/VR.2016.7504705

[33] Kourosh Darvish, Luigi Penco, Joao Ramos, Rafael Cisneros, Jerry Pratt, Eiichi Yoshida, Serena Ivaldi, and Daniele Pucci. 2023. Teleoperation of Humanoid Robots: A Survey. *IEEE Transactions on Robotics* 39, 3 (June 2023), 1706–1727. https://doi.org/10.1109/TRO.2023.3236952 Conference Name: IEEE Transactions on Robotics.

[34] K. Dautenhahn, S. Woods, C. Kaouri, M.L. Walters, Kheng Lee Koay, and I. Werry. 2005. What is a robot companion - friend, assistant or butler?. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1192–1197. https://doi.org/10.1109/IROS.2005.1545189 ISSN: 2153-0866.

[35] Maartje M. A. de Graaf, Somaya Ben Allouch, and Tineke Klamer. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior* 43 (Feb. 2015), 1–14. https://doi.org/10.1016/j.chb.2014.10.030

[36] Edo de Wolf and Jamy Li. 2020. Double Trouble: The Effect of Eye Gaze on the Social Impression of Mobile Robotic Telepresence Operators. In *Social Robotics (Lecture Notes in Computer Science)*, Alan R. Wagner, David Feil-Seifer, Kerstin S. Haring, Silvia Rossi, Thomas Williams, Hongsheng He, and Shuzhi Sam Ge (Eds.). Springer International Publishing, Cham, 357–368. https://doi.org/10.1007/978-3-030-62056-1_30

[37] Eric Deng, Bilge Mutlu, and Maja J. Mataric. 2019. Embodiment in Socially Interactive Robots. *Foundations and Trends® in Robotics* 7, 4 (Jan. 2019), 251–356. https://doi.org/10.1561/2300000056 Publisher: Now Publishers, Inc..

[38] Lorin Dole and Wendy Ju. 2019. Face and Ecological Validity in Simulations: Lessons from Search-and-Rescue HRI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3290605.3300681

[39] Mihai Duguleana, Florin Grigorie Barbuceanu, and Gheorghe Mogan. 2011. Evaluating Human-Robot Interaction during a Manipulation Experiment Conducted in Immersive Virtual Reality. In *Virtual and Mixed Reality - New Trends (Lecture Notes in Computer Science)*, Randall Shumaker (Ed.). Springer, Berlin, Heidelberg, 164–173. https://doi.org/10.1007/978-3-642-22021-0_19

[40] Roberta Etzi, Siyuan Huang, Giulia Scurati, Shilei Lyu, Francesco Ferrise, Alberto Gallace, Andrea Gaggioli, Alice Chirico, Marina Carulli, and Monica Bordegoni. 2019. Using Virtual Reality To Test Human-Robot Interaction During a Collaborative Task. https://doi.org/10.1115/DETC2019-97415

[41] Bo Feng, Siyue Li, and Na Li. 2016. Is a Profile Worth a Thousand Words? How Online Support-Seeker's Profile Features May Influence the Quality of Received Support Messages. *Communication Research* 43, 2 (March 2016), 253–276. https://doi.org/10.1177/0093650213510942 Publisher: SAGE Publications Inc.

[42] Naomi T. Fitter, Megan Strait, Eloise Bisbee, Maja J. Mataric, and Leila Takayama. 2021. You're Wigging Me Out!: Is Personalization of Telepresence Robots Strictly Positive?. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Boulder CO USA, 168–176. https://doi.org/10.1145/3434073.3444675

[43] Sanket Gaurav, Zainab Al-Qurashi, Amey Barapatre, George Maratos, Tejas Sarma, and Brian D. Ziebart. 2019. Deep Correspondence Learning for Effective Robotic Teleoperation using Virtual Reality. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE, Toronto, ON, Canada, 477–483. https://doi.org/10.1109/Humanoids43949.2019.9035031

[44] Carlos Girard, Diego Calderon, Ali Lemus, Victor Ferman, and Julio Fajardo. 2020. A Motion Mapping System for Humanoids that Provides Immersive Telepresence Experiences. In *2020 6th International Conference on Mechatronics and Robotics Engineering (ICMRE)*. 165–170. https://doi.org/10.1109/ICMRE49073.2020.9065155

[45] Mar Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI* 5 (2018). https://www.frontiersin.org/articles/10.3389/frobt.2018.00074

[46] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors* 53, 5 (Oct. 2011), 517–527. https://doi.org/10.1177/0018720811417254

[47] Claudia Hendrix and Woodrow Barfield. 1996. The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments* 5, 3 (Jan. 1996), 290–301. https://doi.org/10.1162/pres.1996.5.3.290

[48] Karen Gay Hickenbottom. 2022. *Almost Like Being There: Embodiment, Social Presence, and Engagement Using Telepresence Robots in Blended Courses*. Ed.D. Seattle Pacific University, United States – Washington. https://www.proquest.com/docview/2656207725/abstract/E0857EFE92E74636PQ/1 ISBN: 9798426815629.

[49] Donald M. Hilty, Karan Randhawa, Marlene M. Maheu, Alastair J. S. McKean, Richard Pantera, Matthew C. Mishkind, and Albert "Skip" Rizzo. 2020. A Review of Telepresence, Virtual Reality, and Augmented Reality Applied to Clinical Care. *Journal of Technology in Behavioral Science* 5, 2 (June 2020), 178–205. https://doi.org/10.1007/s41347-020-00126-x

[50] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. Association for Computing Machinery, New York, NY, USA, 119–125. https://doi.org/10.1145/142750.142769

[51] Andrea B. Hollingshead, Joseph E. Mcgrath, and Kathleen M. O'Connor. 1993. Group Task Performance and Communication Technology: A Longitudinal Study of Computer-Mediated Versus Face-to-Face Work Groups. *Small Group Research* 24, 3 (Aug. 1993), 307–333. https://doi.org/10.1177/1046496493243003 Publisher: SAGE Publications Inc.

[52] Jane Hwang, Sangyup Lee, Sang Chul Ahn, and Hyoung-gon Kim. 2008. Augmented robot agent: Enhancing co-presence of the remote participant. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. 161–162. https://doi.org/10.1109/ISMAR.2008.4637346

[53] K. Inoue, S. Nonaka, Y. Ujiie, T. Takubo, and T. Arai. 2005. Comparison of human psychology for real and virtual mobile manipulators. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* 73–78. https://doi.org/10.1109/ROMAN.2005.1513759 ISSN: 1944-9437.

[54] Fatima El Jamiy and Ronald Marsh. 2019. Distance Estimation In Virtual Reality And Augmented Reality: A Survey. In *2019 IEEE International Conference on*

Electro Information Technology (EIT). 063–068. https://doi.org/10.1109/EIT.2019.8834182 ISSN: 2154-0373.

[55] Brennan Jones, Yaying Zhang, Priscilla N. Y. Wong, and Sean Rintel. 2021. Belonging There: VROOM-ing into the Uncanny Valley of XR Telepresence. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 59:1–59:31. https://doi.org/10.1145/3449133

[56] Myeongul Jung, Jejoong Kim, Kyungsik Han, and Kwanguk (Kenny) Kim. 2022. Social Telecommunication Experience with Full-Body Ownership Humanoid Robot. *International Journal of Social Robotics* 14, 9 (Nov. 2022), 1951–1964. https://doi.org/10.1007/s12369-022-00922-w

[57] Ben Kenwright. 2012. Inverse Kinematics – Cyclic Coordinate Descent (CCD). *Journal of Graphics Tools* 16, 4 (Oct. 2012), 177–217. https://doi.org/10.1080/2165347X.2013.823362 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/2165347X.2013.823362.

[58] Hisham Khalil, Enrique Coronado, and Gentiane Venture. 2021. Human Motion Retargeting to Pepper Humanoid Robot from Uncalibrated Videos Using Human Pose Estimation. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. 1145–1152. https://doi.org/10.1109/RO-MAN50785.2021.9515495 ISSN: 1944-9437.

[59] Hanseob Kim, Myungho Lee, Gerard J. Kim, and Jae-In Hwang. 2021. The Impacts of Visual Effects on User Perception With a Virtual Human in Augmented Reality Conflict Situations. *IEEE Access* 9 (2021), 35300–35312. https://doi.org/10.1109/ACCESS.2021.3062037 Conference Name: IEEE Access.

[60] Hongki Kim, Kil-Soo Suh, and Un-Kon Lee. 2013. Effects of collaborative online shopping on shopping experience through social and relational perspectives. *Information & Management* 50, 4 (June 2013), 169–180. https://doi.org/10.1016/j.im.2013.02.003

[61] Jinwook Kim, Dooyoung Kim, Bowon Kim, Hyunchul Kim, and Jeongmi Lee. 2023. Holobot: Hologram based Extended Reality Telepresence Robot. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 60–64. https://doi.org/10.1145/3568294.3580043

[62] Kangsoo Kim, Gerd Bruder, and Greg Welch. 2017. Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3139131.3139151

[63] Kangsoo Kim, Divine Maloney, Gerd Bruder, Jeremy N. Bailenson, and Gregory F. Welch. 2017. The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. *Computer Animation and Virtual Worlds* 28, 3-4 (2017), e1771. https://doi.org/10.1002/cav.1771 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cav.1771.

[64] Kangsoo Kim, Ryan Schubert, Jason Hochreiter, Gerd Bruder, and Gregory Welch. 2019. Blowing in the wind: Increasing social presence with a virtual human via environmental airflow interaction in mixed reality. *Computers & Graphics* 83 (Oct. 2019), 23–32. https://doi.org/10.1016/j.cag.2019.06.006

[65] Lawrence H. Kim, Veronika Domova, Yuqi Yao, and Pablo E. Paredes. 2022. Effects of a Co-Located Robot and Anthropomorphism on Human Motivation and Emotion across Personality and Gender. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 155–162. https://doi.org/10.1109/RO-MAN53752.2022.9900694 ISSN: 1944-9437.

[66] Maori Kobayashi, Kanako Ueno, and Shiro Ise. 2015. The effects of spatialized sounds on the sense of presence in auditory virtual environments: A psychological and physiological study. *Presence: Teleoperators and Virtual Environments* 24, 2 (May 2015), 163–174. https://doi.org/10.1162/PRES_a_00226

[67] Natasa Koceska, Saso Koceski, Pierluigi Beomonte Zobel, Vladimir Trajkovik, and Nuno Garcia. 2019. A Telemedicine Robot System for Assisted and Independent Living. *Sensors* 19, 4 (Jan. 2019), 834. https://doi.org/10.3390/s19040834 Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

[68] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. 2013. A Review of Mobile Robotic Telepresence. *Advances in Human-Computer Interaction* 2013 (April 2013), e902316. https://doi.org/10.1155/2013/902316 Publisher: Hindawi.

[69] Kaiko Kuwamura, Takashi Minato, Shuichi Nishio, and Hiroshi Ishiguro. 2012. Personality distortion in communication through teleoperated robots. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. 49–54. https://doi.org/10.1109/ROMAN.2012.6343730 ISSN: 1944-9437.

[70] Daniel Lanzoni, Andrea Cattaneo, Andrea Vitali, Daniele Regazzoni, and Caterina Rizzi. 2023. Markerless Motion Capture and Virtual Reality for Real-Time Ergonomic Analysis of Operators in Workstations with Collaborative Robots: a preliminary study. In *Advances on Mechanics, Design Engineering and Manufacturing IV (Lecture Notes in Mechanical Engineering)*, Salvatore Gerbino, Antonio Lanzotti, Massimo Martorelli, Ramón Mirálbes Buil, Caterina Rizzi, and Lionel Roucoules (Eds.). Springer International Publishing, Cham, 1183–1194. https://doi.org/10.1007/978-3-031-15928-2_103

[71] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 2006. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies* 64, 10 (Oct. 2006), 962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002

[72] Myungho Lee, Nahal Norouzi, Gerd Bruder, Pamela J. Wisniewski, and Gregory F. Welch. 2021. Mixed Reality Tabletop Gameplay: Social Interaction With a Virtual Human Capable of Physical Influence. *IEEE Transactions on Visualization and Computer Graphics* 27, 8 (Aug. 2021), 3534–3545. https://doi.org/10.1109/TVCG.2019.2959575 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[73] Min Kyung Lee and Leila Takayama. 2011. "Now, i have a body": uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 33–42. https://doi.org/10.1145/1978942.1978950

[74] Wen-Ying Lee, Mose Sakashita, Elizabeth Ricci, Houston Claure, François Guimbretière, and Malte Jung. 2021. Interactive Vignettes: Enabling Large-Scale Interactive HRI Research. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. 1289–1296. https://doi.org/10.1109/RO-MAN50785.2021.9515376 ISSN: 1944-9437.

[75] Rui Li, Marc van Almkerk, Sanne van Waveren, Elizabeth Carter, and Iolanda Leite. 2019. Comparing Human-Robot Proxemics Between Virtual Reality and the Real World. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 431–439. https://doi.org/10.1109/HRI.2019.8673116 ISSN: 2167-2148.

[76] Siyue Li, Bo Feng, Na Li, and Xuan Tan. 2015. How Social Context Cues in Online Support-Seeking Influence Self-Disclosure in Support Provision. *Communication Quarterly* 63, 5 (Oct. 2015), 586–602. https://doi.org/10.1080/01463373.2015.1078389 Publisher: Routledge _eprint: https://doi.org/10.1080/01463373.2015.1078389.

[77] Jean-Luc Lugrin, Maximilian Wiedemann, Daniel Bieberstein, and Marc Erich Latoschik. 2015. Influence of avatar realism on stressful situation in VR. In *2015 IEEE Virtual Reality (VR)*. 227–228. https://doi.org/10.1109/VR.2015.7223378 ISSN: 2375-5334.

[78] Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. 2021. Pixel Codec Avatars. https://doi.org/10.48550/arXiv.2104.04638 arXiv:2104.04638 [cs].

[79] Dave Madole and Durand Begault. 1995. 3-D Sound for Virtual Reality and Multimedia. *Computer Music Journal* 19, 4 (1995), 99. https://doi.org/10.2307/3680997

[80] Karthik Mahadevan, Mauricio Sousa, Anthony Tang, and Tovi Grossman. 2021. "Grip-that-there": An Investigation of Explicit and Implicit Task Allocation Techniques for Human-Robot Collaboration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445355

[81] Ali Ahmad Malik, Tariq Masood, and Arne Bilberg. 2020. Virtual reality in manufacturing: immersive and collaborative artificial-reality in design of human-robot workspace. *International Journal of Computer Integrated Manufacturing* 33, 1 (Jan. 2020), 22–37. https://doi.org/10.1080/0951192X.2019.1690685 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/0951192X.2019.1690685.

[82] Martina Mara, Jan-Philipp Stein, Marc Erich Latoschik, Birgit Lugrin, Constanze Schreiner, Rafael Hostettler, and Markus Appel. 2021. User Responses to a Humanoid Robot Observed in Real Life, Virtual Reality, 3D and 2D. *Frontiers in Psychology* 12 (2021). https://www.frontiersin.org/articles/10.3389/fpsyg.2021.633178

[83] Elias Matsas and George-Christopher Vosniakos. 2017. Design of a virtual reality training system for human–robot collaboration in manufacturing tasks. *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11, 2 (May 2017), 139–153. https://doi.org/10.1007/s12008-015-0259-2

[84] Elias Matsas, George-Christopher Vosniakos, and Dimitrios Batras. 2017. Effectiveness and acceptability of a virtual environment for assessing human–robot collaboration in manufacturing. *The International Journal of Advanced Manufacturing Technology* 92, 9 (Oct. 2017), 3903–3917. https://doi.org/10.1007/s00170-017-0428-5

[85] James Mccroskey, Virginia Richmond, and John Daly. 1975. The Development of a Measure of Perceived Homophily. *Human Communication Interaction* 1 (June 1975), 323–332. https://doi.org/10.1111/j.1468-2958.1975.tb00281.x

[86] Brian E. Mennecke, Janea L. Triplett, Lesya M. Hassall, and Zayira Jordan Conde. 2010. Embodied Social Presence Theory. In *2010 43rd Hawaii International Conference on System Sciences*. 1–10. https://doi.org/10.1109/HICSS.2010.179 ISSN: 1530-1605.

[87] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction (HRI '09)*. Association for Computing Machinery, New York, NY, USA, 69–76. https://doi.org/10.1145/1514095.1514110

[88] Kristine Nowak and Frank Biocca. 2003. The Effect of the Agency and Anthropomorphism on Users' Sense of Telepresence, Copresence, and Social Presence

in Virtual Environments. *Presence Teleoperators & Virtual Environments* 12 (Oct. 2003), 481–494. https://doi.org/10.1162/105474603322761289

[89] Catherine S. Oh, Jeremy N. Bailenson, and Gregory F. Welch. 2018. A Systematic Review of Social Presence: Definition, Antecedents, and Implications. *Frontiers in Robotics and AI* 5 (2018). https://www.frontiersin.org/article/10.3389/frobt.2018.00114

[90] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 741–754. https://doi.org/10.1145/2984511.2984517

[91] Niklas Osmers, Michael Prilla, Oliver Blunk, Gordon George Brown, Marc Janßen, and Nicolas Kahrl. 2021. The Role of Social Presence for Cooperation in Augmented Reality on Head Mounted Devices: A Literature Review. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3411764.3445633

[92] Minna Pakanen, Paula Alavesa, Niels van Berkel, Timo Koskela, and Timo Ojala. 2022. "Nice to see you virtually": Thoughtful design and evaluation of virtual avatar of the other user in AR and VR based telexistence systems. *Entertainment Computing* 40 (Jan. 2022), 100457. https://doi.org/10.1016/j.entcom.2021.100457

[93] Alexis Paljic. 2017. Ecological Validity of Virtual Reality: Three Use Cases. In *New Trends in Image Analysis and Processing – ICIAP 2017 (Lecture Notes in Computer Science)*, Sebastiano Battiato, Giovanni Maria Farinella, Marco Leo, and Giovanni Gallo (Eds.). Springer International Publishing, Cham, 301–310. https://doi.org/10.1007/978-3-319-70742-6_28

[94] Tomislav Pejsa, Julian Kantor, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2016. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. 1716–1725. https://www.microsoft.com/en-us/research/publication/room2room-enabling-life-size-telepresence-in-a-projected-augmented-reality-environment/

[95] 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). https://www.frontiersin.org/articles/10.3389/frobt.2021.634520

[96] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173620

[97] Irene Rae and Carman Neustaedter. 2017. Robotic Telepresence at Scale. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 313–324. https://doi.org/10.1145/3025453.3025855

[98] Irene Rae, Leila Takayama, and Bilge Mutlu. 2012. One of the gang: supporting in-group behavior for embodied mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 3091–3100. https://doi.org/10.1145/2207676.2208723

[99] Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. In-body experiences: embodiment, control, and trust in robot-mediated communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1921–1930.

[100] Irene Rae, Leila Takayama, and Bilge Mutlu. 2013. The influence of height in robot-mediated communication. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 1–8. https://doi.org/10.1109/HRI.2013.6483495 ISSN: 2167-2148.

[101] Stephen Reysen. 2005. Construction of a New Scale: The Reysen Likability Scale. *Social Behavior and Personality: An International Journal* 33, 2 (2005), 201–208. https://doi.org/10.2224/sbp.2005.33.2.201 Place: New Zealand Publisher: Society for Personality Research.

[102] Taehyun Rhee, Stephen Thompson, Daniel Medeiros, Rafael dos Anjos, and Andrew Chalmers. 2020. Augmented Virtual Teleportation for High-Fidelity Telecollaboration. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (May 2020), 1923–1933. https://doi.org/10.1109/TVCG.2020.2973065 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[103] Hela Ridha-Mahfoudhi and Nguyen-Thong Dang. 2019. Real Time Point Cloud Self-Avatar With A Single RGB-D Camera. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19)*. Association for Computing Machinery, New York, NY, USA, 1–2. https://doi.org/10.1145/3359996.3365041

[104] SoftBank Robotics. [n. d.]. Feature Update: Telepresence Enabled by Pepper \textbar SoftBank Robotics. Retrieved 2022-06-07 from https://www.softbankrobotics.com/emea/en/blog/news-trends/feature-update-telepresence-enabled-pepper

[105] Jennifer Ann Rode. 2018. On Becoming a Cyborg: A Reflection on Articulation Work, Embodiment, Agency and Ableism. In *Breaking Down Barriers*, Pat Langdon, Jonathan Lazar, Ann Heylighen, and Hua Dong (Eds.). Springer International Publishing, Cham, 239–249. https://doi.org/10.1007/978-3-319-75028-6_21

[106] Ofir Sadka, Jonathan Giron, Doron Friedman, Oren Zuckerman, and Hadas Erel. 2020. Virtual-reality as a Simulation Tool for Non-humanoid Social Robots. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–9. https://doi.org/10.1145/3334480.3382893

[107] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. Association for Computing Machinery, New York, NY, USA, 141–148. https://doi.org/10.1145/2696454.2696497

[108] Susanne Schmidt, Oscar Javier Ariza Nunez, and Frank Steinicke. 2019. Blended Agents: Manipulation of Physical Objects within Mixed Reality Environments and Beyond. In *Symposium on Spatial User Interaction (SUI '19)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3357251.3357591

[109] Alexander P. Schouten, Tijs C. Portegies, Iris Withuis, Lotte M. Willemsen, and Komala Mazerant-Dubois. 2022. Robomorphism: Examining the effects of telepresence robots on between-student cooperation. *Computers in Human Behavior* 126 (Jan. 2022), 106980. https://doi.org/10.1016/j.chb.2021.106980

[110] Elena Márquez Segura, Michael Kriegel, Ruth Aylett, Amol Deshmukh, and Henriette Cramer. 2012. How Do You Like Me in This: User Embodiment Preferences for Companion Agents. In *Intelligent Virtual Agents (Lecture Notes in Computer Science)*, Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker (Eds.). Springer, Berlin, Heidelberg, 112–125. https://doi.org/10.1007/978-3-642-33197-8_12

[111] Azadeh Shariati, Mojtaba Shahab, Ali Meghdari, Ali Amoozandeh Nobaveh, Raman Rafatnejad, and Behrad Mozafari. 2018. Virtual Reality Social Robot Platform: A Case Study on Arash Social Robot. In *Social Robotics (Lecture Notes in Computer Science)*, Shuzhi Sam Ge, John-John Cabibihan, Miguel A. Salichs, Elizabeth Broadbent, Hongsheng He, Alan R. Wagner, and Álvaro Castro-González (Eds.). Springer International Publishing, Cham, 551–560. https://doi.org/10.1007/978-3-030-05204-1_54

[112] Kyoung Wan Cathy Shin and Jeonghye Han. 2016. Children's Perceptions of and Interactions with a Telepresence Robot. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*. IEEE Press, Christchurch, New Zealand, 521–522.

[113] Soyeon Shin, Dahyun Kang, and Sonya S. Kwak. 2022. Telepresence Robot for Isolated Patients in the COVID-19 Pandemic: Effects of Socio-relationship and Telecommunication Device Types on Patients' Acceptance of Robots. In *Social Robotics (Lecture Notes in Computer Science)*, Filippo Cavallo, John-John Cabibihan, Laura Fiorini, Alessandra Sorrentino, Hongsheng He, Xiaorui Liu, Yoshio Matsumoto, and Shuzhi Sam Ge (Eds.). Springer Nature Switzerland, Cham, 263–276. https://doi.org/10.1007/978-3-031-24670-8_24

[114] John Short. 1976. The social psychology of telecommunications /. Place: London : Publisher: Wiley,.

[115] David Sirkin and Wendy Ju. 2012. Consistency in physical and on-screen action improves perceptions of telepresence robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction (HRI '12)*. Association for Computing Machinery, New York, NY, USA, 57–64. https://doi.org/10.1145/2157689.2157699

[116] Eliot R. Smith, Selma Šabanović, and Marlena R. Fraune. 2021. Human-robot Interaction Through the Lens of Social Psychological Theories of Intergroup Behavior. *Technology, Mind, and Behavior* 1, 2 (Feb. 2021). https://tmb.apaopen.org/pub/vlke08m8/release/1

[117] Brett Stoll, Samantha Reig, Lucy He, Ian Kaplan, Malte F. Jung, and Susan R. Fussell. 2018. "Wait, Can You Move the Robot?": Examining Telepresence Robot Use in Collaborative Teams. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 14–22. ISSN: 2167-2148.

[118] Susan Straus and Joseph McGrath. 1994. Does the Medium Matter? The Interaction of Task Type and Technology on Group Performance and Member Reactions. *The Journal of applied psychology* 79 (March 1994), 87–97. https://doi.org/10.1037//0021-9010.79.1.87

[119] Hidenobu Sumioka, Shuichi Nishio, and Hiroshi Ishiguro. 2012. Teleoperated android for mediated communication : body ownership, personality distortion, and minimal human design. 32–39.

[120] Ryo Suzuki, Hooman Hedayati, Clement Zheng, James L. Bohn, Daniel Szafir, Ellen Yi-Luen Do, Mark D. Gross, and Daniel Leithinger. 2020. RoomShift: Room-scale Dynamic Haptics for VR with Furniture-moving Swarm Robots. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376523

[121] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–33. https://doi.org/10.1145/3491102.3517719

[122] Leila Takayama and Janet Go. 2012. Mixing metaphors in mobile remote presence. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. Association for Computing Machinery, New York, NY, USA, 495–504. https://doi.org/10.1145/2145204.2145281

[123] Leila Takayama and Caroline Pantofaru. 2009. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 5495–5502. https://doi.org/10.1109/IROS.2009.5354145 ISSN: 2153-0866.

[124] Fumihide Tanaka, Toshimitsu Takahashi, Shizuko Matsuzoe, Nao Tazawa, and Masahiko Morita. 2014. Telepresence robot helps children in communicating with teachers who speak a different language. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*. Association for Computing Machinery, New York, NY, USA, 399–406. https://doi.org/10.1145/2559636.2559654

[125] Ilaria Torre, Alexis Linard, Anders Steen, Jana Tumová, and Iolanda Leite. 2021. Should Robots Chicken? How Anthropomorphism and Perceived Autonomy Influence Trajectories in a Game-theoretic Problem. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 370–379. https://doi.org/10.1145/3434073.3444687

[126] Tzung-Cheng Tsai, Yeh-Liang Hsu, An-I. Ma, Trevor King, and Chang-Huei Wu. 2007. Developing a Telepresence Robot for Interpersonal Communication with the Elderly in a Home Environment. *Telemedicine and e-Health* 13, 4 (Aug. 2007), 407–424. https://doi.org/10.1089/tmj.2006.0068 Publisher: Mary Ann Liebert, Inc., publishers.

[127] Katherine M. Tsui, James M. Dalphond, Daniel J. Brooks, Mikhail S. Medvedev, Eric McCann, Jordan Allspaw, David Kontak, and Holly A. Yanco. 2015. Accessible Human-Robot Interaction for Telepresence Robots: A Case Study. *Paladyn, Journal of Behavioral Robotics* 6, 1 (Jan. 2015). https://doi.org/10.1515/pjbr-2015-0001 Publisher: De Gruyter Open Access.

[128] Tadele Belay Tuli, Tesfaye Olana Terefe, and Md Mamun Ur Rashid. 2021. Telepresence Mobile Robots Design and Control for Social Interaction. *International Journal of Social Robotics* 13, 5 (Aug. 2021), 877–886. https://doi.org/10.1007/s12369-020-00676-3

[129] Josca van Houwelingen-Snippe, Jered Vroon, Gwenn Englebienne, and Pim Haselager. 2017. Blame my telepresence robot joint effect of proxemics and attribution on interpersonal attraction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 162–168. https://doi.org/10.1109/ROMAN.2017.8172296 ISSN: 1944-9437.

[130] Gina Venolia, John Tang, Ruy Cervantes, Sara Bly, George Robertson, Bongshin Lee, and Kori Inkpen. 2010. Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1049–1058. https://doi.org/10.1145/1753326.1753482

[131] Ana M Villanueva, Ziyi Liu, Zhengzhe Zhu, Xin Du, Joey Huang, Kylie A Peppler, and Karthik Ramani. 2021. RobotAR: An Augmented Reality Compatible Teleconsulting Robotics Toolkit for Augmented Makerspace Experiences. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445726

[132] 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 (May 2023), 2596–2606. https://doi.org/10.1109/TVCG.2023.3247072 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[133] Vincent Weistroffer, Alexis Paljic, Philippe Fuchs, Olivier Hugues, Jean-Paul Chodacki, Pascal Ligot, and Alexandre Morais. 2014. Assessing the acceptability of human-robot co-presence on assembly lines: A comparison between actual situations and their virtual reality counterparts. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, Edinburgh, UK, 377–384. https://doi.org/10.1109/ROMAN.2014.6926282

[134] Lawrence R. Wheeless and Janis Grotz. 1977. The Measurement of Trust and Its Relationship to Self-Disclosure. *Human Communication Research* 3, 3 (1977), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2958.1977.tb00523.x.

[135] Steve Whittaker. 2002. Theories and Methods in Mediated Communication. *The Handbook of Discourse Processes* (Aug. 2002).

[136] Luc Wijnen, Paul Bremner, Séverin Lemaignan, and Manuel Giuliani. 2020. Performing Human-Robot Interaction User Studies in Virtual Reality. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 794–794. https://doi.org/10.1109/RO-MAN47096.2020.9223521 ISSN: 1944-9437.

[137] Boram Yoon, Hyung-il Kim, Gun A. Lee, Mark Billinghurst, and Woontack Woo. 2019. The Effect of Avatar Appearance on Social Presence in an Augmented Reality Remote Collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 547–556. https://doi.org/10.1109/VR.2019.8797719 ISSN: 2642-5254.

[138] Jingxin Zhang, Nikolaos Katzakis, Fariba Mostajeran, and Frank Steinicke. 2019. Localizing Teleoperator Gaze in 360° Hosted Telepresence. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, 1265–1266. https://doi.org/10.1109/VR.2019.8797819

A Appendix

A.1 AR Simulation Framework for AvatARoid Evaluation

Why Simulation? We aimed for an initial evaluation of AvatARoid's design, comparatively assessing how different user representations of the teleoperator impact the on-site user's perception of the teleoperator. We decided to implement a simulation-based evaluation system using a virtual robot, instead of a physical robot for the following three reasons. First, we wanted to single out the effects coming solely from variations in the user representation of the teleoperator; using a physical robot could cause a confound of implementation factors. Second, implementing a fully functional AvatARoid system presents significant financial and technical challenges, given its complex system architecture. Third, potential modifications due to technical issues during evaluation could introduce biases, as seen in a prior study where changing robots midway through the study, following a breakdown, confounded the results [129]. In our study context, the simulation approach can offer a controlled, cost-effective, and robust framework for evaluating varying user representations comparatively.

Simulation in HRI Studies: Methods and Their Validity. Interactive 3D simulations offer alternatives to traditional text- and video-based approaches for large-scale HRI studies [74] and have recently been used to assess how co-located robots affect human motivation [65]. Moreover, advancements in VR offer increasingly immersive and realistic simulations across various HRI domains, including control of prosthetics arm in assistive robotics [27], collaboration with robots and manipulators in manufacturing [39, 40, 70, 80, 81, 83, 84, 133], and high-stakes scenarios like search and rescue [13].

Studies validating the use of immersive simulations for Human-Robot Interaction (HRI) evaluation with virtual robots have demonstrated favorable results in terms of acceptability. For instance, a study evaluating children's engagement and proxemic preferences with the humanoid robot Arash through both physical and virtual versions found comparable social experiences [111]. Another study, comparing the effects of 2D, 3D, VR, and physical interaction modes with a humanoid robot, suggested that immersive environments could feasibly replicate the experience of verbal interaction with an actual robot in HRI studies [82]. The validity of VR simulations extends to assessing social interactions between humans and non-humanoid robots [106]. Moreover, studies comparing the acceptability of human-robot collaboration with VR and real robotic manipulators indicated that VR can facilitate preliminary assessments through questionnaires [39, 93, 133]. Another study evaluating participants' sense of security while collaborating with physical and virtual mobile robotic arms found similar security ratings between VR and physical robotic arms [53].

While validated as a tool for assessing social interactions between humans and various types of robots [14, 16, 106, 125], it is important to consider factors such as distance compression in VR HMDs [54], realism of robots [75] and participants sense of presence [38] as studies have indicated potential differences in participant experiences when interacting with real and virtual humanoid robots accounting to these factors. For instance, a study involving a humanoid robot approaching participants found increased discomfort and a desire for greater personal distance when interacting with the virtual robot compared to its physical counterpart [75]. These differences were attributed to compressed distance perception in VR and the lower realism quality of the virtual robot model. Additionally, participants' sense of presence plays a vital role; a study suggested that a diminished sense of presence in VR made virtual robots less "social" than real robots [136].

Addressing these limitations, we believe immersive technologies can provide rich, interactive, and cost-effective platforms for evaluating novel designs like AvatARoid. Building on the findings from existing studies using immersive simulation for HRI studies, we utilize an AR approach for simulating telepresence robots, highlighting AR's ability to integrate virtual entities into real-world environments, offering an advantageous context for robotic telepresence interactions.

Although simulating a robotic telepresence system with a virtual robot limits the tangible presence and physical manipulation abilities afforded by a physical robot, we believe simulation provides a cost-effective and controlled experiment environment for our purpose of identifying differences generated solely from different visual representations of the teleoperator when augmented to a robot. Moreover, immersive simulations with virtual robots have been shown to emulate verbal interaction experiences with actual robots in HRI studies [82].

To make this simulation plausible, we need to address certain requirements to make the robot appear realistic and present in the user's real environment.

Requirements. Immersive simulations must fulfill these requirements to effectively evaluate robot-mediated telepresence scenarios.

(1) R1: Simulate the teleoperator's and the robot's presence in real physical environment — To make the simulation more tailored towards robotic telepresence scenarios, we need to augment the virtual robot in the real physical environment.
(2) R2: Visual and Behavioral Realism of the Robot — We need to make the virtual robot appear and behave in a realistic manner. A humanoid robot's motion must account for its mechanical limitations, such as the range and speed of its joint movements [82] for behavioral plausibility of the robot's simulation. Lower realism of the virtual robot might result in a sense of discomfort and a reduced sense of co-presence with the robot.
(3) R3: Account for compressed distance perception in VR HMDs — We need to account for the sense of compressed distance reported when using mixed reality headsets which leads to difference in proxemic preferences between real and virtual robots.
(4) R4: Sense of Co-presence with Virtual Robot —Achieving a sense of presence is vital for validating any simulated HRI study [38]. For our study, since we augment the physical environment with a virtual robot, we aim to maximize the sense of co-presence with a virtual robot, where users feel they share a physical space with another entity [64].

Our Approaches. The overall approach we took with our simulation framework is to implement the robot-mediated telepresence setup with a virtual humanoid in an immersive AR simulation.

(1) A1 for R1: Use AR as an immersive technology for our simulation — AR seamlessly blends virtual and real worlds, offering a middle ground that maintains the physical context while introducing virtual elements. Unlike VR, AR does not isolate the user from their surroundings and enables the crucial sensation of being in a real environment. Therefore, we propose a novel approach using AR simulations, which we believe is more tailored for telepresence systems.
(2) A2 for R2: Use a realistic 3D model of the robot and animate it respecting its motion constraints — We used a realistic 3D model of a commercial social robot. The robot model was scaled to an approximate human height of 1.65 meters as we envision AvatARoid to ensure a human-like presence for teleoperators and on-site users.We used manufacturer-provided joint specifications to simulate the virtual robot's motion realistically. Using the IK solver package, we confined the virtual robot's arm and head movements to realistic ranges by modeling its joint and speed limitations. The robot's locomotion was animated and slowed down to make the robot appear slow while entering or exiting.
(3) A3 for R3: Strategically place the virtual robot — We positioned the robot at 3 m during the course of interaction, slightly further than the accepted social distance, to account for the distance compression often reported in Mixed Reality studies [54].
(4) A4 for R4: Implement Mechanisms to enhance co-presence with virtual robots — We implemented three mechanisms (discussed in detail in A.2) to increase the sense of co-presence with virtual robot in our simulation. These also help achieve R2 by making the robot more realistic.

A.2 Mechanisms to Enhance Co-presence with Virtual Robot of AR Simulation

We employ three mechanisms previously identified to augment users' sense of co-presence with a virtual entity in AR.

Displaying spatial awareness: A virtual entity's awareness of and interaction with real-world spaces and objects contributes to the shared presence [32, 63, 64, 95]. We designed our virtual robot to display spatial awareness; for instance, the virtual robot used real doors for entering and exiting rooms. The experimenter also acknowledges the robot's presence in the room. The experimenter instructing the participants during study sessions also avoided the robot while moving towards it and looked at it while addressing the teleoperator. We also carefully designed the narrative of our study task, including the word "couch" in the word-guessing game (see section 4.3 for details), allowing the teleoperator (and hence the virtual robot) to point toward an actual couch in the room.

Blending with the real world: Studies indicate that visually and physically plausible visuals and behaviors—like occlusion or avoiding passing through real-world objects—enhance the sense of co-presence with virtual agents [59, 62]. For blending with real-world lighting, we used a virtual directional light in the game engine used for developing the simulation to light the robot, matching the position, color, and intensity of the real-world light in the room. This directional light in the game engine facilitated the lighting and casting of shadows for the virtual robot to blend in with the lighting of the real environment. We used a noise shader for the virtual robot's material in the engine to simulate a grainy effect matching the surrounding environment, as seen through the passthrough AR HMD's display. The headset did not have sensors to provide depth information of the scene, so we used an API provided by the manufacturer to create the room geometry manually and used custom shaders to provide occlusion from the real-world objects (Figure 11(a)) and cast shadow using the virtual robot on the real room floor (Figure 11(b)).

Realistic auditory cues: As auditory cues (such as spatialized audio) have been well established as a crucial factor affecting presence in virtual environments [47, 66, 79], a generic servo audio effect simulated a robotic sound when the robot or its arm moved. The game engine's built-in audio spatialization feature was used to spatialize teleoperator voice and robot-generated sounds.

A.3 Evaluation of simulation: AR Simulation Could be a Viable Tool to Evaluate Robot-mediated Telepresence Systems

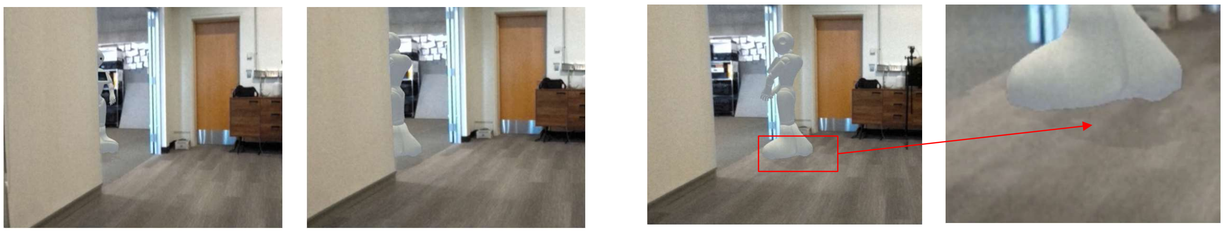
We found empirical evidence supporting the credibility of our simulation approach. Most participants (39 out of 48) rated the co-presence of the robot at four or higher on a seven-point Likert scale (M = 4.87, See Figure 12). In post-study interviews, The visual integration of the robot into the participants' actual environment was consistently noted as a crucial element for this enhanced co-presence. Participants highlighted the act of a "robot coming in the room," and the robot being "visually embedded in the room" made them feel that the "interaction was happening in this room."

In our study, participants highly valued spatial awareness and visual blending for an enhanced sense of co-presence with the AR robot. The robot's realistic entry and exit through doors, strategically obscured by the wall, effectively conveyed a tangible physical presence. Participants mentioned they felt like the robot had a physical presence when "the wall worked" and "the robot appeared behind the wall". The experimenter's demonstration of spatial awareness towards the robot further heightened the perception of sharing the space. Participants expressed in their interview that when the experimenter "walked behind the robot and around it," it made them feel like "the robot was more real".

The realism of the robot's motion and appropriate scaling also emerged as significant factors in maintaining a sense of co-presence. Participants commented that "it moved like a real robot" and highlighted the importance of the robot's realistic size, stating the scale as "realistic" and not "uncanny."

Audio cues provided interesting mixed results, while some participants commented audio cues helped foster co-presence, stating robot movements, sound and sound localization made the interaction feel like it was "happening in this room." Conversely, some participants did not notice the audio spatialization, stating the sound emanated from the headset rather than the robot. This might be because the robot was primarily static, except when it entered or exited, failing to provide enough spatial cues to localize sound.

These findings validate AR as an effective tool for HRI and telepresence scenarios that require seamless spatial integration.



a) Virtual robot occluded by real wall

b) Virtual robot casting shadow on the room floor

Figure 11: The robot blending with the real environment. a) Virtual robot being occluded by real wall. b) Virtual robot casting shadow on the real room floor

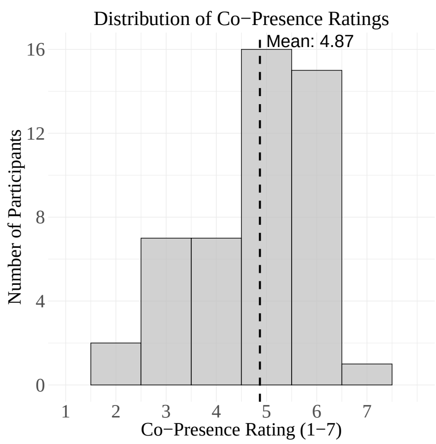


Figure 12: Histogram of participants frequency distribution for different co-presence scores

A.4 Demographics Survey

(1) What is your name?
(2) What is your email address?
(3) With which gender do you most identify?
- Man
- Woman
- Transgender Man / Trans Man
- Transgender Woman / Trans Woman
- Non-Binary
- Two-spirit
- Not Listed
- Prefer not to answer

(4) If you identify with gender not listed above, please add in this field (leave blank if prefer not to answer)
(5) What is your educational qualification? (Ex. High School, Bachelor's, Master's, etc, leave blank if not applicable)
(6) What subject are/were your studies in? (Leave blank if not applicable)
(7) What is your occupation?
(8) What is your age?
- 19-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65 or older
- Prefer not to answer

(9) What is your cultural background? Select all that apply.
- African
- European
- East Asian
- South Asian
- Southeast Asian
- Hispanic or Latinx
- Middle Eastern
- First Nations or Indigenous (Please specify below)
- Other (Please specify below)
- Prefer not to answer

(10) Please specify your cultural background here (Leave blank if prefer not to answer).

A.5 "Get to know You" task questions

(1) How are you?
(2) What do you do? Are you a student or are you working? Or neither.
(3) What initially attracted you to this field of work or study?
(4) What is something that you enjoy about what you do?
(5) Are there any hobbies or projects you are passionate about currently?
(6) What are some of your long-term goals or aspirations, either professionally and personally?
(7) What is the best piece of advice you've ever received?

A.6 Post Study Questionnaire

A.6.1 Perceived Embodiment of the Teleoperator. We used the perceived embodiment of the teleoperator using 6 Likert-scale statements (Cronbach's $\alpha = 0.7$), each paired with a seven-point ranking scale to indicate the level of agreement ranging from 1(strongly disagree) to 7(strongly agree). The statements were adopted from a proposed self-avatar embodiment questionnaire in VR [45]. The wording was changed to reflect the perceived embodiment of the teleoperator by the on-site users.

(1) I felt like the other participant was a real person.
(2) I felt like I was communicating with a robot.
(3) I felt like the robot's body was the other participant's body.

(4) I felt like the robot's movements were caused by the other participant's movements.
(5) I felt like the robot was moving by itself.
(6) I felt like the other participant was located where I saw the robot.

A.6.2 Social Presence of the Teleoperator. We assessed social presence using six semantic differential items (Cronbach's α = 0.82), adopting the measure used in [88]. This questionnaire measures a user's perception of the medium's capacity to have authentic social interaction closer to face-to-face interaction, with a higher score meaning the medium's ability to present the other as a social entity.

(1) To what extent did you feel able to assess your partner's reactions to what you said? [Able to assess reactions —- not able to assess reactions]
(2) To what extent was this like a face-to-face meeting? [A lot like face to face —- not like face to face at all]
(3) To what extent was this like you were in the same room with your partner? [A lot like being in the same room —- not like being in the same room atall]
(4) To what extent did your partner seem "real"? [Very real —- not real at all]
(5) How likely is it that you would choose to use this system of interaction for a meeting in which you wanted to persuade others of something? [Very likely —- not likely at all]
(6) To what extent did you feel you could get to know someone that you met only through this system? [Very well —- not at all]

A.6.3 Impression of the Teleoperator. For the impression of the teleoperator, we used four questionnaires to measure on-site users' impressions regarding trustworthiness, perceived similarity, likeability, and infrahumanization towards the teleoperator.

Trustworthiness. For trustworthiness (from [134]), we used 14 trust-related bipolar adjective pairs (Cronbach's α = 0.91) each scaled 1-7 and averaged to get a final score with minimum of 1 and maximum of 7.

(1) Trustworthy - Untrustworthy
(2) Confidential - Divulging
(3) Benevolent - Exploitative
(4) Safe - Dangerous
(5) Candid - Deceptive
(6) Not Deceitful - Deceitful
(7) Straightforward - tricky
(8) Respectful - Disrespectful
(9) Considerate - Inconsiderate
(10) Honest - Dishonest
(11) Reliable - Unreliable
(12) Faithful - Unfaithful
(13) Sincere - Insincere
(14) Careful - Careless

Likeability. For likeability, we used [101], which contains 11 Likert-scale statements (Cronbach's α = 0.88). Each item was scaled from 1-7 which was averaged to get the final score from 1-7.

(1) This person is friendly.
(2) This person is likable.
(3) This person is warm.
(4) This person is approachable.
(5) I would ask this person for advice.
(6) I would like this person as a coworker.
(7) I would like this person as a roommate.
(8) I would like to be friends with this person.
(9) This person is physically attractive.
(10) This person is similar to me.
(11) This person is knowledgeable.

Homophily. Perceived similarity consisted of nine adjective pairs (Cronbach's α = 0.83) in a semantic differential scale (1-7) adopted from [85]. Scores were averaged to get a final score from 1-7.

(1) Doesn't think like me: Thinks like me
(2) Behaves like me: Doesn't behave like me
(3) Similar to me: Different from me
(4) Unlike me: Like me
(5) Treats people like Ido: Doesn't treat people like I do
(6) Looks similar to me: Looks different from me
(7) Different size than I am: Same size I am
(8) Appearance like mine: Appearance unlike mine
(9) Doesn't resemble me: Resembles me

Infrahumanization. For infrahumanization, we had participants select 8-10 emotions from 16 emotions chosen from prior work [98], which categorized and validated them as either non-uniquely or uniquely human emotions. The more non-uniquely human emotions they selected, the more they felt the teleoperator was less capable of uniquely human emotions

(1) Attraction
(2) Desire
(3) Excitement
(4) Pleasure
(5) Agitation
(6) Anger
(7) Fear
(8) Rage
(9) Regret
(10) Disappointment
(11) Compassion
(12) Love
(13) Hope
(14) Admiration
(15) Bitterness
(16) Enthusiasm

Co-presence with the Robot. To verify that the participants felt present with the simulated virtual robot, we employed a co-presence questionnaire to measure the extent to which the participants thought they were in the same space as the virtual robot. We used the questionnaire (Cronbach's α = 0.77) designed in previous study [64], which contains 3 Likert-scale statements with a 7-point ranking scale and four items on a semantic differential scale. The final two items are subtracted and their differential taken as a single measure similar in [64].

(1) I perceived that I was in the presence of the robot in the room with me

(2) I felt the robot was watching me and was aware of my presence.
(3) I would feel startled if the robot came closer to me.
(4) To what extent did you have a sense of being with the robot? [Not at all —- Very Much]
(5) To what extent was this like you were in the same room with the robot? [Not at all —- Very Much]
(6) I felt I was in the .... space. [Virtual —- Physical]
(7) I felt the robot was in .... space. [Virtual —- Physical]

A.7 Post Study Interview Questions

(1) In the questionnaire, you rated the [specific aspect] as [rating]. Can you share more about what led to that rating? What was going through your mind when you answered [specific question]?
(2) What were your thoughts on the visual representation of Andy? Did you have difficulties identifying the robot as Andy? Did you feel like Andy was controlling the robot? Could you recognize Andy's gestures or movements or body languages through the robot? Could you elaborate?
(3) How aware were you of the robot versus Andy? Can you share moments when Andy felt really 'human' or more robot-like? If so, can you describe these moments?
(4) If you could change anything about this setup, what would that be?
(5) Did you feel like you were in the same room as the robot? Were there moments when you remembered you were interacting with a virtual robot or when it felt like a real robot?
(6) How did you feel about the AR experience? How much did the visual quality interfere or distract you? How much did the control devices interfere with the performance? How much did the sounds of the environment involve you? Could you identify the sound source and where the sound was coming from?
(7) Can you tell us a bit about your past experience/familiarity with Robots?
(8) Can you tell us a bit about your past experience/familiarity with AR/VR devices?
(9) On a scale of 1 to 4, 1 being completely AI and 4 being completely Human, how much would you rate Andy to be an AI or a human? (find out if it affected their rating)
(10) On a scale of 1 to 4, 1 being a member of the research confederate and 4 being another participant, how much would you rate Andy to be a confederate or another participant? (find out if it affected their ratings)