A Qualitative Comparison Between Augmented and Virtual Reality Collaboration with Handheld Devices

Jens Müller, Johannes Zagermann, Jonathan Wieland, Ulrike Pfeil, and Harald Reiterer
HCI Group, University of Konstanz
{jens.mueller,johannes.zagermann,jonathan.wieland,ulrike.pfeil,harald.reiterer}@uni-konstanz.de

ABSTRACT
Handheld Augmented Reality (AR) displays offer a see-through option to create the illusion of virtual objects being integrated into the viewer's physical environment. Some AR display technologies also allow for the deactivation of the see-through option, turning AR tablets into Virtual Reality (VR) devices that integrate the virtual objects into an exclusively virtual environment. Both display configurations are typically available on handheld devices, raising the question of their influence on users’ experience during collaborative activities. In two experiments, we studied how the different display configurations influence user experience, workload, and team performance of co-located and distributed collaborators during a spatial referencing task. A mixed-methods approach revealed that participants’ opinions were polarized towards the two display configurations, regardless of the spatial distribution of collaboration. Based on our findings, we identify critical aspects to be addressed in future research to better understand and support co-located and distributed collaboration using AR and VR displays.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Figure 1: Co-located and distributed collaborators with different configurations of handheld Augmented Reality (AR) displays.

CCS CONCEPTS
- Human-centered computing → Human computer interaction (HCI).

KEYWORDS
co-located collaboration, distributed collaboration, augmented reality, virtual reality, handheld devices

ACM Reference Format:

1 INTRODUCTION
Augmented Reality (AR) applications aim at providing the user a positive experience through the seamless combination of physical and virtual worlds. AR applications have proven beneficial in many domains such as education [7, 15], medical [1], and computer-aided instruction [17]. In addition, research has identified several domains in which collaboration can benefit from AR technology [6, 21, 23, 30, 46, 51, 52]. The overall goal of collaborative AR is "to augment the face-to-face collaborative experience, or to enable remote people to feel that they are virtually co-located" [33]. Collaborative activities involve basic manipulation tasks, such as the selection and positioning of objects [9]. The coordination of these tasks may involve spatial referencing, which can be defined as "the communication and confirmation of an object’s location" [40]. Successful spatial referencing, however, requires...
the collaborators to generate effective spatial references [12]. Co-located collaborators can benefit from a seamless combination of virtual objects and the physical environment, because their environment offers them effective reference points. For distributed collaborators, however, the seamless combination of real and virtual worlds can become problematic, because their environments are individual and therefore do not provide effective physical references to coordinate their actions. Some AR display technologies allow for the deactivation of the see-through option and for the integration of virtual objects into an exclusively virtual environment. Thereby, the physical environment becomes less significant and the virtual objects are integrated into a shared, virtual environment. Both display configurations are typically available on handheld AR devices, raising the question of their influence on users’ experience during collaborative activities. More specifically, we address the following study questions:

- **Q1**: Do different display configurations influence collaborators’ experiences?
- **Q2**: Do different display configurations influence collaborators’ workloads?
- **Q3**: Do different display configurations influence team performance?

## 2 RELATED WORK AND HYPOTHESES

Our work studies how collaborators experience different display configurations of AR tablets, where the display configurations vary in the degree of shared visual information and artificiality. We refer to three areas of related work: examples of collaborative AR, AR as shared space technology, and different AR display configurations.

### Collaborative AR

The goal of collaborative AR is “to augment the face-to-face collaborative experience, or to enable remote people to feel that they are virtually co-located” [33]. That reflects the qualities of collaboration in relation to the spatial condition of collaboration: “Face-to-face conversation provides a richness of interaction seemingly unmatched by any other means of communication. It is also apparent that living and working near others, whether that be in the same house, adjacent offices, or the same city, affords certain opportunities for interaction that are unavailable to those not co-located” [24]. Several examples have demonstrated the value of AR technologies to support both co-located and distributed collaboration (for a detailed discussion see Lukosch et al. [33]). The TransVision system [47], which allows for collaborative design using handheld AR displays, is an early example of AR-based co-located collaboration. Using the Studierstube [51] system, co-located users can view virtual 3D visualizations using AR glasses. A follow-up study by Reitmayr et al. [45] investigated the possibilities of collaborative workspaces using mobile AR devices. Butz et al. [10] introduced EMMI, a collaborative AR system that allows co-located users to interact with shared virtual information using 3D widgets and physical objects. The VITA system [3] represents a multi-modal AR environment for archaeologists to explore offsite visualizations of an archaeological dig. The MagicBook [7] represents a transitional interface that allows readers of a physical book to vary the degree of immersion during the reading experience in three steps: Physical (the physical book is visible), AR (virtual objects appear on the book pages), and VR (users can immerse themselves into the virtual world of the current page). The Construct3D system [27], a continuation of the Studierstube system, represents a tool to construct three-dimensional geometries. It uses a head-mounted display (HMD) and a two-handed 3D interaction tool for personal interaction with virtual objects. More recent research by Benko et al. [4] introduced a collaborative, projection-based AR system that creates the illusion of virtual objects using cameras and projectors, without requiring additional AR displays to view the virtual objects.

The value of AR technology to support distributed collaboration has been demonstrated in several systems. For example, the WearCom system [6] enables distributed conferencing where users can see remote collaborators as virtual avatars being integrated into their physical environment. Höllerer et al. [25] presented a collaborative, mixed-space system, where a remote user can interact with an outdoor user wearing a HMD by placing spatially registered information into the collaborator’s visual environment. A similar system has been presented by Stafford et al. [50]. Minatani et al. [36] developed a system that allows distributed collaborators face-to-face interaction by tracking their upper body and visualizing it into the other collaborator’s environment. Room2Room [41] is a telepresence system which establishes life-size, co-present interaction between two remote participants using projection mapping. Poleman et al. [43] introduced an AR system to support collaborative investigation of crime scenes. Here, an on-site investigator wears an HMD. A remote expert can then access the on-site environment and guide his colleague. Similarly, the system presented by Datcu et al. [13] supports exchange of context-related information between distributed collaborators in the security domain.

### AR as Shared Space Technology

AR applications can be classified depending on the proportions of real-world components and virtual components that are presented to the viewer’s display. Thus, AR applications can be located on a “virtuality continuum” [35], which ranges from the real environment to a virtual environment. Similarly, Benford et al. [2] use the term “artificiality” to express “the ratio of physical to synthetic information” that is presented in a shared space.
They also relate artificiality to the concept of transportation, which "concerns the degree to which users are transported into some new space or remain in their local space" [2]. Because AR displays integrate synthetic information into the viewers physical environment, AR users remain in their local space. In contrast, Virtual Reality technologies provide users an exclusively virtual environment, giving the impression of leaving behind "their local space and entering into some new remote space in order to meet with others" [2]. For distributed collaborators a virtual environment may therefore be more beneficial, because it can be shared with collaborators and thereby increase the sense of being co-located. In addition, collaborators can benefit from "key spatial properties such as containment, topology, movement, and a shared frame of reference" [2], because they help in establishing a "reciprocity of perspective" [2]. With a shared frame of reference, for example, collaborators may "reason about mutual orientation and gaze direction and may be able to spatially reference shared objects (e.g., by pointing or by using spatial language such as 'over there' [2]). Co-located collaborators using AR displays as shared-space technology can benefit from a high degree of spatiality, because their physical environments provide them with spatial properties such as a shared frame of reference (e.g., the surrounding walls, the floor, and the ceiling) or reference points (e.g., landmark objects in the environment).

The benefits of providing key spatial properties and shared visual information have been well-established [16, 18, 19]. For example, Fussel et al. [16] studied dyads performing a repair task, who were either co-located or distributed. Distributed workers were either connected via audio-video or audio only. Results demonstrated the value of a shared visual workspace and showed that co-located collaborators were more efficient than distributed collaborators due to shared spatial properties. Gergle et al. [18] investigated the usefulness of shared visual information in relation to the duration of visual updates. They conclude that delaying visual updates reduces the benefits of shared visual information and decreases performance. Also, they suggest that shared visual information is more beneficial when the task is visually complex. Gauglitz et al. [17] studied whether spatial annotations support spatial referencing during remote assistance tasks. Using their system, the on-site collaborator uses an AR tablet, and the remote user uses a desktop computer which allows for the placement of virtual annotations in the on-site collaborator’s physical environment. Their study showed that the use of spatial annotations benefits performance. The benefits of virtual landmark objects for spatial tasks were shown for co-located [37] and distributed [38] AR collaboration.

Different AR Display Configurations. Several studies investigated the effect of different configurations of AR displays on collaboration. Kiyokawa et al. [28] studied how display technologies influence communication behaviors of co-located users using AR displays. Devices included optical, stereo- and mono-video, and immersive HMDs. Here, the optical see-through device required the least communication and was the preferred display configuration. Chastine et al. [12] studied co-located and distributed collaborators’ “inter-referential awareness,” which they define as “the ability for one participant to refer to a set of objects, and for that reference to be understood” [p. 207]. In an experiment, dyads had to perform a construction task collaboratively. Study conditions differed regarding the spatial dispersion of collaboration (co-located vs. distributed), whether the model to be built was virtual or physical, and the display technology used (none vs. AR vs. VR display). They conclude that collaborators generate different references depending on the spatial dispersion of collaboration suggesting that the environment should provide sufficient reference points. Krichenbauer et al. [31] conducted an experimental comparison between AR and VR configurations in a single-user object manipulation task. They found performance benefits in favor of the AR configuration with no differences in the participants’ reported levels of comfort. Similarly, Boud et al. [8] found evidence of reduced task completion of AR over VR in an assembly task. Research by McGill et al. [34] addressed the issue of HMDs isolating the viewer from their real-world context during VR experiences. They suggest adding real-world elements into the viewer’s VR to reduce this isolation. Gugenheimer et al. [22] addressed the isolation issue from the perspective of bystanders (typically excluded from the viewer’s experience). They propose extending elements of the VR into their physical environment, to enable bystanders interacting and becoming part in the VR viewer’s experience.

Hypotheses

During co-located collaboration, the integration of virtual content into the viewers’ physical environment supports a natural way to interact with virtual content [5]. In addition, co-located collaborators can benefit from their physical environment because it provides a high degree of key spatial properties and effective references to facilitate the coordination of joint actions. As found in a prior study [28], we therefore hypothesize that

\[ H_1 \] Co-located collaborators prefer the AR display configuration.

As opposed to the co-located condition, distributed collaborators do not benefit from the spatial properties of their physical environments for the purpose of coordinating their joint actions, because it does not provide effective spatial properties, such as effective reference points and frames of reference. Therefore, distributed collaborators presumably
engage more with the virtual environment than with their physical environment. We therefore hypothesize, that

$H_2$ Distributed collaborators prefer the VR display configuration.

During co-located collaboration, the key spatial properties on the collaborators’ AR displays are the same as the ones they see off-screen, collaborators do not need to distinguish between them during the task (e.g., when exchanging spatial information), resulting in lower cognitive effort:

$H_3$ Co-located collaborators’ workload is lower with the AR display configuration.

As opposed to the co-located condition, distributed collaborators do not benefit from the spatial properties of their physical environments but have to actively distinguish the virtual (and effective) references from the physical (and ineffective) references, resulting in a higher workload:

$H_4$ Distributed collaborators’ workload is lower with the VR display configuration.

Considering co-located collaboration, and assuming $H_3$, collaborators can process and exchange spatial information more effectively using the AR configuration, resulting in an improved team performance:

$H_5$ Co-located collaborators’ team performance is higher with the AR display configuration.

As opposed to the co-located condition, and assuming $H_4$, collaborators can process and exchange spatial information more effectively using the VR configuration, resulting in an improved team performance:

$H_6$ Distributed collaborators’ team performance is higher with the VR display configuration.

3 STUDY DESIGN

We conducted two lab experiments that differed in the spatial dispersion of collaboration (co-located / distributed). In each experiment, the display configuration represented the within-group factor, with AR and VR being the conditions (Fig. 2). The factor display configuration was counterbalanced.

Study Environments and Task

Two rooms of our research facility served as distributed workspaces for our participants. The rooms differed in size. In experiment 1 (co-located), collaborators used the smaller room. Within each room, a walkable 3D volume of 3.5 m $\times$ 2 m $\times$ 2 m was allotted as shared work space. Regardless of the study condition, one virtual landmark object was placed in each lower corner of the 3D volume (a floor plant, an armchair, a shelf, and a vending machine) and a virtual ceiling lamp in the top center of the shared workspace. In each room, user interactions were audio/video-recorded. For handheld AR displays, each user was provided with an AR tablet with a mass of 370 g and a screen resolution of 1920 $\times$ 1200 pixels on a 7.02” display with 323 ppi [14]. A virtual viewing frustum visualizing collaborator’s current field of view (position and rotation) was shown as a trade-off decision between providing awareness cues and avoiding additional complexity (e.g., showing eye-gaze directions [29, 42]). In experiment 2 (distributed), participants were able to communicate via Teamspeak 3 [20], which was installed on the AR tablets.

We adapted a spatial navigation task [53] as follows: We distributed 20 white, virtual cubes with an edge length of 20 cm within a walkable 3D volume in the collaborators’ environment. Virtual cubes represented the task objects and the 3D volume the shared workspace. We used a question mark on each surface as default texture, symbolizing the “covered” state (Fig. 3). Each task object could be “uncovered” by centering its 3D representation on the tablet display and touching a button. When a task object was uncovered, its texture was replaced by a symbol. We used symbols from the Wingdings and Webdings fonts. The 20 task objects consisted of 10 unique pairs with the same symbol. In each turn, collaborators were allowed to uncover one task object. If the two uncovered task objects formed a pair, they were removed from the shared workspace, otherwise they were automatically covered after 4 seconds. Participants were instructed...
to find all matches as quickly and with as few attempts as possible. We pointed out that they should coordinate their actions well and try to memorize the symbol of each cube.

**Study Procedure**

Participants were welcomed in the corridor near the two study environments. In experiment 1 (co-located), participants were brought to the same room, in experiment 2 (distributed), each participant was brought to a separate room. Once the participants had arrived in their study room(s), they received a welcome letter, an informed consent form, and a demographic questionnaire. Afterwards, participants were provided with a printed description of the task. Participants were then asked to perform a demo task to familiarize themselves with the tablets and tasks. In the demo task, we reduced the number of cubes to 2 pairs, used dedicated test coordinates, and as symbols we used the numbers “1” and “2” to keep this phase short and to minimize interference with the following study task. Participants performed the demo task under the same study condition as in the following, first study iteration. We encouraged participants to familiarize themselves with the system until they felt comfortable with the procedure. Collaborators then started the study task (approximately 10 minutes) in their assigned study condition. Then, participants were asked to answer workload- and experience-related questionnaires. The task and assessments were repeated in the respective other study condition using different coordinates and symbols for the task objects to avoid carry-over effects. After that, a concluding, semi-structured group interview was conducted. In experiment 2 (distributed), one participant was brought to the room of the other for the group interview. Each session took 40 minutes and participants were compensated for their time.

**Dependent Variables and Analysis**

**User experience** was operationalized by measuring participants’ perceived *social presence*, based on the sub-dimensions “social presence – actor within medium” and “spatial presence” of the Temple Presence Inventory questionnaire [32]. To deepen our understanding of users’ experiences and better discuss the questionnaire responses, we conducted a concluding, semi-structured interview. When asked if the participant would like to perform the task again, we inquired participants’ preferred display configuration and the reasons for their choice. In our interviews, we also elicited experienced benefits and disadvantages for each display configuration both in general and in relation to the task. Data from the interviews was analyzed for similar statements in the participants’ responses. Responses for the two experiments (co-located and distributed) were analyzed separately according to the following affinity diagramming approach: First, similar responses were clustered for each participant individually (intra-participant cluster), then these clusters were labeled with a summarizing statement. Second, resulting clusters were analyzed for similar labels and clustered again across all participants in the respective experiment (inter-participant cluster). For each of the resulting final clusters, the number of containing sub-clusters (participants) was determined.

**Workload** was measured using the NASA TLX [39] and **team performance** was measured in task completion time. Implicit **spatial memory** was measured in the number of trials it took the collaborators to find all matches.

We report findings for the two experiments separately. 16 different participants (sample size according to [11]) took part in each experiment. To indicate the display configuration we use subscript $A_i$ for the condition using the AR display configuration and subscript $V_i$ for the condition using the VR display configuration. To analyze for statistical significance, we used the non-parametric Wilcoxon signed-rank test and refer to the median ($Mdn$) if the normal distribution was not given and for ordinal data (including interquartile range, $IQR$). Otherwise, Student’s $t$-test was used and mean values ($M$) are reported (including standard deviation, $SD$). Statistically significant differences assumed an alpha < .05.

4 RESULTS OF EXPERIMENT 1 (CO-LOCATED)

16 participants (9 female, 7 male), between 23 to 41 years ($M = 28.31$, $SD = 5.19$), formed 8 dyads. Members of 6 dyads reported knowing each other. 10 participants were students, 5 research assistants, and 1 was an employee. 6 indicated prior use of tablets with an average usage frequency of 2.17, on a scale from 1 (very infrequently) to 5 (very frequently). 3 participants reported prior experiences with AR applications and 4 indicated prior experiences with VR applications.

**User Experience**

Analysis of **Social Presence** items did not reveal statistical differences between the study conditions. Analysis of **Spatial Presence** items showed that participants rated the dimension *comeToPlace* (“How much did it seem as if the objects and people you saw/heard had come to the place you were?”) higher in the AR condition (Fig. 4). In addition, when asked whether their “experience seemed more like looking at the events/people on a movie screen (+3) or more like looking at the events/people through a window (+1)” (screen–window, semantic differential), responses from the AR condition indicate a significantly higher tendency towards “Through a window” (*comeToPlace*: $Mdn_A = 0.50$, $IQR_A = 2.25$, $Mdn_V = -2.00$, $IQR_V = 2.00$, $Z = -2.55$, $p < .05$; screen–window: $Mdn_A = -0.50$, $IQR_A = 2.00$, $Mdn_V = -2.00$, $IQR_V = 2.00$, $Z = -2.25$, $p < .05$).
Findings related to user experience result from a qualitative analysis of interview data. 9 participants preferred the AR condition and 7 the VR condition. When asked about the positive experiences associated with the AR condition, participants mentioned general positive sensations (n=9) that this condition evoked, for example that “it was easier to solve the task” (#a), that it was “more enjoyable” (#a), “more pleasant” (#a), and “more appealing” (#3b). Also, participants enjoyed having more reference points (n=13) provided by the physical environment when working in the AR condition. These reference points were found to improve orientation (n=6) in the environment, e.g., “the mental split was smaller” (#3b), “you didn’t have to stare into the display all the time” (#2b), and “physical navigation was easier” (#2a). Physical reference points on the display also helped to better assign the task objects (n=5). Also, physical references were used to better coordinate actions (n=2). One participant stated: “When I saw the real objects, I sometimes found it easier to coordinate ourselves ... we actually always discussed with the real objects, it was cool when there were more objects in the room” (#7a). One participant (#3) valued the visibility of the collaborator stating it to be “pleasant” and “helpful” to see the other collaborator on the display.

Reporting negative experiences associated with the AR condition, participants mentioned they felt distracted (n=4) by the physical objects of their environment, and stated that “the many objects were irritating” (#4a), and “fewer objects might have been more beneficial” (#6a). Also, one participant “found it confusing to see [the collaborator] on [his] display” (#4b), and another one stated that “it seemed a bit odd, e.g., when [the collaborator] stood in that plant” (#6b).

When asked about the positive experiences associated with the VR condition, participants mentioned generally positive sensations (n=4) that this condition evoked, for example that it was “easier to solve the task” (#1a), that it looked nicer” (#2b) and was “more homogeneous” (#1b). One participant (#2a) mentioned that he "could become immersed more easily” when working in the VR condition. In addition, participants valued the tidiness and reduction of information (n=6) in the VR condition, claiming that this made it easier to focus on the important aspects. One participant said: “What was important were the spatial relations of the virtual objects and not the environment. We actually talked about virtual things only — plant, shelf, etc.” (#4a) Also, the reduction of objects in the VR condition “made it easier to spot the cubes, because [participants weren’t] distracted” (#3a). One participant stated that “fewer reference points means you focus more intensely on them” (#6a) and another added that “when there were fewer reference points, I could remember them well” (#6b). As negative experiences associated with the VR condition, participants mentioned they felt distracted (n=4) by the physical objects of their environment, and stated that “the many objects were irritating” (#4a), and “fewer objects might have been more beneficial” (#6a). Also, one participant “found it confusing to see [the collaborator] on [his] display” (#4b), and another one stated that “it seemed a bit odd, e.g., when [the collaborator] stood in that plant” (#6b).

When asked about the positive experiences associated with the VR condition, participants mentioned generally positive sensations (n=4) that this condition evoked, for example that it was “easier to solve the task” (#1a), that it looked nicer” (#2b) and was “more homogeneous” (#1b). One participant (#2a) mentioned that he "could become immersed more easily” when working in the VR condition. In addition, participants valued the tidiness and reduction of information (n=6) in the VR condition, claiming that this made it easier to focus on the important aspects. One participant said: “What was important were the spatial relations of the virtual objects and not the environment. We actually talked about virtual things only — plant, shelf, etc.” (#4a) Also, the reduction of objects in the VR condition “made it easier to spot the cubes, because [participants weren’t] distracted” (#3a). One participant stated that “fewer reference points means you focus more intensely on them” (#6a) and another added that “when there were fewer reference points, I could remember them well” (#6b). As negative experiences associated with the VR condition, participants mentioned they felt distracted (n=4) by the physical objects of their environment, and stated that “the many objects were irritating” (#4a), and “fewer objects might have been more beneficial” (#6a). Also, one participant “found it confusing to see [the collaborator] on [his] display” (#4b), and another one stated that “it seemed a bit odd, e.g., when [the collaborator] stood in that plant” (#6b).
condition, one participant mentioned difficulties in orientation due to a "lack of sufficient objects in the room" (#3b).

5 RESULTS OF EXPERIMENT 2 (DISTRIBUTED)

16 participants (7 female, 9 male), between 19 to 37 years of age ($M = 27.81, SD = 5.12$), formed 8 dyads. Members of 7 dyads reported knowing each other. 6 participants were students, 2 assistant lecturers, 7 research assistants, and 1 chief financial officer. 10 participants indicated prior use of tablet computers, with an average usage frequency of 3.11 on a scale from 1 (very infrequently) to 5 (very frequently). 5 participants reported prior experiences with AR applications and 6 indicated prior experiences with VR applications.

User Experience

Analysis of Social Presence items revealed that the rating of item seeing/hearing ("How often did you have the sensation that people you saw/heard could also see/hear you?") was significantly higher in the AR condition ($M_{AR} = 2.00$, $IQR_{AR} = 1.00$, $M_{VR} = 2.00$, $IQR_{VR} = 0.50$, $Z = -2.23$, $p < .05$; Fig. 6). An analysis of Spatial Presence items did not reveal statistical differences between the two study conditions.

Workload

On average, participants’ overall task load was 40.83 in the AR condition ($IQR_{AR} = 18.54$) and 39.66 in the VR condition ($IQR_{VR} = 11.66$). There were no significant differences between the two study conditions regarding task load.

Team Performance

Team performance was measured in task completion time and the number of trials. In the AR condition, it took the dyads 388.37 seconds ($SD_{AR} = 121.32$ s) to complete the task and in the VR condition 368.25 seconds ($SD_{VR} = 103.73$ s) without significant differences between the two conditions. In the AR condition, it took the dyads 18.25 trails ($SD_{AR} = 3.01$) to find all matches and 17 trails ($SD_{VR} = 2.07$) in the VR condition. Results did not differ significantly.

User Experience Data From Concluding Interview

Findings related to user experience result from a qualitative analysis of interview data. 8 participants preferred the AR condition and 8 preferred the VR condition. When asked about their positive experiences associated with the AR condition, participants mentioned generally positive sensations (n=3) that this condition evoked, for example that “it felt more immersive, more natural”(#10a) and “seemed more real”(#14a). Further, it was mentioned that “the experience of being inside was stronger” (#14b) and that working in the AR condition is “not a huge step from everyday life” (#10a). Also, participants valued the display of the physical environment as it supported a better orientation (n=8) than the VR condition, because “there were so many more elements I could refer to” (#9a) and participants “could use the environment [to] work a lot with relations” (#16a). One participant explains: “Sometimes you could refer to the real background. I searched for reference points, in my case I included objects from the real world” (#11a). Also, participants reported that the visibility of the physical environment on the tablet gave them the opportunity to avoid collisions (n=3) with physical objects, because “it was easier to see obstacles” (#16b). One participant explains: “I prefer seeing the real objects, so that I do not run into them, or collide with them” (#15b).

Reporting negative experiences associated with the AR condition, participants mentioned the problem of being in different physical environments (n=8), causing them to “wonder, whether [the collaborator] moves at a different scale, because [his] room seemed to be larger” (#14a). One participant described a situation when his collaborator referred to a door, he “didn’t know [the collaborator’s] room so I didn’t know that there was a door, but in my room there was no door” (#9a). They mentioned that seeing different physical environments on their tablets was confusing, stating that “it took me a minute to realize, that I couldn’t use [the physical objects] in our conversation.” (#9b) and “if [my collaborator] says this chair, I don’t see it” (#10b). One participant explains that “at first I thought I’d have more reference points, [...] but then I realized that they were of no use, because only I could see them” (#14b). Reporting how to overcome this obstacle, #12b explained that “it confused me because we saw different real things, so I totally shut up about the things that were real and focused on things that were not real” (#12b).

When asked about the positive experiences associated with the VR condition, participants mentioned, that the virtual environment provided them with helpful constraints (n=3). Participants stated that they “felt that the virtual room was smaller” (#11b) than the physical room, and thus they “knew that beyond [the virtual walls] there could be no more objects”(#9a). Some participants also mentioned a higher


degree of immersion and estimation possibilities (n=3) as positive aspects of the VR condition, reporting that "it seemed easier to estimate where things were [and] it was easier to estimate the depth of the virtual room" (#13a). Concerning a higher degree of immersion, participants reported to be "way more in that world [as] it really allowed me to dive in...I think we performed better" (#11a). Furthermore, participants valued the tidiness and reduction of information (n=3) in the VR condition, stating that "in this (virtual) room there was less distraction, less going on" (#14b). The "clear structure [was perceived] less distracting" (#15a) and "provided a better overview" (#16a). Also, participants mentioned benefits in relation to the unity of environments (n=2), stating that "[we] see the same things" (#12b) in the VR condition. In comparison to the AR condition, "it was easier if you’re in the same room" (#14a) and participants realized that conversations like "...next to my jacket...which jacket?...at the table...what table?" (#14a) did not occur in the VR condition.

As negative experiences of the VR condition, participants reported the problem of collision (n=3), described as "I ran into a chair, because I focused on the display" (#11a) and "I was afraid to run into something, because you don’t really see what’s in front of you" (#13a). Two participants mentioned problems in unifying the two environments (n=2), stating that in the VR condition they "had to think first where to go" (#16b). One participant explained that "it was not as if I was always entirely in a different world. But the difference was more extreme if I shifted my focus away from the display and saw the real environment" (#14a).

6 DISCUSSION
The discussion is structured following the research questions.

Q1 Do different display configurations influence collaborators’ experiences?

According to H1, co-located collaborators prefer the AR display configuration as the integration of virtual objects into their physical environment facilitates natural collaboration [28]. Being co-located, collaborators can use spatial references in their physical environment. In experiment 1, participants gave higher ratings regarding spatial presence for the AR condition in two dimensions, showing that the AR condition is more closely related to the sensation of spatial presence. Yet, participants indicated no clear preference of co-located collaborators towards a display configuration (AR: 9 / VR: 7), suggesting that a higher sensation of spatial presence does not necessarily correlate with user preferences.

Results from the concluding interview (Table 1) show that most participants associated the AR condition with overall positive attributes such as “easier”, “enjoyable”, “pleasant”, and “appealing”. Additionally, some valued the presence of physical reference points being displayed on the tablets. Others, however, stated that the physical objects were distracting (e.g., “confusing” or “irritating”). These participants perceived the VR condition as being more “homogeneous”, providing a better overview due to the reduced information and its tidiness, allowing them to better focus on the task at hand. These reports suggest that participants were able to engage with the display configurations differently: While some considered the task to happen in the physical world, appreciating the natural collaboration possibilities (e.g., face-to-face communication) including the physical environment as part of the overall experience, others interpreted the task as predominantly digital – focusing on virtual aspects only, seeking to engage with the virtual world to be able to better focus on the virtual task objects. Thus, referring to Q1, participants experienced the display configurations differently but not indicating a clear preference towards the AR condition. Instead, results of experiment 1 indicate that participants’ preferences were polarized towards either of the two display configurations. H1 therefore needs to be rejected.

According to H2, distributed collaborators prefer the VR condition due to more effective references. In experiment 2, participants gave higher rating regarding social presence for the AR condition in one dimension, showing that during distributed collaboration the AR condition was more closely related to the sensation of social presence. This raises the question of what leads to an increase of social presence? During distributed collaboration, collaborators in the VR condition could see a shared virtual environment on their tablets and an individual physical environment. However, with the AR condition, they could see their individual physical environments both, on and off-screen. This suggests that a higher social presence is influenced by seeing the individual physical environment on the tablet (AR condition) rather than having a shared yet virtual environment on the tablets.

Participants indicated no clear preference of distributed collaborators towards a display configuration (AR: 8 / VR: 8). Similar to experiment 1, a higher sensation of social presence...
does not necessarily correlate with user preferences. Results from the concluding interview for experiment 2 (Table 2) show that participants reported the AR condition to be more immersive, real, or raising the sensation of “being inside.” AR also facilitated a better orientation and prevented possible collisions with physical objects. However, participants raised concerns that were related to the physical environment and that this was the cause for confusion, given that the physical environments were individual. Contrasting to this, participants valued the VR condition as providing “unification of environments.” Further positive attributes include the tidiness of the environment and the provision of spatial constraints. The risk of collision and the unification of environments were seen as negative aspects. Two findings related to the “unity” of environments need further elaboration:

Participants in experiment 2 realized the problems that the individual physical environment caused and stated it as a negative aspect of the AR condition. The non-unification of the physical environment appeared to be problematic in the AR condition. Exactly this non-unification of the physical environment was reduced when participants could engage with the VR, as it provided them a shared, hence “uniform” environment. However, this advantage was only reserved for those who could engage with the virtual environment provided in the VR condition. For participants whose attention was on both, the tablet (virtual environment) and off-screen (physical environment) another unification problem occurred: Those who could not fully engage with the virtual environment but also paid attention to their physical environment, automatically tried to match the two environments.

These results show that the visibility of the physical environment is important for individuals, yet it may not be beneficial to coordinate collaborative efforts, leading to a trade-off between these aspects. The ability to engage with the virtual environment (presumably beneficial in a distributed setting) may highly depend on the proportion of provided virtual and physical information and thus correlate with the field of view. Thus, referring to Q1, participants experienced the display condition differently but not did not generally prefer the VR condition. Instead, results show that participants’ preferences were polarized to either of the two display configurations. H2 therefore needs to be rejected. For future research, the question arises whether and how presence-related measures influence users preferences. For that matter, other conceptualizations of “presence” and associated questionnaires, should be taken into consideration.

Q2 Do different display configurations influence collaborators’ workloads?

According to H3 in experiment 1 (co-located), the workload is lower for collaborators using the AR condition and according to H4 in experiment 2 (distributed), the workload is lower when collaborators use the VR condition. Workload results between the two display conditions, however, did not differ significantly. H3 and H4 thus need to be rejected. Given that participants experienced the same display configuration very differently and that some engaged more with the physical and others with the virtual world may explain that workload did not differ between the display configurations, yet it might be related to the task at hand. Future research is necessary to analyze design aspects of AR and VR configurations that influence the perceived workload and to include suitable real-time workload metrics (e.g., using eye-tracking technology [54]).

Q5 Do different display configurations influence team performance?

According to H5, co-located collaborators (experiment 1) perform better when using the AR condition, whereas H6 suggests that distributed collaborators (experiment 2) perform better when using the VR condition. Regardless of the spatial dispersion of collaboration, all performance-related measures of experiment 1 and 2 were in favor of the VR condition. In experiment 1 (co-located), it took dyads fewer trials to find matches. It could have been easier for collaborators to recall task objects’ positions and identify them more accurately. Also, it could have been easier to exchange spatial information due to less misinterpretations. These aspects might also be related as the second assumption (accurate communication) requires the first assumption (accurate knowledge of the position of task objects).

This argument is supported by qualitative data that shows that collaborators appreciated the tidiness and reduced amount of visual information of the virtual environment in both experiments. Future studies should investigate how much visual information is sufficient or can be disadvantageous.

Thus, referring to Q4, co-located collaborators did not benefit from the AR display configuration: Regarding the number of trials, team performance in experiment 1 was even lower in that condition. H5 thus needs to be rejected. Team performance of distributed collaborators in experiment 2 did not benefit from the VR condition as hypothesized. H6 therefore needs to be rejected, too. The fact that team performance related measures were in favor of the VR condition gives rise to questions of whether a larger sample size would have yielded a statistical difference in favor of VR.

7 IMPLICATIONS FOR FUTURE RESEARCH

With the goal of improving handheld, AR-based collaboration, we investigated how the display configuration (AR vs. VR) influences co-located and distributed collaborators’ experiences, workload, and team performance. Results show that:

- participants’ ratings related to social presence and spatial presence are in favor of the AR condition,
participants’ experiences and preferences were polarized towards the two display configurations,
participants’ workload did not differ significantly between the study conditions,
team performance only differed in the numbers of trials in favor of the VR condition (co-located).

These findings suggest that there is no generally preferred configuration of handheld displays due to participants’ polarized preferences. Rather than providing implications for designers, this leads to open questions for future research: As preferences differ, collaborators could potentially be provided with their individually preferred configuration. Thus, cases where their preferences differ from each other imply a “mixed-space collaboration” [21]. Such mixed-space settings may be beneficial for asymmetric collaboration such as remote assistance (e.g., during navigation instructions [21]), but for symmetric scenarios, mixed-space collaboration may not be suitable (e.g., when collaborators work co-located). Future studies should therefore investigate the question:

\textbf{Qf}_1 How suitable is mixed-space collaboration with handheld devices for symmetric collaboration?

Participants who preferred the AR configuration justified their choice by indicating that they had difficulties aligning the virtual room as provided by the VR configuration with their physical environment. However, participants who favored the VR configuration indicated that they felt distracted by the physical environment being displayed in the AR configuration. They indicated that it would keep them from concentrating on the virtual objects. Thus, for cases where task-related requirements do not determine the display configuration, a transitional interface, such as the MagicBook [7] could be offered. This would enable users to dynamically adjust the display configuration and “transition between different levels of reality” [49, p. 418] according to their needs and preferences. This raises the following question:

\textbf{Qf}_2 Are there benefits of transitional, handheld displays for co-located and distributed collaboration over static AR or VR configurations?

However, it should be considered that transitional interfaces may result in mixed-space collaboration (when collaborators use different configurations at the same time).

A symmetric collaboration task was chosen for the study. Results therefore primarily apply to symmetric settings. But collaboration occurs in diverse contexts involving different requirements. Thus, the choice of display configuration may be in favor of the specific nature of the task. In addition, the nature of the task may have influenced what people experience, what is presented on the display, and what they see when their gazes shift off-screen. Thus, the relation of tasks and how the task-related properties influence both individual experiences and team-based performance should be studied, addressing the following research question:

\textbf{Qf}_3 How do different types of tasks influence individual experiences and team-based performances?

Finally, a limitation of the study refers to the generalizability of the findings. The experiments focused on handheld devices as tools to support collaboration. They provide a special kind of AR and VR experience, as AR and VR content is limited to a very narrow field of view. This implies that the large amount of visual off-screen information, such as the physical environment and, in the co-located condition, the other collaborator, may have influenced participants’ preference (see e.g., [26, 44, 48]). Thus, polarized opinions towards the AR and VR configuration may not necessarily apply to displays that cover a larger amount of the visual field.

\section{8 CONCLUSION}

This study investigated the influence of display configuration (AR and VR) of handheld AR devices on co-located and distributed collaboration – focusing on user experience, workload and team performance. Findings show that participants’ experiences and preferences were polarized towards the two display configurations: Participants favoring the AR configuration valued this condition in terms of the physical environment and its numerous referencing options being displayed on their tablets and the integration of the task objects therein. In addition, these participants negatively mentioned that the VR configuration would cause them to mentally align its virtual environment with their physical environment. Participants favoring the VR in contrast, valued this condition in terms of its reduction of visual information and the ability to focus on the task and the associated objects. Performance benefits could be observed when participants were co-located and used the VR configuration. Participants’ subjective evaluation of workload, however, did not differ significantly between the two display conditions. These findings indicate that, regardless of the spatial distribution of collaboration, there is no distinct display configuration to be chosen in collaborative scenarios. Findings, however, not only give rise to research questions to be addressed to understand co-located and distributed collaboration with handheld AR devices; they also recommend using transitional interfaces where users “no longer require different systems for AR and VR, but could transition between different levels of reality at will.” [49, p. 418]

\section*{ACKNOWLEDGMENTS}

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projekt-ID 251654672 – TRR 161 (Project C01).
REFERENCES


2–14.


