# Joint Action in Collaborative Mixed Reality: Effects of Immersion Type and Physical Location

Iana Podkosova*
yana.podkosova@tuwien.ac.at
VR&AR Research Unit, TU Wien
Vienna, Austria

Francesco De Pace*
francesco.pace@tuwien.ac.at
VR&AR Research Unit, TU Wien
Vienna, Austria

Hugo Brument*
hugo.brument@tuwien.ac.at
VR&AR Research Unit, TU Wien
Vienna, Austria

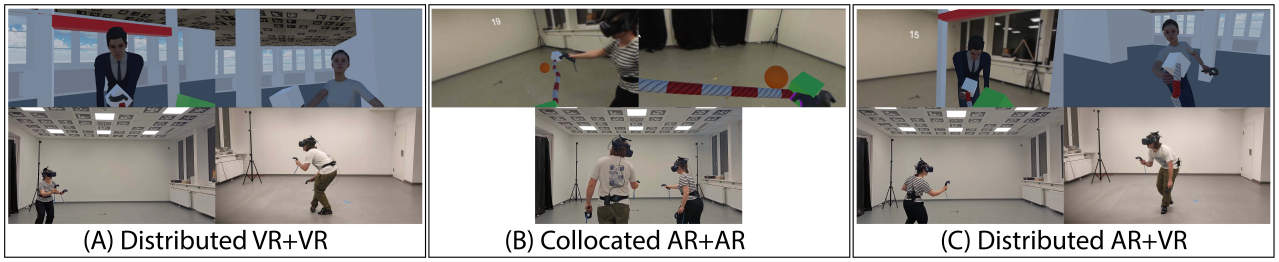


**Figure 1: Overview of some conditions from our experiment. (A) Participants were in distributed setup and immersed in VR while performing the Gate task. (B) Collocated setup and immersed in AR while performing the Fruit task. (C) Distributed setup where one participant was immersed in AR and the other in VR performing the Gate task.**

## ABSTRACT

Understanding how people effectively perform actions together is fundamental when designing Collaborative Mixed Reality (CMR) applications. While most of the studies on CMR mostly considered either how users are immersed in the CMR (e.g., in virtual or augmented reality), or how the physical workspace is shared by users (i.e., distributed or collocated), little is known about how their combination could influence user's interaction in CMR. In this paper, we present a user study (n=23) that investigates the effect of the mixed reality setup on the user's immersion and spatial interaction during a joint-action task. Groups of two participants had to perform two types of joint actions while carrying a virtual rope to maintain a certain distance: (1) Gate, where participants had to pass through a virtual aperture together and (2) Fruit, where participants had to use a rope to slice a virtual fruit moving in the CMR. Users were either in a distributed or collocated setup, and either immersed in virtual or augmented reality. Our results showed that users' proxemics was altered by the immersion type and location setup, but also the user's subjective experience. These results contribute to the understanding of joint action in CMR and they are discussed to improve the design of CMR applications.

## CCS CONCEPTS

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

*All authors contributed equally to this paper.



*SUI '23, October 13–15, 2023, Sydney, NSW, Australia*

ACM ISBN 979-8-4007-0281-5/23/10.
https://doi.org/10.1145/3607822.3614541

## KEYWORDS

Collaborative Mixed Reality, Joint Action, Spatial Interaction

**ACM Reference Format:**
Iana Podkosova, Francesco De Pace, and Hugo Brument. 2023. Joint Action in Collaborative Mixed Reality: Effects of Immersion Type and Physical Location. In *The 2023 ACM Symposium on Spatial User Interaction (SUI '23), October 13–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3607822.3614541

## 1 INTRODUCTION

Collaborative Mixed Reality (CMR) systems, in which two or more users are immersed with interfaces such as Augmented Reality (AR) and Virtual Reality (VR), are a vibrant area of research. Several definitions have been given to the term "Mixed Reality" [48]. Among them, MR can be defined as a system that uses multiple VR and AR interfaces at the same time [44]; this paper follows this definition. CMR applications enable users to engage in collaborative tasks such as training, remote assistance, or maintenance [55] by interacting with Virtual Environments (VEs). This paper focuses on a particular synchronous type of collaborative task: joint action. It refers to tasks involving two or more people coordinating their actions to produce a change in the environment [45] (e.g., moving an object, assembling furniture, dancing in synchronization).

Joint actions are complex to study as they are comprised of individual actions (e.g., lifting and moving one end of a heavy piano) that must be combined to achieve a collective goal (such as moving the piano across the room) [53]. Yet, the study of joint actions in CMR is vital if we want to gain insights into complex processes of individual and shared agency (i.e., the feeling of generating and controlling actions and their effects) [27] and understand how users interact and coordinate their actions. At the same time, knowledge

of joint actions is necessary to design CMRs in ways that guarantee the best user experience in terms of comfort, immersion, and performance.

Although several studies have focused on joint actions in Real Environments (REs) [52], VR [18] or AR [58] systems, they have remained relatively unexplored within the context of MR systems. Yet, many factors can influence the immersive experience of CMR systems. To start with, the type of immersive interface provided to the user (VR or AR) can change the way they perceive the other person through an avatar and may influence the judgment of the agency during the joint action. Further, collaborating users can be connected remotely or share the same physical space. The physical location of users may influence joint action in terms of proxemics and body movements. Finally, the choice of remote or collocated setup and immersive interface likely influence user experience in combination, primarily because the choice of location constricts the type of user representation within CMR.

The objective of this paper is to investigate how immersion type (VR or AR) and the physical location of users (distributed or collocated) influence joint actions in CMR. To our knowledge, we report the first study of joint action with two users in CMR. To study the effect of immersion type in depth, we investigate how joint action in MR might differ from that in multi-user VR and multi-user AR. We tested all combinations of individual immersion (VR+VR, AR+AR, VR+AR, AR+VR) to differentiate between the effects of immersion type and the symmetry or asymmetry between individual immersions while also varying the physical setup (collocated vs distributed). We analyzed two types of joint action (walking through an aperture and cutting fruit while holding a virtual rope) that differ in terms of spatial and temporal demand. Task performance and spatial metrics were evaluated alongside subjective questionnaire responses. Our results contribute to the understanding of joint action in CMR and are discussed with respect to the design of CMR, which could improve users' experience.

## 2 RELATED WORK

### 2.1 Collaboration and User Embodiment in MR

Collaboration in MR has been studied extensively [44], often with a focus on affordances and roles of users on different interface ends. This includes the immersion (e.g., VR or AR), the physical setup (e.g., distributed or collocated), and the interaction timing (e.g., synchronous or asynchronous) [9]. Piumsomboon et al. proposed different proximity cues (FoV frustum, eye gaze, and head-direction ray) to improve MR collaboration [35] finding that a combination of FoV frustum and head-direction ray was beneficial for the effectiveness of collaboration. VR users emerged as leaders in the collaboration; presumably because of the limited field of view (FoV) of the AR display used in the study (Microsoft HoloLens). Discussing implications, the authors advise against the choice of MR interfaces that induce disparity between the users. In the study of Pan et al. [33], AR users emerged as leaders in collaboratively editing a virtual planet in AR-desktop and AR-VR-without-Body setups, with the leadership effects emerging in 3D but not 2D interactions. These results are in line with an earlier study in a mixed setup in which the most immersed user emerged as a leader [47]. Mueller et al. discuss how different handheld interfaces influence user experience, workload, and group performance [31]. The authors compare AR-AR vs. VR-VR conditions for a collaborative search task, considering collocated and distributed setups. Contrary to the author's hypothesis that the VR setup would have been preferred in the distributed setup due to a lack of common spatial references, the outcomes indicate that the users' preferences were not affected by the setups. Moreover, the social presence was reported higher in AR than VR, independently of the setup.

User embodiment is one of the most important design decisions for CMR and has often been researched. It is known that tracking users' heads, hands, and feet improves the sense of embodiment (discussed in detail in [20]) and spatial presence [10] and that having a self-avatar helps with spatial judgments [30]. Various specific user representations have been proposed for CMR. For example, Piumsomboon et al. [36] proposed to miniaturize the remote user's avatar to reduce the negative effects of narrow FoV in AR. De Pace et al. [7] compared the use of abstract metaphors vs. avatar representations in an assembly task with audio and no-audio conditions. While the combination of avatar and no audio enhanced the sense of presence for the remote user, the effective task completion appeared to rely significantly on the use of abstract metaphors. A similar study regarding the importance of speech interfaces for CMR can be found in [23]. Yu et al. compared the use of an avatar based on point cloud volumetric reconstruction and a virtual human avatar of the AR user in a telepresence scenario. Despite the low quality (noise, partially missing features) of the point cloud avatar, it scored better in terms of co-presence, behavioral realism, and humanness [59]. A similar study confirmed the results about superior co-presence achieved with a volumetrically reconstructed avatar [43], with the authors advising against full-body avatars unless they have very realistic animations. Finally, Piumsomboon et al. [37] proposed a *Giant-Miniature* (i.e., local AR and remote VR) MR collaborative system that supports 360 video sharing and tangible interaction. Their research suggests that optimal positioning of the 360-degree camera at the user's shoulder height enables the remote VR user to experience the local AR user's environment. Additionally, the avatar representation of the remote Miniature VR user enhances collaborative interactions.

### 2.2 Joint Action in VR

The study of joint action suggests that we naturally coordinate our actions with other people [45]. It is proposed that the success of joint action depends on knowing what others perceive or don't perceive, what they will do through action observation and what they should do, and aligning their own actions with those of another person in time and space [45]. The spatial and temporal alignment of multiple persons' actions can involve a few body parts (e.g. only hands or arms) or the entire body.

Research on joint actions that are restricted to hands and arms has often focused on pick-and-place tasks in VR [1, 25], analyzing and modeling the choice of passing and not passing an object to the partner and predicting user behavior in several joint pick-and-place tasks. The decision of whether to pass the object or not was found to be primarily influenced by user-target distance [25]. Bunlon et al. [6] demonstrated that the partner's hand appearance (robotic

human-like) did not influence the effectiveness of the studied joint action in a VE. Wang et al. [56] proposed a collaborative model that describes the cooperative behavior of a human dyad when pushing a virtual object using a haptic interface, finding that the dyad achieves the best performance when the leader takes more responsibility than the follower.

Prior research that is most relevant to our work focuses on joint action involving large, full-body movements. In one of the first experiments on this type of joint action in VR, Streuber et al. [49] analyzed the extent to which two users optimize their walking behavior while walking individually and jointly connected by a ladder. The task effort was shown to be split equally between the leader and the follower, thus suggesting the existence of a virtual joint body. Later, Tarr et al. [50] investigated how synchronized movements affect social attitudes and behaviors within groups made of users embodied as avatars and virtual agents. The results showed that participants moving in synchrony reported higher levels of social closeness to the agents than those moving in non-synchrony.

Joint actions in VR between human-human and human-agent conditions have been also explored for crossing roads in [17, 18]. The experiments showed that the users largely treated the virtual agents in the same way as real human partners when jointly crossing roads, independently of whether the partner was acting in a safe or risky modality. Buck et al. [4] analyzed the behavior of pairs of participants jointly passing through an aperture in the RE and in VR. By involving a huge and representative sample size (of more than 100 participants), the authors demonstrated that during real conditions, gender greatly affected the entering order, whereas no such effect was produced in VR. Evaluating triadic jumping in the real and VR conditions, Naito et al. [32] found no differences in the execution order between the RE and VR.

## 2.3 Proxemics in Collaborative VEs

Our focus is on collaborative work in MR with tasks that require sharing of the virtual, and sometimes also the physical space; therefore, we consider how people regulate shared space. The proxemics theory explores the influence of space and distance on interpersonal relationships [14] and studies how people perceive and use their spatial organization cues to mediate interactions. Recent works indicate that proxemics plays an important role in interactions, including studies on robotics [21], VR [3, 29], and improving interaction for users with visual impairments [16].

Studies of collision avoidance in VR are of particular interest to this work. Bühler et al. [5] analyzed how pedestrians regulate interpersonal distances in real and virtual conditions, showing that users maintained greater distances from others in VR. Similar outcomes can be found in [42]. Podkosova et al. [39] presented a study on collision avoidance in the real world and in VR; the physical location of users immersed in VR varied between a collocated and a distributed setup. Users kept further apart from each other and walk slower in the collocated VR condition than in distributed VR and the real scenario. Ríos et al. [41] observed similar effects; also, when animations and sounds were played in sync with avatar movements, users were closer to each other in VR.

# 3 USER STUDY

## 3.1 Study Design and Tasks

Our experiment was conducted with pairs of users and had a mixed 2 x 4 design. **Physical Setup** (**Distributed** vs **Collocated**) was a between-subject factor: each pair of participants either shared or did not share the physical workspace. **Group Immersion** was a within-subject factor with four levels that resulted from the combinations of the immersion type of each participant from a pair. These levels are **VR+VR** when both participants were immersed into VR; **AR+AR**, when both participants were immersed into AR; **VR+AR** and **AR+VR**, when one of the participants was in VR and the other one in AR. The latter two conditions are identical if seen on the group level and represent a typical MR setup. We opted for a full factorial design to be able to distinguish between the effects of group immersion and individual immersion (see 3.3.4); therefore we distinguish between **AR+VR** and **VR+AR** in our analysis. Each pair of participants experienced all four levels of **Group Immersion**, counter-balanced with Latin Square.

Participants were asked to perform two types of collaborative joint action tasks in the VE while holding a virtual *Rope*. The rope was introduced to strengthen the necessity of joint action and emphasize the collaborative nature of the tasks. To hold the rope, each participant from a pair could grab one end of the rope marked with a cube. Each participant would always hold their allocated cube, colored in green in their application view. The rope could stretch until a certain maximum distance ($1.2m$), meaning that users had to stay close enough to each other to perform the tasks. If a user tried to stretch the rope too far, this user's cube snapped from their hand and the user had to grab the cube again. We designed two different joint action tasks - **Gate Task** and **Fruit Task**.

In **Gate Task**, participants had to pass through a virtual aperture together (Figure 1). The aperture's width had two levels: **Narrow Gate** where participants could not pass simultaneously ($0.70m$) and **Wide Gate** where participants could pass side-by-side ($1.40m$). A virtual arrow on the ground under the gate indicated the direction in which to cross the gate. Participants were instructed to avoid colliding with the gate.

In **Fruit Task**, participants had to use the rope they were holding to slice a fruit that was moving towards them (Figure 1). To do it, participants had to make the rope collide with the fruit. However, the fruit could only be cut if the rope was sufficiently stretched (to the length of at least $0.85m$, threshold empirically set during the development); when this happened, participants saw two halves of the sliced fruit falling to the ground. If the rope was too limp when it collided with the fruit, the fruit disappeared but was not counted as successfully cut, and no fruit halves were visible. This way, participants from a pair had to achieve just the right balance in how far they stretched the rope to cut the fruit in order to avoid the rope from being snapped.

Participants performed 9 **Fruit** and 12 **Gate** tasks in each experimental block corresponding to one condition of Group Immersion. The sequence of tasks in each experimental block was randomized, with **Gate** and **Fruit** tasks mixed together. Prior to each task, both participants from a pair had to take specific starting positions, by standing on a pair of marked spots displayed in the environment. Starting positions were introduced to ensure that the distance that

participants needed to cover to approach a gate or a fruit was comparable between multiple task repetitions. The starting positions formed a 5x5 m large square within the environment. The gate in every **Gate** task appeared in the middle of this square, and the fruit in every **Fruit** task was moving towards the center of the square from the left, right, or opposite of where the participants were standing. The first pair of starting positions in a block was fixed to one side of the square, while all subsequent starting positions were randomized. This spatial arrangement and the randomization of the task order were designed to prevent the impression of repetitiveness that might develop after several tasks. Figure 2 shows users at reposition and during **Gate** and **Fruit** tasks in **VR+VR** condition, and various user views are shown in Figure 1.

The embodiment was achieved with either a video see-through of a user's own body or with full-body virtual avatars (Figure 3) animated with Inverse Kinematics (IK). A virtual avatar was always used as a self-avatar when a participant was immersed in VR, and a video see-through view of their own body was always visible when they were immersed in AR. In the latter case, the rendering of the self-avatar was disabled but the avatar object was used to trigger events in the VE. The representation of the collaboration partner as seen by each participant from a pair depended on **Physical Setup**. In the **Distributed** condition, the other user was always seen as a fully animated virtual avatar. In the **Collocated** condition, the other user was seen as a virtual avatar in VR and as a video see-through view of the other user in AR. We used two male and two female virtual avatars.

## 3.2 Hypotheses

Based on our analysis and previous research, we formulate hypotheses for the influence of immersion, physical setup, and their combinations on task performance (HTP), spatial behavior (HSp), and co-presence (HCoPr). Previous work indicates that collocated users might be faster in performing spatial tasks than remotely connected users due to common spatial references [31]. Therefore, we hypothesize **[HTP1]**: Task performance will be better for **Collocated** groups than for **Distributed** groups. Further, **[HTP2]**: **Group immersion** will lead to differences in task performance. Physical collocation leads to more careful spatial behavior in previous works [38, 39]; in accordance with these results, **[HSp1]**: **Collocated** groups will display more careful spatial behavior than **Distributed** ones. Furthermore, users of previous studies often kept larger interpersonal distances in the real world compared to VR [5, 39]. Since AR provides similar spatial references to the real world, we propose **[HSp2]**: VR will lead to more careful spatial behavior than AR. Due to the effects of physical proximity, we propose **[HCoPr1]**: Co-presence will be higher in **Collocated** compared to **Distributed** groups. Since real user representation leads to higher co-presence than avatar-based one in previous work [60], we formulate **[HCoPr2]**: AR will lead to higher co-presence than VR for **Collocated** groups.

## 3.3 Measures

*3.3.1 Spatial analysis and task performance.* During each task, we recorded positions and orientations of users' HMDs, controllers, and all trackers as well as timestamps in every frame. The start and end time of the task was recorded as well. A task started when the task object (the gate to go through or the fruit to slice) was spawned and ended when the task object disappeared and the starting positions for the next task were displayed in the environment. Several types of events were recorded as well: rope losses (when a user stops pressing the trigger to hold the rope) and rope snaps (when a user stretches the rope too much and it snaps away from their hand), successful slicing of the fruit, user collision with the gate, users entering and exiting the gate. We then computed the following metrics to evaluate joint action.

Regarding the **Fruit Task**, *Percentage of cut fruits* is the ratio between the number of trials in which the fruit was sliced successfully to the number of all fruit trials. *Duration* of **Fruit Task** was also computed. *Time of fruit cut* is the time at which the fruit was successfully sliced since the beginning of the task trial. *Distance walked to cut fruit* is the distance (averaged between two users from a group) that users walked from their starting position before slicing the fruit.

Regarding **Gate Task**, we computed *Duration* of **Gate Task** and *Time in gate* as performance metrics, which corresponds to the time it took a pair of participants to pass the gate (starts when the first user enters the gate and ends when the second user leaves the gate). *Number of gate collisions* is the number of all collisions with gates during one experimental block. Regarding proxemics of the joint action, *Average player distance in gate* is the average distance between two users from a pair while they are crossing the gate. This metric is calculated by computing the distance between two users in all frames in which at least one of them is in the gate and taking the average value. *Average head rotation difference in gate* is the average angle (along the up-axis) between the forward vectors of users' HMDs while they are crossing the gate. This metric reflects how much participants from a pair looked at each other while crossing the gate. To compute this metric, we calculate the angle between the HMDs' forward vectors in every frame while at least one player is in the gate and compute the average value. If players are looking in approximately the same direction, the difference angle will be close to 0°. If users are looking at each other, the angle will be close to 180°. This way, the closer the average head rotation difference to 180°, the more frequently users looked at each other while in the gate. *Average pelvis rotation difference in gate* is the average angle between forward vectors of user avatars; pelvises in the gate. This metric is calculated in the same way as *average head rotation difference in gate* but by taking the forward vector of the pelvis bone in each user's avatar instead of the forward vector of the HMD. This metric reflects the spatial orientation of the pair of users while they are crossing the gate, independently from head movements.

**Fruit Task** and **Gate Task**, *Number of rope snaps* is the number of times a user stretched the rope too much so that it snapped away from their hand. This metric characterizes the ease of spatial coordination since it depends on the distance between two rope ends held by users. *Number of rope losses* is the number of times a user stopped holding the rope by pressing the trigger button.

*3.3.2 Subjective metrics.* Participants filled in post-block questionnaires addressing their subjective perception of embodiment (short embodiment questionnaire, pESQ [11]), workload (NASA TLX [15]),

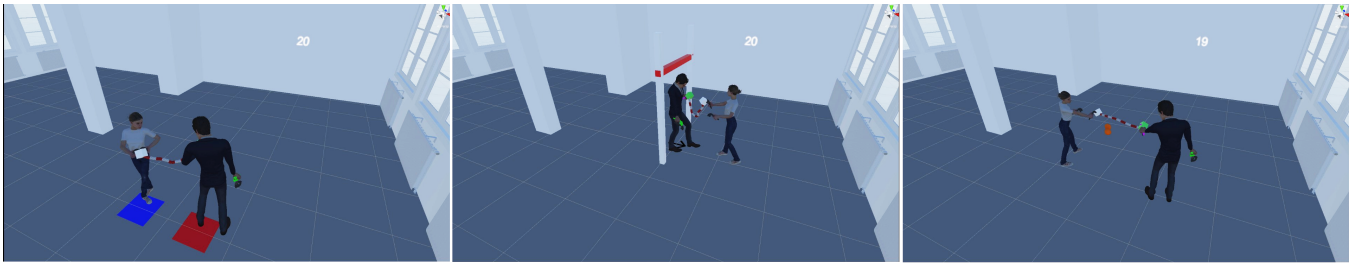

**Figure 2: Test views from Unity3D editor: taking starting positions (left); going through a gate (middle); cutting a fruit (right).**

**Figure 3: The four avatars used for the experiments.**

presence, co-presence, and collaboration (from existing studies [33]). The post-block questions are presented in Table 1. At the end of the experiment, participants answered questions about the relative ease of VR and AR for each task (*Ease-Fruit* and *Ease-Gate*) and their preferred immersion interface (*Pref-Setup*) and provided free-form comments.

*3.3.3 Simulator Sickness Questionnaire.* We administer the Simulator Sickness Questionnaire (SSQ) before the experiment and after each block and compute the pre- and post-block SSQ score accordingly to the methodology described in [19]. We also computed a delta SSQ score for each scale (i.e., the post-block score minus the pre-block score) to gain insights into cybersickness variations after each block.

*3.3.4 Group and individual analysis.* To analyze the effect of immersion type in detail, we distinguish between **Group Immersion** (**VR+VR**, **AR+AR**, **VR+AR** and **AR+VR**) and **Individual Immersion**, which accounts for the immersive interface (VR or AR) each user from a pair had in each of the four group conditions. Since VR and AR were repeated twice for each user, we account for the group setup in the individual measure to perform repeated-measures statistical tests. In the result, our condition **Individual Immersion** has four levels: **AR**, when the target user was in AR and their partner also in AR; **AR-M**, when the user was in AR and their partner in VR (M for "Mixed" group setup); **VR** and **VR-M**.

Some of the metrics described above reflect joint action and can be calculated on the group level only. Other metrics reflect individual actions or experiences and are calculated individually for each participant from a pair. Individual metrics are: *Number of collisions with gate*, *Number of rope losses*, *Number of rope snaps*, and all questionnaire items. All other described metrics related to **Fruit Task** and **Gate Task** are group metrics.

## 3.4 Participants and Apparatus

The user study took place in a large room (12x12 m); 46 users (13 female and 33 male, 23.32±6.8, mean±SD) participated. 12 pairs of participants (24 users) were assigned to the **Distributed** condition of **Physical Setup**, 11 pairs (22 users) to the **Collocated** one. Most of the participants had a strong experience with video games and had experienced VR or AR at least once; however with a low experience of HMD-based AR. Participants signed an informed consent form and were naive to the purpose of the experiment. The study conformed with the standards of the Declaration of Helsinki. All participants finished the experiment without any withdrawal.

The large room in which the experiment took place was divided into two sub-rooms (7x6 m each) with a thick curtain. These two sub-rooms were used as individual workspaces for the participants in the **Distributed** condition. In the **Collocated** condition, one of the sub-rooms was used as the shared workspace. We used HTC Vive Pro HMDs with Vive Wireless Adapter[1] to enable participants to freely walk inside the workspace. Each user was equipped with two HTC Vive controllers and three HTC Vive trackers to track the users' hands, feet, and pelvis, respectively. Three HTC Vive base stations were installed in each sub-room, thus providing reliable coverage of the workspace.

We developed the collaborative experimental platform with Unity3D (2019.4.3f1) and Photon Networking for Unity3D (PUN2) library for the networking functionality. The AR view was implemented with Vive SRWorks SDK for Unity3D[2] that uses two front-facing RGB cameras of the HTC Vive Pro to create a stereo background image of the environment for the video see-through effect (resolution was 480p with an average latency of 200ms). Each user application ran at a frame rate of at least 90Hz. The virtual environment consisted of a virtual replica of the experimental room. The virtual rope was designed with the ObiRope asset for Unity3D, and we used Mixamo avatars to provide virtual user embodiment. The IK solution used to animate the avatars was taken from the AvatarGo project [40].

## 3.5 Procedure

At the beginning of the experiment, each pair of participants was assigned to either **Distributed** or **Collocated** setup. Each pair performed four blocks of the experiment corresponding to four conditions of **Group Immersion**. Before starting the experiment, the participants were introduced to the user study, gave their written

[1]https://www.vive.com/eu/accessory/wireless-adapter/
[2]https://developer-express.vive.com/resources/vive-sense/srworks-sdk/

**Table 1: Post-block questionnaire including the following dimensions: embodiment (pESQ), presence (Pr), co-presence (CoPr), collaboration (Col), workload (TLX).**

| | |
|---|---|
| **Embodiment** - pESQ (5 Point Likert Scale) | |
| pESQ1 | Overall, I felt as if my body was located where I saw the virtual body to be. |
| pESQ2 | The movements of the virtual body were caused by my movements. |
| pESQ3 | I felt like my body was actually there in the environment. |
| pESQ4 | I felt like my bodily movements occurred within the environment |
| pESQ5 | I felt like the environment affected my body. |
| **Presence** - Pr (7 Point Likert Scale) | |
| Pr1 | There was a sense of being "really there" inside the current environment. |
| Pr2 | There were times during the experience when the real world of the laboratory in which the experience was really taking place, was forgotten. |
| **Co-Presence** - CoPr (7 Point Likert Scale) | |
| CoPr1 | The experience was more like working with other people rather than interacting with a computer |
| CoPr2 | During the time of the experience, I felt there was a sense of being with the other person. |
| CoPr3 | The experience resembled being together with another person in a real-world setting. |
| CoPr4 | During the time of the experience, I forgot about the other person and concentrated on the task as if I was the only one. |
| **Collaboration** - Col (5 Point Likert Scale) | |
| Col1 | I could understand what my partner was trying to accomplish by looking at their body movements. |
| Col2 | I enjoyed the experience in a similar manner to a previous real meeting that was enjoyable. |
| **Nasa-TLX** - TLX (10 Point Likert Scale) | |
| TLX-Mental | How mentally demanding was the task? |
| TLX-Physical | How physically demanding was the task? |
| TLX-Temporal | How hurried or rushed was the pace of the task? |
| TLX-Performance | How successful were you in accomplishing what you were asked to do? |
| TLX-Effort | How hard did you have to work to accomplish your level of performance? |
| TLX-Frustration | How insecure, discouraged, irritated, stressed, and annoyed were you? |

consent to participate in the experiment, and filled out a demographics questionnaire (age, gender, amount of experience playing video games, and exposure to VR and AR) and pre-test SSQ. Then, the participants were equipped with the hardware and performed a training scene for around 5 minutes. In the training scene, they could get familiar with the avatars, the rope behavior, and the fruit and gate task performed once each. The training scene was always done in the **VR+VR** condition of **Group Immersion**. After the training scene, the first study block started.

Each block contained the following steps: (1) Users calibrated their avatars to their height and body dimensions. To do it, each participant took a T-pose posture, pressed the trigger button to spawn their self-avatar, then stepped "inside" this avatar and aligned the position of their feet, hands, and head with the model. On the second trigger press, the alignment was confirmed and the participant was embodied. (2) Once the calibration was done, users could see each other in the VE (in the **Collocated** setup, users immersed in AR could see their interaction partners from the start of the scene). Users could grab the rope by touching their end cube with a controller and holding the trigger button. (3) The first pair of starting positions appeared and the users walked to stand on them. When they reached the starting positions, the first task started. (4) Users performed the task, after which a new pair of starting positions appeared. This was repeated until all tasks of the block were done. (5) Users removed their HMDs and completed the post-block SSQ and the post-block questionnaire using two dedicated laptops. After the last block, participants filled out the post-experiment questionnaire and were debriefed about the purpose of the study. Participants were not aware that time was recorded, thus they could interact naturally without rushing to complete the tasks. Moreover, they could talk to each other in both setups without using any kind of device.

## 4 RESULTS

This section reports the results of statistical tests related to our metrics described above. For normally-distributed metrics (that were assessed using the Shapiro-Wilk test), we performed a Mixed analysis of variance (ANOVA) with repeated-measures factors specified separately for each metric below and **Physical Setup** as the between-subject factor in all cases. Greenhouse-Geisser adjustments to the degrees of freedom were applied when the sphericity assumption was violated. For metrics with distributions deviating from normal, we used the non-parametric Aligned Rank Transform (ART) test [57].Post-hoc analysis was based on pairwise t-tests with Bonferroni corrections when the distribution of the dependent variables was normal or the procedure for multifactor contrast tests presented in [8]. In the interest of brevity, we report only statistically significant findings including size effect with eta-square value. When relevant, some non-statistically significant findings are reported. Table 2 sums up the main results presented in this section. For further details, please refer to the following subsections.

**Table 2: Main results found during the experiment.**

| Metric | Result |
|---|---|
| **Fruit** | |
| *Percentage of fruit cut* | **AR+AR** > **VR+VR** |
| *Trial duration* | **VR+VR** > **AR+AR** |
| **Gate** | |
| *Player distance in gate* | **Distributed** > **Collocated**; **Narrow** > **Wide** |
| *Time in Gate* | **Narrow** > **Wide** |
| *Head rotation difference in gate* | **Collocated** > **Distributed**; **Wide** > **Narrow** |
| *Pelvis rotation difference in gate* | **Collocated** > **Distributed**; **Narrow** > **Wide** |
| **Rope** | |
| *Number of rope losses* | **VR** > **AR**; **Fruit** > **Gate** |
| *Number of rope snaps* | **Fruit** > **Gate** |
| **Questionnaire** | |
| *Embodiment* | **AR** > **VR** |
| *Co-presence* | **Collocated** > **Distributed**; **AR** > **VR** |
| *Nasa-TLX-Effort* | **Distributed** > **Collocated** |

### 4.1 Spatial analysis and task performance

*4.1.1 Fruit Task.* The reported metrics were analyzed with Mixed ANOVA with **Group Immersion** as the repeated-measures factor. We found a statistically significant effect of **Group Immersion**

on the *Percentage of cut fruits* ($F_{3,57} = 4.39$, $p < 0.05$, $\eta_p^2 = 0.19$). Specifically, participants cut a higher percentage of fruits in **AR+AR** ($M = 76.0\%$; $SD = 15.87\%$) than in **VR+VR** ($M = 60.35\%$; $SD = 20.16\%$) as observed in pairwise comparisons. The average trial *Duration* was higher in **VR+VR** ($M = 6.002$sec; $SD = 0.22$sec) than in **AR+AR** ($M = 5.16$sec; $SD = 0.16$sec), in the post-hoc of Mixed ANOVA with a significant effect of **Group Immersion** ($F_{3,54} = 5.82$, $p < 0.05$, $\eta_p^2 = 0.24$). Figure 4 illustrates these results.

*4.1.2 Gate Task.* The reported metrics were analyzed with Mixed ANOVA with **Gate Width** and **Group Immersion** as the repeated-measures factors. **Gate Width** ($F_{1,17} = 17.62$, $p < 0.001$, $\eta_p^2 = 0.51$), **Physical Setup** ($F_{1,17} = 20.17$, $p < 0.001$, $\eta_p^2 = 0.54$) and the interaction of **Gate Width** x **Physical Setup** ($F_{1,17} = 4.99$, $p < 0.05$, $\eta_p^2 = 0.23$) had a statistically significant effect for *Average player distance in gate*. Specifically, participants in the **Distributed** setup ($M = 1.64$m; $SD = 0.06$m) were further apart from each other than participants in the **Collocated** setup ($M = 1.28$m; $SD = 0.06$m), and further apart in **Narrow Gate** ($M = 1.55$m; $SD = 0.04$m) than in **Wide Gate** ($M = 1.38$m; $SD = 0.05$m). The difference of *Average player distance in gate* between **Wide Gate** and **Narrow Gate** is larger for **Distributed** groups (**Narrow Gate** $M = 1.41$m; $SD = 0.06$m, **Wide Gate** $M = 1.15$m; $SD = 0.07$m) than for **Collocated** ones (**Narrow Gate** $M = 1.68$m; $SD = 0.06$m, **Wide Gate** $M = 1.60$m; $SD = 0.07$m). For *Time in Gate*, **Gate Width** had statistically significant effect ($F_{1,19} = 67.12$, $p < 0.001$, $\eta_p^2 = 0.78$) - it took participants longer to pass through **Narrow Gate** ($M = 3.42$sec; $SD = 0.16$sec) than through **Wide Gate** ($M = 2.45$sec; $SD = 0.15$sec).

For **Gate Width** ($F_{1,17} = 561.62$, $p < 0.001$, $\eta_p^2 = 0.97$), **Physical Setup** ($F_{1,17} = 12.69$, $p < 0.05$, $\eta_p^2 = 0.43$) and the interaction **Gate Width** x **Physical Setup** ($F_{1,17} = 10.53$, $p < 0.05$, $\eta_p^2 = 0.38$) had significant effect for *Average head rotation difference in gate*. Users looked at each other much more in **Wide Gate** ($M = 112.49°$; $SD = 4.75°$) compared to **Narrow Gate** ($M = 19.92°$; $SD = 1.10°$), and more in **Collocated** setup ($M = 73.83°$; $SD = 4.06°$) than in **Distributed** ($M = 51.67°$; $SD = 3.46°$). The difference between **Narrow Gate** and **Wide Gate** is smaller for **Distributed** groups (in **Narrow Gate** $M = 16.14°$; $SD = 1.43°$, **Wide Gate** $M = 87.19°$; $SD = 5.71°$) than for **Collocated** ones (**Narrow Gate** $M = 23.70°$; $SD = 1.68°$, **Wide Gate** $M = 123.95°$; $SD = 6.69°$).

For *Average pelvis rotation difference in gate*, **Gate Width** ($F_{1,17} = 8.91$, $p < 0.05$, $\eta_p^2 = 0.34$), **Physical Setup** ($F_{1,17} = 6.56$, $p < 0.05$, $\eta_p^2 = 0.12$) and the interaction **Gate Width** x **Physical Setup** ($F_{1,17} = 7.15$, $p < 0.05$, $\eta_p^2 = 0.30$) have a significant effect. Users were more oriented towards each other in **Collocated** setup ($M = 116.63°$; $SD = 5.02°$) than in **Distributed** ($M = 74.20°$; $SD = 4.28°$), and in **Narrow Gate** ($M = 102.82°$; $SD = 4.38°$) compared to **Wide Gate** ($M = 88.01°$; $SD = 3.87°$). The difference between **Narrow Gate** and **Wide Gate** is much larger for **Distributed** groups (**Narrow Gate** $M = 88.24°$; $SD = 5.68°$, **Wide Gate** $M = 60.16°$; $SD = 5.02°$) than for **Collocated** ones (**Narrow Gate** $M = 117.40°$; $SD = 6.66°$, **Wide Gate** $M = 115.86°$; $SD = 5.88°$) where the average values are nearly the same.

*Number of gate collisions* was low in all conditions (Median = 2, summed over all gate trials) and was not affected by any condition. *Duration* of **Gate Task** was not different in any condition either ($M = 6.72$sec; $SD = 2.32$sec). We did not observe any influence of **Individual Immersion** on leader-follower behavior. In the majority of groups, one of two users was the first one to go through the gates in most cases. The discussed proxemics-related effects of **Gate Task** are shown in Figure 4.

*4.1.3 Rope losses and rope snaps.* We used the ART test with **Task Type** (**Fruit** vs **Gate**) and **Individual Immersion** as repeated-measures factor to analyze them on the individual level. For *Number of rope losses*, we found a significant effect of **Individual Immersion** ($F_{3,126} = 7.64$, $p < 0.001$, $\eta_p^2 = 0.15$), **Task Type** ($F_{1,42} = 35.93$, $p < 0.001$, $\eta_p^2 = 0.46$).Post-hoc showed that there were more rope losses per trial in **VR** ($M = 2.0$; $SD = 3.1$) than in **AR** ($M = 1.0$; $SD = 2.4$) and in **Gate Task** ($M = 1.9$; $SD = 3.0$) than in **Fruit Task** ($M = 0.7$; $SD = 1.4$). For *Number of rope snaps*, we found a significant effect of **Task Type** ($F_{1,42} = 147.00$, $p < 0.001$, $\eta_p^2 = 0.77$), where post-hoc tests showed that there were more rope snaps per trial in **Fruit Task** ($M = 2.5$; $SD = 1.9$) than in **Gate Task** ($M = 0.6$; $SD = 1.0$).

## 4.2 Subjective questionnaires

*4.2.1 Post-block questionnaire.* Figure 5 shows box-plots for all scores of post-block questions. Regarding the pESQ, we found a significant effect of **Individual Immersion** on pESQ1 ($F_{3,132} = 0.07$, $p < 0.05$, $\eta_p^2 = 0.06$), where post-hoc showed that scores were higher in **AR** ($M = 4.3$; $SD = 0.7$) than in **VR** ($M = 3.9$; $SD = 1.0$). This was also observed on pESQ2 ($F_{3,132} = 4.28$, $p < 0.01$, $\eta_p^2 = 0.09$), where in post-hoc analyses scores were higher in **AR** ($M = 4.6$; $SD = 0.6$) than in **VR** ($M = 4.3$; $SD = 0.8$), pESQ3 ($F_{3,132} = 5.76$, $p < 0.001$, $\eta_p^2 = 0.11$) where scores were higher in **AR** ($M = 4.3$; $SD = 0.8$) than in **VR** ($M = 3.9$; $SD = 0.7$), and pESQ4 ($F_{3,132} = 3.89$, $p < 0.05$, $\eta_p^2 = 0.08$) where post-hoc analyses did not show differences between levels of **Individual Immersion**. For pESQ5, we found a significant effect of **Physical Setup** ($F_{1,44} = 4.08$, $p < 0.05$, $\eta_p^2 = 0.08$), where in post-hoc analyses scores were higher for **Collocated** groups ($M = 3.5$; $SD = 1.3$) than for **Distributed** ones ($M = 2.6$; $SD = 1.5$).

Regarding Presence and Co-Presence, there was no significant effect of **Physical Setup** ($F_{1,44} = 0.45$, $p$ ==0.49) or **Individual Immersion** ($F_{3,132} = 0.80$, $p$ ==0.50) on Pr1. However, we found an effect of **Individual Immersion** on Pr2 ($F_{3,132} = 29.98$, $p < 0.001$, $\eta_p^2 = 0.40$), where post-hoc showed higher scores in **VR** ($M = 5.2$; $SD = 1.7$) than in **AR** ($M = 3.2$; $SD = 1.9$). We found a significant effect of **Physical Setup** ($F_{1,44} = 3.41$, $p < 0.05$, $\eta_p^2 = 0.07$), **Individual Immersion** ($F_{3,132} = 3.70$, $p < 0.05$, $\eta_p^2 = 0.07$) and an interaction effect ($F_{3,132} = 2.95$, $p < 0.05$, $\eta_p^2 = 0.06$) on CoPr2. In post-hoc analyses, scores were higher in the **Collocated** setup ($M = 6.2$; $SD = 1.1$) than in the **Distributed** one ($M = 5.6$; $SD = 1.2$), higher in **AR** ($M = 6.0$; $SD = 1.2$) than in **VR** ($M = 5.6$; $SD = 1.1$), and that highest scores were reached in **Collocated VR** ($M = 6.09$; $SD = 0.9$) and lowest in **Distributed AR** ($M = 5.4$; $SD = 1.1$). We found a significant effect of **Physical Setup** ($F_{1,44} = 11.88$, $p < 0.01$, $\eta_p^2 = 0.21$) and **Individual Immersion** ($F_{3,132} = 5.79$, $p < 0.001$, $\eta_p^2 = 0.11$) on CoPr3. In post-hoc analyses, scores were higher in the **Collocated** setup ($M = 5.3$; $SD = 1.4$) than in the **Distributed** one ($M = 4.0$; $SD = 1.4$), and higher in **AR** ($M = 4.9$; $SD = 1.6$) than in **VR** ($M = 4.4$; $SD = 1.5$).

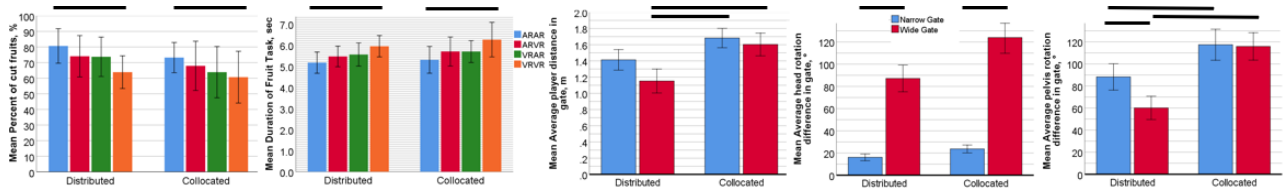


**Figure 4: The most prominent results of spatial analysis. Left: average percentage of cut fruits and fruit trial time. Right: average player distances and head and pelvis rotation differences in the gate. The black bars indicate pairwise comparisons.**

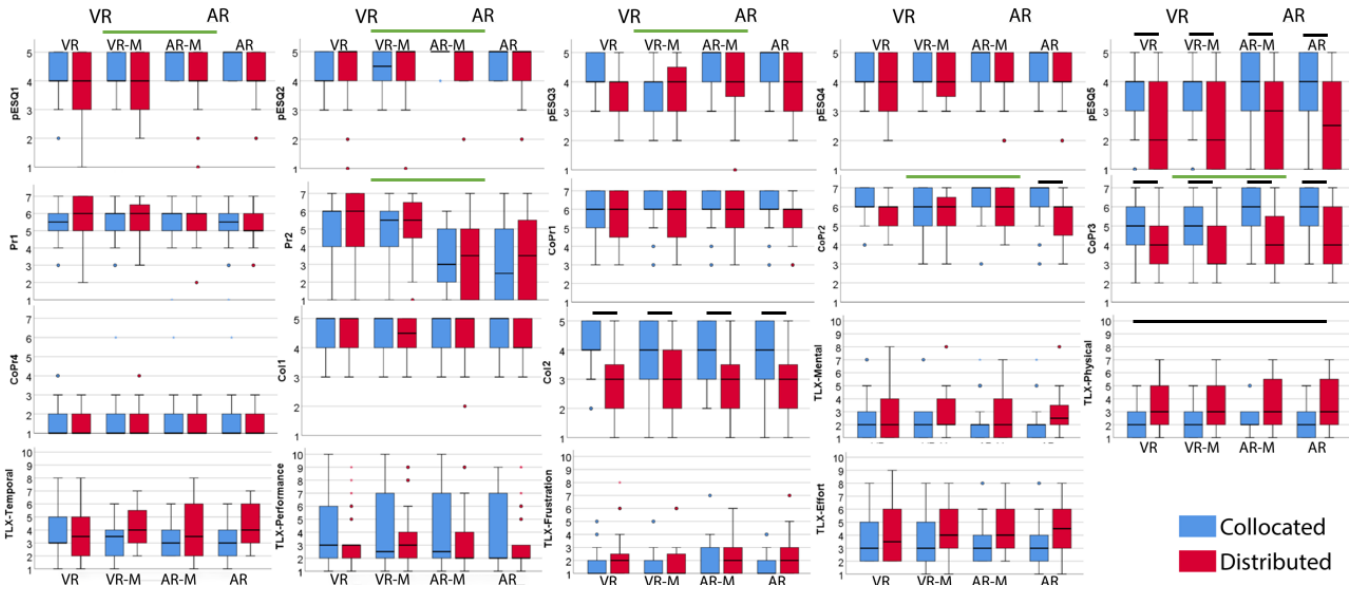


**Figure 5: Boxplot for each question from the post-block questionnaire (Table 1) per Individual Immersion and Physical Setup. The black bars indicate pairwise comparisons and the green bar an effect of the Individual Immersion.**

Regarding collaboration, we did not find any significant effect of **Physical Setup** ($F_{1,44} = 0.21$, $p$ ==0.64) or the **Individual Immersion** ($F_{3,132} = 0.80$, $p$ ==0.49) on Col1. However, we found a significant effect of **Physical Setup** ($F_{1,44} = 10.51$, $p < 0.01$, $\eta_p^2 = 0.19$) on Col2, where post-hoc analyses showed higher scores for **Collocated** groups ($M = 3.8$; $SD = 1.1$) than for **Distributed** ones ($M = 2.9$; $SD = 1.1$).

Regarding Nasa-TLX scores, there was a significant effect of **Physical Setup** ($F_{1,44} = 8.18$, $p < 0.01$, $\eta_p^2 = 0.15$) on TLX-Physical, with higher scores in **Distributed** setup ($M = 3.6$; $SD = 1.8$) than **Collocated** ($M = 2.1$; $SD = 1.1$). We found an interaction effect of **Physical Setup** x **Individual Immersion** ($F_{3,132} = 2.94$, $p < 0.03$, $\eta_p^2 = 0.06$) on TLX-Effort, where scores were the highest for **Distributed AR** ($M = 4.5$; $SD = 2.0$) and the lowest for **Collocated VR** ($M = 3.5$; $SD = 1.7$).

*4.2.2 Post Experiment Questionnaire.* Figure 6 shows the distribution of answers for the post-questionnaires. A chi-square test on *Pref-Setup* showed that answer distributions for **Collocated** and **Distributed** groups were independent from each other ($\chi^2(8.63) = 2$, $p < 0.05$). For *Ease-Gate* and *Ease-Fruit*, answer distributions of

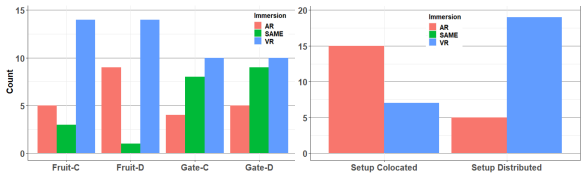


**Figure 6: Post-questionnaire histogram for the Ease-Fruit, Ease-Gate, and Setup-Pref per Immersion and Physical Setup.**

**Collocated** and **Distributed** groups were dependent in the chi-square test ($\chi^2(2) = 0.08$, $p > 0.05$ for *Ease-Gate*, $\chi^2(2) = 2.05$, $p > 0.05$ for *Ease-Fruit*).

*4.2.3 Simulator Sickness Questionnaire.* Table 3 reports the average and the standard deviation of delta SSQ scores for each scale, grouped by **Individual Immersion**. We found a significant effect of **Individual immersion** on oculomotor scale $\chi^2(3) = 6.6$, $p < 0.05$. However, post-hoc did not show any difference between the conditions ($p > 0.05$).

**Table 3: SSQ scores per individual immersion for each scale.**

| | Nausea | Disorientation | Oculomotor | Total |
|---|---|---|---|---|
| VR | 1.45±11.00 | 1.10±18.32 | 3.95±14.08 | 5.40±82.31 |
| VR-M | 3.11±10.25 | 2.11±28.14 | 3.79±14.35 | 18.41±124.39 |
| AR | 3.94±13.42 | 4.53±23.85 | 9.88±16.75 | 30.80±106.01 |
| AR-M | 4.52±23.25 | 2.11±24.54 | 5.93±17.86 | 10.11±103.06 |

# 5 DISCUSSION

## 5.1 Effects of Physical Setup

We did not observe any differences in task performance metrics between **Collocated** and **Distributed** groups. The average task duration for both **Fruit Task** and **Gate Task** was independent of whether users shared the physical space or were distributed, contrary to our expectation of faster completion times in **Collocated** setup that was reported in previous research [31]. Users collided with gates very infrequently in all conditions - only about twice during 12 gate tasks. It appears that users had sufficient spatial references to pass through the gate in all setups. The performance of fruit slicing was not affected by **Physical Setup** either. Our hypothesis **[HTP1]** has to be rejected. Although task performance itself was not affected, we found some influence of **Physical Setup** on the workload measures. The physical workload was judged slightly higher in the **Distributed** setup. These results need further investigation and we would suggest focusing on two research axes: (1) task performance might depend on how well people know each other [46] and (2) task performance could be influenced by using virtual agents in CMR [24].

Participants were more careful in the **Collocated** setup than in the **Distributed** one: they kept larger distances to each other while crossing the gate, looked at each other, and were rotated towards each other more during **Gate Task**. More careful spatial behavior in the **Collocated** setup is also reflected in the differences in metrics related to the width of the gate. To cross a **Narrow Gate**, two strategies were employed. Frequently, the first user to cross the gate would walk forward, and rotate to look at the second user after stepping through the gate. Alternatively, the first user could start crossing the gate backward, while looking at the collaborator and the rope. With **Wide Gate**, participants sometimes chose to walk side by side, which resulted in them being closer to each other while oriented in the same direction. This lead to differences in **Average player distance** and **Average pelvis rotation difference** between **Narrow Gate** and **Wide Gate** in the **Distributed** condition. These differences were much smaller in the **Collocated** condition, showing that users choose the same, safer technique of going one after another to cross both types of gates. All these results confirm **[HSp1]**, in line with previous research, but additional factors such audio-visual cues [28] could also influence spatial behavior between participants and should be investigated.

The **Collocated** setup scored higher on two co-presence items, with users judging it to provide more of a sense of being with another person. In addition, one of the collaboration items (*Col1*) scored higher in the **Collocated** setup and the reported enjoyment was slightly lower in the **Distributed** setup. **[HCoPr1]** is confirmed. **Physical Setup** had an effect on user preferences concerning the type of immersion: participants in **Distributed** groups preferred VR to AR, while in the **Collocated** groups AR was preferred., We cannot offer a definitive explanation of this difference in the immersion type preference; but one plausible interpretation is that AR allowed a better understanding of the spatial arrangement and therefore more security, which was needed in the **Collocated** setup. We analyzed the chosen immersion preference with the help of user comments with both positive and negative feedback. In free-form comments regarding VR, the main positive remark is that VR looked "clean" (i.e., without noise and smooth), more immersive, and the full body avatar helped to understand motion. The negative comments were that sometimes the avatar IK pose was not entirely correct and that some users were reluctant to move fast to cut the fruit because they could not see the real physical boundaries of the workspace. Regarding AR, the main positive comment was that it provided good spatial judgments. However, it felt more exhausting to act in AR because of the blurry rendering of the video see-through. The blurriness of AR has indeed resulted in slightly higher oculomotor *SSQ* results, most probably due to its higher latency and lower quality, and the rest frame theory [26] (i.e., the discrepancy between the RE and the virtual fruit motion could cause instability of representation between stationary and moving objects).

## 5.2 Effects of Immersion Type

During the **Fruit task**, the percentage of cut fruit was higher and the fruits were cut faster in **AR+AR** than **VR+VR**. We did not see any striking effect of the mixed setup; mixed **Group Immersion** lead to results that were in between **VR+VR** and **AR+AR**. We suggest that AR provided better spatial references to move and guide the other person to cut the fruit, resulting in the worst performance in **VR+VR**. Yet, users found it was easier to cut the fruit in **VR** as shown in Figure 6, revealing an inconsistency between the objective of users and their subjective experience. In the **Gate task** no performance differences in terms of completion times or number of collisions were found. We then partially confirm **[HTP2]**, where **AR+AR** groups performed better than **VR+VR groups** in the **Fruit task**. Previous work showed similar results for single-user tasks [2, 22] but we are aware that AR may always lead to higher task performance than VR. The interesting result in our study is that asymmetric setups (**VR+AR**) were never providing the highest task performance but also never the worst. Future work could investigate more about how performance could be improved in asymmetric setups by checking whether the task performance is affected by the **VR** or **AR** user.

The number of times when the rope snapped out of the user's hand and when the user simply left the rope can be attributed to the difficulty of the task. *Number of rope snaps* is a measure of spatial coordination between two users; it is not surprising that they occurred more frequently when users needed to coordinate quickly in the **Fruit Task**. The significance of *Number of rope losses* is not as evident; they can either be a result of tiredness from having to carry the rope around all the time or a manifestation of the mental load of the task (users have to pass through the gate while *also* carrying the rope). Naturally, there were fewer rope losses in **Fruit Task**, where user attention was focused on holding the rope in a good way. **Gate Task**, on the other hand, took longer and required a more sophisticated trajectory with the focus on another object in

the scene (the gate), so participants were more likely to both forget about the rope and to change hands while holding it. However, it is not so clear why more rope losses occurred in **VR** than in **AR**. It might be a sign of greater workload or difficulty in managing the spatial arrangements of bodies and held objects in VR.

Similarly to **[HSp1]**, we wanted to investigate in **[HSp2]** which immersion type will provoke more careful spatial behavior. There was no effect regarding head and pelvis rotations on **Group immersion**, as well as the distance between players, showing that spatial behavior was similar in all combinations of VR and AR immersion. Our assumption that spatial behavior in AR would be similar to that of the real world and thus different from VR did not hold; we thus reject **[HSp2]**. One reason could be the low resolution of the AR immersion that did not reproduce a close enough real-life situation. There was no effect of immersion on leadership behavior in crossing the gate. While previous work found that the most immersed user usually emerges as the leader, our result is not surprising since our setup in fact does not make one or another user more or less immersed: the large FoV of HTC Vive Pro ensures that virtual objects are seen as clearly for the AR user as for the VR user. While the quality of AR was definitely inferior to that of VR, it did not result in any disadvantage for the AR user.

Higher scores for *Pr2* question were reported in **VR** than in **AR**. This was expected, as **VR** fully blocks the RE whereas **AR** provides a video see-through of the RE with a few virtual objects (a rope, gate, and fruit). It would be difficult to forget about the real location in this setup. Regarding co-presence, *AR* scored better than *VR* for the sense of being with the other person. However, no differences were observed regarding collaboration between AR and VR. These results are similar to a recent study [13], where authors argue that if feeling physically present with a teammate is essential, AR could be preferred over VR. Our results confirm **[HCoPr2]** at least partially, co-presence being higher in **AR** than in **VR**.

We decided to choose a short questionnaire over gold standard ones [34] to assess embodiment, as our main research question was not to focus on the embodiment entirely but rather on having insights about how the differences in the setup may impact embodiment. In the pESQ, pESQ1, and pESQ3 assess users' self-location, pESQ2 and pESQ4 users' agency and pESQ5 body ownership. Our results showed that participants felt higher self-location and agency in **AR** than in **VR**, whereas body ownership was similar across both immersion types. A real but low-quality representation of users with latency still yielded higher embodiment than the virtual avatars, a result similar to previous findings [60]. We suggest that the lack of personalization of our avatar may have also yielded lower scores in VR immersion [54]. According to our embodiment results, designers of CMR should consider choosing AR for users for whom self-location and agency are most important. It is worth noticing that the employed IK avatar solution is not fully accurate in computing the elbow's pose and thus, the sense of embodiment could have been negatively affected in the setups that allowed the users to see the virtual avatars. Yet, participants did not necessarily notice or complain about it.

We are aware that evaluating co-presence, presence, and embodiment in AR compared to VR is a difficult topic. For example, there is a distinction between presence in AR and the extensively studied notion of presence in VR, which raises uncertainties about the reliability of the conventional presence questionnaire when applied to AR scenarios [12]. While several works discussed this concern, few questionnaires for measuring presence in the spectrum of MR have been proposed [51]. Thus, further work and research should be considered to understand better how the difference between AR and VR immersion could impact users' experience in CMR.

## 6 LIMITATIONS AND FUTURE WORK

The analyses of individual immersion and location setup provided interesting insights about how such factors could influence users' collaboration in VEs. However, future work should investigate additional research to address a few limitations in our current work that we will describe hereafter. First, we are aware that our sample has some limitations: we did not reach a gender balance, and users mainly were familiar with each other since they were recruited as groups. However, our results still provide interesting insights into collaboration in VEs since we mainly collaborate with people we know. Further experiments should be conducted to understand better how gender and levels of users' familiarity (known vs. unknown people) could influence collaboration in MR. Our experiment only uses four avatars (two males and two females). Yet, the absence of avatar personalization could diminish users' sense of embodiment and identification with their virtual representations. Future work will consider users' ability to customize their avatars with features such as appearance or clothing to foster stronger body ownership, thus enhancing user immersion in the virtual space. In addition, future works will have to consider expanding the range of tasks and environments beyond spatial exploration. Incorporating tasks that involve cognitive load or assessing interactions in asynchronous settings would help researchers to understand the cognitive demands and social dynamics of collaboration in VEs. This also includes analyzing users' vision dysfunctions and the changes in the language spoken to assess whether they have an impact on task performance and embodiment. Considering the apparatus, while our study focused on consumers' HMD, we know that the see-through AR provided by the HTC Vive is not the best in terms of resolution and framerate. Thus, future work should consider assessing alternative AR devices such as the Zedmini or the Varjo XR-3 to see if the quality of the AR rendering could also affect user experience.

## 7 CONCLUSION

Factors that are important to CMR such as type of immersion and physical location have not been studied in conjunction with joint action. This paper proposed a study investigating the influence of these factors on task performance, spatial behavior of users, and their subjective perceptions in two types of joint action tasks. The main outcomes indicate that AR leads to better performance than VR for joint tasks where the temporal aspect is important. Moreover, independently of the immersion type, users perform joint actions more carefully in the collocated setup than in the distributed one. However, this is only a first step toward better understanding collaboration with MR setups during joint action tasks. This work opens new perspectives on how the interaction between the physical workspace is shared or not by users and the way they are immersed in the VE should be considered when designing CMR applications.

# REFERENCES


[1] Diana Babajanyan, Gaurav Patil, Maurice Lamb, Rachel W Kallen, and Michael J Richardson. 2022. I Know Your Next Move: Action Decisions in Dyadic Pick and Place Tasks. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 44.

[2] Andrew C Boud, David J Haniff, Chris Baber, and SJ Steiner. 1999. Virtual reality and augmented reality as a training tool for assembly tasks. In *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*. IEEE, 32–36.

[3] Lauren E Buck, Soumyajit Chakraborty, and Bobby Bodenheimer. 2022. The impact of embodiment and avatar sizing on personal space in immersive virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2102–2113.

[4] Lauren E Buck, John J Rieser, Gayathri Narasimham, and Bobby Bodenheimer. 2019. Interpersonal affordances and social dynamics in collaborative immersive virtual environments: Passing together through apertures. *IEEE transactions on visualization and computer graphics* 25, 5 (2019), 2123–2133.

[5] Marco A Bühler and Anouk Lamontagne. 2018. Circumvention of pedestrians while walking in virtual and physical environments. *IEEE transactions on neural systems and rehabilitation engineering* 26, 9 (2018), 1813–1822.

[6] Frédérique Bunlon, Jean-Pierre Gazeau, Floren Colloud, Peter J Marshall, and Cédric A Bouquet. 2018. Joint action with a virtual robotic vs. human agent. *Cognitive Systems Research* 52 (2018), 816–827.

[7] Francesco De Pace, Federico Manuri, Andrea Sanna, and Davide Zappia. 2019. A comparison between two different approaches for a collaborative mixed-virtual environment in industrial maintenance. *Frontiers in Robotics and AI* (2019), 18.

[8] Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*. 754–768.

[9] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (2019), 81–98.

[10] James Coleman Eubanks, Alec G Moore, Paul A Fishwick, and Ryan P McMahan. 2020. The effects of body tracking fidelity on embodiment of an inverse-kinematic avatar for male participants. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 54–63.

[11] James Coleman Eubanks, Alec G Moore, Paul A Fishwick, and Ryan P McMahan. 2021. A Preliminary Embodiment Short Questionnaire. *Frontiers in Virtual Reality* 2 (2021), 647896.

[12] Adélaïde Genay, Anatole Lécuyer, and Martin Hachet. 2021. Virtual, real or mixed: How surrounding objects influence the sense of embodiment in optical see-through experiences? *Frontiers in Virtual Reality* 2 (2021), 679902.

[13] Moinak Ghoshal, Juan Ong, Hearan Won, Dimitrios Koutsonikolas, and Caglar Yildirim. 2022. Co-located Immersive Gaming: A Comparison between Augmented and Virtual Reality. In *2022 IEEE Conference on Games (CoG)*. IEEE, 594–597.

[14] Edward T Hall. 1966. *The hidden dimension*. Vol. 609. Anchor.

[15] Sandra G. Hart. 1986. *NASA Task Load Index (TLX). Volume 1.0; Computerized Version.* Technical Report. NASA.

[16] Tiger F Ji, Brianna Cochran, and Yuhang Zhao. 2022. VRBubble: Enhancing Peripheral Awareness of Avatars for People with Visual Impairments in Social Virtual Reality. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–17.

[17] Yuanyuan Jiang, Elizabeth E O'Neal, Pooya Rahimian, Junghum Paul Yon, Jodie M Plumert, and Joseph K Kearney. 2018. Joint action in a virtual environment: Crossing roads with risky vs. safe human and agent partners. *IEEE transactions on visualization and computer graphics* 25, 10 (2018), 2886–2895.

[18] Yuanyuan Jiang, Elizabeth E. O'Neal, Pooya Rahimian, Junghum Paul Yon, Jodie M. Plumert, and Joseph K. Kearney. 2019. Joint Action in a Virtual Environment: Crossing Roads with Risky vs. Safe Human and Agent Partners. *IEEE Transactions on Visualization and Computer Graphics* 25, 10 (2019), 2886–2895. https://doi.org/10.1109/TVCG.2018.2865945

[19] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

[20] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The Sense of Embodiment in Virtual Reality. *Presence: Teleoperators and Virtual Environments* 21, 4 (11 2012), 373–387. https://doi.org/10.1162/PRES_a_00124 arXiv:https://direct.mit.edu/pvar/article-pdf/21/4/373/1625283/pres_a_00124.pdf

[21] Kim Klüber and Linda Onnasch. 2023. Keep your Distance! Assessing Proxemics to Virtual Robots by Caregivers. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 193–197.

[22] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2017. Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics* 24, 2 (2017), 1038–1048.

[23] Lucie Kruse, Joel Wittig, Sebastian Finnern, Melvin Gundlach, Niclas Iserlohe, Oscar Ariza, and Frank Steinicke. 2023. Blended Collaboration: Communication and Cooperation Between Two Users Across the Reality-Virtuality Continuum. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 54, 8 pages. https://doi.org/10.1145/3544549.3585881

[24] Christos Kyrlitsias and Despina Michael-Grigoriou. 2022. Social interaction with agents and avatars in immersive virtual environments: A survey. *Frontiers in Virtual Reality* 2 (2022), 786665.

[25] Maurice Lamb, Rachel W Kallen, Steven J Harrison, Mario Di Bernardo, Ali Minai, and Michael J Richardson. 2017. To pass or not to pass: Modeling the movement and affordance dynamics of a pick and place task. *Frontiers in psychology* 8 (2017), 1061.

[26] Joseph J LaViola Jr. 2000. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin* 32, 1 (2000), 47–56.

[27] Janeen D Loehr. 2022. The sense of agency in joint action: An integrative review. *Psychonomic Bulletin & Review* 29, 4 (2022), 1089–1117.

[28] Daniel Medeiros, Rafael Dos Anjos, Nadia Pantidi, Kun Huang, Maurício Sousa, Craig Anslow, and Joaquim Jorge. 2021. Promoting reality awareness in virtual reality through proxemics. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 21–30.

[29] Daniel Medeiros, Rafael K dos Anjos, Daniel Mendes, João Madeiras Pereira, Alberto Raposo, and Joaquim Jorge. 2018. Keep my head on my shoulders! why third-person is bad for navigation in vr. In *Proceedings of the 24th ACM symposium on virtual reality software and technology*. 1–10.

[30] Betty J. Mohler, Heinrich H. Bülthoff, William B. Thompson, and Sarah H. Creem-Regehr. 2008. A Full-Body Avatar Improves Egocentric Distance Judgments in an Immersive Virtual Environment. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization* (Los Angeles, California) *(APGV '08)*. Association for Computing Machinery, New York, NY, USA, 194. https://doi.org/10.1145/1394281.1394323

[31] Jens Müller, Johannes Zagermann, Jonathan Wieland, Ulrike Pfeil, and Harald Reiterer. 2019. A Qualitative Comparison Between Augmented and Virtual Reality Collaboration with Handheld Devices. In *Proceedings of Mensch Und Computer 2019* (Hamburg, Germany) *(MuC'19)*. Association for Computing Machinery, New York, NY, USA, 399–410. https://doi.org/10.1145/3340764.3340773

[32] Ayana Naito, Kentaro Go, Hiroyuki Shima, and Akifumi Kijima. 2022. Synchrony in triadic jumping performance under the constraints of virtual reality. *Scientific Reports* 12, 1 (2022), 12417.

[33] Ye Pan, David Sinclair, and Kenny Mitchell. 2018. Empowerment and embodiment for collaborative mixed reality systems. *Computer Animation and Virtual Worlds* 29, 3-4 (2018), e1838.

[34] Tabitha C Peck and Mar Gonzalez-Franco. 2021. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality* 1 (2021), 575943.

[35] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The Effects of Sharing Awareness Cues in Collaborative Mixed Reality. *Frontiers in Robotics and AI* 6 (2019). https://doi.org/10.3389/frobt.2019.00005

[36] 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 *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173620

[37] Thammathip Piumsomboon, Gun A Lee, Andrew Irlitti, Barrett Ens, Bruce H Thomas, and Mark Billinghurst. 2019. On the shoulder of the giant: A multi-scale mixed reality collaboration with 360 video sharing and tangible interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–17.

[38] Iana Podkosova and Hannes Kaufmann. 2018. Co-Presence and Proxemics in Shared Walkable Virtual Environments with Mixed Colocation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (Tokyo, Japan) *(VRST '18)*. Association for Computing Machinery, New York, NY, USA, Article 21, 11 pages. https://doi.org/10.1145/3281505.3281523

[39] Iana Podkosova and Hannes Kaufmann. 2018. Mutual Collision Avoidance during Walking in Real and Collaborative Virtual Environments. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (Montreal, Quebec, Canada) *(I3D '18)*. Association for Computing Machinery, New York, NY, USA, Article 9, 9 pages. https://doi.org/10.1145/3190834.3190845

[40] Jose Luis Ponton, Eva Monclús, and Nuria Pelechano. 2022. AvatarGo: Plug and Play self-avatars for VR. In *Eurographics 2022 - Short Papers*, Nuria Pelechano and David Vanderhaeghe (Eds.). The Eurographics Association. https://doi.org/10.2312/egs.20221037

[41] Alejandro Ríos, Marc Palomar, and Nuria Pelechano. 2018. Users' Locomotor Behavior in Collaborative Virtual Reality. In *Proceedings of the 11th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Limassol, Cyprus) *(MIG '18)*. Association for Computing Machinery, New York, NY, USA, Article 15, 9 pages.

https://doi.org/10.1145/3274247.3274513

[42] Ferran Argelaguet Sanz, Anne-Hélène Olivier, Gerd Bruder, Julien Pettré, and Anatole Lécuyer. 2015. Virtual proxemics: Locomotion in the presence of obstacles in large immersive projection environments. In *2015 IEEE virtual reality (vr)*. IEEE, 75–80.

[43] Prasanth Sasikumar, Soumith Chittajallu, Navindd Raj, Huidong Bai, and Mark Billinghurst. 2021. Spatial perception enhancement in assembly training using augmented volumetric playback. *Frontiers in Virtual Reality* 2 (2021), 698523.

[44] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. A Survey on Synchronous Augmented, Virtual, AndMixed Reality Remote Collaboration Systems. *ACM Comput. Surv.* 55, 6, Article 116 (dec 2022), 27 pages. https://doi.org/10.1145/3533376

[45] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. 2006. Joint action: bodies and minds moving together. *Trends in cognitive sciences* 10, 2 (2006), 70–76.

[46] Natalie Sebanz and Günther Knoblich. 2021. Progress in joint-action research. *Current Directions in Psychological Science* 30, 2 (2021), 138–143.

[47] Mel Slater, Amela Sadagic, Martin Usoh, and Ralph Schroeder. 2000. Small-Group Behavior in a Virtual and Real Environment: A Comparative Study. *Presence: Teleoper. Virtual Environ.* 9, 1 (feb 2000), 37–51. https://doi.org/10.1162/105474600566600

[48] Maximilian Speicher, Brian D Hall, and Michael Nebeling. 2019. What is mixed reality?. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[49] Stephan Streuber, Astros Chatziastros, Betty J Mohler, and Heinrich H Bülthoff. 2008. Joint and individual walking in an immersive collaborative virtual environment. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. 191–191.

[50] Bronwyn Tarr, Mel Slater, and Emma Cohen. 2018. Synchrony and social connection in immersive virtual reality. *Scientific reports* 8, 1 (2018), 3693.

[51] Alexander Toet, Tina Mioch, Simon NB Gunkel, Omar Niamut, and Jan BF van Erp. 2021. Assessment of Presence in Augmented and Mixed Reality: Presence in Augmented and Mixed Reality. In *Conference on Human Factors in Computing Systems, CHI 2021*. ACM SigCHI.

[52] Robrecht PRD van der Wel, Cristina Becchio, Arianna Curioni, and Thomas Wolf. 2021. Understanding joint action: Current theoretical and empirical approaches. , 103285 pages.

[53] Cordula Vesper, Stephen Butterfill, Günther Knoblich, and Natalie Sebanz. 2010. A minimal architecture for joint action. *Neural Networks* 23, 8-9 (2010), 998–1003.

[54] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1643–1652.

[55] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Xiangyu Zhang, Shuxia Wang, Weiping He, Yuxiang Yan, and Hongyu Ji. 2021. AR/MR Remote Collaboration on Physical Tasks: A Review. *Robotics and Computer-Integrated Manufacturing* 72 (2021), 102071. https://doi.org/10.1016/j.rcim.2020.102071

[56] Yiwei Wang, Pallavi Shintre, Sunny Amatya, and Wenlong Zhang. 2022. Bounded Rational Game-theoretical Modeling of Human Joint Actions with Incomplete Information. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 10720–10725.

[57] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[58] Jacob Young, Tobias Langlotz, Matthew Cook, Steven Mills, and Holger Regenbrecht. 2019. Immersive telepresence and remote collaboration using mobile and wearable devices. *IEEE transactions on visualization and computer graphics* 25, 5 (2019), 1908–1918.

[59] Kevin Yu, Gleb Gorbachev, Ulrich Eck, Frieder Pankratz, Nassir Navab, and Daniel Roth. 2021. Avatars for teleconsultation: Effects of avatar embodiment techniques on user perception in 3D asymmetric telepresence. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (2021), 4129–4139.

[60] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth. 2021. Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence. *IEEE Transactions on Visualization & Computer Graphics* 27, 11 (nov 2021), 4129–4139. https://doi.org/10.1109/TVCG.2021.3106480