We Can Work It Out: Comparing Symmetric and Asymmetric Interaction Paradigms for Co-Located AR Co-Manipulation

Bibek Khattri [1], **and Maite Frutos-Pascual**[1], **and Ian Williams** [1]

[1] College of Computing, Birmingham City University, Birmingham, UK

Abstract

With recent advancements in technology, augmented reality (AR) systems are now capable of providing practical shared experiences offering new ways to work, socialize, and play together. While previous research has explored systems for asynchronous collaboration, studies focusing on real-time co-located synchronous collaboration where two or more users perform the same task together in tandem remain limited. Additionally when users are working in shared collaborative environments together, the level of symmetry to their movements (i.e., symmetric motions in their paired coordination) and the requirements for asymmetry (i.e., individualize motions within the collaboration) is presently underexplored. Therefore while highly immersive, realistic collaborative XR systems can be created, it is unclear how the symmetry provided by the system, notably in the collaborative interaction paradigms, should be designed to optimize user experience and task performance. To address this, we present a formal user study of two common AR interaction paradigms which offer varying levels of symmetry in a paired task. We explore **Grasp** (symmetric) and **AirTap** (asymmetric) within a bi-manual, co-located manipulation task where users (n = 32) in pairs (n = 16) complete a collaborative AR docking task. To illustrate the level of successful collaboration and user experience, we report on multiple metrics, notably task

Corresponding author: Bibek Khattri, bibek.khattri@mail.bcu.ac.uk

completion time, spatial accuracy, errors (drops in interaction), team workload and usability. Furthermore to explore user opinion and negate the over reliance on post study user questionnaires, we apply a sentiment analysis process directly to the communication between user pairs during the collaborative task. This supports a richer exploration of user positive or negative sentiment between the **Grasp** (symmetric) and **AirTap** (asymmetric) conditions.

Results demonstrate that the level of symmetry offered by the interaction paradigm (**Grasp** and **AirTap**) presents significant differences in collaborative performance, team workload and usability while maintaining comparable task spatial accuracy. Task performance is found to be lower with grasping interactions which required greater symmetry between user motions and actions and resulted in longer completion times and higher collaborative effort.

Sentiment of the paired user communication further highlighted a more reliable collaboration in the asymmetric (**AirTap**) condition. These findings align to prior work assessing asymmetry in interaction and therefore support the suggestion that symmetric interaction methods (**Grasp**), while intuitive, may hinder coordination and increase cognitive demands paired collaboration. We recommend that AR researchers and designers of collaborative virtual environments consider introducing levels of controlled asymmetry in interaction paradigms to better support user performance.

Introduction

Collaboration is a union; an act of engaging together in joint and interdependent activities by two or more individuals with the aim to achieve a shared goal (Terveen, 1995). Recognized as a vital skill in the 21st century (Lamb, Maire, & Doecke, 2017), collaboration has been integrated into various domains to foster the exchange of ideas, knowledge, and actions in teams and improving working experience.

Owing to the recent advancements in immersive spatial computing systems, namely Augmented Reality (AR) and Virtual Reality (VR), fully collaborative virtual environments (CVE) can exist which present a valuable way to foster new social interactions across many domains notably, education (Cumberbatch, Olatunji, & Robila, 2023), e-commerce (McLean & Wilson, 2019) and training (Mortara et al., 2014).

However, despite recent technological advancements, many immersive applications are often limited to single-user (Kim, Billinghurst, Bruder, Duh, & Welch, 2018). Additionally, CVE are commonly only available as remote or spatially separated experiences (Schäfer, Reis, & Stricker, 2022), thus lacking the potential for users to collaborate in one location to work, play and socialize. This means that while XR technology offers a promising future for a user, there is presently a limit to the availability of shared co-located CVE (i.e., where two or more users occupy the same physical and virtual space). Conversely, much research has focused on the nuances of single-user experiences in XR, often exploring the interaction methods (Spittle, Frutos-Pascual, Creed, & Williams, 2023) (Argelaguet & Andujar, 2013), workload (Maag et al., 2023) and experience (Slater & Wilbur, 1997) of one user. While some work has explored co-located CVE, assessing the challenges (Irlitti, Smith, Von Itzstein, Billinghurst, & Thomas, 2016), interaction methods (Duval, Lécuyer, & Thomas, 2006) or shared working (Renevier & Nigay, 2001), the extent of this work in comparison is limited.

Recent CVE research has explored cooperative manipulation i.e., the combined multiple user interaction of a space or virtual object, in support of understanding the individual manipulation of the virtual scene (Pinho, Bowman, & Freitas, 2008). However, this has largely focused on situations where each user individually interacts, manipulates or adjusts the scene elements separate from one another (Sereno, Wang, Besançon, McGuffin, & Isenberg, 2022). Limited work has explored co-dependent manipulation, whereby each user must combine their inputs on the same scene element to perform some form of cooperative symmetrical manipulation (i.e.,

translation, rotation or adjustment) (Basdogan, Ho, Srinivasan, & Slater, 2000). Therefore, the development of co-dependent CVE and the nuances to the design of these symmetrical manipulation systems remain largely underexplored. A key question therefore remains when designing co-manipulation systems, notably "how different interaction paradigms and their inherent symmetry affects the user's ability to collaborate effectively?".

In this work, we define interaction symmetry as the degree to which users must coordinate their spatial positions and temporal gestures in parallel to perform a shared task in XR. With this definition, it can be determined that common interaction techniques such as **Grasp** and **AirTap** differ fundamentally in the degree of interaction symmetry they impose. **Grasp** requires users to synchronize spatial positions and hand motions closely, leading to high symmetry and potentially increased cognitive and physical demand (Frutos-Pascual, Creed, & Williams, 2019). In contrast **AirTap** allows greater spatial flexibility and asymmetry, possibly reducing coordination effort but introducing new challenges for collaboration. While **Grasp** and **AirTap** are widely used in single-user AR context (Frutos-Pascual et al., 2019), (Spittle et al., 2023), their effects on collaborative co-dependent manipulation remain unexplored, particularly in a co-located setting. This leaves a gap in the understanding how specific interaction methods shape co-dependent performance, workload and communication in real-time, co-located AR.

To address this, we present an evaluation of co-dependent symmetrical manipulation in collaborative AR. We explore the asymmetry offered by two common interaction techniques, notably **Grasp** (symmetrical) and **AirTap** (asymmetrical) and how the asymmetry provided by the interaction method influences the co-dependent manipulation of objects. We illustrate effectiveness and potential problems when employing these interaction methods in bi-manual co-dependent interactions between pairs of users. We applied a rigorous experiment methodology reporting on measures of task completion time, spatial performance and interaction consistency. We also report on the communication and sentiment of users in the interaction conditions,

drawing from the work of Slater et al. (2022) who calls for the use of user sentiment analysis to be used within immersive and spatial interface studies alongside measures of user team workload and experience.

The rest of this paper is structured as follows: The Related Work section, details the current literature in collaborative XR, whle the Methodology section presents the experimental design for our study, detailing the Task, Participants, Apparatus, Protocol, Metrics. Then in the Results we present the outcomes from the experiment, quantifying the asymmetry differences between the interaction paradigms in terms of collaborative performance (Completion Time: Overall Time, Accuracy: Spatial Accuracy and Accuracy: Object Drops), and Communication. We present an exploration of the results in the Discussion section illustrating the findings of the experiment. Finally, we present the Limitations and Future Work for the study alongside the Conclusion and Recommendations.

Related Work

Collaborative Virtual Environments

Collaborative Virtual Environments (CVE) are a simulated shared experience designed to engage users and stimulate their sensory perceptions (Churchill & Snowdon, 1998) . This definition highlights the important characteristics of shared environments, notably to support engagement, interaction and connection between users in a symmetrical, shared experience. Work by Johansen (1988) also emphasized the key components of a CVE, categorizing them into the “when” and “where” dimensions of a collaborative environment. This was positioned within the space-time matrix and covered the physical setup of the CVE as either “co-located” (i.e., sharing the same place) or “remote” (i.e., sharing different spaces).

These important distinctions on the attributes of CVE led to seminal work by Margery, Arnaldi, and Plouzeau (1999) who proposed the framework for cooperative manipulation in CVE. This framework presented three primary increasing levels of cooperation within a virtual environment. These levels of CVE cooperation still apply today in the design of AR and VR (Margery et al., 1999) cooperative systems.

Level 1 (Margery et al., 1999) covers the most basic form of user cooperation, notably where users in the CVE can perceive each other and communicate cooperatively. Level 2 extends on these distinctions defining a CVE where users can act on and change elements of the scene individually. This level of user interaction is analogous to single-user interactions in AR/VR (Argelaguet & Andujar, 2013) (Spittle et al., 2023) but where the CVE supports multiple "single-user experiences". A further distinction is made at Level 2 in two sub-levels. Sub level 2.1 covers predefined or animated interactions, notably where a user can trigger an interaction or where interactions are a result of some predefined animations, and sub-level 2.2 which covers the spectrum of dynamic single-user interactions namely, touch, gaze and the spectrum of interaction paradigms offered from immersive technology (Argelaguet & Andujar, 2013) (Spittle et al., 2023). Level 3 then concludes the framework and covers a CVE that supports the simultaneous multi-user manipulation of the same scene objects. Again two sub levels are presented which offer a nuanced variation in the paired user cooperation. Sub level 3.1 supports independent cooperation i.e., one user controlling an object attribute (i.e., the object position) and another user controlling another attribute (i.e., the object color). Finally, sub level 3.2 presents the level of cooperative actions in the CVE and covers co-dependent interactions which are a result of the combined input of two or more users. This final sublevel would be akin to two users moving, lifting or manipulating a object in the CVE, working in symmetry and can therefore be comparable to the "piano movers problem" (Lengyel, Reichert, Donald, & Greenberg, 1990) a

common paradigm for co-dependent user interaction. This classic problem highlights the challenges of moving an object from one position to another within a constrained space.

Collaborative virtual environments play an important role in experiencing, enhancing or exploring situations which may not be viable in reality. CVE also offer exciting opportunities to present new methods for enhancing communication and understanding among users (Gutwin & Greenberg, 1999) and new models for symmetric co-working. However, most research has focused on representations of CVE which are at level 3.1 or below. Therefore commonly where the immersive or spatial computing technology is used as a means to enhance levels of collaborative awareness (Piumsomboon, Clark, Billinghurst, & Cockburn, 2013), (Chen et al., 2021), loss of physicality in remote collaboration (Ens et al., 2019), communication between collaborators (Kiyokawa et al., 2002) or to support thinking and communication (Brereton & McGarry, 2000). However, limited work has explored the challenges in level 3.2 of Margery et al's (Margery et al., 1999) framework, specifically when considering co-dependent and co-located environments (Sereno et al., 2022) and symmetrical collaborative tasks like the "piano movers problem" (Lengyel et al., 1990). Under these circumstances co-dependent performance could be effected and more work is needed to understand the nuances of level 3.2 in CVE.

Therefore, understanding collaboration and user experience in CVE requires a multidimensional examination of performance, teamwork and experiential factors such as the perception of presence between collaborators. Presence is commonly defined as the psychological sense of "*being there*" within a mediated environment (Slater & Wilbur, 1997). When a strong sense of presence is experienced, users tend to respond to the virtual environment in ways that resemble real-world behavior (Osmers et al., 2021). Similarly, IJsselsteijn, De Ridder, Freeman, and Avons (2000) describe presence as the sense of being located within a physical environment. In collaborative environments, this concept is often discussed in relation to social presence, which can be interpreted as the "*sense of being together and communicating with another*" (Biocca,

Harms, & Burgoon, 2003). Unlike broader concepts such as engagement or general user experience, presence refers to the perceptual sense of being located within and sharing the mediated environment. This perceptual experience has been associated with a range of outcomes including increased enjoyment, improved skill acquisition, and enhanced task performance (Von der Pütten et al., 2012). Although presence is not directly measured in this study the realism and interaction symmetry of the **Grasp** and **AirTap** techniques may influence aspects of spatial and social presence in collaborative AR settings.

Interactions

In all levels of Margery et al. (1999) framework, interactions are paramount in CVE (Schäfer et al., 2022) to support users in completing collaborative tasks and being aware of coworkers sharing mutual symmetrical experiences. Although interaction paradigms in single-user virtual environments are widely studied (Spittle et al., 2023), there is an increased interest in CVE interactions which focus on understanding the challenges of team working, cooperation, communication and knowledge sharing (Feng, He, Zhang, Billinghurst, & Wang, 2023), (Khattri, Chanrittia, Fritos-Pascual, & Williams, 2025). Fundamental research on synchronous collaboration sought to address ways to enhance engagement between users while minimizing mistakes in tasks (Kiyokawa et al., 2002) and proposed that co-location enhances dialogue exchange and reduces collaborative task error (Kiyokawa et al., 2002). This research supports the notion of collaboration as a union between partners and how collaboration can foster increased exchange of ideas. However, much research on synchronous collaborative performance has commonly focused on interaction challenges associated with cross modality devices (Piumsomboon et al., 2013) (i.e., VR, AR or head worn and handheld devices). Findings have highlighted how multiple interaction methods (i.e., gaze) can support increased awareness,

intention and attention in CVE and more recently how co-located and distributed environments differ in task completion and user communication (Podkosova, De Pace, & Brument, 2023).

Regarding the specifics of interaction in CVE, seminal work by Ruddle, Savage, and Jones (2002) defined common interaction techniques into two distinct groups, namely symmetric integration of movement, where users perform the same action (i.e., translate, lift, rotate), and asymmetric integration of motion (i.e., one user may lift an object while another rotates). Ruddle et al. (2002) also explored the methods to combine the simultaneous interactions of users in CVE, to support a "human like" variation in their individual motions and thus obtain an improved final motion in the task. In a comparison of an averaging technique (i.e., the unweighted sum of each user's interaction) and an intersection technique (i.e., the minima of each user's interaction), results highlighted comparability in task completion time, although the averaging method was superior for different actions (i.e., when moving an object in a non-linear direction). This technique of combining user inputs in CVE for symmetric interactions was also approached by Basdogan et al. (2000), who rather than computing an average or intersection of the cooperating user motion, used the dot product of each user's interaction to improve translation of a virtual object along a predefined trajectory (a ring on a wire).

Additional work has also explored how simultaneous interactions in CVE can inherently designed to support a level of tolerance (or asymmetry) to human variation in motion and thus improve interaction and collaboration. The SkeweR method by (Duval et al., 2006) explored manipulations based on Ruddle's motion combinations (Ruddle et al., 2002), however rather than computing an average, or intersection of user motion, the SkeweR method explored crushing points which can exist at the intersection with the surface of the object. Additionally, Riege, Holtkamper, Wesche, and Frohlich (2006) method supported more asymmetry in symmetrical interaction tasks via the projection of dynamic rays which bent depending on the individual user

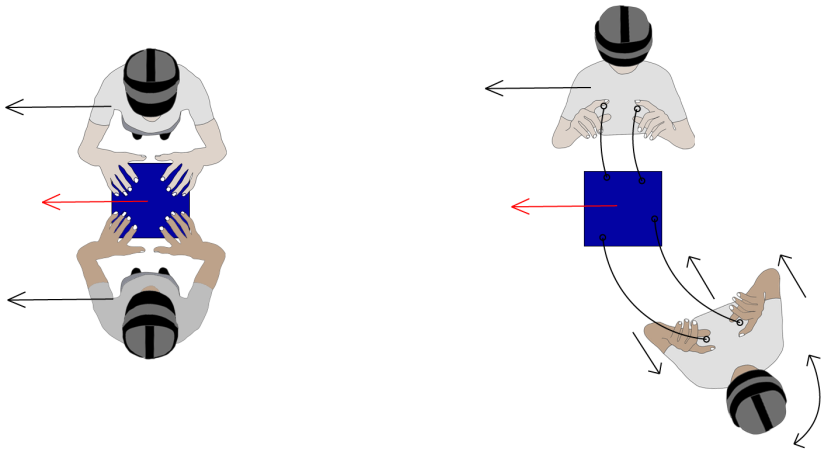
position. Both methods, SkeweR and Bent Pick Ray, support a tolerance to user motion and an inherent increase in the level of asymmetry in symmetrical interaction.

More recently in single-user interaction, the commonly supported interaction paradigms for AR and VR hardware have been explored, notably when considering **Grasp** and **AirTap** and for object manipulation (Frutos-Pascual et al., 2019). Results have illustrated that the **Grasp** method is perceived as more natural and usable, and offers improved accuracy in single-user translation tasks. However, the **AirTap** method gave improved performance for users in object rotation tasks. Although applied in single-user studies, and not considering the elements of CVE, when considering the interaction paradigm used an immersive environment, the **Grasp** and **AirTap** offer a differing level of asymmetry inherent in their design. Specifically the **AirTap** offers more variation (asymmetry) to how a user can interact in an object manipulation task than the **Grasp** which is more constrained and therefore more symmetrical across users (i.e., in their spatial position, orientation or motion).

For cooperative symmetric interaction from (Ruddle et al., 2002), (Duval et al., 2006), (Riege et al., 2006), **Grasp** and **AirTap** could therefore support different levels of inherent user asymmetry in a collaborative task (i.e., users moving faster or slower than each other or moving in different trajectories of motion). Additionally if considering bi-manual interactions, whereby each user must manipulate an object using both hands (Hough, Williams, & Athwal, 2015), this level of asymmetry could be of key importance.

Figure 1 presents an example where two users are performing a bi-manual cooperative symmetric manipulation task in a CVE. In this example the users performing the **Grasp** interaction (Figure 1[a]) require more symmetry in their coordination and combined interaction to translate the object. However, users performing the **AirTap** interaction (Figure 1[b]) are afforded more asymmetry in the interaction and therefore can have more variation in their individual approach to

perform the same translation (i.e., their motion, hand placement and position). This variation of inherent asymmetry in the interaction method could result measurable variations in task approach, task accuracy and therefore affect the design of users working in CVE.



(a) (b)

Figure 1: Bi-manual cooperative symmetrical interaction with two common interaction paradigms, **Grasp** (a) and **AirTap** (b). Here although the translation interaction is symmetrical the different interaction methods offer varying levels of asymmetry. With the **Grasp** interaction (a) the asymmetry is low and both users are required to interact and translate the object in the direction of symmetrical movement. In the **AirTap** method (b) there is an increased level of asymmetry to complete the same translation. Users can directly map their movement in the direction of the translation or individually move their arms or even rotate in location to complete the translation.

While our study focuses on symmetric bi-manual interaction, prior work by (Wieland, Zagermann, Müller, & Reiterer, 2021) has reported on co-manipulation techniques but in

collaborative handheld AR, revealing performance differences in their symmetrical method (i.e., their composition method was the slowest and was seen as confusing for participants to perform). Additionally Grandi, Debarba, Nedel, and Maciel (2017) reported how task performance in CVE is affected by the system providing the interaction and highlighted how users adapted different techniques in planning (i.e., collaboratively preparing an approach to the task) and executing (i.e., completing and actioning the task). Therefore, while the work of (Grandi et al., 2017) and (Wieland et al., 2021) do not explicitly explore **Grasp** and **AirTap** on co-dependent task, their results do offer clear direction and justification for further exploration into the potential differences that interaction methods may have on XR collaboration.

Measuring Collaborative Performance

Measures for evaluating single-user interactions in AR are well-established, with commonly spatial performance, task completion time (Frutos-Pascual et al., 2019) or system usability (Brooke, 1996) being commonplace. However, the evaluation of symmetrical collaboration within CVE has mostly focused on performance at a system level (Kozachek, Yaqin, Prasad, & Wang, 2023), typically focusing on objective measurable factors relating to the efficiency, accuracy, and usability of the system rather than the human-centric aspect of collaboration (such as communication or team working).

Surveys from Swan and Gabbard (2005) have provided a framework dividing evaluation of performance into areas of perception, performance and collaboration. However, owing to the nature of collaboration as a union between individual partners, each contributor can influence the outcome of the collaborative task. Kiyokawa et al. (2002) argued for a multi-dimensional approach to understand collaboration in immersive interfaces which should involve evaluating performance, process and subjective measures, thus providing a richer insight into the impact of users working synchronously. Aligned to this key work has sought to address this in the form of

the Team Workload Questionnaire (TWLQ) (Sellers, Helton, Näswall, Funke, & Knott, 2014) which measures an individual team members contribution in collaborative tasks. TWLQ was developed based on NASA TLX (Hart & Staveland, 1988) and wider team workload theory, and has since proven useful as a viable measurement co-working in CVE with (Khattri et al., 2025). Complimentary to other metrics of system performance, notably the System Usability Scale (SUS) (Brooke, 1996), TWLQ offers an individual per participant measure of perceived workload and has been used as a measure of team workload in human robot cooperation (Shuggi, 2018), crisis response (Judek, Verhaegen, Belo, & Verdel, 2018) and autonomous vehicle handover control (Sellers, Helton, Näswall, Funke, & Knott, 2015).

However, more recently, Nguyen and Bednarz (2020) highlighted how evaluations in collaborative environments, should not be limited to system level evaluations, but should also to consider the human factors in collaboration such as group dynamic, communication, awareness and the perception of presence between collaborators. While one could argue that TWLQ offers this evaluation, Slater et al. (2022) also highlights how post-experience questionnaires (Podkosova et al., 2023), (Gugenheimer, Stemasov, Frommel, & Rukzio, 2017) such as the TWLQ can impose bias in findings or miss elements of importance from the evaluation. Thus, they propose a call to action which requests the integration of objective measures (i.e., physiological state (Dey, Chen, Zhuang, Billinghurst, & Lindeman, 2018)) and sentiment analysis (B. Liu, 2022) to uncover richer findings not possible from standard post task questionnaire responses. Recently work by Khattri et al. (2025) has illustrated the potential correlation of sentiment driven approaches to the TWLQ, specifically for co-working in brainstorming tasks.

Methodology

We present an experiment on bi-manual cooperative manipulation in CVE, considering the asymmetry within two primary independent variables, namely the symmetrical and asymmetrical

interaction paradigms of **Grasp** and **AirTap** (see Figure 3). We present measures of task accuracy and task completion time, while also considering task measures of the communication between cooperating pairs of users. We finally report on team workload via the TWLQ (Sellers et al., 2014) to explore how the asymmetry in **Grasp** and **AirTap** influence perceived team workload.

Task

In both the **Grasp** and **AirTap** conditions, participants performed a docking task where the worked together to select, translate, release an object in synchrony. Docking tasks are common in XR studies to assess interaction methods (Frutos-Pascual et al., 2019), however they are usually presented as a solo user task (Frutos-Pascual et al., 2019). We present a collaborative docking task, analogous to a shared physical activity (Mendes, Caputo, Giachetti, Ferreira, & Jorge, 2019) known as as the "piano movers problem" (Lengyel et al., 1990). This task requires paired users to spatially coordinate their movements and actions as a partnership in the CVE.

We apply two different interaction paradigms (**Grasp** and **AirTap**), both standard, pre-existing input techniques in literature that are commonly supported by commercial AR platforms such as the Hololens 2. In the **Grasp** condition (see Figure 3[a]) participants are required to simultaneously **Grasp** a cuboid object ($30\ cm \times 20\ cm \times 20\ cm$) by placing both of their hands on any surface to select and lift it. For the **AirTap** condition (see Figure 3[d]), participants were required to perform a bi-manual **AirTap** on the object surface to select the object and lift it. For the **AirTap** condition we use the Bent Pick Ray method by Riege et al. (2006), therefore each user saw a projection of dynamic ray which bent depending on their individual user position in the task.

$$P_r = \big(\frac{P1_1, P1_2}{2} + \frac{P2_1, P2_2}{2}\big)/2 \tag{1}$$

For both conditions users were required to perform the docking task by simultaneously translating the object to the target position (see Figure 4). In both **Grasp** and **AirTap** conditions we apply Ruddle et al. (2002) method for unweighted averaging of the inputs (See Figure 2[a]), shown to offer good performance in these types of tasks. Since we are considering docking tasks based on translation only we do not calculate orientation from the user input. Instead we consider only the unweighted average of the user's bi-manual position (see Figure 6[f] and Figure 6[a]) and therefore calculate the unweighted average of both user inputs using Equation 1 (see Figure 2[b]). Where $P1_1$ and $P1_2$ are the bi-manual hand positions of user 1, and $P2_1$ and $P2_2$ are the bi-manual hand positions of user 2. P_r represents the unweighted average of the bi-manual user position information and therefore the object motion.

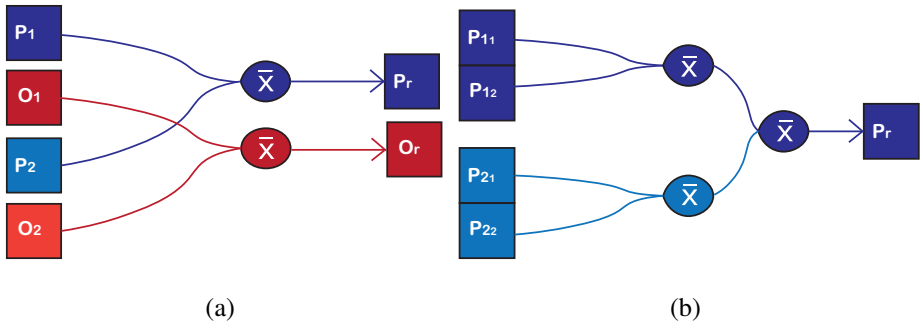
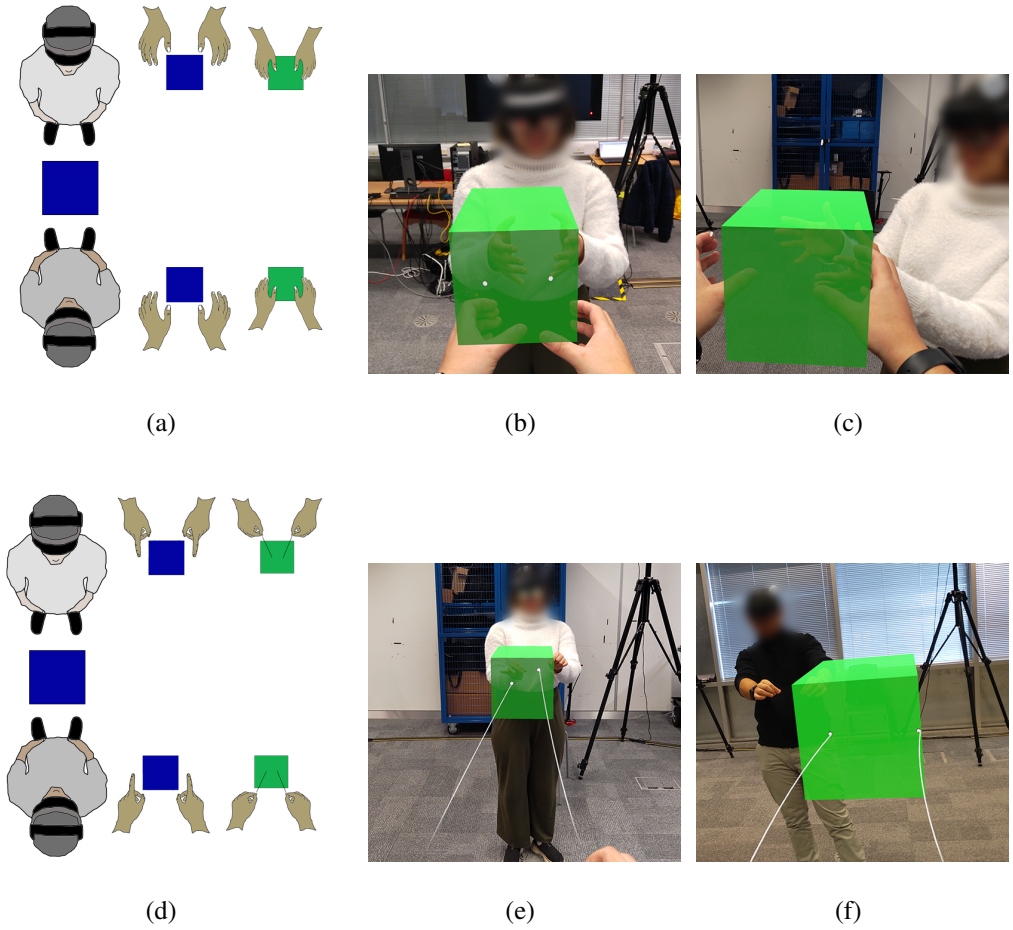


Figure 2: Cooperative Symmetrical Interaction model of Ruddle et al. (2002) replicated from Duval et al. (2006) (a) and adapted for bi-manual Cooperative Symmetrical Interaction (b). (a) Considers single inputs from each users position (P1, P2) and orientation (O1, O2). (b) Considers bi-manual cooperation and position information from both users (P1,P2) and both hands simultaneously ($P1_1$, $P1_2$ and $P2_1$, $P2_2$). An unweighted average of user inputs control the position (Pr) or Orientation (Or).

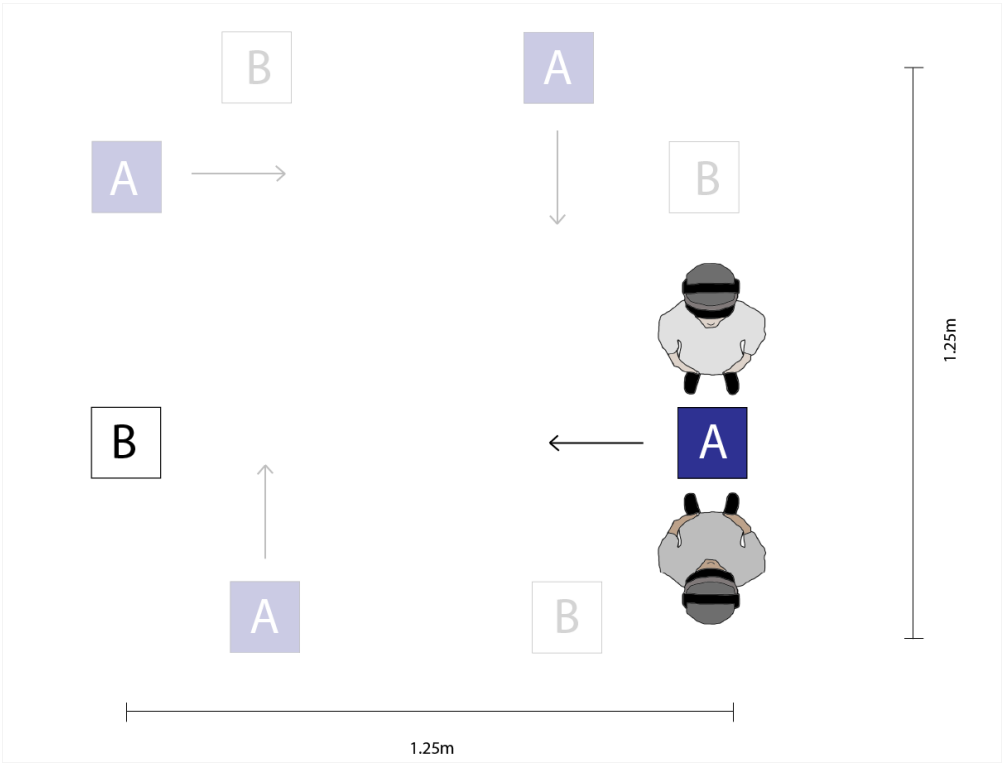
Four different translation tasks were presented to users in the X and Y directions (see Figure 4[a]) over three different interaction heights ($1\ m, 1.25\ m$ and $1.5\ m$). Each task required a direct X or Y translation to dock at a target location positioned $1.25\ m$ from the starting location (see Figure 4[a]). Three repeats for each translation task (position and height) resulted in 36 ($4 \times 3 \times 3$) translations for each interaction method (**Grasp** and **AirTap**) and with 72 tasks (36×2) gave 1152 interactions recorded during the study (72 interactions $\times$ 16 pairs). The primary interaction conditions were counterbalanced with half of the user pairs starting with the **Grasp** condition and half starting with the **AirTap** condition. Target positions were also randomized order across the $\pm X, Y$ and height conditions to mitigate learning bias and order effects. To support feedback within the interaction for both conditions the cube changed color to Green when a successful paired selection was complete (see Figure 3).



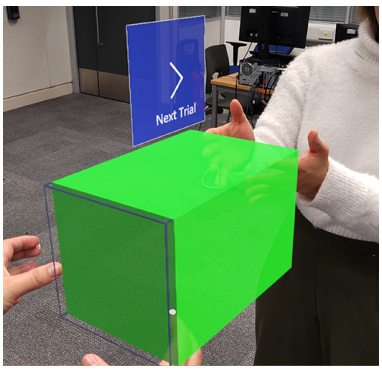
(a) (b) (c)

(d) (e) (f)

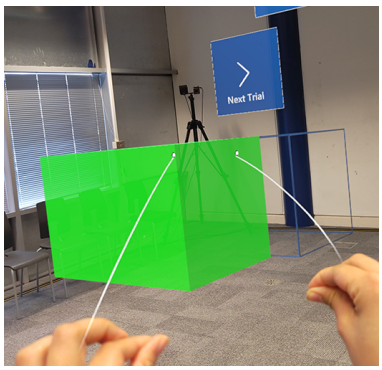
Figure 3: Simultaneous AR interactions within the study. The **Grasp** condition (a) requires two users to bi-manually **Grasp** the virtual object to perform a selection and manually translate the object (b), (c). The **AirTap** condition requires users to bi-manually **AirTap** the virtual object (d) and perform the manual translation of the object (e),(f)



(a)



(b)



(c)

Figure 4: The simultaneous interaction task of the study. Paired users were tasked to translate a virtual object (a) over a distance of 1.25 m from a starting position (A) to a docking position (B). The same docking task was performed in the **Grasp** (b) and **AirTap** (c) conditions and repeated over 4 different directions and three heights.

Participants

A within subjects design was applied where 32 participants (21 male, 10 female and 1 non-binary), aged 19 to 49 (M=29.09, SD =7.35) were recruited from the university student and staff population and allocated into pairs (n = 16). To avoid biases or variations within participants, pairs were allocated whom had no prior experience of collaborative working together. Partnered pairs reported their level of acquaintance with 6 participants having less than 3 months of familiarity, but had not worked on any form of collaborative tasks together, and 24 had never met their partner before.

Participants were randomly assigned into pairs without balancing for XR experience and no specific matching based on XR expertise was conducted. Overall, 13 self declaring as novices, 15 declaring average experience and 4 as experts. Additionally, their experience with collaborative immersive technology was reported, with 23 reporting as novices and 9 as average users. To ensure feasibility in the study user's height and arm length were measured prior to each experiment. These measurements were not to categorize users, or to support wider statistical evaluation, but to ensure interaction feasibility within the study conditions. Participants' arm lengths ranged from 59 cm to 81 cm (M = 71.47, SD = 5.29), height measurements ranged from 150 cm to 193 cm (M = 173.5, SD = 10.68).

All Participants were required to pass an Ishihara test (Ishihara, 1951) to exclude color blindness and their visual acuity was measured via a Snellen chart (Snellen, 1862). No participants who suffered from color blindness or with visual acuity of < 0.80 were included in the analysis. Prior to the test participants completed a standardized consent form and no financial compensation was provided. All data captured, stored and processed in alignment with the institutional ethical review board.

Apparatus

A custom experimental framework for the HoloLens 2 [1] using Unity 2021.3.1 [2] and the Mixed Reality toolkit [3] was developed. Photon Unity Networking (PUN2) [4] was used for developing the shared, multi-user experience with C# used as the scripting language. For the environment alignment between paired Hololens 2 devices, a 3D-printed rig was used to ensure consistent initialization of the origin position at the start of each condition. This was to ensure that the virtual object was continually presented in the same location for the paired users. The origin position was selected to comply with the device manufacturer's recommendations. A Sony ZV-E10 camera and wireless Lavalier Microphones (K39) were used to capture participants' video and audio.

Protocol

Prior to the experiment participants completed a demographic questionnaire. Visual acuity was assessed using a Snellen chart (Snellen, 1862) and color blindness evaluated using an Ishihara test (Ishihara, 1951). Participants wore any prescribed visual correction devices throughout the test. Following the methods of Khattri et al. (2025), we included activities to build rapport between pairs and to support collaboration and establish familiarization among pairs. Participating pairs underwent an icebreaker activity where they were given two minutes to introduce themselves and discuss hobbies interests and detail their current employment of study activities. Following the ice breaker the test coordinator performed the setup procedure to ensure accurate origin location (see Section Apparatus).

[1] https://learn.microsoft.com/en-us/hololens/

[2] https://unity.com/

[3] https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/

[4] https://www.photonengine.com/pun

Participants then underwent a series of training (see Figure 5) where they were presented with task protocols and had to collaboratively translate a capsule object using the **Grasp** and **AirTap** conditions (see Figure 5[c]). Participants completed the training task until they felt comfortable with the tasks and their partnership (see Figure 5[b]). They were also given the opportunity to ask questions and seek help from the test coordinator at this stage. Once training was complete which lasted no more than 5 minutes, the participants were presented with the main experimental task.

Prior to the main study, gesture tracking accuracy was validated under typical lighting and occlusion conditions and was determined to be within the manufacturers guidelines. In the main experiment and at the beginning of each trial, a start button was presented at the center of the interactive space for participants to acknowledge they were ready to complete the trial. Once clicked the interactive object (a cube) and target position (a cube wireframe) were presented. A virtual tracker for the number of trials completed was displayed outside of the interactive space (see Figure 4[b] and 4[c]). The trials were repeated until all 36 translations had been completed successfully. Throughout all trials, users were requested to complete the trials as quickly and as accurately while openly communicating within their pairs to support improved task completion and team-working. After each condition (**Grasp** or **AirTap**) participants filled out the post-experiment questionnaires (see Section Metrics).



(a) (b) (c)

Figure 5: Scenes from the training and ice-breaker phases, where paired participants familiarized themselves with each other and the experimental tasks. (a–b) Users reviewed instructions displayed in the shared environment. (c) A representative docking task involving a capsule object

Metrics

Within this work, we explore the influence of asymmetry in co-dependent symmetrical interaction with the **Grasp** and **AirTap** interaction methods. We report on metrics related to time and spatial accuracy alongside communication performance and team workload.

Completion Time: Overall Time Overall task completion time is defined as the time taken from the start of the collaborative interaction to the confirmation of placement in the docking task.

Completion Time: Human action cycle To support further analysis we also consider the time taken for completion of each phase of interaction detailed in the *"human action cycle"* by Norman (1984). These phases cover, 1) *forming the intent*, 2) *selecting the action*, 3)*Executing the*

outcome and 4) *Evaluating the outcome*. Each phase is determined as follows. 1) *forming intent* is recorded as the period of consultation and planning between and occupies the period from the selection of the start trial button up to the initial approach of object selection. 2) *selecting the action* covers the period of completing a successful selection of the cuboid object from both user pairs in the **Grasp** and **AirTap** condition. 3) *executing the outcome* is the period of translating the cuboid object from the starting location to the target location, including releasing the object. 4) *evaluating the outcome* is the period of each pair confirming completion of the task and agreeing on successful accuracy in the task.

Accuracy: Spatial Accuracy Following the methodology of Frutos-Pascual et al. (2019) we define spatial accuracy as the *euclidean distance* (in cm) between the interactive object and the target location at the end of each trial.

Accuracy: Object Drops In addition to spatial accuracy measured at task completion, we introduce a measure of collaborative interaction accuracy which is measured throughout the interactive (executing the outcome) phase of the task. In a real-world collaborative translation task like the "piano movers problem" (Lengyel et al., 1990) where users must synchronize their shared interaction around a single object (i.e., lifting, moving and releasing a piano), a potential exists for one, or both users, to misplace their hands on the object, to involuntarily release their grip on the object, or to present problems due to motion asymmetry in their translation (i.e., one user moving faster, slower or in a different direction than their partner). Each of these would result in them potentially "dropping" the object (the piano) and thus causing an error in the interaction. To mimic, this possibility we introduce the potential of a virtual *"drop"*, notably an error within the *executing the outcome* phase (the translation) in the docking task. A drop is triggered where one, or both users releases or disengages with the interaction before task

completion. Gesture tracking reliability was verified during pilot testing and the devices calibrated prior to each study condition, therefore drops were logged as a failure in a user forming the required gesture in the **AirTap** condition (see Figure 6[h] and 6[j]) due to effects such as "Gorilla Arm", moving out of view of the headset sensors, or because the user's hands moved too far from the virtual object in the **Grasp** condition (see Figure 6[c] and 6[e]). In both situations the object changed color from green to red when a drop occurred (see Figure 6) and users were required to re-engage with the cooperative interaction from that position.

We capture the total number of drops in the *execution the outcome* interaction phase and excluded other phases, to account for the intentional releases and re-adjustments made in these phases. Thus, in the **Grasp** condition when one or both users either release at least one hand from grasping the interactive object a drop is registered. Likewise in the **AirTap** condition where one or both users disengages from the air tap or causes an error from motion a drop is registered.

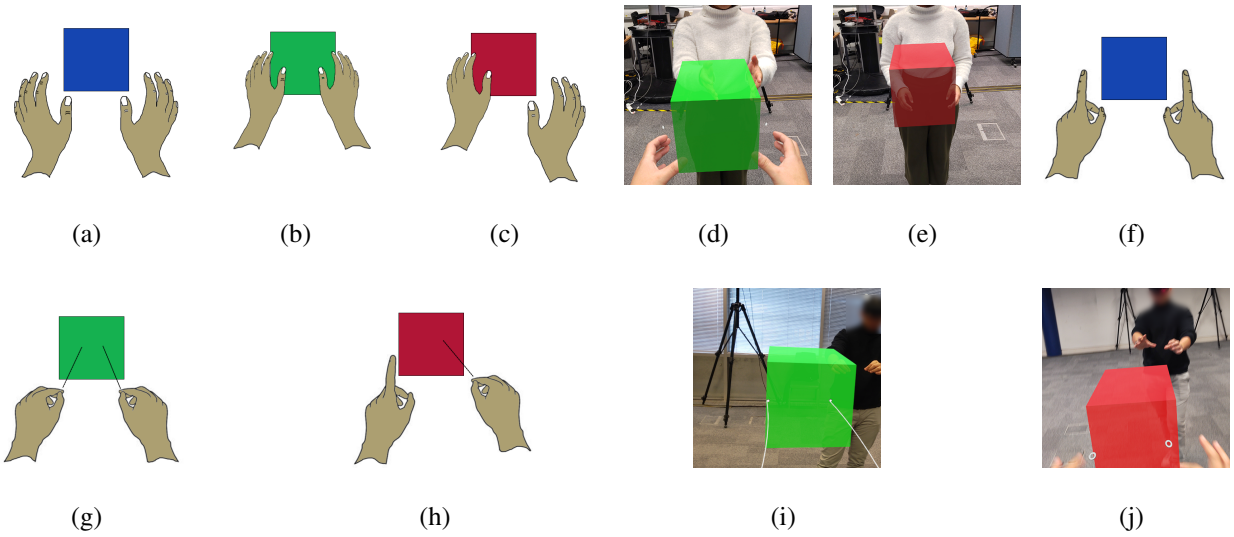


Figure 6: Interactive states within the task. Users were presented with a cuboid object in both the **Grasp** (a) and **AirTap** (f) conditions. Successful bi-manual interactions from both users in the **Grasp** (b,d) and **AirTap** (g,i) conditions selected the object which turned green and supported translation. If either user disengaged with the object at any point within the trial either by moving their hands from the object in the **Grasp** condition (c,e) or by releasing the **AirTap** gesture (h,j) a drop would be recorded with the object turning red.

Communication Aligned to the requirements of Slater et al. (2022) and to avoid evaluations based solely on task performance measures or post test questionnaires, we explore the extent of communication within the task itself. For this users were instructed to openly communicate within their assigned pairs to support task completion, to discuss their strategy, approach and to improve their team working.

Slater et al. (2022) presents the viability of using sentiment analysis as a tool for objective performance of experience in VR. While this illustrates the use of sentiment analysis on written

bodies of text conducted post experience, they call for the use of sentiment analysis within interactive studies, specifically to avoid biases from post task questions and to support richer exploration and triangulation of results. Recent work from Khattri et al. (2025) has illustrated the potential of sentiment analysis as a measure of collaborative performance in CVE based brainstorming tasks and the correlation this has to teamwork using the TWLQ.

Here we adopt the methodology of Khattri et al. (2025) and apply the VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto & Gilbert, 2014) to determine emotional tone (positivity, negativity) and subjectivity in participant pairs spoken conversation throughout the trials of the study. As VADER was originally developed for social media text, its ability to operate effectively on short, informal statements means its application within paired conversations is suitable and aligns to the work of Khattri et al. (2025). While more advanced sentiment processes are possible, notably BERT (Devlin, Chang, Lee, & Toutanova, 2019) or RoBERTa (Y. Liu et al., 2019) which provide increased interpretation of sentiment, they typically require fine-tuning on large, domain-specific datasets and are therefore unfeasible to apply in this study.

Positive and negative (Verma, 2022) sentiment is determined via a corpus of positive words (i.e., “accomplish” , “amazed” ,“aspire” ,“awesome”) and negative words (i.e., “afraid”, “annoying”, “baffled”, “frustrating”). We also applied standard stop word removal using a predefined lexicon (Hu & Liu, 2004) of common stop words (i.e., “I”, “and”, “stop”, “like”, “yes”).

We evaluate the Polarity (*PS*) score of the conversation using Equation 2, where *PW* is the total positive words in the conversation and *NW* is the total negative words. *TW* is the total overall words in the conversation and λ is a constant set to 0.000001. *PS* is on a scale of -1 to +1, where a -1 indicates negativity, +1 indicates positivity and 0 indicates neutrality.

$$PS = \frac{|PW| - |NW|}{(|PW| + |NW|)} + \lambda \tag{2}$$

We also evaluate the Subjectivity Score (*SS*) of the conversation given in Equation 3, where *PW* is the total positive words in the conversation and *NW* is the total negative words. *TW* is the total overall words in the conversation and λ is a constant set to 0.000001. *SS* ranges from 0 to 1 where a 1 indicates a high level of subjectivity.

$$SS = \frac{(|PW| + |NW|)}{|TW|} + \lambda \tag{3}$$

Team Workload Following prior work which verified the viability of the Team workload questionnaire (TWLQ) (Sellers et al., 2014) for CVE (Khattri et al., 2025) (Shuggi, 2018), (Judek et al., 2018) (Sellers et al., 2015), we apply TWLQ as a measure of collaboration in both the **Grasp** and **AirTap** conditions. TWLQ is composed of team workload and the team-task balancing dimensions, notably *Communication:* How much communication activity was required (e.g., discussing, negotiating, sending and receiving messages, etc.), *Coordination:* How much coordination activity was required (e.g., correction, adjustment, etc.), *Team Performance:* How much did the task require you to monitor your partner's performance, *Team Support:* How difficult was it to provide and receive support (providing guidance, helping team members, providing instructions, etc.) from team members?, *Team Dissatisfaction:* How irritated and annoyed were you with your partner? and *Team Emotion:* How emotionally demanding was working in the team (e.g., anger, joy, disappointment).

Hypothesis

Based on previous research we formulate hypothesis for the influence of the asymmetry offered by the interaction methods of **Grasp** and **AirTap** in symmetrical interactive CVE. We present hypothesis on time, task performance, communication and team working.

Although prior work illustrated that an interaction method can impact completion time for spatial tasks (Schultheis, Jerald, Toledo, Yoganandan, & Mlyniec, 2012) and that improved task accuracy can be achieved via interaction methods which offer more control (Grandi et al., 2017), (Grandi, Debarba, & Maciel, 2019), paired user simultaneous interactions have not been explored. Furthermore, (Ruddle et al., 2002) illustrated comparable levels of completion time and workload for differing implementations of interactive symmetry and somewhat comparable task performance. This illustrates the potential similarity which could be found in the **Grasp** and **AirTap**, but as yet is unexplored. In addition, Wieland et al. (2021) found significant differences in completion time and user workload depending on the interaction type, where a separation method was fastest and least demanding, a composition method was slower and more frustrating, and accuracy/participation remained comparable across techniques.

Additionally, considering communication in CVE, prior research has explored synchronous collaboration and the level of communication between users (Kiyokawa et al., 2002) (Podkosova et al., 2023) and findings illustrate correlations between the level of communication and the interactive task space. However, while results demonstrate that the physical proximity of users relates to more exchanges in dialogue between collaborative users, work has not explored this within a symmetrical interactions on a single object.

Therefore we present the following null hypothesis: $H0_1$: There is no difference in the completion time in co-dependent collaborative object docking tasks between the **Grasp** and **AirTap** interaction methods

$H0_2$: There is no difference in spatial accuracy in co-dependent collaborative object docking tasks between the **Grasp** and **AirTap** interaction methods

$H0_3$: There is no difference in the level of communication between paired users when performing co-dependent collaborative object docking tasks using the **Grasp** and **AirTap** interactions methods

$H0_4$: There is no difference in the level of team workload between paired users when performing co-dependent collaborative object docking tasks using the **Grasp** and **AirTap** interactions methods

Results

All data presented were analyzed using IBM SPSS Statistics (Version 29.0.2.0). For the quantitative analysis, normality of data is assessed using the Shapiro-Wilk test for data samples < 50 and Kolmogorov-Smirnov when data samples > 50, both with a significance level of $p < .05$. Pairwise comparisons are conducted between the two interaction conditions, **Grasp** and **AirTap**, across dependent variables such as completion time (s), accuracy, and communication. Where violation of normality is observed we report on pairwise comparisons using the Wilcoxon Signed Ranks Test and where normality is observed the Paired *t*-Test. The study adhered to the significance level of $p < .05$ with a 95% confidence interval. For the conversation and sentiment analysis, voice-recordings were transcribed using (OpenAI, 2022) Whisper.

Interaction	Median (IQR)	Mean (SD)	95% CI (LL, UL)	Skewness	Kurtosis
Overall Time (s)					
Grasp	27.55 (17.51)	31.98 (17.16)	30.57, 33.38	2.20	7.76
AirTap	17.51 (12.18)	22.64 (16.97)	21.25, 24.03	3.43	18.83
Completion Time: The Human Action Cycle (s)					
Grasp - Intent	5.43 (2.69)	6.09 (3.06)	5.84, 6.34	3.22	17.52
Grasp - Select	2.39 (2.53)	3.95 (4.71)	3.46, 4.33	3.58	17.41
Grasp - Execute	10.66 (9.91)	13.88 (10.98)	12.98, 14.78	3.10	15.72
Grasp - Evaluate	5.20 (6.41)	8.04 (8.36)	7.36, 8.73	3.01	11.55
AirTap - Intent	6.64 (3.85)	7.77 (4.31)	7.41, 8.12	3.22	17.66
AirTap - Select	1.09 (0.70)	2.39 (3.61)	2.09, 2.69	3.48	15.17
AirTap - Execute	2.38 (1.47)	4.24 (6.29)	3.72, 4.75	5.91	46.06
AirTap - Evaluate	4.56 (5.56)	8.21 (10.62)	7.34, 9.08	3.25	12.87
Accuracy: Spatial Accuracy (cm)					
Grasp	0.98 (1.13)	1.43 (1.38)	1.31, 1.54	2.54	8.84
AirTap	0.83 (0.92)	1.26 (1.24)	1.16, 1.36	2.72	9.4
Accuracy: Drops					
Grasp	14.00 (4)	14.50 (4.48)	12.11, 16.89	-0.028	0.563
AirTap	0.0 (2)	1.56 (3.79)	-0.46, 3.58	3.350	12.017
Communication					
Grasp - Total words	563.50 (494)	515.19 (278.08)	367.01, 663.37	-0.190	-1.262
Grasp - PS	0.350 (0.449)	0.381 (0.28)	0.229, 0.533	0.431	-0.355
Grasp - SS	0.302 (0.020)	0.035 (0.018)	0.025, 0.045	1.31	1.36
AirTap - Total words	340.50 (340)	442.63 (230.79)	319.64, 565.61	1.057	0.252
AirTap - PS	0.464 (0.574)	0.412 (0.330)	0.236, 0.588	-0.135	-0.965
AirTap - SS	0.305 (0.021)	0.034 (0.015)	0.026, 0.043	1.324	2.18
TWLQ					
Grasp	57.91 (12.92)	58.255 (12.15)	53.87, 62.63	-0.261	0.439
AirTap	42.92 (20.63)	40.98 (15.10)	35.54, 46.43	0.086	-0.700
SUS					
Grasp	72.50 (28.1)	68.28 (19.13)	61.38, 75.18	-0.57	0.104
AirTap	87.5 (17.5)	82.81 (14.0)	77.75, 87.86	-1.30	2.35

Table 1: Presents the descriptive statistics, comparing **Grasp** and **AirTap**. This includes measures for *completion time*, *spatial accuracy*, *communication*, *team workload (TWLQ)* and *system usability (SUS)*. For each measure, *median(IQR)*, *mean(SD)*, *95 percent confidence intervals (CI)*, *skewness* and *Kurtosis* are reported, offering a comprehensive statistical overview of the differences in the conditions.

Completion Time: Overall Time

Completion time data were significantly non-normal for **Grasp**, $p < .001$, and for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test illustrated a significant result, $Z = -11.80$, $p < .001$, with a medium effect size, $r = -0.492$, and **AirTap** ($\mathrm{Mdn} = 17.51\,\mathrm{s}$) being faster overall than **Grasp** ($\mathrm{Mdn} = 27.55\,\mathrm{s}$).

Completion Time: Human Action Cycle

As a significant difference was observed in the overall time for completion between the **Grasp** and **AirTap** we present independent analysis of completion time within the four stages of the Human Action Cycle (Norman, 1984). This supports a richer understanding of the temporal nuances of the interaction when considering the *Intent, Selection, Execution and Evaluation* times.

Human Action Cycle: Forming the Intent

Normality tests indicated a significant non-normal distribution of data for **Grasp**, $p < .001$, and for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test showed a significant difference, $Z = -8.84$, $p < .001$, with a medium effect size, $r = -0.368$. **Grasp** ($\mathrm{Mdn} = 5.43\,\mathrm{s}$) was faster overall than **AirTap** ($\mathrm{Mdn} = 6.64\,\mathrm{s}$). This suggests that users were taking less time to consider their approach to the docking task in the **Grasp** condition.

Human Action Cycle: Selection

Normality tests indicated non-normal distributions for **Grasp**, $p < .001$, and for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test illustrated a significant difference, $Z = -10.47$, $p < .001$, with a medium effect size, $r = -0.436$. **AirTap** ($\mathrm{Mdn} = 1.09\,\mathrm{s}$) was faster than **Grasp**

($\mathrm{Mdn} = 2.39\,\mathrm{s}$), illustrating potential challenges in bi-manual paired **grasping** during the *selection* phase.

Human Action Cycle: Execution the outcome

Normality tests indicated non-normal distributions for **Grasp**, $p < .001$, and for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test again illustrated significant differences between **Grasp** and **AirTap**, $Z = -18.42$, $p < .001$, with a large effect size, $r = -0.768$. Descriptive data further clarified this difference, with user pairs in the **Grasp** ($\mathrm{Mdn} = 10.66\,\mathrm{s}$) condition taking almost four times longer than **AirTap** ($\mathrm{Mdn} = 2.38\,\mathrm{s}$) to complete the *execution* phase of the docking task.

Human Action Cycle: Evaluation the outcome

Normality tests again illustrated a non-normal distribution for **Grasp**, $p < .001$, and for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test showed a statistically significant difference, $Z = -2.902$, $p = .004$, with a small effect size, $r = -0.121$. Descriptive data indicated that the time difference between **AirTap** ($\mathrm{Mdn} = 4.56\,\mathrm{s}$) and **Grasp** ($\mathrm{Mdn} = 5.20\,\mathrm{s}$) was small, $< 0.7\,\mathrm{s}$, suggesting practical similarity in the *evaluation* phase.

Accuracy: Spatial Accuracy

Considering spatial accuracy, we present results of target accuracy as a comparison to task completion time. Normality tests indicated non-normal distributions for **Grasp**, $p < .001$, and for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test showed a significant difference, $Z = -3.954$, $p < .001$, with a small effect size, $r = -0.165$. Although differences in spatial accuracy were observed, descriptive data indicated that the practical difference in target distance between **AirTap** ($\mathrm{Mdn} = 0.83\,\mathrm{cm}$) and **Grasp** ($\mathrm{Mdn} = 0.98\,\mathrm{cm}$) was small, $< 0.2\,\mathrm{cm}$, suggesting statistically significant but practically negligible differences in accuracy.

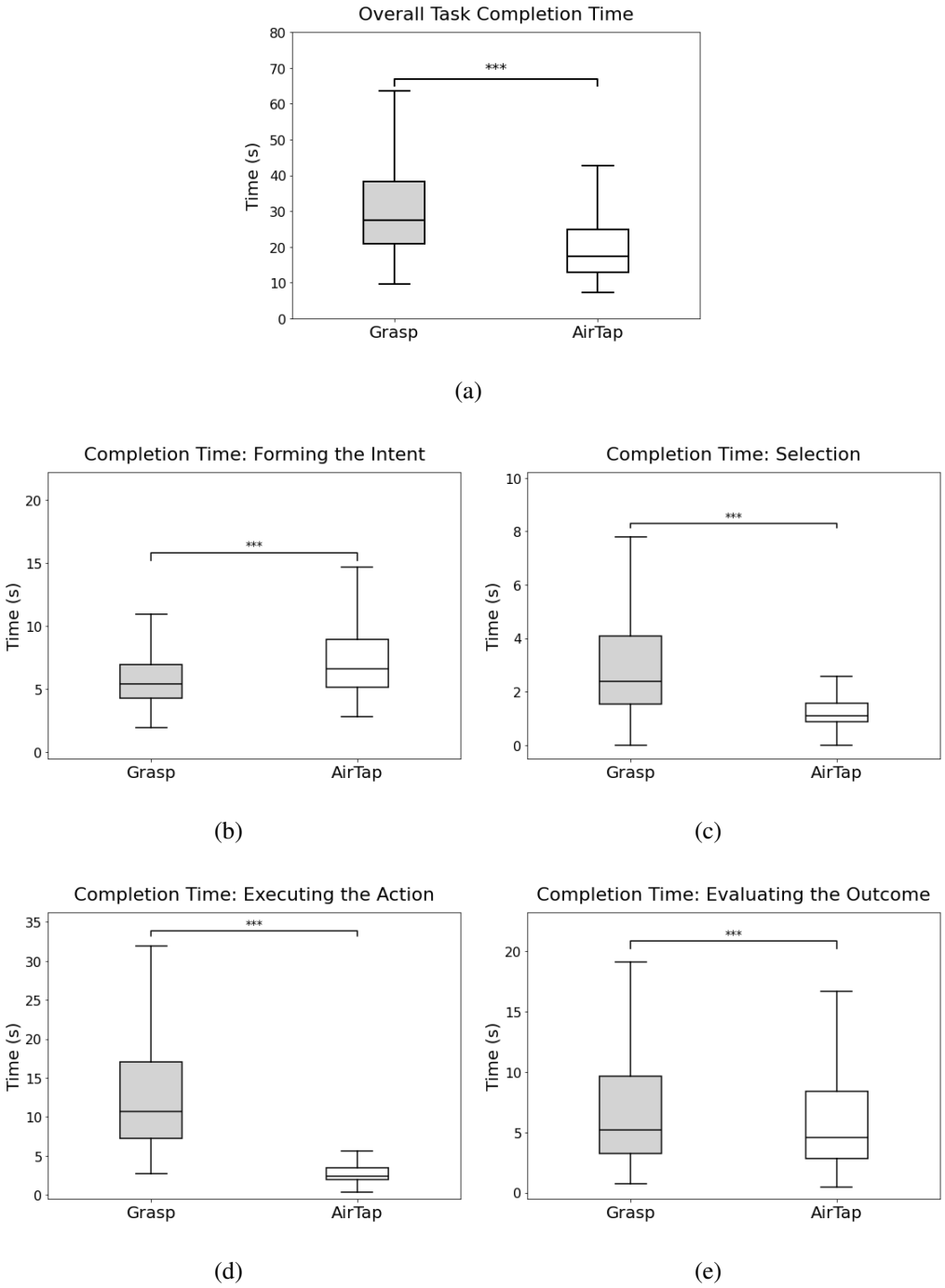
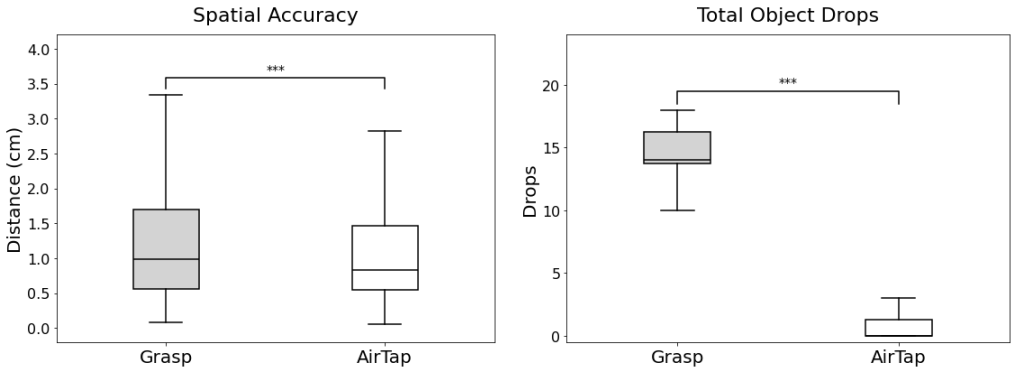


Figure 7: Box plots presenting data for the **Grasp** and **AirTap** conditions against overall completion time (a) and each phase of the human interaction cycle, forming intent (b), selecting (c), executing the action (d) and evaluating the outcome (e). Statistical significance is indicated with $* * * = p < 0.01$



(a) (b)

Figure 8: Box plots presenting data for the **Grasp** and **AirTap** conditions against accuracy as measures of spatial accuracy (a) and total number of drops (b). Statistical significance is indicated with $* * * = p < 0.01$

Accuracy: Object Drops

To present a more nuanced evaluation of collaborative task accuracy that complements spatial accuracy, we report object drops within the *execution* phase of the study, where paired users translated the object. Normality tests indicated a non-normal distribution for **AirTap**, $p < .001$. A Wilcoxon Signed Rank Test revealed a significant difference, $Z = -3.467$, $p < .001$, with a large effect size, $r = -0.867$. **AirTap** ($\mathrm{Mdn} = 0.00$) resulted in few drops within the interaction, whereas **Grasp** ($\mathrm{Mdn} = 14.00$) posed challenges for users, with considerably more drops recorded.

Communication

To complement the tasked based performance analysis, we also present a semantically oriented evaluation of the conversation between user pairs when completing the docking task. We report on the overall communication between user pairs and compare differences in overall total words and polarity and subjectivity of the verbalization.

Total Words

Regarding the total number of words uttered across all pairs of users, a non-normal distribution was observed for the **AirTap** condition, $p = .022$. A Wilcoxon Signed Rank Test indicated no significant difference, $Z = -0.672$, $p = .501$. However, the **Grasp** condition ($\mathrm{Mdn} = 563.5$) showed a higher number of words than **AirTap** ($\mathrm{Mdn} = 340.50$), suggesting a practical requirement for some user pairs to communicate more in this condition.

Subjectivity Score (SS)

Regarding subjectivity score (SS), a non-normal distribution was observed for the **Grasp** condition, $p = .035$. A Wilcoxon Signed Rank Test found no significant difference, $Z = -0.155$, $p = .877$, indicating no difference in the subjectivity of the communication. Descriptive data confirmed this, with **AirTap** ($\mathrm{Mdn} = 0.305$) and **Grasp** ($\mathrm{Mdn} = 0.302$) showing very similar subjectivity scores, < 0.004.

Polarity Score

Complementary to the subjectivity scores, the polarity score (PS) indicated normal distributions for both **Grasp**, $p = .200$, and **AirTap**, $p = .200$. A paired *t*-test found no significant difference, $t(15) = -0.489$, $p = .632$, with a small effect size, $d = -0.2553$. This indicates comparable polarity between conditions, with **AirTap** ($\mathrm{M} = 0.465$) showing slightly higher positivity than **Grasp** ($\mathrm{M} = 0.350$), a difference of 0.114.

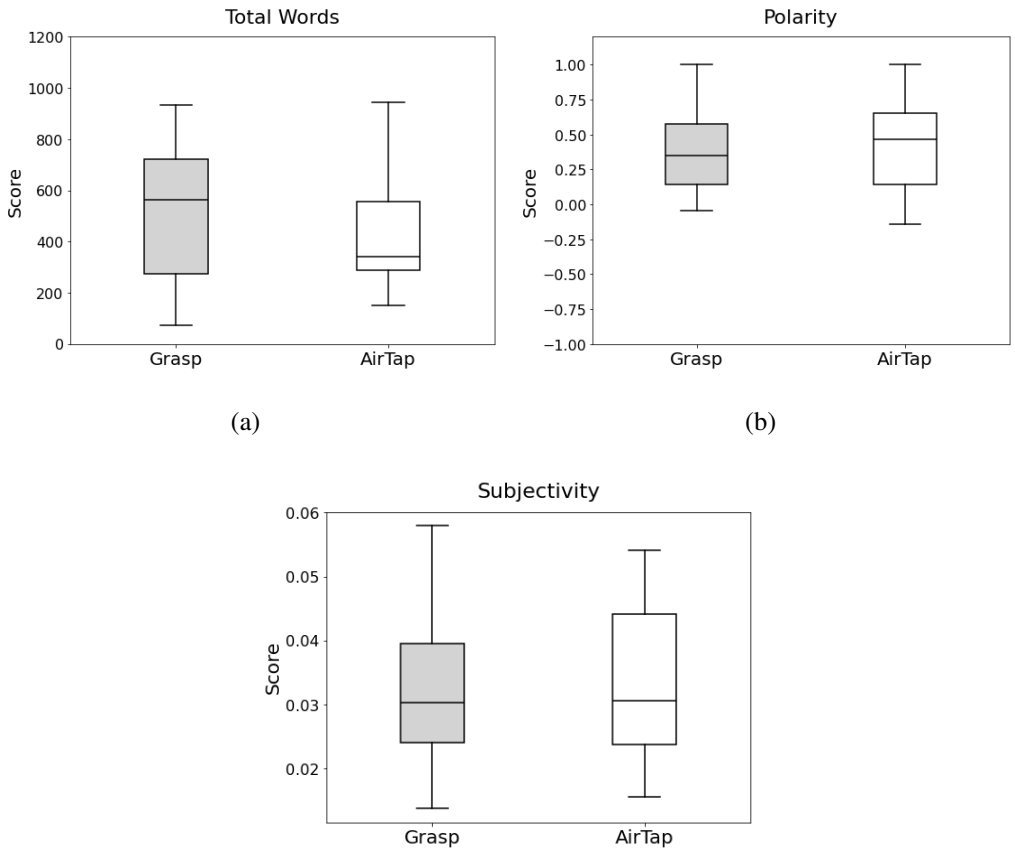


Figure 9: Box plots presenting data for the **Grasp** and **AirTap** conditions against communication within the task with Total number of words (a), Polarity Score (b) and Subjectivity Score (c)

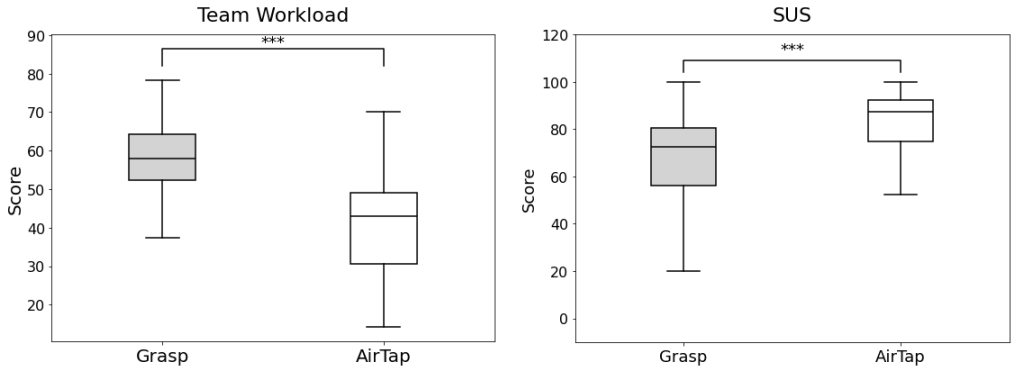
Team Workload

Considering team workload (TWLQ), normal distributions were observed for both conditions, **Grasp**, $p = .219$, and **AirTap**, $p = .629$. A paired t-test showed a significant difference, $t(31) = 7.78$, $p < .001$, with a large effect size, $d = 1.375$. Descriptive statistics further indicated

that **AirTap** ($\mathrm{M} = 40.98$) had a lower TWLQ score than **Grasp** ($\mathrm{M} = 58.25$), potentially reflecting a lower perceived task difficulty when using the bi-manual **Grasp** interaction.

SUS

In contrast to TWLQ, normality tests indicated a significant deviation from normality for the SUS data in the **AirTap** condition, $p = .032$. A Wilcoxon Signed Rank Test found a significant difference, $Z = -3.961$, $p < .001$, with a large effect size, $r = -0.70$. This difference was reflected in higher SUS scores for **AirTap** ($\mathrm{Mdn} = 87.5$) compared to **Grasp** ($\mathrm{Mdn} = 72.5$). These results align with the trend observed in the TWLQ findings, suggesting increased perceived task difficulty in the **Grasp** condition, although both **AirTap** and **Grasp** demonstrated acceptable SUS scores.



(a) (b)

Figure 10: Box plots presenting data for the **Grasp** and **AirTap** conditions against team workload (TWLQ) (a) and SUS (b). Statistical significance is indicated with $* * * = p < 0.01$

Discussion

This study examined how interaction symmetry influences collaborative performance in co-located AR docking tasks. We compared two AR interaction techniques, **Grasp** (symmetric) and **AirTap** (asymmetric), within a bi-manual collaborative manipulation task. Four hypotheses were formulated relating to completion time, accuracy, communication and perceived team workload. The results suggest that within the context of the task examined, the **AirTap** interaction was associated with faster task completion, fewer interaction drops, lower perceived team workload and higher usability scores compared to **Grasp**, while communication patterns between conditions remained broadly similar. These observations suggest that interaction techniques requiring greater symmetry between collaborators may introduce additional demands in co-dependent manipulation tasks. Similar findings have been reported in prior work on natural interaction techniques in immersive systems, where higher levels of interaction realism do not always lead to improved performance or usability outcomes (Kangas, Kumar, Mehtonen, Järnstedt, & Raisamo, 2022; Navarro & Sundstedt, 2019). In the following sections, we examine these findings in more detail, discussing how interaction symmetry and realism may influence collaborative performance across completion time (including phases of the human action cycle), accuracy, communication and perceived team workload.

Completion Time

The results for collaborative performance in Table 1 and Section highlight significant differences between the **Grasp** and **AirTap** conditions on the overall completion time. The **AirTap** interaction paradigm resulted in user pairs performing faster docking tasks compared to the **Grasp** interaction. This finding does not align with the work of Ruddle et al. (2002) and the faster completion time illustrates a potential increase due to the ease of use of **AirTap** as a method for collaborative selection, translation and releasing of objects owing to the increased asymmetry in

this method. The **Grasp** interaction, although a more natural form of interaction when compared to human collaborative situations, presented higher average completion times ($+30\%$), a finding which aligns with the work of (Frutos-Pascual et al., 2019) and (Wieland et al., 2021).

The results further suggest that interaction techniques requiring higher realism and greater symmetry between collaborators, (as seen in the **Grasp** interaction), may introduce additional coordination challenges during collaborative manipulation. The **Grasp** interaction which mimics the grasping of a real object and is therefore often considered a natural and intuitive interaction metaphor, commonly described as a "gold standard" approach for object manipulation (LaViola Jr, Kruijff, McMahan, Bowman, & Poupyrev, 2017). However, performing this technique required collaborators to closely synchronize their actions, which, in the context of this study, appeared to contribute to longer completion times. This observation may reflect a potential trade-off between interaction realism and task efficiency, where more natural interaction techniques do not necessarily result in improved performance. Similar findings have been reported in prior work on natural interaction in immersive systems, where higher levels of realism do not always lead to improved performance (Kangas et al., 2022). It is important to note that in other collaborative AR applications, such as training or simulation scenarios where interaction realism is critical, longer task completion times may represent an acceptable trade-off for increased naturalness. These findings therefore highlight the importance of carefully considering the degree of interaction symmetry when designing collaborative AR interaction techniques. This interpretation is broadly consistent with prior research on asymmetric collaborative systems and setups, which suggests that differences in interaction capabilities between collaborators can reduce coordination demands and support more efficient task performance (Grandi et al., 2019).

Understanding the variation in overall completion time supports a comparison of potential task efficiency in paired working, however to explore underlying temporal differences in the **Grasp**

and **AirTap** interaction paradigms, we presented time analysis of the four distinct phases within the *"human action cycle"* (Norman, 1984).

Forming intent is the first phase in the "human action cycle" (Norman, 1984) when completing an interactive task. Results presented in Section Completion Time: Human Action Cycle illustrate how the **Grasp** interaction provided paired users with the shortest times in this phase and a moderate significant effect when compared to **AirTap**. This shorter interaction time meant that users started the interaction phase of the task quicker in the **Grasp** interaction, potentially attributable to this interaction being more natural and analogous to grasping real objects in paired working. Additionally the lower time spent "forming intent" could also be attributable to the reduced asymmetry in the **Grasp** interaction, therefore meaning users spent less time reorienting themselves in the interactive space and simply navigated to the object to be translated. Prior work suggesting that users' actions are influenced by their ability to act upon objects within reachable space (Costantini, Ambrosini, Tieri, Sinigaglia, & Committeri, 2010). In this context, the graspable object may have helped users form their intent more quickly. Analysis of the dialogue between users in this phase of interaction showed that in the **Grasp** condition they often discussed the location of the object rather than varying their approach to the task: *"... so same task, same object","... that is the target and here is the object"* - Pair 1 , or *"... next trial"* - Pair 8 or *"... there, the target is behind you"* - Pair 15. However for the **AirTap** condition they often discussed their strategy and spatial positioning more: *" ... so the box is here and the target there"* - Pair 1 or *" ... first Let me come to the object"* - Pair 10 or *" ... you move a little bit back of the object!"* - Pair 12, therefore illustrating more asymmetry in approach and longer times planning their interaction.

While results presented in the Completion Time: Human Action Cycle section illustrate a faster performance for the **Grasp** when "forming intent", for an interaction in all other phases, notably *selection, execution* and *evaluation* the **AirTap** constantly resulted in quicker completion times

(see Table 1). Significant differences were reported for phase time completion against the **AirTap** and the **Grasp** interaction in all phases with a moderate (*selection*), strong (*execution*) and small (*evaluation*).

In the *selection* phase the slower completion time (see Table 1 and the Completion Time: Human Action Cycle section) in the **Grasp** interaction can be attributed to difficulties observed from the user pairs in this phase. Commonly in the **AirTap** condition users selected the object in synchrony, usually at first attempt. In the **Grasp** condition multiple selection attempts were common owing to the reduced asymmetry in this approach to achieve correct hand placement on the object.

In the *execution* phase, when paired users are translating the object, the **AirTap** condition resulted in a more consistent interaction with fewer errors (drops) than the **Grasp** condition (see the Accuracy: Object Drops section). In the **Grasp** interaction the requirement for a higher level of temporal synchrony and therefore more symmetry between user actions and motions often resulted in one or both users losing hand contact with the object. This lack of contact to the object triggered a break in the interaction (a drop) requiring users to select ("pick up") the object and therefore increasing execution time. This observation is consistent with prior work on joint action, which suggests that tasks involving shared control of an object require participants to align their actions in both time and space, and that small mismatches in movement can disrupt task performance (Knoblich, Butterfill, & Sebanz, 2011). Within the context of this study, the symmetric nature of the **Grasp** interaction may have increased sensitivity to such mismatches, contributing to longer execution times. Analysis in the conversation data in this phase illustrated that for the **Grasp** condition user pairs often verbalized when a drop occurred : *"... Drop. Drop ... that's a good plan actually to say drop"* - Pair 5, and *" ... it's here, it stopped"* - Pair 10, or verbalized their feelings about task complexity : *" ... one thing I will say is that it does require*

teamwork", *" . . . as long as we move our hands together we should be good"* - Pair 6, again highlighting the level of symmetry in this cooperative manipulation task.

Considering the *evaluation* phase a significant difference was observed between the **AirTap** and **Grasp** conditions. However, the effect size was reported to be small ($r = -0.121$) and the measured median difference in temporal performance was $< 0.7\ s$. Therefore and potential practical impact to this time difference may be negligible in CVE. This finding could be expected and it would be expected that in this final "evaluation" phase there is less impact from the interaction conditions (i.e., users are confirming their task success). Within the communication data, users commonly verbally agreed and discussed their shared perception of task accuracy: *" . . . as long as there's 90 percent accuracy . . . "* - Pair 12, or their target success : *" . . . let's stop trial here"* - Pair 7. Furthermore the differences in spatial accuracy as observed in the Accuracy: Spatial Accuracy section illustrate a small median difference in end target position ($< 0.20\ cm$), therefore proposing similarity in the end accuracy and evaluation process.

Considering the results observed for completion time we therefore reject the null hypothesis $H0_1$ *there is no difference in the completion time in co-dependent collaborative object docking tasks between the **Grasp** and **AirTap** interaction methods* when considering both the overall completion time and the interaction stages of the human action cycle.

Accuracy

Considering task accuracy we presented two primary metrics, notably spatial accuracy against end target position, and interaction accuracy as object drops during the translation task.

Spatial accuracy is commonplace as a success measure in docking tasks (Frutos-Pascual et al., 2019) and is often used to complement temporal accuracy. In the Accuracy: Spatial Accuracy section we report on an observed statistical difference in spatial accuracy between **Grasp** and

AirTap. However, owing to the small effect size and the similarity in measured median difference to target ($< 0.2\ cm$), the practical impact of this reported difference could be negligible. No observations supported this difference with users collaboratively confirming the end target placement at the end of each task. This finding also aligns with the reported similarity in temporal performance in the *evaluation* phase of the interaction (see the Completion Time: Human Action Cycle).

Owing to the collaborative nature of this work and to complement the measures of spatial and temporal accuracy, we present a measure of collaborative interaction accuracy based on object "drops". Drops are presented as a temporal break in the interaction during the *execution* phase of the interaction, resulting from one or both users temporarily breaking the interaction or a failure in the collaborative interaction.

Results presented in the Accuracy: Object Drops section illustrate a significant difference in drops between the **AirTap** and **Grasp** conditions with a large effect size and more drops occurring in the **Grasp** condition (see Table 1). In both conditions paired users translated the object to the target location while they moved in synchrony and in some cases this collaborative symmetric locomotion presented challenges. These were most prevalent in the **Grasp** condition, which required users to locomote using more timing accuracy and symmetry in their motion (see Fig 1[a). Therefore, if on user translated faster, or with more asymmetry in their motion (i.e., moved sidewards, stopped, or slowed down) this may cause problems for their partner to match, resulting in a break (drop) in the interaction. Analysis of the communication data between user pairs also highlighted this and communication often related to the cooperative strategy for the **Grasp** condition : *" . . . just move it slowly, just don't be fast, Okay, lets move it bit slowly"* - Pair 2. Similar findings although to a lesser extent were also observed in the **AirTap** condition, where users commonly discussed their task approach and how they utilized the asymmetry of the

interaction method by remaining more static in the translation : *"... you don't have to move a lot"* - Pair 1 or*"... I think we do not need to walk as much on this"* - Pair 2.

Together, these findings suggest that the degree of symmetry in collaborators' movements may influence task accuracy within the context of the collaborative docking task studied. The **Grasp** interaction, while providing a more natural and realistic form of manipulation, required tightly synchronized movement between collaborators, making the interaction more sensitive to small deviations in motion. Previous researchers have shown these small mismatches can have negative affect in task performance (Knoblich et al., 2011). In contrast, the **AirTap** interaction allowed greater independence in user movement and positioning, appeared to reduce coordination constraints and enabling more stable object translation during the docking task. Considering these results observed for spatial accuracy we therefore reject the null hypothesis $H0_2$ *there is no difference in spatial accuracy in co-dependent collaborative object docking tasks between the* ***Grasp*** *and* ***AirTap*** *interaction methods* when considering both the spatial accuracy and the object drops.

Communication

To complement task performance and present a richer understanding of collaborative performance, we present results based on a sentiment analysis of user utterances within the task. The Protocol section details how users were encouraged to communicate openly during the task and to vocalize their thoughts, concerns, strategy and to improve their team working. Inspired by the work of Slater et al. (2022) who call for the use of sentiment based processes in immersive studies, we applied the methodology of sentiment analysis following Khattri et al. (2025). We evaluated the number of words within the communication (total words) alongside the sentiment as the polarity (from negative to positive) and subjectivity of the conversation (see Table 1).

Results presented in the Communication section illustrated no significant difference in the total words, polarity or subjectivity between **Grasp** and **AirTap**. However, some differences were observed in the descriptive data (see Table 1), notably more words on average in the **Grasp** condition (563) than the **AirTap** condition (340). This finding illustrates a partial alignment with the work of Kiyokawa et al. (2002) who highlighted that more communication is conducted between users if the task space is centered between them. This closeness in proximity of paired users alongside more symmetry to their movements is evident in the **Grasp** condition, resulting in more communication between them. This finding also partially aligns to results from Podkosova et al. (2023) who highlighted a relationship to how much people engage in communication based on the physical proximity of the users in co-located setups. However, we also acknowledge that the potential perceived difficulty in the task, specifically in the **Grasp** condition, owing to the increased symmetry in the interaction, and that this may also be a factor in the observed increased communication between users. The observations of the communication data from the study, highlighted how users most commonly discussed task based operations (i.e., target location, starting location) rather than explicit challenges. Communication was therefore largely focusing on success in trials: *" . . . move it back a bit"* - Pair 2, the approach to the interaction: *" . . . same object, virtual walk"* - Pair 1, and the current perception of the task: *" . . . that was easy"* - Pair 2. These utterance were commonly vocalized on a trial by trial bases, which could be therefore be attributable to similarity in the polarity (**Grasp**: 0.35, **AirTap**: 0.46) and subjectivity (**Grasp**:0.30, **AirTap**: 0.30) scores and the recorded sentiment.

Together, these observations suggest that communication in the task primarily was to support and coordinate actions between collaborators. This is broadly consistent with prior work in CVE, which suggest that verbal communication often functions as a mechanism for coordinating joint action between collaborators. (Billinghurst, Kato, Kiyokawa, Belcher, & Poupyrev, 2002), (Podkosova et al., 2023). Although a realistic counterpart for natural real world interaction and

higher symmetry in the **Grasp** condition, participants more frequently discussed positioning and movement strategies in order to maintain the interaction. In contrast, users were able to adjust their movements more independently object translation within the collaborative docking task, the greater flexibility afforded by the **AirTap** interaction reduced the need for continuous verbal coordination.

Considering the results observed for communication we therefore accept the null hypothesis $H0_3$ *There is no difference in the level of communication between paired users when performing in co-dependent collaborative object docking tasks using the **Grasp** and **AirTap** interactions methods* when considering the overall total number of words and the sentiment polarity and subjectivity measures.

Team Workload

The results for Team Workload were assessed using the Team Workload Questionnaire (TWLQ) (Sellers et al., 2014) as the perceived team working from each user in the pairs (see Section Metrics. Results in the Team Workload section illustrate a strong statistical difference in TWLQ between the **Grasp** and **AirTap** conditions, with **AirTap** having a lower level of perceived team workload than **Grasp**. This lower workload could be attributable to an increase in the asymmetry of interaction and therefore a reduction in user perception of task difficulty for the **AirTap** condition. This difference may be related to the design of the **Grasp** interaction used in the collaborative docking task. In this setup, both collaborators were required to simultaneously grasp and manipulate the object during translation. As a result, small deviations in movement or timing frequently required users to adjust their positioning and actions in order to maintain the interaction. These repeated adjustments may have contributed to the higher perceived workload reported in the **Grasp** condition.

This interpretation is also broadly consistent with the objective performance measures reported earlier, where the **Grasp** condition requiring more teamwork between users to complete the task accurately. While there were no reported differences in the sentiment of the conversation between users, there were some statements of task difficulty for the **Grasp** condition: *" . . . you lost me"* - Pair 8, or requirement of symmetry: *' . . . if you stay put, it will be difficult for me to do it"* - Pair 2, or ' . . . *try moving closer, bring you hands here it might be easier to do that"* - Pair 14 so there is a partial alignment with some user vocalised sentiment and TWLQ.

Complementary to this, we reported on SUS (Brooke, 1996) to determine any obvious variations in standardised system usability. Results highlighted a statistical difference in usability of the **Grasp** and **AirTap** conditions with the **AirTap** having a higher SUS score (Mdn 87.5) than **Grasp** (Mdn 72.5). This illustrates an alignment in the usability and team workload for the two conditions although illustrating that both **Grasp** and **AirTap** present above "good" levels of usability. This alignment can act as further confirmation that the level of asymmetry in the interaction method can influence perceived teamwork and usability in bi-manual cooperative symmetric manipulation tasks in CVE. Considering this we reject the null hypothesis $H0_4$*: there is no difference in the level of team workload between paired users when performing co-dependent collaborative object docking tasks using the* ***Grasp*** *and* ***AirTap*** *interactions methods*.

Taken together, these findings suggest that interaction symmetry may influence collaborative performance within the co-located AR docking task examined in this study. In the experimental setup used here, the **Grasp** interaction required both collaborators to simultaneously grasp and manipulate the object. This design meant that small deviations in hand position or timing could interrupt the interaction and result in object drops, which may explain the higher workload and longer completion times observed in this condition. In contrast, the **AirTap** interaction allowed users to adjust their movements more independently when translating the object, which was associated with faster completion times and fewer drops. Communication between collaborators

Metric	Z/t Value	p Value	Effect Size	Differences(*Mdn*)	Result and Suggestion
Time: Overall	-11.80	$< .001$	-0.492	AirTap faster by 10.04s	**AirTap** $<$ **Grasp**
Time: Intent	-8.84	$< .001$	-0.368	Grasp faster by 1.21s	**Airtap** $>$ **Grasp**
Time: Selection	-10.47	$< .001$	-0.436	AirTap faster by 1.30s	**AirTap** $<$ **Grasp**
Time: Execution	-18.42	$< .001$	-0.768	AirTap faster by 8.28s	**AirTap** $<$ **Grasp**
Time: Evaluation	-2.902	.004	-0.121	AirTap faster by 0.64s	**AirTap** $<$ **Grasp**
Accuracy: Spatial	-3.954	$< .001$	-0.165	AirTap accurate by 0.2 cm	**AirTap** $>$ **Grasp**
Accuracy: Drops	-3.467	$< .001$	-0.867	AirTap fewer drops	**AirTap** $<$ **Grasp**
Total Words	-0.672	.501	N/A	Grasp more by 223 words	**AirTap** $=$ **Grasp**
Subjectivity Score	-0.155	.877	N/A		**AirTap** $=$ **Grasp**
Polarity Score	-1.022	.307	N/A		**AirTap** $=$ **Grasp**
TWLQ	-4.750	$< .001$	-1.187	AirTap lower	**AirTap** $<$ **Grasp**
SUS	-3.961	$< .001$	-0.70	AirTap higher	**AirTap** $>$ **Grasp**

Table 2: Overall experiment comparison of **Grasp** vs **AirTap**, focusing on *completion time*, *accuracy*, *communication*, *team workload* and *usability*. Median differences are derived from the descriptive statistics in Table 1.

also appeared to reflect these task requirements, with participants more frequently discussing positioning and movement strategies when performing the **Grasp** interaction. Together, these observations suggest that interaction techniques requiring higher symmetry and greater interaction realism may introduce additional challenges for collaborative manipulation in co-dependent AR tasks such as the docking scenario examined in this study.

Limitations and Future Work

We presented a robust formal evaluation of the asymmetry within the common interaction paradigms of **Grasp** (symmetrical) and **AirTap** (asymmetrical) in bi-manual cooperative manipulation, however we acknowledge some limitations to this work. Firstly we've explored cooperative manipulation in a direct single object translation task and while this is representative of the "piano movers problem" (Lengyel et al., 1990), a more complex task involving multiple

objects, combinations of translation, rotation (Frutos-Pascual et al., 2019) and virtual obstacles (Ruddle et al., 2002) may present further unknown challenges cooperative manipulation. Such tasks might impose different demands on coordination and symmetry, particularity in the context of training, creative work or simulation.

Secondly we have explored two common interaction paradigms only with each being a co-dependent bi-manual interaction (i.e., both users using the same interaction technique, either **Grasp** or **AirTap**). While our results offer value to the future design of CVE, further work could explore asymmetric combinations of hybrid and dynamic interactions methods i.e., users collaborating with different interaction methods depending on their roles. This combination of interaction methods would potentially present greater asymmetry in collaborative tasks (Pinho et al., 2008). Future work could also investigate the asymmetry in the system setups of CVE (Gardeli & Vosinakis, 2023) with the aim to better understand their potential impact on collaborative performance and communication in immersive CVEs while further complimenting work exploring levels 3.2 and 3.1 of Margery et al. (1999) framework.

Finally, we have reported on novel findings from the use of conversation and communication analysis as a performance measure of co-working. However, although building on prior CVE analysis methods (Khattri et al., 2025), we report on only one lexicon based method of sentiment analysis notably VADER (Hutto & Gilbert, 2014). As this technique for sentiment analysis does not consider valance shifting, statements like *"this is not very easy"* or *"this is not that difficult"* will return incorrect polarity and could influence the findings. Therefore more sophisticated methods, such as BERT (Devlin et al., 2019) or RoBERTa (Y. Liu et al., 2019), could be employed in future and future work should explore the quality and suitability of alternative sentiment analysis process for performance analysis and user experience. Additionally, we recommend that further work moves beyond utterance based sentiment analysis to analyze more complex inter task user dynamics such as, leadership, team support, turn-taking and conversation

distribution and support a more informed design of CVE. Lastly, we would welcome a replication of this study across different task types which offer unique challenges not found in our study (i.e., paired assembly tasks or paired co-manipulation of multiple scene elements). This wider exploration would help create a more generalized set of findings to inform the design of immersive collaborative systems.

Conclusion and Recommendations

The aim of this study was to examine the influence in collaboration from the level of asymmetry provided by cooperative symmetrical interactions when performing collaborative translation tasks. Users (n = 32) in pairs (n = 16) were tasked to collaboratively perform docking tasks in a tandem bi-manual interaction. We formally explored the level of asymmetry offered in two different interaction paradigms notably, **Grasp** (symmetric), where users apply a direct object grip to manipulate the object against **AirTap** (asymmetric) where the interaction is controlled by a remote gesture.

To overcome the reported problems and over reliance on post task questionnaires to assess immersive experiences, we present an in depth analysis using metrics of time, spatial accuracy, drops in interaction and team workload. Our findings show that while the **Grasp** interaction method supports natural coordinated spatial manipulation, it also increases task time, object drops, and team workload. This is likely due to the symmetry of this method requiring users to align their individual motions more closely. In contrast the **AirTap** interaction method allowed participants to collaborate with increased asymmetry in their individual approach and motion resulting in faster completion times, lower error rates and a lower measured team working demand.

Furthermore, following the recent call to action of Slater et al. (2022) and following the the methodology of Khattri et al. (2025) we also explore a sentiment-driven evaluation of communication within the task. Results showed similarity in sentiment of the **Grasp** and **AirTap** conditions. However, on average more communication was uncovered in the **Grasp** condition and users spoke more when the interaction methods required more closer symmetrical working.

Our findings challenge an assumption that symmetry and naturalness in XR interaction are inherently beneficial for collaborative systems. Instead, we argue for the importance of asymmetry in the interaction design for paired working in co-dependent scenarios, though this should be balanced against a potential reduction in team communication and realism. We therefore suggest that future researchers and designers of collaborative AR systems explore more personalized or adaptive interaction scheme that consider the need for controlled symmetry against task goals and personal user preferences.

References

Argelaguet, F., & Andujar, C. (2013). A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, *37*(3), 121–136. https://doi.org/10.1016/j.cag.2012.12.003

Basdogan, C., Ho, C.-H., Srinivasan, M. A., & Slater, M. (2000, December). An experimental study on the role of touch in shared virtual environments. *ACM Trans. Comput.-Hum. Interact.*, *7*(4), 443–460. https://doi.org/10.1145/365058.365082

Billinghurst, M., Kato, H., Kiyokawa, K., Belcher, D., & Poupyrev, I. (2002). Experiments with face-to-face collaborative ar interfaces. *Virtual Reality*, *6*, 107–121.

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, *12*(5), 456–480.

Brereton, M., & McGarry, B. (2000). An observational study of how objects support engineering design thinking and communication: implications for the design of tangible media. In *Proceedings of the SIGCHI conference on*

human factors in computing systems (pp. 217–224).

Brooke, J. (1996). Sus: A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4–7.

Chen, L., Liu, Y., Li, Y., Yu, L., Gao, B., Caon, M., … Liang, H.-N. (2021). Effect of visual cues on pointing tasks in co-located augmented reality collaboration. In *Proceedings of the 2021 ACM symposium on spatial user interaction* (pp. 1–12).

Churchill, E. F., & Snowdon, D. (1998). Collaborative virtual environments: an introductory review of issues and systems. *Virtual Reality*, *3*, 3–15.

Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? an investigation about affordances in space. *Experimental Brain Research*, *207*(1), 95–103.

Cumberbatch, I., Olatunji, J., & Robila, S. A. (2023). Using extended reality technology in science education. In *2023 IEEE long island systems, applications and technology conference (LISAT)* (p. 1-6). https://doi.org/10.1109/LISAT58403.2023.10179579

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dey, A., Chen, H., Zhuang, C., Billinghurst, M., & Lindeman, R. W. (2018). Effects of sharing real-time multi-sensory heart rate feedback in different immersive collaborative virtual environments. In *2018 IEEE international symposium on mixed and augmented reality (ISMAR)* (pp. 165–173).

Duval, T., Lécuyer, A., & Thomas, S. (2006). Skewer: a 3d interaction technique for 2-user collaborative manipulation of objects in virtual environments. In *3d user interfaces (3DUI'06)* (pp. 69–72).

Ens, B., Lanir, J., Tang, A., Bateman, S., Lee, G., Piumsomboon, T., & Billinghurst, M. (2019). Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, *131*, 81–98.

Feng, S., He, W., Zhang, X., Billinghurst, M., & Wang, S. (2023). A comprehensive survey on ar-enabled local collaboration. *Virtual Reality*, *27*(4), 2941–2966.

Frutos-Pascual, M., Creed, C., & Williams, I. (2019). Head mounted display interaction evaluation: Manipulating

virtual objects in augmented reality. In *Human-computer interaction – Interact 2019* (pp. 287–308). Cham: Springer International Publishing.

Gardeli, A., & Vosinakis, S. (2023). Exploring the co-manipulation of physical and virtual objects in tangible mobile augmented reality for collaborative learning. In *Proceedings of the 2nd international conference of the ACM greek SIGCHI chapter* (pp. 1–8).

Grandi, J. G., Debarba, H. G., & Maciel, A. (2019). Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE conference on virtual reality and 3d user interfaces (VR)* (pp. 127–135).

Grandi, J. G., Debarba, H. G., Nedel, L., & Maciel, A. (2017). Design and evaluation of a handheld-based 3d user interface for collaborative object manipulation. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 5881–5891).

Gugenheimer, J., Stemasov, E., Frommel, J., & Rukzio, E. (2017). Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 4021–4033).

Gutwin, C., & Greenberg, S. (1999). The effects of workspace awareness support on the usability of real-time distributed groupware. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *6*(3), 243–281.

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Hough, G., Williams, I., & Athwal, C. (2015). Fidelity and plausibility of bimanual interaction in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, *21*(12), 1377-1389. https://doi.org/10.1109/TVCG.2015.2480060

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM sigkdd international conference on knowledge discovery and data mining* (pp. 168–177).

Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th international conference on weblogs and social media (icwsm).*

IJsselsteijn, W. A., De Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence: concept, determinants, and measurement. In *Human vision and electronic imaging v* (Vol. 3959, pp. 520–529).

Irlitti, A., Smith, R. T., Von Itzstein, S., Billinghurst, M., & Thomas, B. H. (2016). Challenges for asynchronous collaboration in augmented reality. In *2016 IEEE international symposium on mixed and augmented reality (ISMAR-adjunct)* (pp. 31–35).

Ishihara, S. (1951). *Tests for colour-blindness.* Nippon Isho Shuppan Company.

Johansen, R. (1988). *Groupware: Computer support for business teams*. The Free Press.

Judek, C., Verhaegen, F., Belo, J., & Verdel, T. (2018). Crisis managers' workload assessment during a simulated crisis situation. In A. Malizia & M. D'Arienzo (Eds.), *Enhancing cbrne safety & security: Proceedings of the SICC 2017 conference* (pp. 325–332). Cham: Springer International Publishing.

Kangas, J., Kumar, S. K., Mehtonen, H., Järnstedt, J., & Raisamo, R. (2022). *Trade-off between task accuracy, task completion time and naturalness for direct object manipulation in virtual reality. multimodal technologies and interaction 6, 1 (2022).*

Khattri, B., Chanrittia, P., Fritos-Pascual, M., & Williams, I. (2025). I've got a feeling: Sentiment analysis for collaborative performance in vr. In *2025 IEEE conference on virtual reality and 3d user interfaces abstracts and workshops (VRW).* https://doi.org/10.1109/VRW66409.2025.00339

Kim, K., Billinghurst, M., Bruder, G., Duh, H. B.-L., & Welch, G. F. (2018). Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). *IEEE Transactions on Visualization and Computer graphics*, *24*(11), 2947–2962.

Kiyokawa, K., Billinghurst, M., Hayes, S. E., Gupta, A., Sannohe, Y., & Kato, H. (2002). Communication behaviors of co-located users in collaborative ar interfaces. In *Proceedings. international symposium on mixed and augmented reality* (pp. 139–148).

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: theory and data. *Psychology of Learning and Motivation*, *54*, 59–101.

Kozachek, D., Yaqin, M. A., Prasad, K., & Wang, S.-M. (2023). Evaluating the outcome of collaborative vr mind

mapping sessions with sentiment analysis and emotional intelligence. In *International conference on human-computer interaction* (pp. 245–263).

Lamb, S., Maire, Q., & Doecke, E. (2017). Key skills for the 21st century: An evidence-based review.

LaViola Jr, J. J., Kruijff, E., McMahan, R. P., Bowman, D., & Poupyrev, I. P. (2017). *3d user interfaces: theory and practice*. Addison-Wesley Professional.

Lengyel, J., Reichert, M., Donald, B. R., & Greenberg, D. P. (1990). Real-time robot motion planning using rasterizing computer graphics hardware. In *Proceedings of the 17th annual conference on computer graphics and interactive techniques* (p. 327–335). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/97879.97915

Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Maag, C., Schömig, N., Naujoks, F., Karl, I., Keinath, A., & Neukum, A. (2023). Measuring workload effects of augmented reality head-up displays using detection response task. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*, 201-219. https://doi.org/10.1016/j.trf.2022.11.010

Margery, D., Arnaldi, B., & Plouzeau, N. (1999). A general framework for cooperative manipulation in virtual environments. In *Virtual environments' 99: Proceedings of the eurographics workshop in vienna, austria, may 31-june 1, 1999* (pp. 169–178).

McLean, G., & Wilson, A. (2019). Shopping in the digital world: Examining customer engagement through augmented reality mobile applications. *Computers in Human Behavior*, *101*, 210–224.

Mendes, D., Caputo, F. M., Giachetti, A., Ferreira, A., & Jorge, J. (2019). A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum* (Vol. 38, pp. 21–45).

Mortara, M., Catalano, C. E., Bellotti, F., Fiucci, G., Houry-Panchetti, M., & Petridis, P. (2014). Learning cultural heritage by serious games. *Journal of Cultural Heritage*, *15*(3), 318–325.

Navarro, D., & Sundstedt, V. (2019). Evaluating player performance and experience in virtual reality game interactions using the htc vive controller and leap motion sensor. In *3rd international conference on human computer interaction theory and applications, hucapp 2019-part of the 14th international joint conference on computer vision, imaging and computer graphics theory and applications, visigrapp 2019, 25 february 2019 through 27 february 2019* (pp. 103–110).

Nguyen, H., & Bednarz, T. (2020). User experience in collaborative extended reality: overview study. In *Virtual reality and augmented reality: 17th eurovr international conference, eurovr 2020, valencia, spain, november 25–27, 2020, proceedings 17* (pp. 41–70).

Norman, D. A. (1984). Stages and levels in human-machine interaction. *International Journal of Man-Machine Studies*, *21*(4), 365–375.

OpenAI. (2022). *Whisper: Automatic speech recognition.* `https://github.com/openai/whisper`. (Accessed: March 09, 2025)

Osmers, N., Prilla, M., Blunk, O., George Brown, G., Janßen, M., & Kahrl, N. (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* (pp. 1–17).

Pinho, M. S., Bowman, D. A., & Freitas, C. M. D. S. (2008). Cooperative object manipulation in collaborative virtual environments. *Journal of the Brazilian Computer Society*, *14*, 53–67.

Piumsomboon, T., Clark, A., Billinghurst, M., & Cockburn, A. (2013). User-defined gestures for augmented reality. In *CHI'13 extended abstracts on human factors in computing systems* (pp. 955–960).

Podkosova, I., De Pace, F., & Brument, H. (2023). Joint action in collaborative mixed reality: Effects of immersion type and physical location. In *Proceedings of the 2023 ACM symposium on spatial user interaction* (pp. 1–12).

Renevier, P., & Nigay, L. (2001). Mobile collaborative augmented reality: the augmented stroll. In *Ifip international conference on engineering for human-computer interaction* (pp. 299–316).

Riege, K., Holtkamper, T., Wesche, G., & Frohlich, B. (2006). The bent pick ray: An extended pointing technique for

multi-user interaction. In *3d user interfaces (3dui'06)* (p. 62-65). https://doi.org/10.1109/VR.2006.127

Ruddle, R. A., Savage, J. C., & Jones, D. M. (2002). Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *9*(4), 285–308.

Schäfer, A., Reis, G., & Stricker, D. (2022). A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems. *ACM Computing Surveys*, *55*(6), 1–27.

Schultheis, U., Jerald, J., Toledo, F., Yoganandan, A., & Mlyniec, P. (2012). Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3d tasks. In *2012 IEEE symposium on 3d user interfaces (3dui)* (pp. 117–124).

Sellers, J., Helton, W. S., Näswall, K., Funke, G. J., & Knott, B. A. (2014). Development of the team workload questionnaire (twlq). In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 58, pp. 989–993).

Sellers, J., Helton, W. S., Näswall, K., Funke, G. J., & Knott, B. A. (2015). The team workload questionnaire (twlq): A simulated unmanned vehicle task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1382-1386. https://doi.org/10.1177/1541931215591231

Sereno, M., Wang, X., Besançon, L., McGuffin, M. J., & Isenberg, T. (2022). Collaborative work in augmented reality: A survey. *IEEE Transactions on Visualization and Computer Graphics*, *28*(6), 2530-2549. https://doi.org/10.1109/TVCG.2020.3032761

Shuggi, S. P. H. J., I.M. (2018). Changes in motor performance and mental workload during practice of reaching movements: a team dynamics perspective. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 236). https://doi.org/10.1007/s00221-017-5136-8

Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., & Oliva, R. (2022). A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, *3*. https://doi.org/10.3389/frvir.2022.914392

Slater, M., & Wilbur, S. (1997, 12). A framework for immersive virtual environments (five): Speculations on the role

of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603-616. 10.1162/pres.1997.6.6.603

Snellen, H. (1862). *Letterproeven, tot bepaling der gezigtsscherpte* (Vol. 1). J. Greven.

Spittle, B., Frutos-Pascual, M., Creed, C., & Williams, I. (2023). A review of interaction techniques for immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, *29*(9), 3900-3921. https://doi.org/10.1109/TVCG.2022.3174805

Swan, J. E., & Gabbard, J. L. (2005). Survey of user-based experimentation in augmented reality. In *Proceedings of 1st international conference on virtual reality* (Vol. 22, pp. 1–9).

Terveen, L. G. (1995). Overview of human-computer collaboration. *Knowledge-Based Systems*, *8*(2-3), 67–81.

Verma, P. (2022). Sentiment analysis using subjectivity and sentiment score. *The Next Normal: Building Agile, Sustainable, Tech-enabled Organizations (TNN–BASTO-2022)*, *55*.

Von der Pütten, A. M., Klatt, J., Ten Broeke, S., McCall, R., Krämer, N. C., Wetzel, R., … Klatt, J. (2012). Subjective and behavioral presence measurement and interactivity in the collaborative augmented reality game timewarp. *Interacting with computers*, *24*(4), 317–325.

Wieland, J., Zagermann, J., Müller, J., & Reiterer, H. (2021). Separation, composition, or hybrid?–comparing collaborative 3d object manipulation techniques for handheld augmented reality. In *2021 IEEE international symposium on mixed and augmented reality (ismar)* (pp. 403–412).