Human-Human Joint Interaction in Cross-Reality Systems: The Impact of Immersion and Synchronicity

Computers & Graphics 140 (2026) 104752

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on RAGI 2026

Human–human joint interaction in Cross-Reality systems: The impact of immersion and synchronicity☆

Roberta Macaluso [a],*, Noemi Di Cicco [a], Alberto Cannavò [a], Adalberto L. Simeone [b], Fabrizio Lamberti [a]

[a] Dipartimento di Automatica e Informatica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Turin, Italy
[b] Department of Computer Science, KU Leuven, Leuven, Belgium

A R T I C L E I N F O

Keywords:
Cross-Reality
Joint interaction
Co-experience
Reality-virtuality continuum

A B S T R A C T

Performing a Joint Interaction (JI) across the Reality–Virtuality (R-V) continuum remains challenging particularly in enabling shared manipulation of tangible real-world entities from distinct realities. To address this gap, this work investigates how interaction synchronicity and immersion level shape Cross-Reality (CR) JI involving shared manipulation of tangible entities by focusing on their impact on aspects pertaining task perceived presence, perceived workload, sense of inclusion/exclusion, empowerment and perceived control. To support this investigation, an exploratory user study examined human–human CR JI scenarios through the development of three interaction probes: a checkerboard, a comic book, and a music player, each re-creating specific types of JI that enables an Augmented Reality (AR) user and a real-world user to engage with a shared, manipulable entity under varying synchronization conditions. The quantitative results suggest that synchronous interaction enhanced perceived presence, while qualitative accounts pointed to a stronger inclusion and connection. Perceived workload showed differences across immersion conditions, with qualitative findings further highlighting the role of interaction constraints and control distribution in shaping participants' experiences. Overall, effective CR JI design requires balancing physical and virtual affordances while enabling meaningful participation across asymmetric realities. These insights are synthesized into initial design implications for future CR JI systems, providing guidance for interaction approaches that support meaningful participation, awareness, and coordination across heterogeneous realities.

1. Introduction

As technology has progressed, the boundaries between different levels of immersion along the Reality–Virtuality (R-V) Continuum [1] have become increasingly fluid, giving rise to interactive systems that allow users to navigate across the continuum and interact with each other from distinct positions along it without being restricted to a single immersion level [2].

This continuum-spanning immersion paradigm is made possible by Cross-Reality (CR) systems [3,4], whose user-experienced manifestations can vary or evolve over time [2]. Auda et al. [2] proposed a categorization of CR systems into three distinct types: *transitional*, enabling dynamic transition along the R-V Continuum [5–8]; *substitutional*, allowing interaction with re-contextualized tools adapted to the user's present actuality [9–12]; *multi-User* – in this paper referred to as Cross-Reality Multi-User (CR-MU) systems – enabling asymmetric multi-user interactions across realities, with "asymmetry" expressing the concept that each user can operate from different positions of the R-V Continuum by using different interfaces, while preserving a coherent shared experience [13–17].

Despite the numerous benefits and the promising results already reported in the literature [18], CR-MU systems still present complex and interrelated challenges, which are systematically summarized in Table 1.

This work specifically focuses on shared object and environment manipulation (indicated with * in the table), since these types of interactions are central to collaborative tasks and remain partially unexplored, as pointed out in recent surveys such as [35]. Within this context, Joint Interaction (JI) is used in the present work to refer to a form of interaction in which users engage with a shared entity or task state, such that their actions and the resulting changes are mutually observable. This notion builds upon previous work on

☆ This article is part of a Special issue entitled: 'RAGI 2026' published in Computers & Graphics.

* Corresponding author.
E-mail addresses: roberta.macaluso@polito.it (R. Macaluso), noemi.dicicco@studenti.polito.it (N. Di Cicco), alberto.cannavo@polito.it (A. Cannavò), adalberto.simeone@kuleuven.be (A.L. Simeone), fabrizio.lamberti@polito.it (F. Lamberti).

https://doi.org/10.1016/j.cag.2026.104752
Received 24 May 2026; Received in revised form 21 August 2026; Accepted 2 September 2026
Available online 9 September 2026

Table 1
Categorization of Cross-Reality Multi-User systems' challenges.

Challenges	References
Handling multimodal misalignment	[16,19,20]
Ensuring adequate social and environmental awareness	[14,15,21–25]
Enabling telepresence	[13–15,17,20,23]
Allowing shared object manipulation*	[14,16,17,19,22,26–29]
Allowing shared environment manipulation*	[20,30–34]

joint action, where shared activities rely on mutual awareness and the continuous coordination of actions among users [18,36,37]. In this domain, the literature reports three main limitations. First, studies on JI have largely concentrated on purely virtual entities [16,21,22,30,31], often exploring asymmetric AR-AR [28,34], VR-VR [21,38], and AR-VR [16,19,29–31] configurations. This means that scenarios involving interaction with real-world entities remain partially unexplored. Second, when interaction with the real world is involved, prior work has primarily focused on tangible interfaces [39] restricted to single-user scenarios [40] or co-located collaborative settings based on shared or replicated physical artifacts [41,42], disregarding JI across heterogeneous realities (i.e., multiple users immersed in different spaces along the R-V Continuum).

Third, even when asymmetric immersion is considered, the studies mainly involved users manipulating separate entities [17,27] or lacked experimental setups for comparing different interaction approaches [32,33].

To address this gap, this work presents an exploratory user study aimed at investigating different types of JI with entities that span multiple levels of immersion along the R-V Continuum and operate respectively on a real artifact and its virtual counterpart. This specific form of JI is later referred to as Cross-Reality Joint Interaction (CR JI).

Considering the wide range and complexity of factors involved in CR JI, the present work narrows the investigation to JI scenarios involving two human users across heterogeneous realities, by focusing on the effects of immersion level and interaction synchronicity across multiple dynamics of CR JI, on task-perceived presence and perceived workload, as well as other qualitatively assessed aspects such as sense of inclusion/exclusion, sense of empowerment, and perceived control. This objective is pursued through a study that employs a mixed-methods approach and integrates quantitative as well as qualitative analyses to address the following research questions:

RQ1 How does CR JI affect the perceived presence within the task?

RQ2 How does CR JI affect perceived workload?

RQ3 How do users perceive and describe their sense of inclusion/exclusion, empowerment, and control within CR JI?

The selection of the above research questions is grounded in the principal effects of collaboration across the R-V Continuum documented in the literature. More specifically, investigating perceived presence (RQ1) is crucial to understand whether differences in immersion level influence users' experience of the JI [21,29]. Perceived workload (RQ2) was examined since manipulating shared assets across heterogeneous realities may introduce significant cognitive load and communication overhead due to perceptual mismatches in users' respective representations of the assets [17,43]. Finally, exploring how users perceive inclusion/exclusion, empowerment, and control (RQ3) provides insights into the power dynamics and role imbalances that may emerge when one user interacts with physical artifacts while another relies on mediated digital representations [21,38,41].

All the dimensions investigated in this study through the above research questions have been evaluated considering variations in immersion level, interaction synchronicity, and JI dynamics

To conduct the study, this work relies on three interaction probes that have been designed to recreate familiar contexts – playing a checkerboard game, co-reading a comic book, and interacting with a shared music player – and involve two users, one manipulating a real-world artifact, the other engaging with its virtual counterpart. The interaction probes were also devised to explore three key dimensions of JI: task coupling (tightly vs. loosely coupled) [44], control distribution (individual, shared, and unilateral) [26,45], and inter-player dynamics (co-activity and competition) [46]. The configuration of these three dimensions guided the design of two main modes of interaction: synchronous and asynchronous. While all the interaction probes were intended to facilitate synchronous engagement, reflecting the inherently co-temporal nature of JI, asynchronous variants were selectively incorporated in cases where the task structure allowed participants to progress independently.

These choices were aimed at supporting an extensive analysis of the considered design space, capable of guiding future implementations of CR JI.

In summary, the main contribution of this work lies in an exploratory user study aimed at investigating the effects of immersion level and interaction synchronicity on perceived presence and task workload, as well as other qualitatively assessed aspects (e.g., inclusion/exclusion, empowerment, and control). The outcomes of the study are then synthesized into design implications for future CR JI systems, highlighting how representations, interaction techniques, and synchronization mechanisms can support meaningful participation, awareness, and coordination across heterogeneous realities.

2. Related works

To ground this work, prior research on shared manipulation and JI in immersive and semi-immersive environments has been analyzed, focusing on asymmetric settings with heterogeneous user configurations. The reviewed configurations range from those in which tasks involve interaction just with virtual entities to those where interaction, while remaining virtually mediated, targets real-world elements. The lack of systematic evaluations of JI in the latter contexts motivates this study, which examines CR JI across the R-V Continuum and calls for a more structured classification of such interactions.

2.1. Shared manipulation in CR-MU systems

As outlined above, research on shared manipulation in CR-MU systems has largely focused on JI within fully virtual tasks in immersive or semi-immersive environments (Fig. 1).

2.1.1. Asymmetric AR settings

Some studies have explored experimental AR setups employing different display technologies to access augmented content. As a matter of example, Friedl-Knirsch et al. investigated Hand-Held, Optical-See-Through, and Video-See-Through displays in a co-located AR data analysis task involving only virtual content, concluding that each device offered different trade-offs in terms of precision, usability, and collaborative engagement [28]. Other studies examined the effectiveness of collaborative workflows in role-asymmetric AR settings. For instance, Numan et al. examined a co-located AR scenario where experts, possessing extended operational privileges, guided non-experts in modifying an outdoor site through interaction with AR objects anchored to the physical environment and accessed via mobile devices [34]. The collaboration was synchronous, allowing all participants to immediately see and respond to each other's changes; however, anchoring virtual objects to physical space did not entail direct interaction with the physical environment.

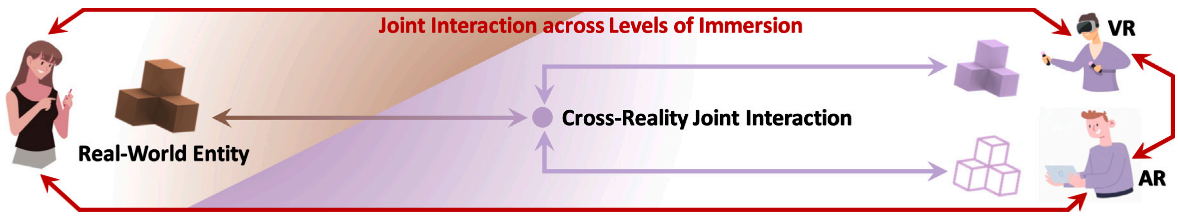


Fig. 1. Cross-Reality real-time Joint Interaction across physical and virtual realms.

2.1.2. Asymmetric VR settings

Experiments have been done also in the context of asymmetric VR settings. For instance, in [38], Ouverson et al. proposed a framework for asymmetric VR experiences in co-located settings, where participants use different devices such as headsets and tablets. The study focused on how technological asymmetries affect communication and collaboration within virtual environments. Similar objectives were pursued in the studies by Drey et al. [21] and Peter et al. [47], both emphasizing that asymmetric VR setups foster role-based interactions. However, none of the examined settings incorporated interaction with physical entities, limiting it exclusively to the virtual environment.

2.1.3. AR-VR settings

Continuing this line of thought, extensive research has explored settings where asymmetry extends beyond devices or roles to include varying levels of immersion along the R-V Continuum. As a first example, Grandi et al. evaluated performance and collaboration in VR-AR, VR-VR, and AR-AR pair-work settings, focusing on 3D object co-manipulation [29]. Another example is provided in [16], where Tserenchimed et al. proposed a viewpoint-sharing method for asymmetric VR/AR collaboration in which, instead of shifting the user's viewpoint, the virtual objects are transformed to match users' perspective. In [14], Huang et al. introduced VirtualNexus, an AR/VR remote collaboration system based on 360° video that employs environment cutouts and virtual replicas to support shared scene manipulation as well as to enhance spatial awareness and coordination. In [30], Cho et al. proposed X-person, a framework for asymmetric VR/AR interaction, enabling users with varying immersion levels to collaborate in a shared environment through dynamic adaptation of perspectives and interaction mappings. In a similar work, Cho et al. introduced XAVE, a cross-platform environment that extends asymmetric collaboration to a broader range of devices and input modalities, providing device-specific mappings to ensure coherent multi-user interaction and extend access to Extended Reality (XR) [31]. Finally, in [19], An et al. proposed a method for aligning heterogeneous XR systems within a shared physical space, thus enabling cross-device collaboration though still restricted to virtual interaction.

The above review shows that, although roles played by users, devices used, or immersion levels involved may be different, interaction in CR-MU systems remained mostly confined to the virtual domain so far, with physical space mainly serving as an anchor.

2.1.4. Virtual-mediated real-world interaction

A parallel line of work has explored the use of physical actuation to manipulate virtual content through tangible interaction [39], where physical artifacts serve as interfaces to control virtual representations. While effective in bridging physical and virtual interaction, these approaches predominantly focus on single-user scenarios [40] or on co-located multi-user interaction around the same or replicated tangible interfaces [41,42]. As a result, they do not explicitly address JI across heterogeneous realities or asymmetric levels of immersion.

Recent works have explored methodologies for both integrating non-immersive users into virtual environments and enabling remote users to participate in real-world interactions. In pursuit of the first goal, Jansen et al. proposed ShARe, an asymmetric co-located AR-VR system combining a head-mounted display (HMD) with a projector to align AR content with planar surface projections, enabling non-HMD users to interact via touch and gestures [22]. The second goal is generally achieved through remote expert-local guided interactions, where local users manipulate physical objects while remote experts supervise through virtual representations of the environment and corresponding actions, with varying degrees of fidelity. For instance, Chang et al. introduced a VR-AR remote JI technique where a VR expert guides a local AR user in assembly via virtual replicas overlaid on their physical counterparts [27], whereas Zaman et al. presented MRMAC, a system for asymmetric JI on LEGO and domino assembly where the so-called remote VR-Travelers access a virtual reference model to guide local AR-Hosts in real time [17]. In both cases, users manipulate separate entities that are not directly linked, such that actions performed on one entity do not affect the other. This limitation is partially overcome by the works by Lee et al. [32,33], who studied local users interacting with the physical environment while remote VR experts guided them via 3D models synchronized through the operator's coordinate system and AR tracking, facilitating real-time context awareness. However, although these works integrated real-world user interaction, the lack of experimental setups for systematically comparing various approaches limits rigorous validation, which would benefit from multiple implementations addressing different methodological aspects to yield robust and generalizable results.

2.2. Dimensions of JI

Understanding JI across different realities requires identifying its key dimensions, which are outlined below through a review of relevant literature.

2.2.1. Degree of task coupling

A JI is inherently influenced by the nature of the underlying task. Specifically, in [46] Jarrasse et al. distinguished between divisible tasks, allowing independent action, and interactive tasks, requiring explicit coordination. Each task can also be agonistic, with shared objectives, or antagonistic, with conflicting goals. This distinction aligns with models in collaborative virtual environments, where tasks are classified by their degree of coupling: loosely coupled tasks show low temporal and spatial interdependence, whereas tightly coupled tasks demand continuous synchronization, joint attention, and rich communication. The PASCAL framework by Gilbert et al. implements this distinction, enabling fluid transitions between coupling modes through spatial context sharing, holographic avatar replication, and parallel objects for action suggestion [44].

2.2.2. Control distribution

Control is a core aspect in JI, determining how actions are initiated, coordinated, and distributed. In virtual environments, it governs access to shared objects by specifying who can manipulate them, when, and how. In [26], Auda et al. investigated how control over virtual objects, mediated by haptic props, affects remote collaboration in VR, comparing single ownership, which limits interaction to locally associated objects, with shared ownership, which allows transferring control between users. Results indicated that restricted control enhances communication and awareness but slows completion, whereas shared control improves efficiency at the cost of reduced social interaction. In [45], Wieland et al. explored control strategies for collaborative 3D object manipulation in AR, comparing separation, where users perform different tasks simultaneously, composition, where both perform the same task, and a hybrid configuration. Results showed that separation improves efficiency through clear division of labor, while the hybrid configuration maximizes flexibility. These works highlighted the role of control in JI but focused on fully virtual and symmetrically immersive settings. In CR-mediated manipulation, interactions span different realities and heterogeneous interfaces, requiring control to be redefined to support coordination and mutual understanding.

2.2.3. Inter-player dynamics

The nature of interaction in joint motor tasks can be described through the taxonomy by Jarrasse et al. [46], which classifies interaction modes based on role distribution, interdependence, and interaction symmetry. Building on it, in [48], Mace et al. identified four main inter-player relationships: co-activity, where individuals act in parallel without significant influence; competition, where one's progress hinders the other's; collaboration, involving shared goals and continuous adaptation; and cooperation, with asymmetric but coordinated roles toward a common objective. This framework provides a solid basis for analyzing human–human dynamics in a JI and for designing systems that support diverse inter-player relationships in both physical and virtual contexts.

2.3. Summary and research gap

The reviewed literature shows that research on shared manipulation and JI in CR-MU systems has developed along two largely separate tracks. On the one hand, studies on asymmetric AR, VR, and AR-VR settings [14,16,19,21,28–31,34,38,47] have systematically varied devices, roles, and immersion levels, but have confined interaction to virtual content, with the physical environment used, at most, as a spatial anchor for virtual objects rather than as a target of direct manipulation. On the other hand, studies addressing virtual-mediated real-world interaction [17,22,27,32,33] have shown that JI can indeed extend to physical elements. However, these studies remain limited to isolated case studies built around a single implementation and a single pair of realities (typically VR-AR), each introducing its own task, metrics, and notion of coupling between users; as a result, they do not offer a shared framework for comparing how CR JI behaves as immersion asymmetry changes, nor do they systematically vary the position of each user along the R-V Continuum [1] while keeping the interaction with real-world elements constant.

The dimensions identified in Section 2.2 (i.e., degree of task coupling [44,46], control distribution [26,45], and inter-player dynamics [46,48]) have similarly been studied and validated only in fully virtual and symmetrically immersive settings, leaving open whether and how they generalize to CR-mediated manipulation of real-world entities.

The present work addresses this gap by systematically evaluating CR JI in tasks that involve direct interaction with real-world elements, rather than purely virtual content or physical objects used only as anchors. Building on the dimensions reviewed above, the proposed exploratory user study leverages a structured classification of such interactions and empirically assesses how CR JI is affected as the R-V configuration of the collaborating users changes.

3. Material and methods

This section outlines the design methodology guiding the development of the three interaction probes and details their technical implementation.

3.1. Interaction probes design

In defining the interaction probes, two layers of design considerations were taken into account: the first concerns the overall experimental design rationale, while the second focuses on the configuration of the JI dimensions outlined above.

3.1.1. Design rationale

The interaction probes were designed according to two main design principles:

- **Immersion Asymmetry:** In line with the state of the art, this work frames asymmetry not only as device heterogeneity [28] but as a divergence in the levels of immersion through which users engage with the shared environment. Each interaction probe situates one user in direct engagement with the real-world entity, while the other participates through its AR counterpart. This setting was purposefully selected, as AR supports a form of interaction in which immersion is conceived not as full presence within a virtual environment, but as an alternative task-oriented paradigm of interaction that simultaneously preserves a sense of belonging and connection to the real space [49,50].
- **Everyday Scenarios:** The interaction probes were designed around familiar, everyday tasks to support inclusive, accessible co-experience without requiring domain-specific expertise.

3.1.2. JI configuration

The dimensions of JI were configured to systematically explore multiple combinations, supporting diverse interaction settings while keeping both users equally engaged and aligned in hierarchical functions.

- **Degree of Task Coupling:** The interaction probes were designed to investigate how CR JI is influenced by the degree of task coupling, defined as the degree of interdependence required between users to complete the activity. CR JI was examined across tasks ranging from *tightly coupled*, where close coordination is essential, to *loosely coupled*, where users operate more independently [44,46].
- **Control Distribution:** To examine how control distribution influences CR JI, the interaction probes were designed to go beyond the assumption of symmetric control only: equal access to a shared object may not guarantee effective coordination, as it can introduce ambiguity or conflict in highly interdependent tasks. Accordingly, distinct control configurations are investigated, which, informed by established strategies discussed in prior works [26,45], were adapted to remain constrained and clearly distinguishable from one another: hybrid solutions were intentionally avoided to enable an unambiguous interpretation of the observed effects. Consequently, three refined control configurations are explored: *individual control*, in which each user manipulates separate components; *shared control*, in which both users jointly manipulate the same entity; and *unilateral control*, in which control is exclusively assigned to a single user.
- **Inter-Player Dynamics:** Previous studies often employed hierarchical role distributions [17,27,32,33] effective for guiding task execution but unsuitable for settings in which the remote or immersed user is required to assume an equally active role and where the main challenge is achieving comparable degrees of engagement. In contrast, this paper examines balanced CR JI by enabling both users to initiate, influence, and sustain joint action regardless of immersion level. To this aim, the inter-player dynamics identified by Mace et al. [48] were examined. *Competition* and *Co-Activity* were selected as they support balanced engaging and allow JI under both tightly and loosely coupled conditions without requiring strict reliance on shared objectives.

3.2. Implemented interaction probes

Based on the design considerations outlined above, the interaction probes developed in this work are described below.

3.2.1. Checkerboard

The Checkerboard system models a **tightly coupled** scenario where JI is essential. Turn-based board games like checkers or chess, common in studies of inter-player dynamics [51], require continuous mutual adaptation: each move influences the opponent's decisions and the game's progression; the antagonistic nature of the task allows for the examination of CR JI in which users experience different levels of immersion, revealing whether such differences lead to perceived advantages or disadvantages during the JI.

This system offers just a synchronous mode, reflecting the tightly coupled nature of the task. Each user controls their own set of pieces: the non-immersed one the physical board and the immersed one its virtual counterpart. This configuration models an **individual control** setup, with authority over separate components but tightly interdependent actions. In terms of inter-player dynamics, the system models *competition*, with control distributed and equal but strategically adversarial. Each user's move aims to counter the opponent, making progress depend on mutual constraints rather than a shared strategy. This setup examines how differences in immersion levels affect users' engagement in an antagonistic task when both participants have equal control.

3.2.2. Comic book

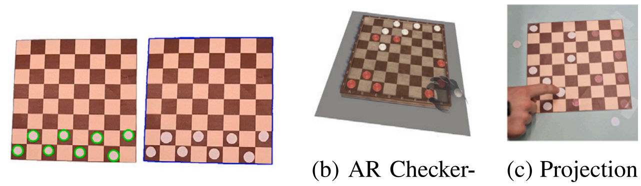
Shared reading among adults has been shown to increase social connectedness, emotional attunement, and well-being [52]. More specifically, in the case of digital comics, the tactile and sensory dimensions significantly influence the reading experience; moreover, active interaction with devices and platforms has been shown to enhance narrative immersion [53]. Building on these insights, shared comic book reading was chosen as a **loosely coupled** task to examine CR JI, where joint actions are non-essential but can enrich the experience.

The system implements two interaction modes, each reflecting a distinct control configuration. In synchronous mode, the non-immersed user manipulates the physical comic while the immersed user passively follows the reading via its virtual counterpart and coordinates verbally with the partner. This configuration corresponds to **unilateral control**. In asynchronous mode, each user interacts independently: the non-immersed one with the physical comic, and the immersed one with the virtual counterpart. This setup allows autonomous progression and represents **individual control**. These configurations supports the investigation of how unilateral versus individual control influences CR JI in low-dependency, asymmetrical settings. In terms of inter-player dynamics, both modes can be characterized as forms of *co-activity*. In the synchronous mode, co-activity manifests as a *leader–follower co-experience*, in which users remain temporally aligned, with one setting the pace and the other adapting to it. By contrast, the asynchronous mode supports a *parallel co-experience*, allowing users to progress independently without continuous coordination.

3.2.3. Music player

The third system considers shared music listening, an activity that, similarly to reading, when shared, gains a strong social dimension, enhancing enjoyment, emotional resonance, and interpersonal bonds [54]. Furthermore, music has the capacity to synchronize affective states and influence neurohormonal responses, such as oxytocin release linked to trust and social connectedness [55], making shared listening an alternative **loosely coupled** task for exploring CR JI where co-experience, though optional, is beneficial.

As in the previous case, two interaction modes are foreseen. In synchronous mode, both the users control the same music playback system, one via the real device and the other via its virtual counterpart, thus modeling **shared control** in which any action instantly affects playback for both. In asynchronous mode, each user interacts independently with their own counterpart, modeling an **individual control** setup. This configuration supports the analysis of how shared versus individual control influences CR JI in low-dependency, asymmetrical settings. Regarding inter-player dynamics, both modes embody *co-activity*: the synchronous mode supports *co-experience* through shared temporal alignment, while the asynchronous mode supports a *parallel co-experience*.



(a) Pieces Detection App (b) AR Checkerboard App (c) Projection Overlay

Fig. 2. Components of the Checkerboard CR system.

3.3. Implementation

This section presents the technical implementation of the three interaction probes, each combining real and virtual components. The latter were implemented through AR applications developed using Unity version 2022.3.21f1 and the Meta XR All-in-One SDK.[1] The AR applications were developed for a Meta Quest 3 headset, with interactions performed using its built-in hand tracking to preserve natural hand movements. The proposed interaction probes do not rely on spatial anchoring of virtual objects onto their physical counterparts. Instead, consistency between physical and virtual representations is maintained through application-level state synchronization, whose implementation is described in the following subsections.

3.3.1. Checkerboard

The Checkerboard system is composed of three key components:

- Pieces Detection App: A Python app executed on a mobile device and developed using OpenCV[2] to detect both the physical white pieces moved by the non-immersed user and the checkerboard edges (Fig. 2(a));
- AR Checkerboard App: A Unity AR application that renders the virtual counterpart of the checkerboard (Fig. 2(b));
- Projection Overlay: The projection of the virtual red pieces, controlled by the immersed user and not materially present onto the physical checkerboard (Fig. 2(c)).

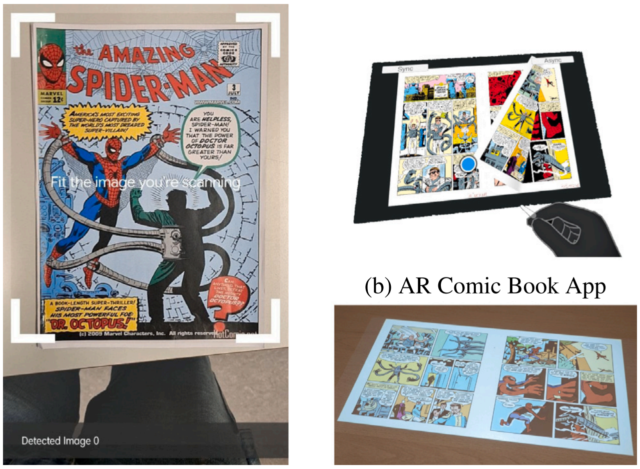
The Pieces Detection App detects the white pieces from a video stream and transmits their positions as JSON via WebSocket to the AR Checkerboard App, which maps them in the 3D space. Meanwhile, as the immersed user moves the virtual red pieces in AR through hand-tracking gestures, their rendering is transmitted via OBS Studio,[3] an open-source software for video capture and live streaming, and projected onto the physical checkerboard through the Projection Overlay.

3.3.2. Comic book

The Comic Book system is composed of three key components:

- Image Detection App: An app executed on a mobile device and developed using Android Studio[4] version 2022.3.1, designed to detect the pages of the physical book (Fig. 3(a));

[1] Meta XR All-in-One SDK: https://tinyurl.com/metaXRAllInOne.
[2] OpenCV: https://opencv.org.
[3] OBS Studio: https://obsproject.com.
[4] Android Studio: https://developer.android.com.



(a) Image Detection App (b) AR Comic Book App (c) Projection

Fig. 3. Components of the Comic Book CR system.

- AR Comic Book App: A Unity AR application enabling access to the 3D virtual counterpart of the book (Fig. 3(b));
- Projection: The projection of the virtual book onto a physical surface (Fig. 3(c)).

The Image Detection App leverages Google's ARCore[5] to detect the pages of the physical comic (Fig. 3(a)) by matching each observed page with a database of labeled, reference pages. By replacing the database and the labels, the system can adapt to any other book. Once a page is detected, its number is sent via WebSocket to the AR Comic Book App using the .NET WebSocketSharp[6] library. In synchronous mode, the AR Comic Book App updates the virtual book to mirror the physical page flips, with the immersed user being limited to passive viewing. In contrast, in asynchronous mode, the immersed user can interact with the virtual book through the AR Comic Book App using the headset's hand-tracking system. Specifically, pinch and tap gestures are used to flip pages, as shown in Fig. 3(b). The view of the AR Comic Book App is streamed via OBS Studio, like for the Checkerboard system, to the projection, thus allowing the non-immersed user to monitor the immersed user's page-flipping interactions (Fig. 3(c)).

3.3.3. Music player

The Music Player system is composed of two key components:

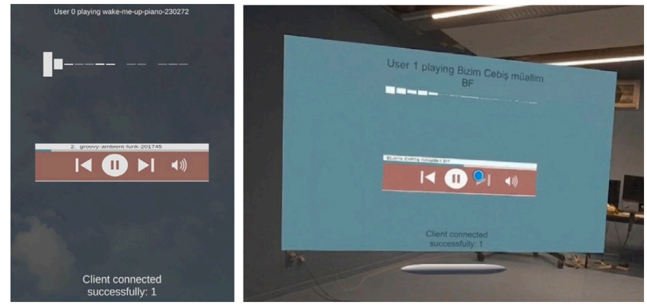
- Music Player App: A mobile Unity app for smartphones, serving as the real-world counterpart (Fig. 4(a));
- AR Music Player App: An AR application implementing its virtual counterpart (Fig. 4(b)).

In synchronous mode, playback is mirrored in real-time between the devices via Server–Client Remote Procedure Calls (RPCs) using Netcode for GameObjects[7] (NGO). Inputs from smartphone buttons and hand-tracking gestures trigger updates across the system, ensuring consistent functionality despite different interaction methods. In asynchronous mode, each user controls their own player independently, with status updates (e.g., "*User paused the track*") shared via RPCs.

[5] ARCore: https://developers.google.com/ar.

[6] WebSocketSharp: https://github.com/sta/websocket-sharp https://github.com/sta/websocket-sharp.

[7] NGO: https://tinyurl.com/NetcodeGO.



(a) Music Player App (b) AR Music Player App

Fig. 4. Components of the Music Player CR system.

4. Evaluation

This section introduces the exploratory user study conducted to examine the effects of immersion level and interaction synchronicity, across multiple types of CR JI, on task perceived presence, perceived workload, and other qualitative aspects, including sense of inclusion/exclusion, empowerment, and perceived control. The following subsections describe the participant demographics, experimental setup, and study procedure, including the evaluation measures, the structure of the experimental sessions, and the tasks participants were instructed to perform.

4.1. Participants

The study involved 18 participants (7 males, 10 females, 1 n.d.) aged 22 to 31 years (M = 25.39, SD = 2.64), recruited via a departmental newsletter announcement. All participants provided informed consent, and the study was approved by the KU Leuven's Ethics Committee (approval no. G-2024-8270-R2(MIN)). Experimental sessions were organized into 9 dyads. To characterize the sample, participants were also asked to rate their previous AR/VR experience (5-point scale, from 'Never Used" to "Frequent User") and familiarity with their dyad partner (5-point scale, from "Complete Strangers" to "Very Familiar"). The resulting scores (AR/VR experience: M = 3.22, SD = 1.11; familiarity with the partner: M = 3.67, SD = 0.87) indicate moderate prior exposure to immersive technologies and that the dyads were not composed of complete strangers.

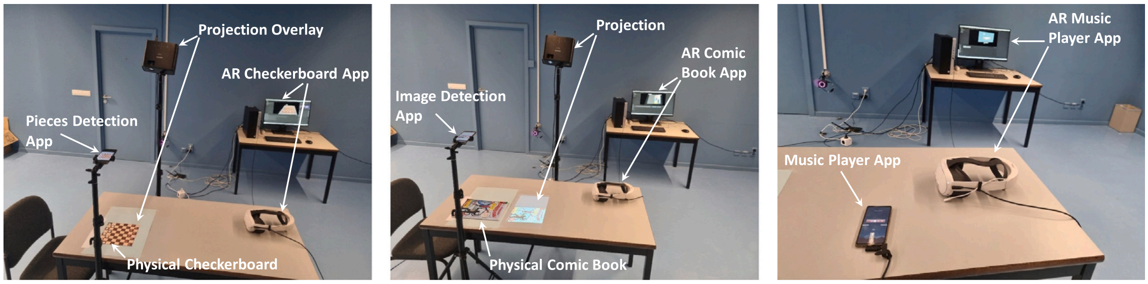
4.2. Experimental setup

The study was conducted in a controlled co-located setup, commonly used to emulate remote JI [56]. This isolated factors such as network latency and communication constraints, facilitated observation of user interactions and behaviors, and enabled pragmatic implementation choices, such as omitting VoIP communication and employing a simplified multi-user architecture, thereby accelerating the implementation process.

The hardware setup across the three interaction probes included the following components: a Meta Quest 3 headset used in tethered mode via Meta Quest Link, with the AR application running on a connected Intel Core i7-14700KF PC (32 GB RAM, GeForce RTX 4060), a Philips Neopix 120 home projector, and a Samsung Galaxy S20 FE smartphone. Each interaction probe employed a combination of these devices according to its requirements, as shown in Fig. 5.

4.3. User study

The user study was structured so that each dyad sequentially interacted with all three interaction probes following a within-dyads design.



(a) Checkerboard (b) Comic Book (c) Music Player

Fig. 5. Experimental setups.

4.3.1. Independent variables

For each interaction probe, the primary independent variable was the *immersion level*: one member of the dyad engaged with the virtual counterpart in the semi-immersed modality (S-IMM), while the other member interacted with the real-world counterpart in the non-immersed modality (N-IMM). For both the Comic Book and Music Player systems, *interaction synchronicity* was also manipulated as an independent variable, with dyads experiencing both synchronous (SYNC) and asynchronous (ASYNC) modes.

4.3.2. Dependent variables

In order to examine the quality of CR JI, this study investigated five principal dependent variables, drawn from the literature on XR JI. *Perceived presence* refers to the extent to which participants feel engaged and situated within the shared task regardless of immersion level. *Perceived workload* reflects the effort required to perceive, process, and respond to information during task performance. *Sense of inclusion/exclusion* captures subjective feelings of being integrated or marginalized within the interaction. *Empowerment* denotes the perceived advantage over others, arising, for non-immersed participants, from the possibility of direct interaction with the original entity and, for immersed participants, from the enhanced features and capabilities provided by immersive technology. Finally, *perceived control* refers to the participant's perceived ability to initiate actions and influence or manage the task's progression. These primary constructs were further refined to consider additional ones emerging from participants' evaluations, such as feedback and awareness, engagement, collaboration, connection and conflict.

4.3.3. Measures

To assess the dependent variables, the study employed both quantitative and qualitative measures: the former were collected via self-report questionnaires, while the latter were gathered through semi-structured interviews at the end of each experimental session.

Quantitative measures. Perceived presence was assessed through items adapted from the Multimodal Presence Scale (MPS) by Makransky et al. [57], grounded in Lee's tripartite model of presence [58]. Items were adapted to fit the different immersion conditions experienced by participants. In particular, items related to self-presence were adapted to capture participants' perceived connection with the shared CR JI activity rather than identification with a virtual body. For example, references to virtual embodiment were replaced with references to the virtual object (e.g., "*my virtual embodiment and my real body became one and the same*" became "*the virtual object and the real object became one and the same*"). Accordingly, perceived presence referred to the extent to which participants felt involved and experienced a sense of "being there" within the JI. The scale assessed whether the CR JI effectively bridged levels of immersion, fostering involvement in the joint activity despite asymmetric access. Responses were recorded on a 5-point Likert scale and aggregated into an overall perceived presence score, consistent with prior work reporting overall perceived presence measures [59,60].

Perceived workload was measured using the NASA Task Load Index (NASA-TLX) [61]. The questionnaires are available as supplemental material.

Qualitative measures. A semi-structured interview was conducted aiming both at providing additional validation of perceived presence and perceived workload, as well as at exploring constructs such as sense of inclusion/exclusion, empowerment, and perceived control. While perceived presence and workload were also assessed through validated quantitative instruments, the remaining constructs were intentionally investigated qualitatively. Given the exploratory nature of the study and the lack of widely adopted validated measures for inclusion/exclusion, empowerment, and perceived control tailored to CR JI contexts, a semi-structured interview was considered better suited to capture participants' subjective experiences. The semi-structured format also allowed predefined questions to be complemented with flexible probing, facilitating the emergence of unanticipated insights. Questions addressed task alignment between physical and virtual objects, mutual involvement, and the extent to which CR JI supported shared presence, while also probing preferences between SYNC and ASYNC, perceived pros and cons of S-IMM and N-IMM, and broader reflections on potential applications. The interview questions are provided in the supplementary material.

4.3.4. Experimental session

Each experimental session followed a structured procedure, shown in Fig. 6, which lasted approximately 2 h and consisted of three main phases: Introductory, Task Execution, and Interview.

For each participant dyad, the session began with a 10 min **Introductory Phase**, presenting key concepts of CR, followed by a brief overview of the test procedure. Participants then provided informed consent and completed a short demographic form including age and gender. During the **Task Execution Phase**, participants interacted with the three interaction probes across the various modalities. As an initial step, one participant was randomly assigned to the S-IMM and equipped with the Meta Quest 3 headset to access the corresponding AR application, while the other participant started with the N-IMM and was provided with the real-word counterpart. The Checkerboard system involved only the SYNC, followed by questionnaire completion. For the Comic Book and Music Player systems, participants experienced both the SYNC and ASYNC in counterbalanced order, completing a questionnaire after each mode. For each modality, participants were given a maximum interaction time of approximately 10 min; however, they were free to conclude the interaction earlier once they felt satisfied with the experience. Between systems, participants took a 10-minute break and alternated immersion levels, ensuring that each one experienced both the S-IMM and N-IMM. This counterbalancing strategy was applied to reduce habituation and minimize confounds such as fatigue or desensitization from prolonged exposure to one modality, while ensuring that all participants experienced both of them. The order

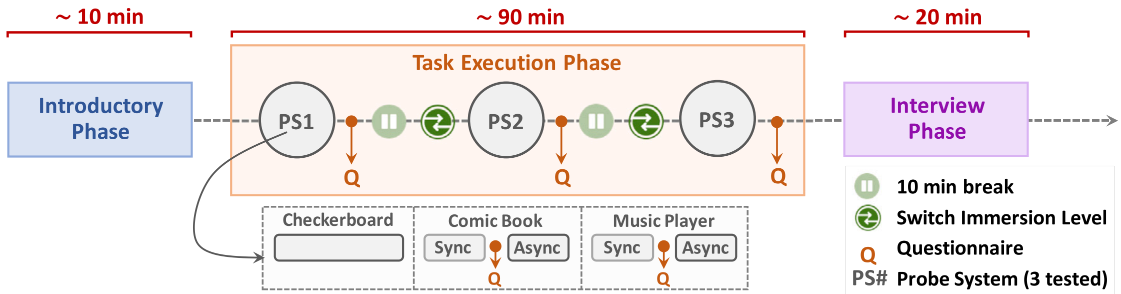


Fig. 6. Timeline and structural workflow of the experimental procedure.

of the interaction probes, however, remained fixed across sessions to ensure a consistent session flow and onboarding process. Specifically, the Checkerboard was presented first because its structured and rule-based nature allowed participants to familiarize themselves with the CR JI paradigm, the asymmetric immersion setting, and the interaction modalities before engaging in less constrained scenarios. The Comic Book and Music Player were subsequently introduced to explore different coordination and control requirements. Although maintaining a fixed order may introduce potential order-related effects, including learning or fatigue effects, this choice represented a methodological trade-off adopted within the exploratory scope of the study. Importantly, the interaction probes were designed as distinct exploratory instances of CR JI rather than directly comparable experimental conditions; therefore, the analyses focused on the effects of immersion level and interaction synchronicity within each probe. This structure also facilitated the collection of comparative insights between participants during the semi-structured interview conducted in the final **Interview Phase**, which lasted approximately 20 min.

4.3.5. Tasks

For each interaction probe, participants were asked to complete specific tasks, which are described below.

Checkerboard: Participants played a simplified variant of the traditional checkers game (Fig. 2), chosen for its low visual and cognitive demands, aiming to limit complexity-related load [62] and reduce potential performance anxiety [63]. Before beginning the task, participants received a brief explanation of the rules and were subsequently instructed to play ten moves each.

Comic Book: Participants interacted with a freely accessible Marvel comic book titled "*Spider-Man versus Doctor Octopus*" (Fig. 3) and were asked to locate specific sentences spoken by characters. The book's concise 25-page format allowed participants to finish the reading within the session's allotted time. To avoid monotony, the task was split into two halves, one read under the SYNC, the other under the ASYNC.

Music Player: Participants engaged with an interface designed to resemble common mobile music apps (Fig. 4) to enhance recognizability and reduce the need for instruction or prior familiarization [64]. The task required participants to listen to tracks and interact with all available buttons. The playlist featured five songs spanning different genres and, like for the Comic Book system, the task was divided into SYNC and ASYNC.

5. Results

This section presents the findings from both quantitative and qualitative evaluations, with the former based on statistical methods and the latter derived from thematic analysis.

5.1. Quantitative results

To obtain the quantitative results, two separate analyses were conducted: one aimed at assessing the impact of immersion level, and the other at assessing the impact of interaction synchronicity, both on perceived presence and perceived workload. Although addressing different independent variables, the two analyses followed a common methodological procedure. Normality was assessed on paired differences using Shapiro–Wilk tests (W and p reported in plots). Most comparisons met the normality assumption; for one comparison this assumption was violated. Accordingly, Paired-Samples T-Tests were used as primary analyses, and a Wilcoxon Signed-Rank test was additionally performed for the non-normal case as a robustness check, yielding consistent results. The first analysis assessed, for each interaction probe, the effect of immersion level by comparing S-IMM and N-IMM participants for each synchronization mode. The second analysis evaluated the effect of interaction synchronicity by comparing the SYNC and ASYNC for each interaction probe excluding the Checkerboard (which featured only the SYNC), and separately for S-IMM and N-IMM participants. All pairwise comparisons were treated as exploratory analyses, and no correction for multiple comparisons was applied. A significance threshold of $p < 0.05$ was adopted for all tests, with significant results indicated using brackets in plots. To evaluate the magnitude of the observed effects, Cohen's d was calculated as a measure of effect size to get insights on the practical significance of results. In addition to the above analyses, potential interaction effects between immersion level and interaction synchronicity were studied by means of Mixed Two-way Repeated Measures ANOVA.

To contextualize the statistical power of the reported comparisons, a post hoc sensitivity analysis was performed with G*Power (v3.1.9.7). The results indicated that, given the achieved sample size and the within-dyads design, the study had 80% power to detect only large effects ($d \geq 1.07$).

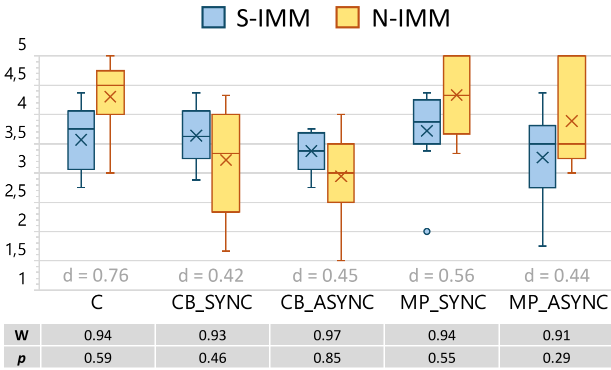
Immersion level. For brevity, the following abbreviations will be used to refer to the interaction probes: C (Checkerboard), CB_SYNC (Comic Book, Synchronous), CB_ASYNC (Comic Book, Asynchronous), MP_SYNC (Music Player, Synchronous), and MP_ASYNC (Music Player, Asynchronous).

As shown in Fig. 7(a), no statistically significant differences in perceived presence were found between the S-IMM and N-IMM conditions for any interaction probe.

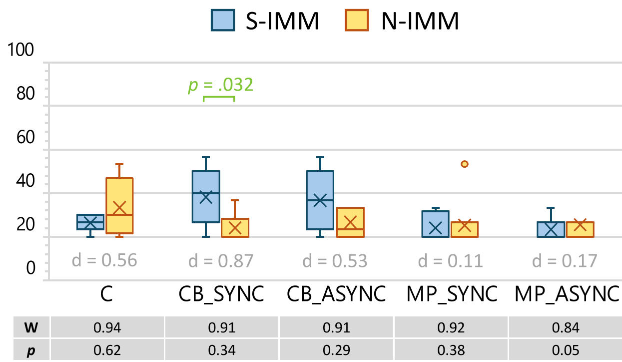
By contrast, Fig. 7(b) reveals a statistically significant difference in perceived workload between the S-IMM and N-IMM only conditions for the CB_SYNC (M = 38.15, SD = 12.81 vs. M = 24.07, SD = 5.96; p = .032; d = 0.87), with higher scores for the S-IMM.

Interaction synchronicity. Also here, the following abbreviations will be used: CB_S-IMM (Comic Book, Semi-Immersed), CB_N-IMM (Comic Book, Not Immersed), MP_S-IMM (Music Player, Semi-Immersed), and MP_N-IMM (Music Player, Not Immersed).

As shown in Fig. 8(a), from the perspective of perceived presence, significant differences between the SYNC and ASYNC were found in the

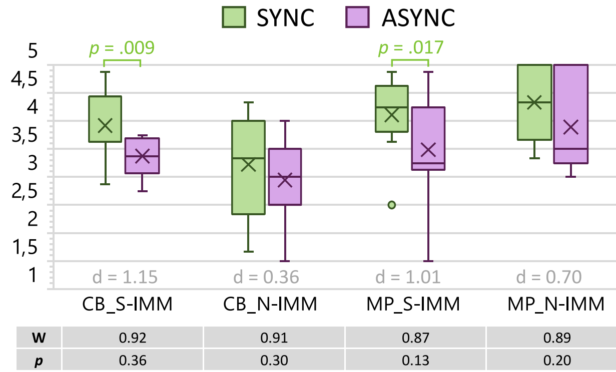


(a) Perceived presence

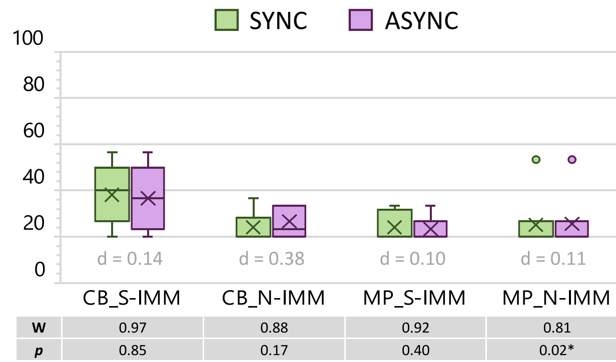


(b) Perceived workload

Fig. 7. Quantitative results comparing S-IMM and N-IMM participants. Each box plot displays the median (horizontal line), mean (×), interquartile range (box edges), minimum and maximum observed values (whiskers), and outliers (∘). Statistically significant differences are highlighted in green.



(a) Perceived presence



(b) Perceived workload

Fig. 8. Quantitative results comparing SYNC and ASYNC modes. Each box plot displays the median (horizontal line), mean (×), interquartile range (box edges), minimum and maximum observed values (whiskers), and outliers (∘). Statistically significant differences are highlighted in green.

CB_S-IMM (M = 3.92, SD = 0.63 vs. M = 3.38, SD = 0.36; p = .009; d = 1.15) and the MP_S-IMM (M = 4.11, SD = 0.71 vs. M = 3.49, SD = 0.97; p = .017; d = 1.01), indicating that SYNC participants reported substantially higher perceived presence in the SYNC compared to the ASYNC. Regarding the other cases, no significant differences were found. Referring to Fig. 8(b), perceived workload remained generally low across all interaction probes in both the SYNC and ASYNC, and no significant differences were found.

Interaction effects. The results of the Two-Way Mixed ANOVA analysis are reported in Table 2. For the Comic Book (CB), no significant main effect of immersion level or interaction synchronicity were observed. No significant interaction effect was found. Conversely, for the Music Player (MP), a highly significant main effect of interaction synchronicity was observed ($F(1,16) = 10.54, p = .005, \eta_p^2 = .40$), indicating that synchronous interaction scenarios yielded substantially higher perceived presence across both immersion groups. The main effect of immersion level trended toward significance ($F(1,16) = 3.54, p = .078, \eta_p^2 = .18$), while the interaction effect between immersion and synchronicity was non-significant.

Regarding workload with the CB, a significant main effect of immersion level was found ($F(1,16) = 7.40, p = .015, \eta_p^2 = .32$), with immersed (IMM) users experiencing a significantly higher workload ($M = 37.41$, $SD = 12.18$) compared to non-immersed (NOT-IMM) users ($M = 25.37$, $SD = 5.86$) regardless of synchronicity. No significant main effect of synchronicity or interaction effect was found. Finally, for the MP, there were no significant main effects for immersion, synchronicity, or their interaction.

5.2. Qualitative results

The qualitative data gathered from 17 semi-structured interviews (participants P1–P17, as participant P18 was unable to complete the interview) were analyzed using Thematic Analysis following the six-phase procedure by Braun and Clarke [65].

The analysis followed a hybrid inductive-deductive approach. An initial set of codes was derived deductively from constructs defined by the evaluation metrics, while additional codes were developed inductively from the data to capture emerging and unanticipated patterns. A codebook was iteratively developed and refined throughout the analysis process.

All transcripts were coded iteratively by a single researcher and assigned codes to meaningful text "segments". The resulting codes were then grouped into candidate themes based on shared semantic meaning. These themes were subsequently reviewed, refined, and consolidated through repeated comparisons with both the coded segments and the original transcripts to ensure internal coherence, consistency with participants' reported experiences, and mitigate interpretive bias. The codebook was continuously refined throughout this process as new patterns emerged. To enhance the transparency of the analysis, representative participant quotations are reported throughout the Results section, allowing readers to assess the relationship between the raw data and the resulting interpretations.

As shown in Fig. 9, the distribution of coded segments across themes is reported descriptively to illustrate thematic recurrence. The first five themes correspond to constructs anticipated by the evaluation metrics, while the remaining themes emerged inductively during the analysis.

Table 2
Summary of two-way mixed ANOVA statistical results for perceived presence and perceived workload.

Outcome variable	Probe	Immersion level	Interaction synchronicity	M (SD)	Significant results
Perceived Presence	CB	S-IMM	SYNC	3.64 (0.46)	M_{imm} : $F(1,16) = 2.38, p = .142, \eta_p^2 = .13$
			ASYNC	3.38 (0.34)	M_{sync} : $F(1,16) = 3.51, p = .080, \eta_p^2 = .18$
		N-IMM	SYNC	3.22 (0.86)	$imm \times sync$: $F(1,16) < 0.01, p = .962, \eta_p^2 < .01$
			ASYNC	2.94 (0.68)	
	MP	S-IMM	SYNC	3.72 (0.68)	M_{imm} : $F(1,16) = 3.54, p = .078, \eta_p^2 = .18$
			ASYNC	3.26 (0.74)	M_{sync} : $F(1,16) = 10.54, p = .005, \eta_p^2 = .40$
		N-IMM	SYNC	4.33 (0.61)	$imm \times sync$: $F(1,16) < 0.01, p = .961, \eta_p^2 < .01$
			ASYNC	3.89 (0.81)	
Perceived Workload	CB	S-IMM	SYNC	38.15 (12.08)	M_{imm} : $F(1,16) = 7.40, p = .015, \eta_p^2 = .32$
			ASYNC	36.67 (13.05)	M_{sync} : $F(1,16) = 0.07, p = .796, \eta_p^2 < .01$
		N-IMM	SYNC	24.07 (5.62)	$imm \times sync$: $F(1,16) = 0.93, p = .349, \eta_p^2 = .05$
			ASYNC	26.67 (6.09)	
	MP	S-IMM	SYNC	24.72 (5.83)	M_{imm} : $F(1,16) = 0.19, p = .671, \eta_p^2 = .01$
			ASYNC	23.33 (4.71)	M_{sync} : $F(1,16) = 0.02, p = .892, \eta_p^2 < .01$
		N-IMM	SYNC	25.19 (10.32)	$imm \times sync$: $F(1,16) = 0.17, p = .684, \eta_p^2 = .01$
			ASYNC	25.56 (10.18)	

Perceived Presence (17 segments) was heightened across all SYNCs: in coordinated music playback (P3), shared comic reading (P1), and most notably within the checkerboard's competitive game (P5: "*Felt part of the game*"; P7: "*Like a real game*"), where turn-taking mechanics generated a "*real interaction*" (P5, P6) and a "*strong connection between real/virtual objects*" (P17). Additionally, AR arm movements promoted tangible co-presence (P9: "*Like interacting with a real person*"), while interface familiarity helped bridge realities (P3: "*Like a real object*"). By contrast, ASYNCs appeared to diminish this sense of presence (P3: "*Felt in a different world*"; P15: "*Low sense of partner's presence*").

Perceived workload findings (11 segments) emerged exclusively in the Comic Book: S-IMM participants reported AR navigation as effortful (P16: "*More effort in AR...have to pinch and point fingers*"; P7: "*Pinch multiple times to turn*"), whereas N-IMM participants found physical interaction more intuitive (P4: "*Easier in real world*"; P12: "*Didn't require much effort*"). Notably, the SYNC introduced pacing considerations, with N-IMM participants feeling compelled to monitor the partner's progress (P3: "*Need to [...] turn the pages a bit slower*"; P5: "*I felt pressure: I didn't know if the other had finished reading*") and S-IMM participants struggling with pace misalignment (P4: "*Couldn't read slower [...] pages flipped faster*"). This coordination burden was alleviated in the ASYNC (P3: "*Don't have to mind other's pace*"; P4: "*Better than the SYNC*").

Exclusion/Inclusion (23 segments) emerged mostly in the Comic Book, where participants related it to the absence/presence of visual cues (P4: "*Couldn't see or experience the other; P8: "Could see everything [...] didn't feel excluded*"). S-IMM participants noted AR isolation (P1: "*I was engrossed with the headset [...] quite isolated*"; P12: "*The headset isolated me*"), particularly in the ASYNC (P5: "*More in my own world*"; P11: "*A bit excluded*"), while the SYNC generally improved inclusion (P1: "*Equally participating*"; P2: "*For the synch it was ok*"). However, ASYNC was found to support individual focus (P11: "*Didn't feel distracting [...] felt like reading together*"; P15: "*Focused on my task*").

Empowerment (12 segments) mostly emerged with N-IMM participants often reporting increased agency through physical interaction (P4: "*Do whatever I want*"; P12: "*Having direct command over the pages*"), which occasionally resulted in creating subordination dynamics, particularly in the SYNC of the Comic Book (P10: "*I dictated the tempo [...] the other had to follow*"). S-IMM participants, on the other hand, expressed mixed perceptions, with some seeing AR as offering expanded possibilities (P15: "*More freedom than reality*", while others feeling constrained by the AR interface (P8).

Perceived control (22 segments) was one of the most discussed themes, and the analysis showed that ASYNCs consistently supported individual autonomy (P6: "*Turn the page whenever I want*"; P10: "*Had more control*"; P13: "*Can do more*"), whereas SYNCs often constrained it, except in the Checkerboard, where the turn-based structure distributed control evenly (P4: "*Play independently without dominance*"). In the Comic Book, some noted that "*when in control, the other person seemed isolated*" (P9) and also emphasized that the unconstrained nature of physical interaction facilitated their agency (P10: "*Can turn three pages at the same time*"; P1: "*Full control over what I was doing*").

Feedback and awareness (21 segments) emerged as another critical factor. SYNCs naturally enhanced mutual awareness (P12: "*Nice to see the other person interacting*") while ASYNCs often impaired it (P6: "*Wonder what the other is doing*"; P12: "*Wasn't aware of the other*"; P13: "*Have to explain what's happening*"), directing participants' attention toward their own tasks (P8: "*Focused on the task rather than the partner*"; P14: "*Busy with my things*"). Only a minority found feedback sufficient (P2: "*The other sees my projection and I'm notified while remaining in my zone*").

Engagement (16 segments) peaked in the Checkerboard, described as "*more real and fun*" (P14). The competitive dynamic enhanced the experience (P1: "*Quite exciting [...] gave me something extra*"; P5: "*More real interaction [...] moves impacted the other*"). SYNCs consistently boosted engagement across the interaction probes (P14: "*Music Player more engaging in SYNC*"; P5: "*Interaction with others increases engagement*"). Manipulation was equally engaging, whether virtual (P3: "*Enjoyed moving virtual pieces*") or physical (P4: "*Easier/more entertaining in the real world*").

Collaboration (15 segments) and **connection** (17 segments) were consistently enhanced by SYNCs. In the Checkerboard, game mechanics made the CR JI more natural (P8: *Easier to interact with the other*"; P13: *Influences the shared state*"; P17: *Actions impact*"). SYNCs also fostered connection across activities, from shared gameplay (P1: *As playing with a real person*") to coordinated music listening (P6: *Shared song created connection*"). By contrast, ASYNCs reduced joint participation and mutual connection (P9: *Each does their own thing*"; P17: *No collaboration*"), and interactions often felt disconnected (P5: *Experiences felt disconnected*"; P15: "*Monitoring without connecting*").

Conflict (7 segments) arose mainly in the Music Player, due to asynchronous audio overlap (P11: "*Distracting, could hear the overlap*"; P16: "*Should use headphones*") and in the Comic Book, where the SYNC created pacing challenges requiring "*negotiation of reading speed*" (P17), mitigated by the ASYNC's independence (P6: "*Turn pages whenever I want*"; P10: "*My own space*").

6. Discussion

The study's combination of qualitative and quantitative methods provided complementary evidence to address the proposed research

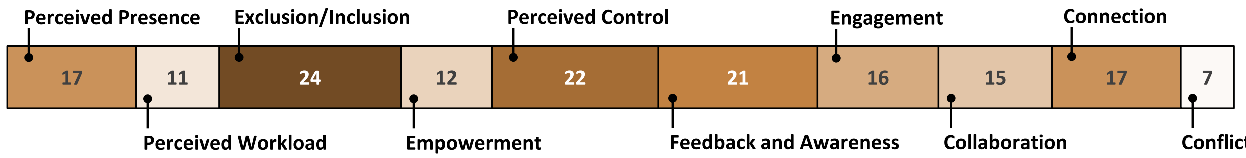


Fig. 9. Theme recurrence in terms of number of segments, with color intensity increasing with occurrence counts.

questions. Given the nature of the study and the number of participant dyads involved, the findings should be interpreted as exploratory observations emerging from the interaction of multiple design dimensions rather than broadly generalizable conclusions. Since immersion level, interaction synchronicity, and probe-specific interaction characteristics were intentionally combined within ecologically meaningful interaction settings, their individual contributions cannot be fully disentangled. Instead, the results highlight relevant insights that motivate future confirmatory investigations.

Building on these considerations, the following subsections discuss the findings with respect to the three research questions, present the resulting design implications for CR JI systems, and outline the limitations of the study and future research directions.

6.1. Discussion of the research questions

The discussion below addresses the three research questions and interprets the main findings of the study by integrating quantitative results with qualitative insights.

RQ1 (How does CR JI affect the perceived presence within the task?): The findings suggest that perceived presence within the CR JI may be more associated with interaction synchrony and task coupling than with immersion level, as no significant differences emerged between the S-IMM and N-IMM.

This interpretation is consistent with prior work on JI involving users in different immersive setups (e.g., VR-AR), suggesting that the experience is shaped not only by the difference in immersion level but also by the dynamics of interaction during the shared task [18,66]. However, this result should be interpreted considering the co-located nature of the experiment. Physical co-presence may have provided additional social and environmental cues, such as direct verbal communication and peripheral awareness of the partner, which have been shown to support users' awareness of others' actions and involvement during collaborative activities [67,68]. These cues may have contributed to support participants' experience of the shared activity independently of the immersion condition. Consequently, the absence of significant differences between the S-IMM and N-IMM should not be interpreted as evidence that immersion asymmetry is irrelevant, but rather as an indication that, under the tested conditions, other factors may have played a more prominent role in shaping perceived presence during the CR JI. Similar effects could potentially emerge in remote CR JI scenarios when adequate mechanisms for preserving social and environmental awareness are provided; conversely, when these forms of awareness are not adequately mediated [69,70], remote settings may amplify the effects of immersion asymmetry by influencing how both users perceive their partner's actions, involvement, and contribution to the shared activity. These considerations suggest that examining the interplay between immersion asymmetry, awareness, and interaction dynamics across different deployment settings represents an interesting direction for future CR JI research.

Nonetheless, within the S-IMM, the SYNCs were associated to higher perceived presence during the CR JI than the ASYNCs; this effect was confirmed qualitatively, with participants describing the SYNCs as fostering collaboration while naturally increasing awareness of ongoing events. This pattern was observed across the SYNCs despite their different interaction characteristics. However, the contribution of probe-specific factors to the observed effects cannot be excluded. In particular, in the Checkerboard, the high interdependence between users' actions and the competitive nature of the task may have contributed to reducing perceived separation between realities, suggesting that tightly coupled interactions may reinforce perceived presence [71], although the specific contribution of competitive dynamics remains intertwined with the characteristics of the task.

RQ2 (How does CR JI affect perceived workload?): Quantitative results showed perceived workload did not differ significantly between SYNC and ASYNC, suggesting that interaction synchronicity alone may not substantially affect workload under the tested conditions. This is consistent with previous studies on asynchronous interaction supported by immersive technologies, which indicate that, although asynchronous interaction can reduce the need for real-time coordination, its impact on user experience and workload depends on the specific task requirements rather than being inherently beneficial or detrimental [72]. The only significant difference emerged when comparing immersion levels in the CB_SYNC condition, where S-IMM participants reported higher workload than N-IMM participants. Qualitative findings provide a possible interpretation of this difference: in the CB_SYNC, S-IMM participants seemed to experience higher workload than N-IMM participants because they had no control over the comic book and often struggled to keep pace with the fully controlling partner. This dependence may have reduced their sense of autonomy by requiring continuous adaptation to the partner's pace, potentially increasing coordination effort, consistent with previous work reporting an inverse relationship between human autonomy and perceived workload in collaborative settings [73]. However, qualitative findings suggest that autonomy alone may not fully explain this difference. Although N-IMM participants retained direct control over the comic book, they were also required to coordinate their actions with their S-IMM partner's pace, indicating that direct control did not necessarily imply complete autonomy within the task. Therefore, the higher workload experienced by S-IMM participants may be related not only to reduced autonomy but also to their limited control over the interaction process, increased dependence on their partner, and the additional perceptual and interaction demands potentially associated with reading a virtual comic book in AR.

Although no significant differences emerged in the remaining conditions, the observed effect sizes provide some exploratory indications. In particular, the CB_ASYNC condition showed a moderate effect size, which should be interpreted cautiously given the limited statistical power of the study. However, qualitative results suggest a possible interpretation of this tendency: while asynchronous interaction allowed participants to navigate independently, some participants perceived virtual manipulation as less direct and flexible than interacting with the physical counterpart. These findings indicate that perceived workload in CR JI may depend not only on immersion and synchronicity, but also on interaction design choices, highlighting the importance of providing adequate control and matching virtual interactions to natural degrees of freedom [74] to minimize frustration and perceived inequity when manipulating 3D entities.

RQ3 (How do users perceive and describe their sense of inclusion/exclusion, empowerment, and control within CR JI?): The qualitative findings suggest that participants' experiences of inclusion/exclusion, empowerment, and perceived control appeared to be related to how interaction capabilities were distributed across realities, consistent with previous observations [30,75,76].

Even in cases where S-IMM participants lacked direct control (e.g. with the Comic Book), the SYNCs still allowed them to remain actively involved in the shared activity, indicating that synchronous CR JI may represent a promising approach to actually bridge realities. By contrast, the ASYNCs appeared to shift participants' attention to individual components, reducing awareness, inclusion, and connection while increasing perceived control and focus on individual outcomes consistent with prior work on control distribution effects [26,45].

Across the interaction probes, directly touching the physical object gave users a stronger sense of control and was often associated with immediacy and unconstrained manipulation [39,77,78], whereas AR interaction elicited more heterogeneous perceptions. While some participants perceived the virtual interface as enabling capabilities beyond those available in the physical world, others experienced it as introducing new forms of restriction, in line with [79]. This suggests that the way interaction capabilities are provided and constrained may play a key role in shaping users' sense of empowerment, rather than virtual interaction itself. Across synchronicity modalities, ASYNCs were generally associated with greater individual autonomy, whereas SYNCs tended to limit perceived control. This latter pattern was particularly evident in the CB_SYNC condition, where the same asymmetry discussed in RQ2 in relation to workload emerged as a key factor shaping participants' perceptions of control. S-IMM participants had limited influence over task progression, as navigation depended on their N-IMM partner. However, N-IMM participants, despite directly manipulating the comic, still needed to adapt to their S-IMM partner's pace. These findings suggest that perceived control in asymmetric interactions appears to depend not only on who can directly manipulate the shared artifact but also on how users can contribute to and influence the shared activity. A different pattern emerged in the Checkerboard, where turn-based mechanics naturally regulated interaction opportunities, making moments of contribution more predictable and easier to coordinate [70]. Consequently, participants rarely described the interaction in terms of dominance or reduced control, indicating that the turn-based structure of the activity may have helped balance opportunities for contribution despite asymmetric interaction capabilities.

These observations suggest that, in CR JI, differences in immersion level may not necessarily translate into perceived inequalities in inclusion, empowerment or control. Rather, users' perceptions appear to depend on whether the interaction design enables both participants to contribute meaningfully despite their asymmetric interaction capabilities. While these considerations require further validation, they point toward design directions such as explicit turn-taking mechanisms, shared decision points, or interaction techniques that allow both users to influence task progression even when only one directly manipulates the shared artifact.

6.2. Design implications

Taken together, the findings discussed above highlight a central design challenge for CR JI systems: managing asymmetry between users across realities. The following implications are preliminary design propositions grounded in the observed cases, whose broader applicability is constrained by the small number of dyads and the probe-specific nature of the study.

Unlike traditional interactive systems, CR JI inherently introduces differences in perception, embodiment, and available interaction capabilities. These asymmetries should not necessarily be removed, as different realities can provide complementary affordances; rather, they should be designed so that differences in interaction capabilities do not prevent users from meaningfully affecting the shared activity.

In this perspective, synchronization in CR JI should not be considered only as a temporal alignment of object states or user actions, but also as a semantic alignment of how changes in the shared environment are understood across realities. The results suggest that simply synchronizing actions may be not sufficient: in the Comic Book, for example, synchronous manipulation ensured consistency between realities but could still generate perceived imbalance when one user's actions determined the progression of the other. Conversely, in the Checkerboard, the structure of the interaction provided clearer and more predictable opportunities to contribute, reducing perceptions of dominance despite asymmetric interaction capabilities.

Therefore, future CR JI systems may benefit from moving beyond the goal of reproducing equivalent interaction experiences across realities and instead focus on supporting meaningful interaction despite heterogeneous capabilities. For example, in scenarios similar to the Comic Book, where one user directly manipulates a physical object while another relies on an augmented representation, the system could provide mechanisms such as previews of intended manipulations, ways to express preferences, or alternative forms of influence over the shared object state. Similarly, when users perform different actions across realities, visual feedback should communicate not only the updated state of the object, but also how and why that state changed, for example by indicating who triggered an action, the partner's current focus, or pending changes. These cues could preserve the flexibility of independent interaction while reducing uncertainty about the partner's activity. More broadly, these observations highlight potential directions for future CR representations, interaction techniques, and synchronization mechanisms that balance consistency across realities with users' awareness and ability to influence the shared environment.

Overall, the findings suggest that effective CR JI design may benefit from focusing less on minimizing asymmetry and more on ensuring that asymmetric interaction capabilities still support meaningful participation, awareness, and coordination across realities. Table 3 summarizes the main design implications emerging from these observations.

6.3. Limitations and future works

Although the current study provides exploratory insights into CR JI, the findings should be interpreted in light of certain limitations.

The relatively small sample size (9 dyads) limited statistical power and generalizability. The sensitivity power analysis indicated that the study was primarily able to detect large effects, suggesting that smaller effects may require further investigation. Future studies with larger samples could provide further evidence supporting these findings and extend the investigation beyond dyadic CR JI scenarios, where increasing the number of simultaneous users may introduce additional challenges related to coordination, mutual awareness, and interaction dynamics beyond those explored in the present study.

The co-located setup enabled controlled comparisons by reducing factors such as network latency and communication delays. However, physical co-presence may have provided additional social cues, such as direct communication and peripheral awareness, which may have influenced presence and inclusion independently of immersion. Future work should investigate remote CR JI scenarios to understand how mediated social awareness, communication constraints, and network factors influence the relationship between immersion asymmetry, interaction dynamics, and user experience.

The interaction probes were designed to capture different CR JI configurations by intentionally varying task characteristics, interaction constraints, and coordination requirements. This approach enabled the exploration of a broader design space, although differences between scenarios may also have contributed to the observed effects alongside the investigated factors. Similarly, the fixed probe order was adopted to ensure a consistent experimental procedure across sessions, but potential learning or fatigue effects cannot be completely excluded. Future studies could extend this investigation through counterbalanced designs and more controlled variations of individual interaction dimensions.

The exploratory nature of the study motivated the use of subjective measures and qualitative interviews to capture participants' experiences. However, integrating objective behavioral measures, such as task

Table 3
Cross-Reality Joint Interaction future design implications emerging from the study.

Challenge	Design implication
Supporting perceived presence and inclusion	Prioritizing synchronous interaction when maintaining a shared sense of participation is more important than maximizing individual autonomy, particularly in tightly coupled activities.
Balancing coordination and autonomy	Adapt the level of synchronization to the coordination requirements of the task rather than enforcing a single interaction strategy across all CR JI scenarios.
Managing asymmetric interaction capabilities	Ensure that all users can meaningfully influence task progression even when only one user directly manipulates the shared artifact or users have different interaction capabilities.
Supporting awareness across realities	Provide awareness cues that communicate not only object state changes, but also who initiated them, the partner's current focus, and pending actions to facilitate coordination across realities.
Designing interaction techniques	Design interaction techniques according to task requirements and complementary affordances of each reality, rather than aiming to reproduce identical interaction experiences across realities.

performance, gaze patterns, movement analysis, or conversational analysis, could provide additional insights into CR JI dynamics. Moreover, the thematic analysis was performed by a single researcher, and this should be taken into account when interpreting the robustness of the qualitative findings.

Finally, the study focused on relatively basic interaction scenarios. Investigating richer CR JI settings could reveal additional interaction and coordination challenges beyond those observed in the present study, including scenarios involving more complex multi-user configurations or interactions between humans and intelligent agents. Finally, the relatively short interaction duration provided insight into initial CR JI experiences, while longer-term adaptation, sustained collaboration, and potential physical discomfort associated with prolonged AR usage remain important directions for future investigation.

7. Conclusion

This study explored how immersion level and interaction synchronicity shape user experience in CR JI. Quantitative findings suggest that synchronous interaction was associated with higher perceived presence, while qualitative accounts pointed to a stronger sense of inclusion and connection. Perceived workload showed differences across immersion conditions, with qualitative findings further highlighting that it appears to be more closely related to interaction constraints and control distribution than to immersion asymmetry alone. Furthermore, the results suggest that effective CR JI should aim to balance the complementary affordances of physical and virtual interaction while ensuring that both users can meaningfully contribute to the shared activity. Although exploratory, these findings provide initial design considerations for future CR systems.

Future research should extend this investigation to larger samples, remote settings, CR scenarios involving more than two simultaneous users or interactions between humans and agents, and longer-term interactions, while incorporating additional behavioral measures to further understand coordination, awareness, and control in CR JI.

CRediT authorship contribution statement

Roberta Macaluso: Writing – original draft, Formal analysis, Conceptualization. **Noemi Di Cicco:** Software, Methodology. **Alberto Cannavò:** Writing – original draft. **Adalberto L. Simeone:** Supervision, Methodology, Conceptualization. **Fabrizio Lamberti:** Writing – review & editing, Supervision, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work, the authors used Gemini and ChatGPT to revise and improve the English language of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research is supported by Internal Funds KU Leuven (C14/20/078).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cag.2026.104752.

Data availability

No data was used for the research described in the article.

References

[1] Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: A class of displays on the reality-virtuality continuum. In: Proc. SPIE telemanipulator and telepresence technologies. 1995, p. 282–92.
[2] Auda J, Gruenefeld U, Faltaous S, Mayer S, Schneegass S. A scoping survey on cross-reality systems. ACM Comput Surv 2023;56(4):1–38.
[3] Cools R, Han J, Esteves A, Simeone AL. From display to interaction: Design patterns for cross-reality systems. IEEE Trans Vis Comput Graph 2025;31(5):3129–39.
[4] Simeone AL, Khamis M, Esteves A, Daiber F, Kljun M, Čopič Pucihar K, Isokoski P, Gugenheimer J. International workshop on cross-reality (XR) interaction. In: Proc. ACM conference on interactive surfaces and spaces. 2020, p. 111–4.
[5] Cools R, Esteves A, Simeone AL. Blending spaces: Cross-reality interaction techniques for object transitions between distinct virtual and augmented realities. In: Proc. IEEE international symposium on mixed and augmented reality. 2022, p. 528–37.
[6] Feld N, Bimberg P, Weyers B, Zielasko D. Simple and efficient? Evaluation of transitions for task-driven cross-reality experiences. IEEE Trans Vis Comput Graph 2024;30(12):7601–18.
[7] Pavavimol T, Ometov A, Valkama M, Thibault M. Transitions between realities: A systematic review on the usage of XR systems for bridging reality and virtuality. In: Proc. ACM international conference on interactive media experiences. 2025, p. 158–74.
[8] Pointecker F, Friedl J, Schwajda D, Jetter H-C, Anthes C. Bridging the gap across realities: Visual transitions between virtual and augmented reality. In: Proc. IEEE international symposium on mixed and augmented reality. 2022, p. 827–36.
[9] Ioannidis P, Eklund L, Løvlie AS. We dare you: A lifecycle study of a substitutional reality installation in a museum space. J Comput Cult Herit 2021;14(3):1–21.
[10] Kari M, Grosse-Puppendahl T, Coelho LF, Fender AR, Bethge D, Schütte R, Holz C. TransforMR: Pose-aware object substitution for composing alternate mixed realities. In: Proc. IEEE international symposium on mixed and augmented reality. 2021, p. 69–79.

[11] Simeone AL, Velloso E, Gellersen H. Substitutional reality: Using the physical environment to design virtual reality experiences. In: Proc. ACM conference on human factors in computing systems. 2015, p. 3307—-3316.

[12] Tao Y, Lopes P. Integrating real-world distractions into virtual reality. In: Proc. ACM symposium on user interface software and technology. 2022, p. 1–16.

[13] Furukawa T, Yamamoto D, Sugawa M, Peiris R, Minamizawa K. TeleSight: Enabling asymmetric collaboration in VR between HMD user and non-HMD users. In: Proc. ACM SIGGRAPH emerging technologies. 2019, p. 1–2.

[14] Huang X, Yin M, Xia Z, Xiao R. VirtualNexus: Enhancing 360-degree video AR/VR collaboration with environment cutouts and virtual replicas. In: Proc. ACM symposium on user interface software and technology. 2024, p. 1–12.

[15] Suma EA, Krum DM, Bolas M. Sharing space in mixed and virtual reality environments using a low-cost depth sensor. In: Proc. IEEE international symposium on VR innovation. 2011, p. 349–50.

[16] Tserenchimed T, Kim H. Viewpoint-sharing method with reduced motion sickness in object-based VR/AR collaborative virtual environment. Virtual Real 2024;28(122):1–12.

[17] Zaman F, Anslow C, Chalmers A, Rhee T. MRMAC: Mixed reality multi-user asymmetric collaboration. In: Proc. IEEE international symposium on mixed and augmented reality. 2023, p. 591–600.

[18] Brument H, De Pace F, Podkosova I. Does mixed reality influence joint action? Impact of the mixed reality setup on users' behavior and spatial interaction. J Multimodal User Interfaces 2025;19:119–38.

[19] An J, Lee JH, Park S, Ihm I. Integrating heterogeneous VR systems into physical space for collaborative extended reality. IEEE Access 2024;12:9848–59.

[20] Pabst M, Rudolph L, Brasch N, Biener V, Eghtebas C, Eck U, Schmalstieg D, Klinker G. MRUnion: Asymmetric task-aware 3D mutual scene generation of dissimilar spaces for mixed reality telepresence. IEEE Trans Vis Comput Graph 2025;31(5):3354–64.

[21] Drey T, Albus P, der Kinderen S, Milo M, Segschneider T, Chanzab L, Rietzler M, Seufert T, Rukzio E. Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In: Proc. ACM conference on human factors in computing systems. 2022, p. 1–19.

[22] Jansen P, Fischbach F, Gugenheimer J, Stemasov E, Frommel J, Rukzio E. ShARe: Enabling co-located asymmetric multi-user interaction for augmented reality head-mounted displays. In: Proc. ACM symposium on user interface software and technology. 2020, p. 459–71.

[23] Schott E, Makled EB, Zoeppig TJ, Muehlhaus S, Weidner F, Broll W, Froehlich B. UniteXR: Joint exploration of a real-world museum and its digital twin. In: Proc. ACM symposium on virtual reality software and technology. 2023, p. 1–10.

[24] von Willich J, Funk M, Müller F, Marky K, Riemann J, Mühlhäuser M. You invaded my tracking space! using augmented virtuality for spotting passersby in room-scale virtual reality. In: Proc. ACM designing interactive systems conference. 2019, p. 487–96.

[25] Yang K-T, Wang C-H, Chan L. ShareSpace: Facilitating shared use of the physical space by both VR head-mounted display and external users. In: Proc. ACM symposium on user interface software and technology. 2018, p. 499–509.

[26] Auda J, Busse L, Pfeuffer K, Gruenefeld U, Rivu R, Alt F, Schneegass S. I'm in control! transferring object ownership between remote users with haptic props in virtual reality. In: Proc. ACM symposium on spatial user interaction. 2021, p. 1–10.

[27] Chang E, Lee Y, Billinghurst M, Yoo B. Efficient VR-AR communication method using virtual replicas in XR remote collaboration. Int J Hum-Comput Stud 2024;190:103304.

[28] Friedl-Knirsch J, Stach C, Pointecker F, Anthes C, Roth D. A study on collaborative visual data analysis in augmented reality with asymmetric display types. IEEE Trans Vis Comput Graph 2024;30(5):2633–43.

[29] Grandi JG, Debarba HG, Maciel A. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In: Proc. IEEE virtual reality and 3D user interfaces. 2019, p. 127–35.

[30] Cho Y, Kang J, Jeon J, Park J, Kim M, Kim J. X-person asymmetric interaction in virtual and augmented realities. Comput Animat Virtual Worlds 2021;32(5):e1985.

[31] Cho Y, Park M, Kim J. XAVE: Cross-platform based asymmetric virtual environment for immersive content. IEEE Access 2023;11:71890–904.

[32] Lee Y, Yoo B. XR collaboration beyond virtual reality: Work in the real world. J Comput Des Eng 2021;8(2):756–72.

[33] Lee Y, Yoo B, Lee S-H. Sharing ambient objects using real-time point cloud streaming in web-based XR remote collaboration. In: Proc. ACM international conference on 3D web technology. 2021, p. 1–9.

[34] Numan N, Brostow G, Park S, Julier S, Steed A, Van Brummelen J. CocreatAR: Enhancing authoring of outdoor augmented reality experiences through asymmetric collaboration. In: Proc. ACM conference on human factors in computing systems. 2025, p. 1–22.

[35] Wang P, Yang H, Billinghurst M, Zhao S, Wang Y, Liu Z, Zhang Y. A survey on XR remote collaboration in industry. J Manuf Syst 2025;81:49–74.

[36] Gutwin C, Penner R, Schneider K. Group awareness in distributed software development. In: Proc. ACM conference on computer supported cooperative work. 2004, p. 72–81.

[37] Sebanz N, Bekkering H, Knoblich G. Joint action: Bodies and minds moving together. Trends Cogn Sci 2006;10(2):70–6.

[38] Ouverson KM, Gilbert SB. A composite framework of co-located asymmetric virtual reality. Proc the ACM Human-Computer Interact 2021;5:1–20.

[39] Ishii H, Ullmer B. Tangible bits: Towards seamless interfaces between people, bits and atoms. In: Proc. ACM conference on human factors in computing systems. 1997, p. 234–41.

[40] Kaimoto H, Monteiro K, Faridan M, Li J, Farajian S, Kakehi Y, Nakagaki K, Suzuki R. Sketched reality: Sketching bi-directional interactions between virtual and physical worlds with AR and actuated tangible UI. In: Proc. ACM symposium on user interface software and technology. 2022, p. 1–12.

[41] Choi KY, Jung T, Harasha NC, Ishii H. Picto: Crafting remote tangible gestures via recordable, replayable, and shareable motions. In: Proc. ACM international conference on tangible, embedded, and embodied interaction. 2025, p. 1–15.

[42] Zuckerman O, Press VS, Barda E, Megidish B, Erel H. Tangible collaboration: A human-centered approach for sharing control with an actuated-interface. In: Proc. ACM conference on human factors in computing systems. 2022, p. 1–13.

[43] Radu I, Joy T, Schneider B. Virtual makerspaces: Merging AR/VR/MR to enable remote collaborations in physical maker activities. In: Proc. ACM conference on human factors in computing systems – extended abstracts. 2021, p. 1–5.

[44] Gilbert D, Bose A, Kuhlen TW, Weissker T. PASCAL – a collaboration technique between non-collocated avatars in large collaborative virtual environments. IEEE Trans Vis Comput Graph 2025;31(5):3525–35.

[45] Wieland J, Zagermann J, Müller J, Reiterer H. Separation, composition, or hybrid? – comparing collaborative 3D object manipulation techniques for handheld augmented reality. In: Proc. IEEE international symposium on mixed and augmented reality. 2021, p. 403–12.

[46] Jarrassé N, Charalambous T, Burdet E. A framework to describe, analyze and generate interactive motor behaviors. PloS One 2012;7(11):e49945.

[47] Peter M, Horst R, Dörner R. VR-guide: A specific user role for asymmetric virtual reality setups in distributed virtual reality applications. In: Proc. workshop virtual and augmented reality of the GI group VR/AR. 2018, p. 83–94.

[48] Mace M, Kinany N, Rinne P, Rayner A, Bentley P, Burdet E. Balancing the playing field: Collaborative gaming for physical training. J Neuroeng Rehabil 2017;14(1):116.

[49] Jung E, Lee Y. Virtualizing a collaboration task as an interactable environment and installing it on real world. 2025, http://dx.doi.org/10.48550/arXiv.2504.03375, arXiv:2504.03375.

[50] Schein KE, Rauschnabel PA, Praxmarer-Carus S, Babin BJ. Unpacking augmentation quality and local presence: Factors that drive effective augmented reality marketing. J Acad Mark Sci 2026;54:49–69.

[51] Liu T, Saito G, Lin C, Saito H. Inter-brain network underlying turn-based cooperation and competition: A hyperscanning study using near-infrared spectroscopy. Sci Rep 2017;7:8684.

[52] Järvholm K, Ohlsson A, Bernhardsson K, Gustafsson AW, Johnsson P, Malmström M, Wistrand J, Forslid T. Shared reading as an intervention to improve health and well-being in adults: A scoping review. Front Psychol 2025;16:1484839.

[53] Berube L, Priego E, Wisdom S, Cooke I, Makri S. "Moving with the story": The haptics of reader experience and response to digital comics. New Rev Hypermedia Multimed 2024;30(1–2):94–113.

[54] Harwood J, Wallace SD. Shared musical activity and perceptions of relationship commitment. Psychol Music 2022;50(6):1760–78.

[55] Savage PE, Loui P, Tarr B, Schachner A, Glowacki L, Mithen S, Fitch WT. Music as a coevolved system for social bonding. Behav Brain Sci 2021;44:e59.

[56] Wang P, Bai X, Billinghurst M, Zhang S, Zhang X, Wang S, He W, Yan Y, Ji H. AR/MR remote collaboration on physical tasks: A review. Robot Comput-Integr Manuf 2021;72:102071.

[57] Makransky G, Lilleholt L, Aaby A. Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. Comput Hum Behav 2017;72:276–85.

[58] Lee KM. Presence, explicated. Commun Theory 2004;14(1):27–50.

[59] Godden E, Groenewegen J, Wheeler M, Pan MK. Social-physical interactions with virtual characters: Evaluating the impact of physicality through encountered-type haptics. 2025, http://dx.doi.org/10.48550/arXiv.2511.05683, arXiv:2511.05683.

[60] Volkmann T, Wessel D, Franke T, Jochems N. Testing the social presence aspect of the multimodal presence scale in a virtual reality game. In: Proc. mensch und computer 2019. 2019, p. 433–7.

[61] Hart SG. NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting. vol. 50, (9):2006, p. 904–8.

[62] Allen K, Brändle F, Botvinick M, Fan JE, Gershman SJ, Gopnik A, Griffiths TL, Hartshorne JK, Hauser TU, Ho MK, et al. Using games to understand the mind. Nat Hum Behav 2024;8(6):1035–43.

[63] Grant DM, White EJ. Influence of anxiety on cognitive control processes. In: Worrell FC, editor. Oxford research encyclopedia of psychology. Oxford University Press; 2016, p. 1–22.

[64] Shen Z, Zhang L, Xiao X, Li R, Liang R. Icon familiarity affects the performance of complex cognitive tasks. I-Perception 2020;11(2):2041669520910167.

[65] Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77–101.

[66] Podkosova I, De Pace F, Brument H. Joint action in collaborative mixed reality: Effects of immersion type and physical location. In: Proc. ACM symposium on spatial user interaction. 2023, p. 1–12.
[67] Piumsomboon T, Dey A, Ens B, Lee G, Billinghurst M. The effects of sharing awareness cues in collaborative mixed reality. Front Robot AI 2019;6:5.
[68] Assaf R, Mendes D, Rodrigues R. Cues to fast-forward collaboration: A survey of workspace awareness and visual cues in XR collaborative systems. Comput Graph Forum 2024;43(2):e15066.
[69] Hindmarsh J, Fraser M, Heath C, Benford S, Greenhalgh C. Object-focused interaction in collaborative virtual environments. ACM Trans Computer-Human Interact 2000;7(4):477–509.
[70] Gutwin C, Greenberg S. A descriptive framework of workspace awareness for real-time groupware. Comput Support Coop Work (CSCW) 2002;11(3):411–46.
[71] Bimberg P, Zielasko D, Weyers B, Froehlich B, Weissker T. Come look at this: Supporting fluent transitions between tightly and loosely coupled collaboration in social virtual reality. IEEE Trans Vis Comput Graphics 2024;31:4833–49.
[72] Mayer A, Combe T, Chardonnet J-R, Ovtcharova J. Asynchronous manual work in mixed reality remote collaboration. In: Proc. international conference on extended reality. 2022, p. 17–33.
[73] Van Dijk W, Baltrusch SJ, Dessers E, De Looze MP. The effect of human autonomy and robot work pace on perceived workload in human-robot collaborative assembly work. Front Robot AI 2023;10:1244656.
[74] Li Z, Buruk OO, Bujić M, Kerouš B, Hamari J. Natural manipulation of virtual reality: The effects of kinesthetic gloves with force feedback on user performance and experience for accurate 3D object manipulation. Int J Hum-Comput Stud 2025;199:103507.
[75] Pan Y, Sinclair D, Mitchell K. Empowerment and embodiment for collaborative mixed reality systems. Comput Animat Virtual Worlds 2018;29(3–4):e1838.
[76] Villanueva A, Zhu Z, Liu Z, Wang F, Chidambaram S, Ramani K. ColabAR: A toolkit for remote collaboration in tangible augmented reality laboratories. Proc the ACM Human-Computer Interact 2022;6:1–22.
[77] Bozgeyikli E, Bozgeyikli LL. Evaluating object manipulation interaction techniques in mixed reality: Tangible user interfaces and gesture. In: Proc. IEEE conference on virtual reality and 3D user interfaces. 2021, p. 778–87.
[78] Yan Q, Adiwangsa M, Ozdowska A. Beyond augmented reality, towards augmented physicality: Gaming with tangible embodied interactions. In: Proc. 36th australasian conference on human-computer interaction. 2024, p. 31–45.
[79] Steffen JH, Gaskin JE, Meservy TO, Jenkins JL, Wolman I. Framework of affordances for virtual reality and augmented reality. J Manage Inf Syst 2019;36(3):683–729.