Comparing Different Approaches for the Collaborative Manipulation of Virtual 3D Objects in Augmented Reality

Master Thesis

presented by
Jonathan Wieland

at the

Universität Konstanz

Human-Computer Interaction Group
Department of Computer and Information Science

1. Supervisor: Prof. Dr. Harald Reiterer
2. Supervisor: Dr. Johannes Fuchs

Konstanz, 2019

HUMAN-COMPUTER INTERACTION GROUP, UNIVERSITY OF KONSTANZ

HCI.UNI-KONSTANZ.DE

Thanks to my advisors Jens Müller and Johannes Zagermann for their great support.

Printed in November, 2019.

Abstract

Comparing Different Approaches for the Collaborative Manipulation of Virtual 3D Objects in Augmented Reality

Applications on hand-held *Augmented Reality* devices can create interactive experiences that enable users to see virtual objects seamlessly integrated into their physical environment. One use case for such applications is interior design, which requires users to place virtual furniture in their homes, relying on 3D object manipulation. Although furnishing a place is typically a collaborative task, IKEA only offers a single-user AR application. Besides, different studies show that a collaborative manipulation of virtual 3D objects can be more efficient and increase user experience. Literature analysis revealed three different approaches: 1) *Separation of DOFs* separates the manipulation of the degrees of freedom of an object between users, 2) *Composition of Users' Action* implements a merge policy to combine users' inputs to one composed manipulation, and 3) *Hybrid Approaches* combine both variants. However, based on related work, it is difficult to decide which of the three approaches one should use as they have not been compared for hand-held AR devices yet.

Consequently, this thesis addresses the question of which of the three approaches a hand-held AR application should implement that enables two users to furnish their homes collaboratively. Therefore, a study prototype was developed that implements all three approaches. In a controlled lab study with 48 participants, the prototype was then used to compare the three approaches in terms of *performance*, *workload*, and *user experience*. While the measures for performance and workload suggest using the implementation for *Separation of DOFs*, participants' preferences were polarized between the implementations for *Separation of DOFs* and the *Hybrid Approach*. Based on the results of the experimental comparison and their discussion, this thesis draws implications for future applications and highlights further research questions to be addressed by future work.

Contents

1

Introduction

Augmented Reality (AR) devices can create an interactive experience that enables users to see virtual objects seamlessly integrated with the real world. While already the application of AR for single-user applications is very promising, the technology's primary benefit is often seen in its potential use for building next-generation collaborative systems [1]. One domain for such systems is interior design and space planning. Microsoft [2] demonstrates this vividly with one of their concept videos that shows how they envision the future of Augmented Reality. The video is about a designer called Penny, who has to set up an empty stockroom as the new flagship shoe store for a group of clients. She uses an Augmented Reality headset to furnish the room with virtual objects (cf. Figure 1.1). Penny is short of time since the clients will arrive earlier as planned, and she has to present her design ideas to them. Luckily, she gets support from her colleagues. Because they also wear Augmented Reality headsets, they have not even to be at the future shoe store but join remotely. Together, they develop new ideas by arranging virtual shoe racks and other fitments seamlessly inside the place. Upon arrival, the client also puts on an Augmented Reality headset and inspects the design of Penny and her colleagues.

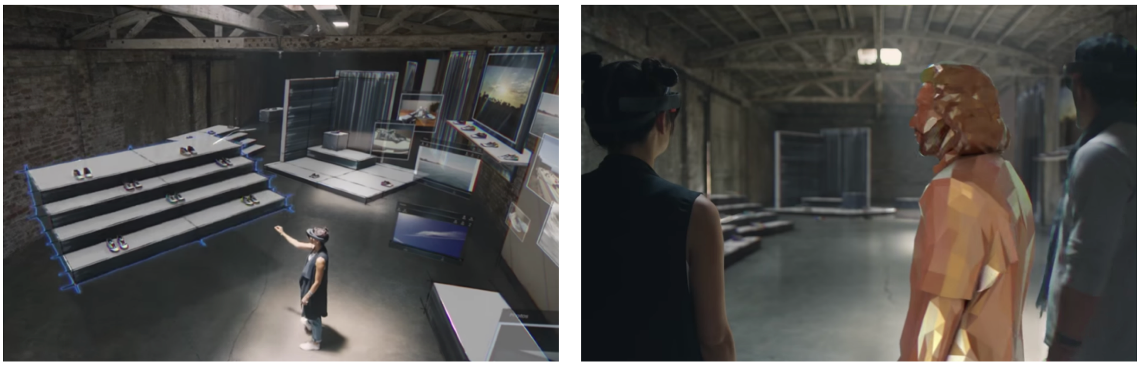


Figure 1.1: One of Microsoft's concept videos [2] shows how they envision the future of interior design using Augmented Reality (AR). Designers are wearing AR headsets that enable collaboration for arranging virtual objects inside a place. *Taken from [2].*

IKEA's Augmented Reality application "Place" [3] already realizes parts of Microsoft's vision. The app enables its users to arrange virtual, one-to-one scaled representations of IKEA's furniture in their homes. Instead of headsets, users only need their smartphones or tablets to run the app, which makes the Augmented Reality experience easily accessible. Figure 1.2 demonstrates the usage of IKEA's Place: Users select a piece of furniture from a digital catalog, move it through their homes, and finally place it. Although Microsoft imagines the future of interior design as collaborative, IKEA's Place is only a single-user application. Given that different studies suggest that a collaborative manipulation of virtual objects can be more efficient [4, 5], this is rather astonishing. Particularly because furnishing a place is a task that people often do collaboratively. A possible reason for IKEA deciding against a multi-user application could be the conflict that two or even more users provoke if they manipulate the same virtual object at the same time: How to combine actions that can have different directions and different magnitudes?

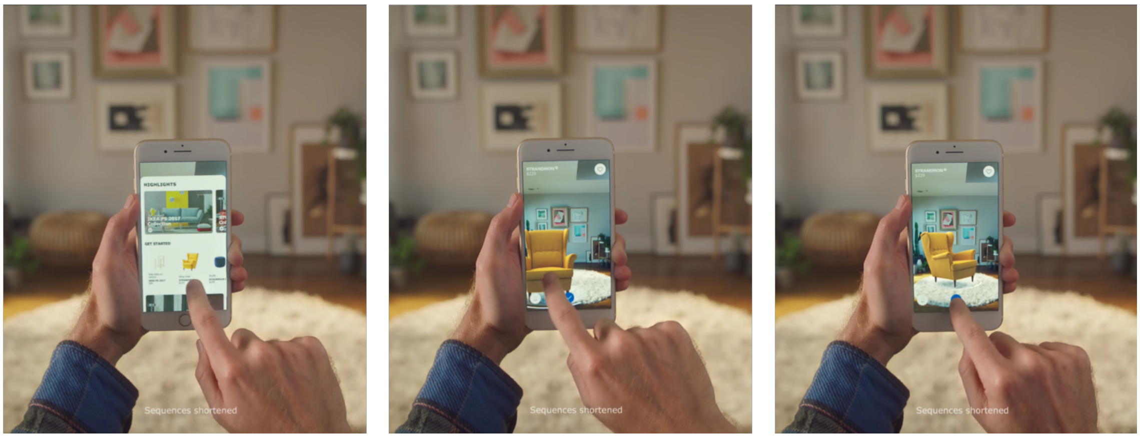


Figure 1.2: IKEA's AR application "Place" [3] enables users to furnish their homes with digital versions of the company's furniture using smartphones or tablets. Therefore, they select them from a digital catalog *(left)*, move them through their homes *(middle)*, and finally place them *(right)*. *Taken from [6].*

The literature describes three approaches to solve this conflict: The first one separates the manipulation of the degrees of freedom (DOFs) of an object between the users and is referred to as *Separation of DOFs* [4, 7]. The second one is called *Composition of Users' Actions*. It combines the inputs of the users to one resulting manipulation and applies it to the object [8, 9]. As a third, *Hybrid Approach*, combinations of both strategies are used [5, 10, 11]. However, the comparison of all three approaches within an Augmented Reality setting on mobile devices has not been widely studied yet. Therefore, this thesis prompts the following question:

> Which of the three approaches (*Separation of DOFs, Composition of Users' Actions, Hybrid Approach*) should a hand-held AR application implement to enable users to furnish a room with digital furniture collaboratively?

To address this question, three specific touch-based implementations of the three approaches were developed and integrated into a study prototype. Similar to IKEA's Place, the prototype enables users to furnish a place with digital pieces of furniture – with the difference that it supports a collaborative manipulation of digital objects by two users. This prototype then was used for an experimental comparison of the three approaches. The results of this comparison are reported, evaluated, and discussed in this thesis.

The thesis is structured as follows (cf. Figure 1.3): The next chapter summarizes the theoretical foundations needed for the view on related work given in Chapter 3. Based on foundations and related work, Chapter 4 derives requirements and introduces the developed interaction concept for the three approaches and their implementation in a study prototype. Chapter 5 describes the experimental comparison and reports associated results. The evaluation and discussion of these results is the subject of Chapter 6. Besides, it draws implications for future applications, discusses the limitations of the study, and provides an outlook on topics to be addressed by future work. The last chapter concludes the thesis with a summary of the most critical aspects.

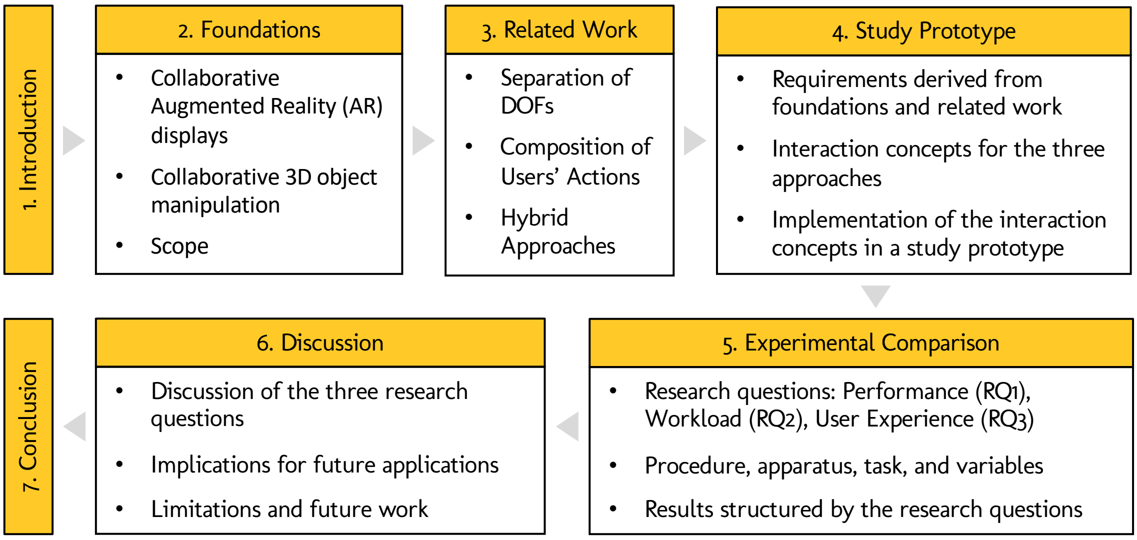


Figure 1.3: The structure of this work.

2 Foundations

This chapter summarizes the theoretical foundations needed for informing the related work, developing interaction concepts for the three approaches, and implementing them in a study prototype. The first two sections cover the definition of Augmented Reality (AR) and how collaboration in it can be classified. The focus of the third section is on 3D object manipulation with hand-held AR displays. It introduces different basic definitions and interaction techniques. The fourth section details the crucial factors for collaborative 3D object manipulation. The chapter ends with a conclusion that refines the scope of this work.

2.1 Augmented Reality (AR) Displays

Many different definitions of *Augmented Reality (AR)* exist. Azuma [12] proposed a widely accepted one that avoids "limiting AR to specific technologies" but still "retains [its] essential components" [12]. In his survey paper published in 1997, he defines AR by reference to three characteristics:

> An Augmented Reality (AR) system *"1) combines real and virtual,"* is *"2) interactive in real-time"* and *"3) registered in 3D"* [12].

In their survey paper form 2015, Billinghurst et al. [13] ascertain that these three characteristics imply the technical requirements of an AR system. At first, the system requires a display that is capable of combining the real with the virtual. Secondly, it needs the computational power to "generate interactive graphics that respond to user input in real-time" [13]. Finally, a registration in 3D presumes "a tracking system that can find the position of the user's viewport and enable the virtual image to appear fixed in the real world" [13]. Nowadays, even hand-held displays like smartphones or tablets meet these requirements. Since those devices have become everyday objects, they make AR easily accessible.

Milgram and Kishino [14] define AR "in the context of other technologies" [13] by ranging it in their often-cited *Virtuality Continuum (VC)*. As shown in Figure 2.1, they span this continuum from solely real environments to entirely virtual ones. Since *Virtual Reality (VR)* displays produce exclusively virtual environments, they have to be located at the rightmost extremum. Between both extrema lies the area of *Mixed Reality (MR)*, which labels devices that present "real world and virtual world objects ... together within a single display" [14]. It consists of two sub-classes between which the authors decide based on the proportion of virtual objects present in the scene: Towards the left extremum of real environments, one finds the class of AR, where the real world is augmented by virtual objects. The example in Figure 2.1 shows a real room to which a virtual armchair is added. Towards the right extremum of virtual environments, the authors align the class of *Augmented Virtuality (AV)*, which augments a virtual environment by real objects. As illustrated in Figure 2.1, this could be a virtual environment to which a live video capture of the user's real hands is added.

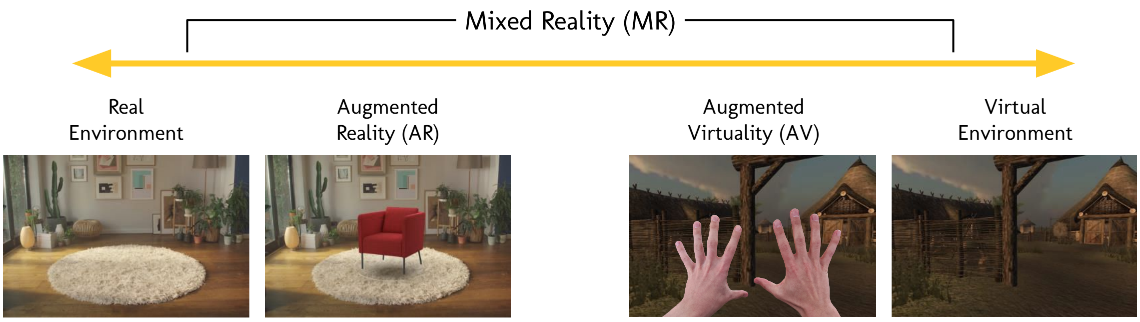


Figure 2.1: The *Virtuality Continuum (VC)* introduced by Milgram and Kishino [14] ranges from solely real environments to entirely virtual ones. The authors define the part of it where both virtual and real objects are presented together on the same display as Mixed Reality (MR). Based on the proportion of virtual objects in the scene, they further divide it into the sub-classes Augmented Reality (AR) and Augmented Virtuality (AV). *Redrawn from [14] and extended with images from IKEA Place [6] and own images.*

Bimber and Raskar [15] further divide the class of AR displays into three different categories "on the optical path in between the observer's eyes and the physical object to be augmented." As shown in Figure 2.2, they base their categorization on the distance from the user's eyes to the display. Starting from the eyes, the first class consists of *head-attached displays*. The probably most prominent example of such a display is Microsoft's HoloLens [16], which presents the AR experience directly in front of the user's eyes. *Hand-held displays* form the next class. Since smartphones and tablets are the typical devices of this class, it is also the most popular one [17]. Hand-held AR devices utilize the back-facing camera to capture the real environment and present it together with the virtual augmentations on their display. The class with the greatest distance to the user's eyes are *spatial displays*, which "detach most of the technology from the user and integrate it into the environment" [15]. Typical displays of this class are stationary ones like desktop displays and those created by projectors [17]. Research presented in this thesis focuses on hand-held devices since they are "the most popular platform for AR to date," widely spread and in contrast to HMDs, therefore, also socially more accepted [17].

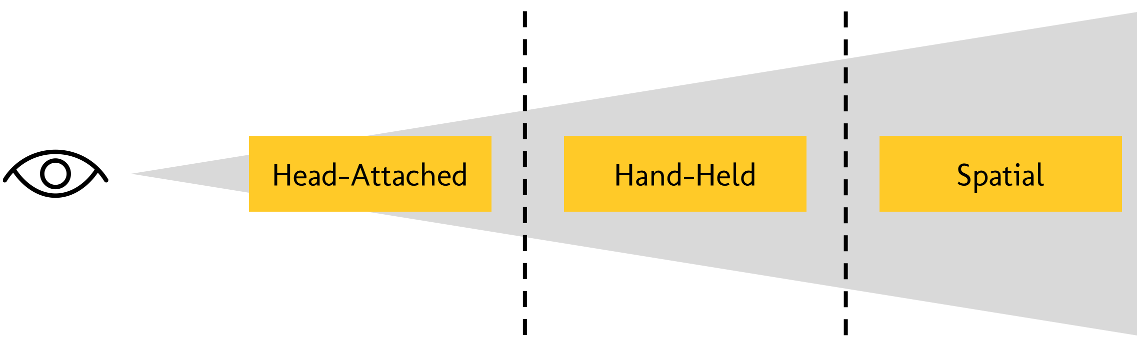


Figure 2.2: Bimber and Raskar [15] identify three different classes of Augmented Reality displays. They base their classification on the distance from the user's eyes to the display and therefore distinguish between *head-attached, hand-held, and spatial* displays. *Redrawn from [17] and slightly adapted.*

2.2 Classifying Collaboration in AR

A useful framework to classify computer-supported collaboration at large is the *Time/Space Groupware Matrix* developed by Johansen [18]. As depicted in Figure 2.3 the matrix consists of the two dimensions *space* and *time*, which subdivide collaborative computer systems into four groups. Regarding the dimension of space, collaboration can either happen *co-located* or *remote*. Co-located collaboration denotes collaboration that takes place at the same, shared location. On the contrary, if collaborators are located in different places, the matrix defines it as remote collaboration. The dimension of time divides these two categories further: If people collaborate at the same time, it is called *synchronous* collaboration, whereas *asynchronous* collaboration happens, if the users are working at different times, i. e., more independently of each other [17].

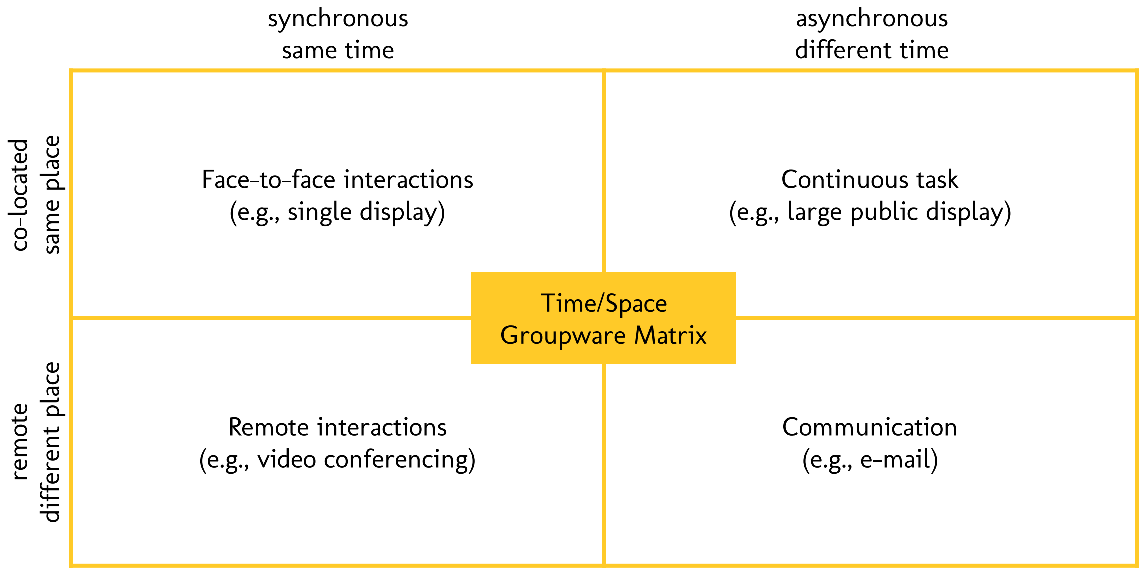


Figure 2.3: The *Time/Space Groupware Matrix. Redrawn from [19].*

According to Schmalstieg and Höllerer [17], AR is used for synchronous collaboration primarily; "asynchronous AR is less frequently utilized." Therefore, most AR applications align themselves in the left half of the matrix. Schmalstieg and Höllerer reason this with the second characteristic of AR – the interactivity of the medium. Consequently, this work investigates a co-located and synchronous collaboration, because it covers collaborative 3D object manipulation for the scenario of furnishing a home.

In addition to the Time/Space Groupware Matrix, Margery et al. [20] propose a three-level framework for classifying cooperation in virtual environments (cf. Figure 2.4). Despite their focus on solely virtual environments, their framework is also applicable to AR environments because those also require users to interact with virtual objects. Systems of the first level enable "users to perceive each other" and provide "ways of communication between the users" [20]. For co-located, synchronous AR, these two characteristics are met: Both users are at the same place and can see and hear each other. A system supports level 2 cooperation if it allows "each user [to] change the scene individually" [20]. The level's central limit "is that users cannot cooperate on the objects present in the scene" [20]; this is that only one user can manipulate an object at a time. Level 3 systems allow this kind of cooperation on an object either in an *independent* way or in a *codependent* way. In this context, "independent" means that the independent properties of the object can be modified simultaneously. For example, one user could move a chair, while the other one changes its color. Codependent, on the other hand, means that both users modify the same or linked properties of the object. Margery et al. [20] note that the latter is "requested to perform cooperative manipulation." This work focuses on level 3 cooperation since its subject is the collaborative manipulation of the same object at the same time.

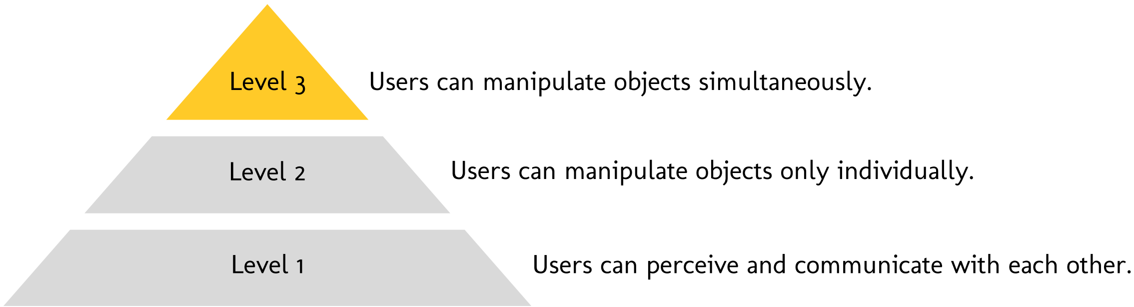


Figure 2.4: Margery et al. [20] define the three levels of cooperation. This work focuses on the third level of cooperation at which users manipulate the same object simultaneously.

2.3 3D Object Manipulation on Hand-held AR Displays

According to LaViola et al. [21], one can decompose every manipulation of a virtual 3D object into the four canonical tasks *selection/release*, *positioning*, *rotation*, and *scaling*. The task selection precedes all other tasks and is, therefore, the most basic one. With it, the user acquires or identifies the object to be manipulated. Its opposite is called release. Since a release is used to revoke a selection, it ends every manipulation. Selection and release count together as one canonical task. Positioning is the task of "changing the 3D position of an object" [21]. During positioning, the user has to define the targeted position of the object. The task rotation is used to change the orientation of an object and requires the user to input the wished rotation angle. Eventually, a user performs the task scaling for varying the size of an object by specifying its new size. Figure 2.5 shows a typical interaction sequence of a 3D object manipulation based on IKEA's Place: At first, the user taps on an object like the armchair to select it. After selection, the object then can be manipulated by positioning, rotating, or scaling it. In the example, the user drags the object to a new position. Manipulation ends with the release of the armchair. Therefore, the user taps elsewhere on the screen.

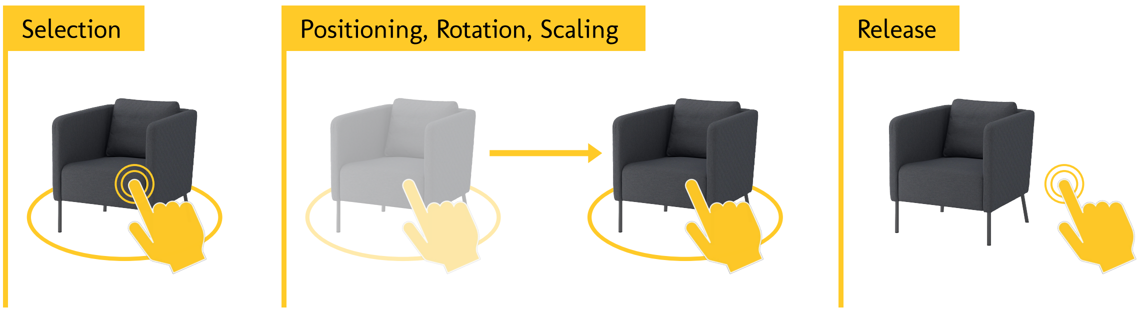


Figure 2.5: A walk-through of an exemplary manipulation. At first, the user selects the object. With its selection, the object is attached to the user, who then can manipulate its position, rotation, and scale. Finally, the user releases the object to end its manipulation. *With images from [22].*

From a more technical point of view, nine *degrees of freedom (DOFs)* define the position, rotation, and scale of a virtual 3D object. As shown in Figure 2.6, during each of the canonical tasks positioning, rotation, and scaling, a user performs transformations on three of the 9 DOFs of an object. Accordingly, a system that supports 6 DOFs typically offers only the tasks positioning and rotation. Systems that support 7 DOFs additionally provide uniform scaling. Independent scaling into all three directions is allowed by systems that provide manipulation techniques for all 9 DOFs. [23]

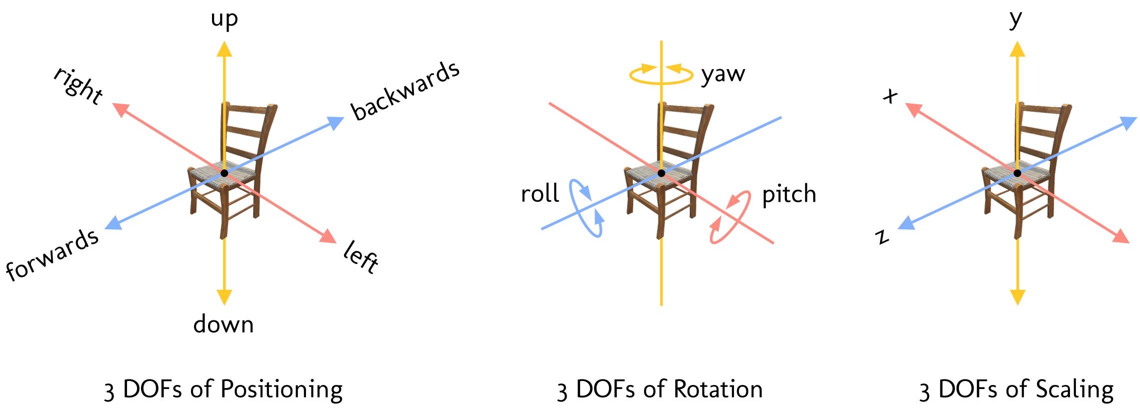


Figure 2.6: With each of the canonical tasks positioning, rotation, and scaling, the user manipulates different *degrees of freedom* (DOFs) of an object.

Hand-held AR devices like smartphones and tablets provide a touch-sensitive screen as input modality. Therefore, it is natural to use touch input for the execution of the canonical tasks and the manipulation of the DOFs of virtual 3D objects on those devices [17]. Because touchscreens are planar surfaces, this creates a problem: The user performs gestures in two dimensions to manipulate 9 DOFs in three dimensions. The scenario of furnishing a room, however, allows reducing the complexity of each manipulation to two dimensions: Pieces of furniture usually stand on the floor and can only be moved on it. Further, they typically only need to be rotated around the yaw axis. Also, scaling is not required. As illustrated in Figure 2.7, the study prototype (cf. Chapter 4), therefore, relies on ray-casting for selection and positioning and uses the established rotate gesture performed with two fingers for rotating objects. IKEA's Place and the AR samples provided by Google [24] also utilize this interaction concept.

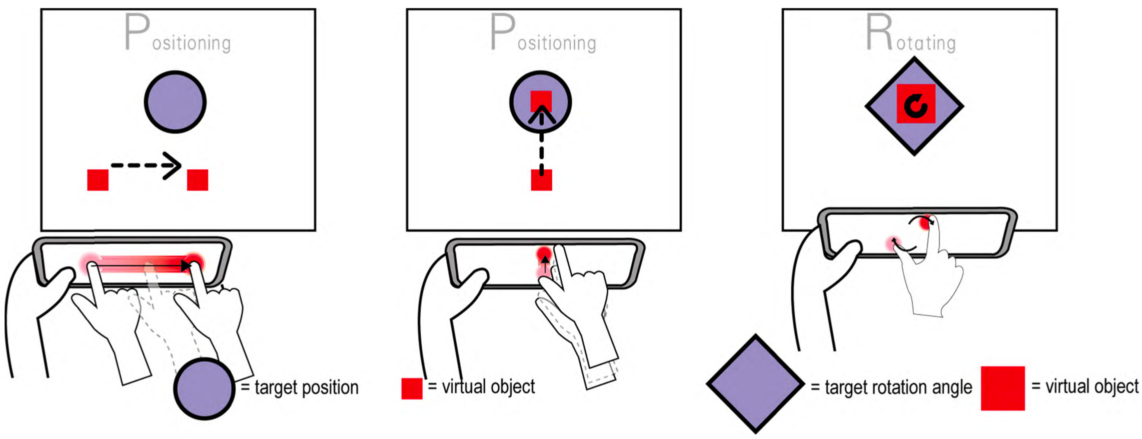


Figure 2.7: Touch-based manipulation of 3 DOFs as used by IKEA's Place and the AR samples provided by Google. *Taken from [25].*

2.4 Crucial Factors for Collaborative 3D Object Manipulation

Ruddle et al. [26] elaborate on three crucial factors to be considered for every systems supporting a collaborative manipulation of 3D objects: 1) Network communications, 2) feedback, and 3) action integration.

The crucial factor of *network communications* deals with different technical problems. Network delays can make it challenging to decide whether actions of different users "should be considered to be simultaneous or successive" [26]. Further, data loss can cause inconsistencies between the states of the virtual objects on the devices of different users. For study prototypes that aim to investigate "the behavioral aspects of interaction," the authors suggest using a "best-case scenario for network communications" to avoid such issues [26].

Feedback is a crucial factor for a collaborative manipulation since it helps users to "understand the intended actions of each other" [26]. For co-located AR environments, especially visual feedback can be helpful that provides information concerning the direction and extend of the manipulations each user is attempting to apply. Remote AR and VR scenarios require additional feedback mechanisms like avatars and communication channels since the users can not see and talk to each other.

Collaborations that involve two or more persons manipulating the same DOF of an object create a conflict. Collaborative approaches for 3D object manipulation, therefore, need to solve this conflict by performing *action integration*. This need for action integration can also be reasoned as a result of the crucial difference between the collaborative manipulation of real and virtual objects. If two persons are carrying a real object together, they are physically connected by it. This physical connection is not present during the collaborative manipulation of virtual objects. Therefore, "rules have to be defined that allow multiple inputs to be integrated" [26]. Those rules can either define a symmetric or an asymmetric action integration. *Symmetric manipulation* "requires two (or more) people to perform actions that are coordinated in all respects, i.e., actions that have the same magnitude and direction as each other, and that are performed at the same time" [26]. On the contrary, "with *asymmetric manipulation*, two users will deliberately make substantially different actions – for example, one person allowing the object to pivot, while the other changes its orientation" [26]. Asymmetric manipulation can also mean that the "users are not treated as equal partners," which relates to the concept of symmetric and asymmetric collaboration [26].

While this distinction between symmetric and asymmetric manipulations is aiming to cover systems that simulate real-world manipulations, the look at systems in other publications mainly reveals three different approaches. The first approach is called *Separation of DOFs* [4, 7] and refers to an asymmetric action integration. It enforces work division to solve the conflict by separating the manipulation of the different DOFs of an object between the users. Systems utilizing the second approach compose the actions of the users to one resulting manipulation. The approach is therefore called *Composition of Users' Actions* [9, 26, 27] and implements a rather symmetric manipulation. Finally, some systems also implement a combination of both variants as a *Hybrid Approach* [5, 10, 11]. The discussion of related work in the next chapter gives examples for all three approaches.

2.5 Conclusion

This work focuses on a comparison of the three probably most common approaches for performing action integration to enable the collaborative manipulation of virtual 3D objects: *Separation of DOFs*, *Composition of Users' Actions*, and their combination as a *Hybrid Approach*. Therefore, it sets the scope on hand-held AR displays and their usage for the real-world scenario of interior design: Users should be able to furnish the same home at the same time with the possibility of manipulating objects simultaneously. The type of collaboration investigated by this work, therefore, can be classified as co-located and synchronous with a level 3 cooperation. Due to the scenario of furnishing a place, this work only considers the canonical tasks selection, positioning, and rotation for 3D object manipulation. Scaling is not taken into account since the virtual objects need to match in size with their real-world counterparts. Since the lifting or lowering of an object is not needed either, only the two other DOFs of the canonical task positioning are investigated. For the same reason, rotations are limited to the yaw axis. Consequently, 3 DOFs are considered in total. Because this work focuses on hand-held devices, a touch-based manipulation of these 3 DOFs is considered. For the implementation of the three approaches in a study prototype, the crucial factors of network communications and feedback need to be addressed additionally.

3

Related Work on Collaborative 3D Object Manipulation

Based on the Seminar to the Master's Project [28], this chapter summarizes related work on collaborative 3D object manipulation. During the literature analysis it turned out, that regarding 3D object the differences between the terms "collaboration" and "cooperation" are not clearly defined [29]. This work generally decides on "collaboration" following most recent works [5, 10, 11]. Nevertheless, in passages that summarize publications that refer to "cooperation," the term is retained. Further, since only a few publications approach an AR setting, also those focusing on virtual environments are included. At first, studies are presented, that investigate a *Separation of DOFs* and afterward, those who focus on a *Composition of User's Actions*. Subsequently, works are introduced that pursue a *Hybrid Approach*. Finally, studies are summarized that compare different approaches for action integration witch each other.

3.1 Separation of DOFs

Motivated by concrete manipulation techniques and tasks, Pinho et al. [4] highlight two different ways to separate the manipulation of an object's DOFs between the users.

The first strategy is to separate the canonical tasks between the collaborators. For example, one user could care about the 3 DOFs of positioning and the other one about the 3 DOFs of rotation. According to the authors, this form of a *Separation of DOFs* can be useful for manipulation techniques that facilitate some of the manipulation tasks but also complicate others. As an example of such a technique, they discuss ray-casting. The *ray-casting technique* [30] attaches a virtual ray to the user's hand or device. To select an object the user needs to point on it with this ray and actuate a trigger. The selected object is then attached to the ray and can be moved and rotated with it. Ray-casting works well for selection and movements around the user. However, since the object is attached to the ray and not to the user's hand, the technique has some drawbacks. Especially, rotations around the vertical yaw-axis of the object are difficult to afford without changing the position of the object. Therefore, Pinho et al. [4] argue to "allow one of the users to control the object's position and the other one to control the object's rotation" to solve this issue.

The second way they highlight for a *Separation of DOFs* is to separate the DOFs of one canonical task between the users. For the canonical task positioning, this could mean that one user moves the object in x and z-direction while the other one moves it in the y-direction. The authors also motivate this way to perform a *Separation of DOFs* with an example: Users are required to shelve a virtual object with the ray-casting technique (cf. Figure 3.1). A user that stands in front of the shelf has no problems to position the object horizontally and vertically before the right division. However, he has difficulties in assessing the depth of the object for proper placement. The perspective of a second user standing on the side allows perceiving the depth of the object more easily. Consequently, he can assist by sliding the object along the ray of the first user into the division (*"Sliding Feature"*).

Figure 3.1: Although the user on the left has no problems to position the object horizontally and vertically before the right shelf, he has problems to assess its depth. Because of his perspective, this comes naturally to the user on the right side who can assist by sliding the object along the ray of the first user. *Taken from [4].*

To examine the different variants of the *Separation of DOFs* approach, Pinho et al. [4, 7] developed a modular software framework for virtual environments that eases the development and exploration of such cooperative manipulation techniques as described in the examples. Thereby, their goal was to develop not only usable but also useful cooperative manipulation techniques that meet the following four requirements, which are cited from [7]:

> **Awareness:** Showing to one user the actions his partner is performing.
>
> **Evolution:** Building cooperative techniques as natural extensions of existing single-user techniques in order to take advantage of prior user knowledge.
>
> **Transition:** Moving between a single-user and a collaborative task in a seamless and natural way without any sort of explicit command or discontinuity in the interactive process, preserving the sense of immersion in the virtual environment.
>
> **Reuse:** Facilitating the implementation of new cooperative interaction techniques, allowing the reuse of existing code.

As single-user manipulation techniques, Pinho et al. [7] included ray-casting and the *simple virtual hand* technique into their framework to address the requirement of *Evolution*. In contrast to ray-casting, the simple virtual hand technique directly maps the user's hand motion to a virtual hand, which is used to grab and manipulate a virtual object. The authors then built cooperative manipulation techniques by combining the single-user techniques in two ways: *Homogeneous cooperative techniques* are "built from the same single-user interaction technique" [7]. The already discussed example in Figure 3.1 shows such a homogeneous cooperative technique. One user utilizes ray-casting to align the object before the right division of a shelf. The second user then slides the object into the shelf along the ray of the first user. *Heterogeneous cooperative techniques*, in contrast, are "built from different single-user techniques" [7], e. g., using the simple virtual hand for positioning and ray-casting for rotation. The authors implemented techniques of both groups using the different combinations of simple virtual hand and ray-casting summarized in Table 3.1.

Interaction Technique A	DOF(s) User A	Interaction Technique B	DOF(s) User B	Comments
SVH	Position	SVH	Rotation	Useful for docking tasks and small adjustments. Good when one user cannot see parts of the object.
SVH	X, Y	SVH	Z	Facilitates precise positioning.
Ray-casting	Position	Ray-casting	Rotation	Useful for rotations that are difficult with ray-casting.
Ray-casting	Position	Ray-casting	Rotation Slide	Useful for distant placement and rotations.
SVH	Rotation	Ray-casting	Position	Useful for rotations that are difficult with ray-casting.
SVH	Rotation, Slide	Ray-casting	Position	Useful for distant placement and rotations

Table 3.1: Pinho et al. [7] combine the two interaction techniques simple virtual hand (SVH) and ray-casting to develop cooperative ones. Each of the users A and B uses one of both interaction techniques to manipulate other DOFs. *Redrawn from [7].*

Their first homogeneous combination of two simple virtual hand techniques allows one user to manipulate the position of an object, whereas the other one controls its orientation. Their second homogeneous cooperative technique for the simple virtual hand lets the first user translate the object along its x- and y-axis. Simultaneously, the second user can only modify the distance between the object and the first user. The authors also homogeneously combined ray-casting in two ways. In the first variant, one user controls the position of the object, and the other one its orientation. The second variant additionally enables the second user to slide the object along the first user's ray. Finally, they implemented two variants for the heterogeneous combinations of simple virtual hand and ray-casting. The first variant provides one user with the simple virtual hand technique to rotate the object. The second user is equipped with a ray-casting technique that implements the sliding feature to position the object. In the second variant, the user with the simple virtual hand techniques also controls the sliding along the other user's ray.

To test these six variants, Pinho et al. [7] conducted a first preliminary study, with 12 participants forming six dyads. Each dyad had to execute one of the following three tasks: During the first task, participants had to place a set of objects on a matching set of platforms. The second task required them to move a couch through a door, with the users positioned on opposite sides of the wall. And as the third task, they had to place objects between some walls. Thereby, one user stood far from these walls and another user next to them (cf.Figure 3.2).

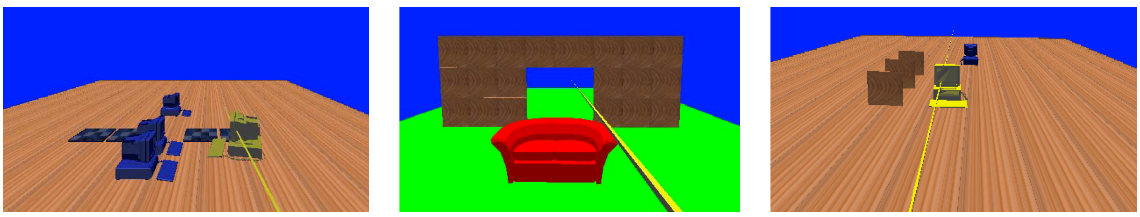


Figure 3.2: In the study of Pinho et al. dyads had to solve one out of three tasks: 1) Place a set of objects on a set of platforms *(left)*. 2) Move a couch through a door *(middle)*. 3) Place a set of objects between some walls *(right)*. *Taken from [7].*

During the task, both users wore an HMD that displayed the virtual environment. A tracking system tracked their heads and one of their hands. In the other hand, participants held a button that was used as a trigger for grabbing and releasing virtual objects. Both participants were seated and faced each other. Further, they were located at different, fixed points in the virtual environment and could no move through it. At first, participants had to perform the task alone. Afterward, they were allowed to help each other, by manipulating the objects only sequentially but not cooperatively. Finally, dyads were asked to solve the task using a cooperative manipulation technique. The observations and subsequent interviews Pinho et al. conducted led them to the following preliminary conclusions:

Preliminary conclusions of the first study from Pinho, Bowman and Freitas
1. Cooperative techniques can provide increased performance and usability in difficult manipulation scenarios. However, single-user manipulation is simpler to use and understand for most manipulation tasks.
2. The use of a cooperative technique is applicable to those situations in which cooperation allows the users to control some DOFs that cannot be controlled with the single-user technique.
3. The ease of learning for cooperative techniques depends more on the individual user than on the technique itself, i.e., those users who learned quickly how to use an individual technique also learned quickly how to use the cooperative one. On the other hand, those who had difficulty in learning the individual technique also took more time to learn the cooperative one.
4. Users adapted to the system and learned the appropriate times to manipulate objects individually and cooperatively. Users had no trouble with the transition between single-user mode and cooperative mode because of our careful design and implementation.

Table 3.2: Based on observations and interviews, Pinho et al. [7] formulate some preliminary conclusions about the collaborative techniques and the *Separation of DOFs*. *Redrawn from [7].*

Due to the preliminary character of their study, Pinho et al. did not provide quantitative results like task completion times or measures for the accuracy of the cooperative techniques compared to the single-user ones. This quantitative data would be required to deepen their conclusions on their suggested variants for a *Separation of DOFs*. However, in terms of task completion times, a subsequent publication by Pinho et al. [4] provides further insights. In this paper, they report the results of another similar study, which consisted of three experiments. The authors recruited 60 participants forming 30 dyads. In each of the three experiments, ten dyads took part. Further, the experiments followed the same procedure as in their previous study. In a first pass, dyads had to perform the task without cooperative manipulation. Afterward, they repeated the task by using the cooperative manipulation technique. For both passes, task completion times were measured. In the first experiment, the virtual environment simulated a classroom (cf. Figure 3.3, left). The participants were positioned in opposite corners of the room and had to place computers on desks. This time participants had to use the *HOMER technique* [30], whose acronym stands for "hand-centered object manipulation extending ray-casting."

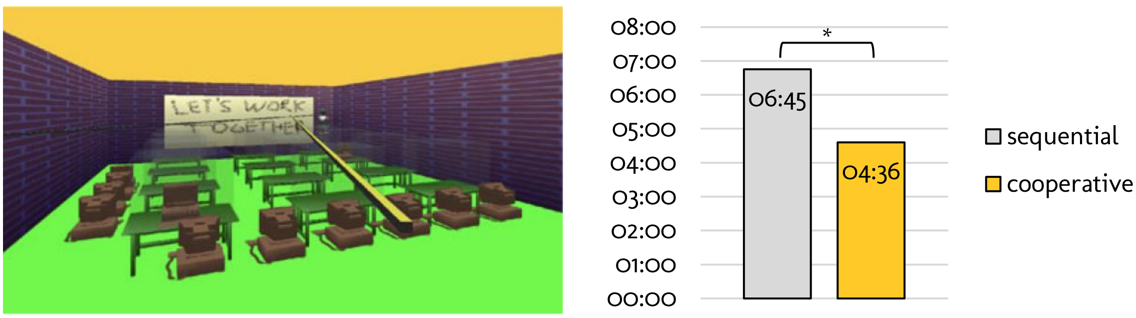


Figure 3.3: During the first experiment of Pinho et al. [4] the users had to position computers on desks inside a classroom. *Taken from [4].*

HOMER combines ray-casting for selection with the simple virtual hand for manipulation. Therefore, “instead of the object being attached to the ray, the user’s virtual hand instantly moves to the object and attaches to it. The technique then switches to the manipulation mode, allowing the user to position and rotate the virtual object” [21]. As cooperative manipulation technique Pinho et al. provided the participants with a homogeneous combination of the technique, which let one participant control the object’s position and the other one its rotation. Their analysis of the task completion times for this experiment revealed that participants were significantly faster in the pass with the cooperative manipulation technique (cf. Figure 3.3, right).

In the second experiment, dyads had to place objects into divisions of a shelf. One participant (P1) was positioned in front of the shelf, whereas the other one (P2) had the view shown in Figure 3.4 on the left. During this experiment, participants used ray-casting with the sliding feature. In the pass without cooperative manipulation, P2 moved the objects in front of the shelf, and P1 sorted them into it afterward. In the cooperative pass, P1 controlled the object’s translation, whereas P2 was responsible for their rotation and the sliding along the ray of P1. Consequently, P1 had only to point into the right division of the shelf, and P2 slid it into the division. Also, for this task, the statistical analysis of task completion times revealed that participants were significantly faster using the cooperative manipulation technique (cf. Figure 3.4, right).

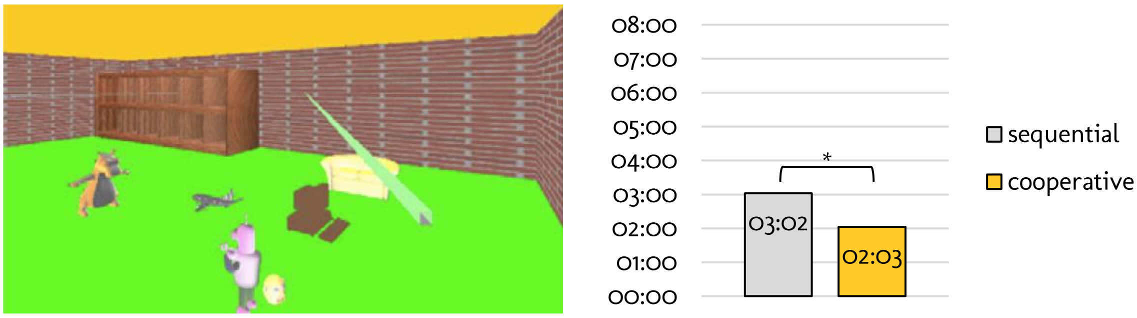


Figure 3.4: The second experiment by Pinho et al. [4] was about large movements and object fitting. Dyads had to place objects into divisions of a shelf. *Taken from [4].*

Their third experiment investigated movement in a cluttered environment. As the task, dyads had to move a couch through an aisle full of obstacles (cf. Figure 3.5, left). One participant (P1) was placed at one end of the aisle and the other one (P2) at its side. For the pass without cooperative manipulation, dyads again had to use the HOMER technique. During the cooperative pass, P1 controlled the translations of the couch and P2 its rotation and sliding along P1’s ray. As for the other two experiments, statistical analysis revealed that participants were significantly faster in the second pass, i. e., with the cooperative manipulation technique (cf. Figure 3.5, right).

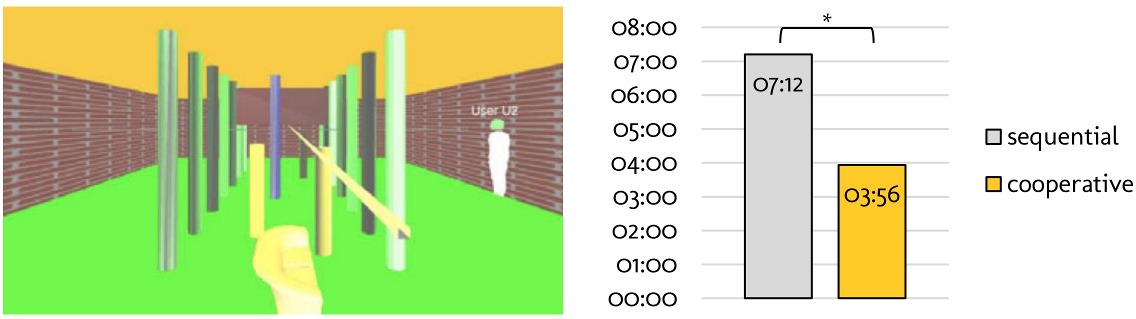


Figure 3.5: The third experiment of Pinho et al. [4] examined movement in a cluttered environment. Dyads had to move a couch through an aisle full of obstacles. *Taken from [4].*

The results of their study let Pinho et al. conclude “that the experiments allow to assert quite confidently that for many tasks in which the insertion of a collaborator (with non-cooperative manipulation) improves task execution, the use of simultaneous cooperative manipulation provides an even greater benefit” [4]. Altogether, both of their works clearly show that the *Separation of DOFs* between the users is a promising approach for collaborative 3D object manipulation. Nevertheless, their studies lack a comparison according to the accuracy of their cooperative techniques, which would be interesting to see. As possible future work, they mention a *Separation of DOFs* between more than two collaborators or remote ones.

3.2 Composition of Users’ Actions

Works that investigate a *Composition of Users’ Actions* allow concurrent manipulations of the same DOFs by implementing a *merge policy* for this symmetric action integration [10]. As shown in Figure 3.6, common merge policies compose the inputs of the collaborators to generate one single transformation by either taking the *sum,* the *mean,* or the *common component* of the users’ inputs [8]. For the computation of the sum, the vectors are added up. For the mean, the two vectors are added up and divided by two. The direction of the common component bisects the angle between the manipulations of the users. The common component’s magnitude is calculated from the dot product. In addition to these three merge policies, some systems use more complex ones to mimic the constraints involved when carrying a real-world object collaboratively. Apart from the computation of the sum, this section presents related work for each of the different merge policies. An example of the sum as a merge policy is given in the next section, which examines the hybrid approaches.

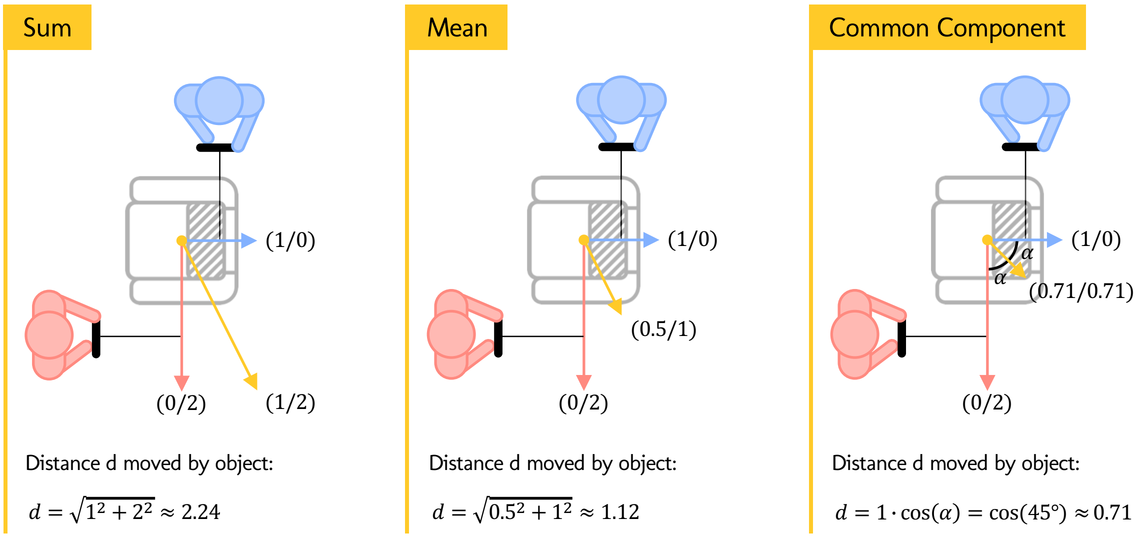


Figure 3.6: Related Work reveals different *merge policies* to perform a *Composition of Users’ Actions*. Some add up the inputs of the users *(left)*. Others calculate their mean *(middle)*. Finally, also the common component is used *(right)*. For the latter, the object’s direction of movement bisects the angle between the manipulations of the two users. The distance it moves is calculated by the dot product.

Ruddle et al. [26] compared the two merge policies that compute the mean and the common component with each other. Therefore, they implemented both of them in a virtual environment (VE) presented on a desktop computer. In an experimental comparison of the two variants, dyads had to solve the so-called *piano movers’ problem*. Therefore, each of the participants used a spatially tracked controller to steer a virtual person holding one side of a large virtual object.

Together, they had to carry this virtual object through two different VEs that were based on parts of a building (cf. Figure 3.7). The VEs represented different scenarios: In the offset VE, participants had to carry the object through two offset doors in consecutive walls. The C-shaped VE required them to carry the object around the corner of an aisle. The object could collide with the walls of the VEs, which means participants were not able to move the object through the walls, the floor, or the ceiling. Participants had to perform the task with both merge policies on two consecutive days. On the first day, they offset VE was presented to them and on the second day, the C-shaped VE.

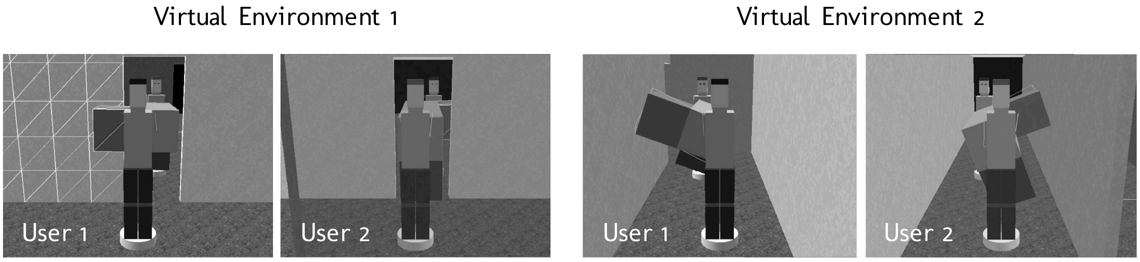


Figure 3.7: During the study of Ruddle et al. [26] participants had to solve the so-called *piano movers' problem* for two different virtual environments (VEs). The offset VE consisted of two offset doors in consecutive walls *(left)*, and the C-shaped VE simulated a corner in an aisle *(right). Taken from [26] and slightly adapted.*

According to overall task completion times, the results of their study showed no statistically significant differences between the two merge policies. However, when they divided the task into different phases, they found differences between the merge policies: For the phases when both participants needed to execute a similar action, they were faster with the variant that used the common component as merge policy. On the other hand, in phases in which participants needed to perform different types of movement, the interaction was quickest with the mean rule. Further, they found out that with the mean as a merge policy, the virtual object was in collision with the VE for significantly more time than with the common component. However, participants remained less time stationary with the object not colliding for the variant with the mean as merge policy. Additionally, the compared the results with a previous experiment during which dyads had to solve the task alone with a single-user mode. It became apparent that with the common component as a merge policy, participants were significantly slower as with the single-user mode. In contrast, the task completion times with the computation of the mean as merge policy were similar to the single-user mode. Consequently, Ruddle et al. [26] conclude that "moderately difficult tasks can be performed cooperatively almost as quickly as when individuals perform them by themselves" – at least with the merge policy that uses the mean.

The *SkeweR* system by Duval et al. [9] tries to simulate the grabbing and carrying of a real-world object in a virtual environment presented by a 3D projector. Consequently, its merge policy is rather complex and introduced the concept of crushing points. SkeweR enables two users to move an object at the same time by "grabbing" any part of it with a virtual cursor. "Each user then manipulates the object by [the grabbed] crushing point, like handling the extremity of a skewer" [9]. Consequently, translation information from the users' motion is sufficient to perform both translations and rotations of the virtual object (cf. Figure 3.8). To evaluate their prototype, they performed some preliminary tests with two users. As the task, participants had to move an object through a 3D maze. Based on this preliminary test, Duval et al. report that their "first observations indicate that the *SkeweR* technique seems very intuitive and natural to use" [9]. Nevertheless, they admit that they would need to "perform a series of experiments to measure if this technique is easy and efficient enough for collaborative manipulations of 3D virtual objects" [9].

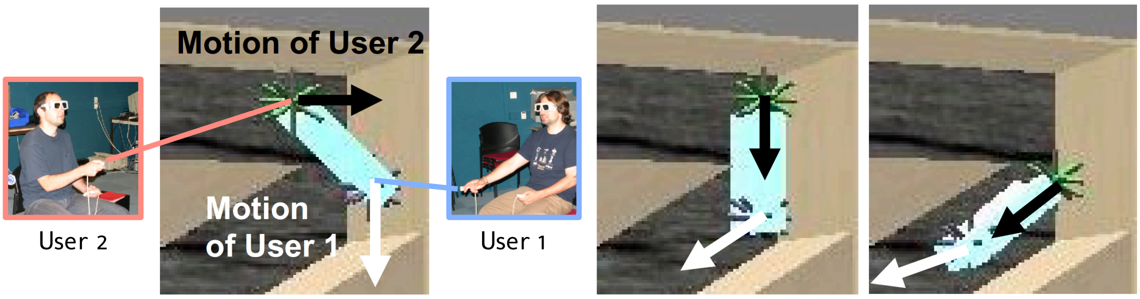


Figure 3.8: In the *SkeweR* system by Duval et al. [9], each of the two collaborators "grabs" and controls one crushing point. Therefore, translations are sufficient to perform also rotations. *Taken from [9] and slightly adapted.*

Riege et al. [27] propose the *Bent Pick Ray* as a further collaborative 3D object manipulation technique for projection-based virtual environments. Their technique is based on ray-casting, which typically provides the user with a straight ray that connects the input device with the object to be manipulated. However, if two users manipulate an object at the same time, one user can drag the object away from the ray of the second user. The system of Riege et al. bends the rays of the users to preserve the connection between their input devices and the object. Figure 3.9 illustrates this with an example: As long as a single user manipulates the object, its ray is always straight. At the moment when the second user also selects the object and starts manipulating it, their rays bend to compensate for movements in different directions. If one user releases the object, his ray gets immediately unbent. The second user's ray gets unbent with further movements. To combine the users' inputs, they investigated two different merge policies. The first one averaged the inputs of the users. However, tests with this merge policy revealed that the activity of both users is unequal most of the time. Consequently, they implemented a weighted averaging, which takes the activity of the users into account. With this weighted averaging, a larger amount of hand movement has a bigger influence on the resulting manipulation.

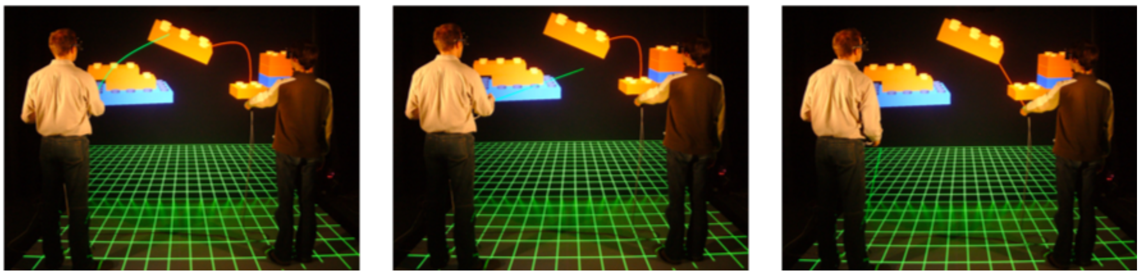


Figure 3.9: The *Bent Pick Ray* technique by Riege et al. [27] is based on ray-casting. To compensate for conflicting movements, it bends the rays of the users *(left)*. As soon as one user releases the object, his ray get immediately unbent *(middle)*. The ray of the other user gets unbent with further movement *(right)*. *Taken from [27].*

Riege et al. did not extensively evaluate their Bent Pick Ray technique. However, the report some first user experiences. Users liked the "transparent way of connecting and disconnecting to an object, independent whether it had already been selected by the collaborating user or not." Further, they valued the "very distinct feedback, generated by the bending of the ray, visualizing two users controlling the movement of an object." Regarding the two different merge policies, the averaging was "intuitively usable and understandable for most users," whereas the weighted averaging "caused some confusion." Users reported that "it was not always clear to them how to move the input device to achieve a desired effect" and that "they felt a loss of control." [27]

3.3 Hybrid Approaches

Related work covered in this section investigates a combination of a *Separation of DOFs* and a *Composition of Users' Actions*. Some systems let the users explicitly switch between the two approaches, and others do it dependent on the users' inputs.

Chenechal et al. [10] propose a VR system with up to four different user roles. As illustrated in Figure 3.10, both the *Separation of DOFs* and the *Composition of Users' Actions* are present between the roles. Further, each role provides users with different devices. The *giant* is provided with a stereoscopic display that presents a global view on the VE. The user that takes over this role can translate the object. Additionally, he shares the possibility to rotate the object with the *ant*. To combine concurrent rotations of giant and ant, the authors implemented a custom merge policy that gives a bigger influence to the giant's actions. The ant sees the VE through an HMD and is located inside the virtual object to be manipulated. According to the authors, "this position enables him to manipulate the object with a fine accuracy" [10]. In addition to the shared control over the rotation, he is responsible for scaling the object. The *helping user* is the third role, which is optional. This user is also provided with an HMD but has a third-person view on the virtual object. Besides, he shares the possibility to scale the object with the ant. To combine concurrent scaling actions of ant and helping users Chenechal et al. decided to use the merge policy that applies the sum of the scaling factors of the different users to the object. Finally, the system allows further users to take over the role of the *spectator*, which can toggle between the views of the other roles. Users that take over this role can support the other users with oral instructions but are not able to intervene in the manipulation.

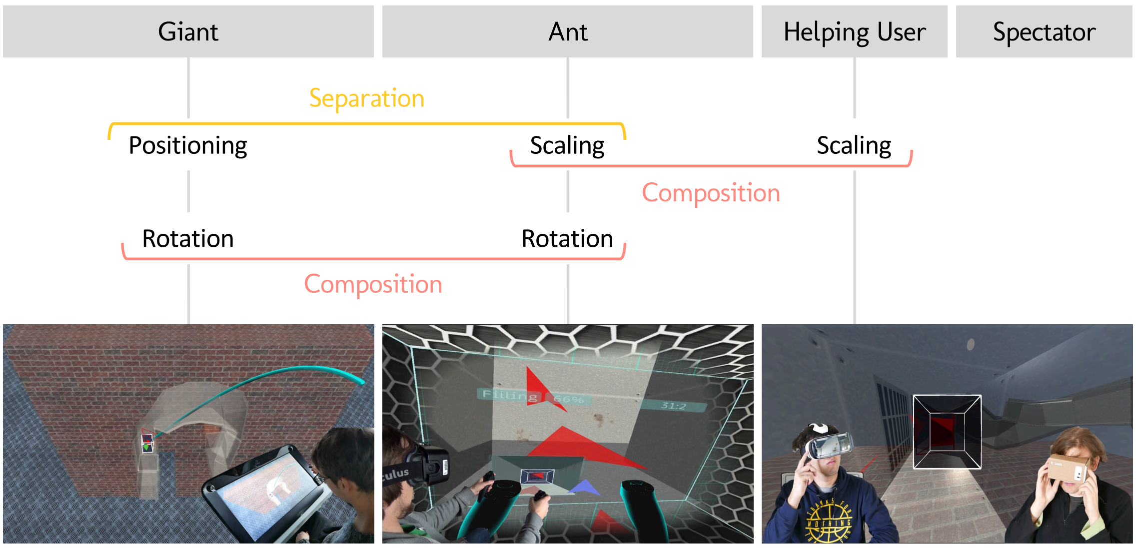


Figure 3.10: The system by Chenechal et al. [10] lets the users take over different roles. Both, a *Separation of DOFs* and a *Composition of Users' Actions* are used. The *giant* has a global view on the scene and is responsible for translating the object to be manipulated *(left)*. Together with the *ant*, he shares the possibility to rotate the object. The ant is located inside the object and can additionally scale it *(middle)*. The *helping user* has a third-person view on the object and shares the possibility to scale the object with the ant *(right)*. The *spectator* can toggle between the views of the other users but is not able to intervene *(right)*. *With images from [10].*

Chenechal et al. evaluated their system with preliminary tests during which users had to maneuver a cube through different labyrinths "while maximizing the courses filling of the object." They conclude that those "user tests show a good efficiency of the different techniques." Nevertheless, they admit that "a formal evaluation should be done in order to confirm the performances of the approach." This formal evaluation especially should "compare [their] proposal with a solution from the state-of-the-art where the collaboration is symmetrical, i.e., with equivalent roles and viewpoints." [10]

Another example for a *Hybrid Approach* is the system of Grandi et al. [5]. As shown in Figure 3.11, they used smartphones with their touchscreen and inertial sensors as the input interface. For positioning, the orientation of the mobile device defines the plane on which the object is translated. The translation itself is performed by sliding gestures on the touchscreen of the smartphone. To rotate an object, the user needs to hold down the volume down button. The virtual object then follows the rotations of the mobile device. Uniform scaling of the virtual object is accomplished by performing the pinch gesture.

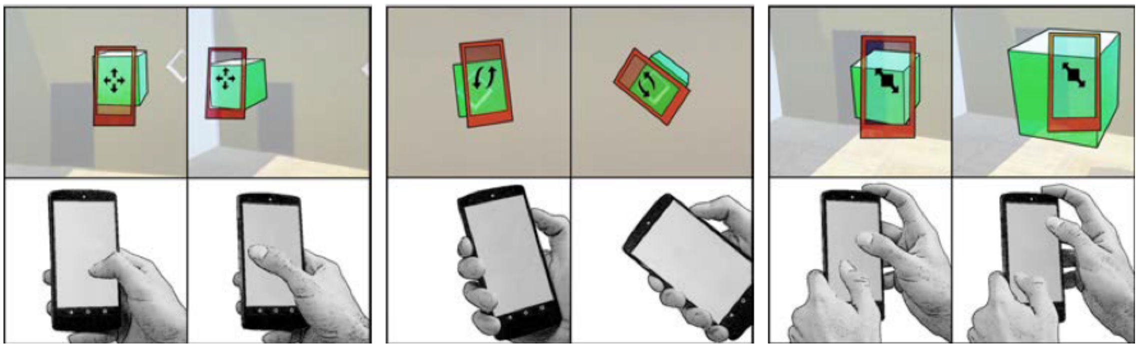


Figure 3.11: The various manipulations that a mobile phone can perform on the selected object: Translations can be applied by touching the mobile phone screen and sliding the point of contact. The object translates on the plane defined by the mobile device's orientation *(left)*. By holding the volume down button and rotating the phone, one can rotate the object likewise *(middle)*. Scaling is applied by touching the screen with two fingers and producing a pinch/spread gesture *(right)*. *Taken from [5].*

It is notable about their work that they "do not impose limits on the number of simultaneous users," which means that "all users have access to all available functions" [5]. During the collaboration, users can decide whether or how they split the DOFs between them or if they perform them simultaneously. To enable the users to do so, the system combines the participants' actions by summing them up as transformation steps, as shown in Figure 3.12. For example, "if two users move the object in opposite directions [by 4 steps], the position of the object will not change", since the same number of positive and negative transformation steps is summed up [5].

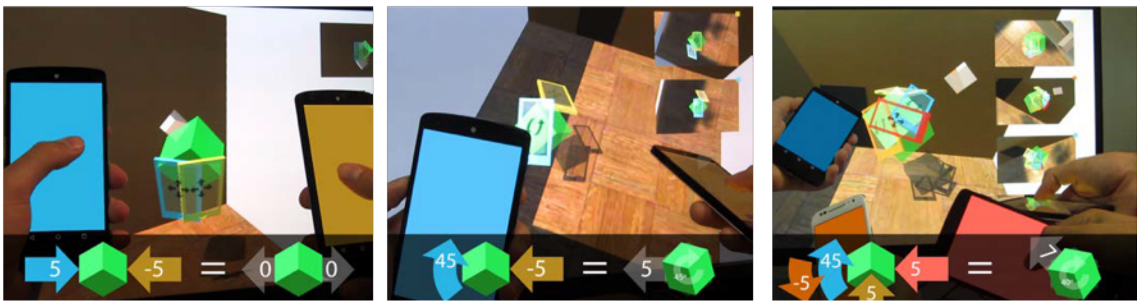


Figure 3.12: The system of Grandi et al. [5] allows an arbitrary number of collaborators to manipulate the same virtual object at the same time: Opposing manipulations of the same DOF cancel each other *(left)*. An example of a transformation merge for the simultaneous manipulation of different DOFs *(middle)*. Concurrent access to the same object by a group of four *(right)*. *Taken from [5].*

In a study, they compared teams composed of one, two, three, and four users with each other, i. e., the group size was the independent variable. As shown in Figure 3.13 on the left side, the virtual environment in which the objects were manipulated was displayed on a big screen, and each user was equipped with a mobile device. Groups had to solve an obstacle crossing task (cf. Figure 3.13, right). Therefore, a cube needed to be manipulated in order to carry it through a sequence of walls with openings while avoiding collisions but occupying the maximum volume possible. The dependent variables they measured were task completion time, accuracy, and transformation actions (i. e., positioning, rotation, scale, and camera rotation). Regarding task completion time, they could not reveal a significant difference. However, they found significant evidence for the tested group range, that the accuracy in completing tasks proportionally increases if groups increase in members. They also found that users tend to specialize in one transformation in larger groups. Finally, their results indicate a correlation between the work division strategy and the error drop: Members in bigger groups less frequently changed the transformation they performed and also made fewer errors. Noteworthy is also that qualitative results valued the comfort of their technique. As future work, they suggest to "focus on different levels of controlled labor division in a way that each user will be responsible for only one specific type of transformation" [5], which corresponds as a last consequence to the *Separation of DOFs* approach. They expect that "by forcing the role division between the peers, the accuracy could be increased for all the groups" [5].

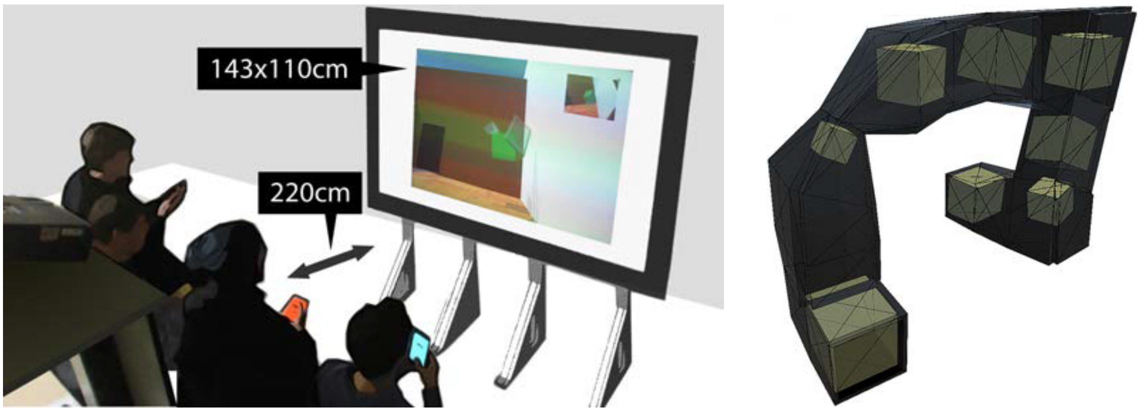


Figure 3.13: In the study by Grandi et al. [5], the participants used smartphones to manipulate objects on a big screen *(left)*. Their task was to carry a cube through a sequence of walls with openings while avoiding collisions *(right). Taken from [5].*

In a follow-up publication, Grandi et al. [11] introduce a modified AR version of the original prototype. The implementation of their *Hybrid Approach* stays the same, but the way how the users interact with the prototype is changed. In this version, manipulation is either performed with touch gestures or with device movements. As shown in Table 3.3, both techniques allow to modify the DOFs of positioning, rotation, and scaling. During a first single-user study, they compared both techniques with a hybrid combination, which enabled the users to switch between both techniques during the interaction. Participants had to solve a series of docking tasks, which means they had "to dock a virtual moving piece, controlled by the user with a similar virtual static piece" [5]. The dependent variables they measured were accuracy and time. When using the hybrid combination, participants were significantly faster as with the solely touch gestures or device movements methods. Furthermore, there was no significant difference in accuracy between the three conditions. Therefore, results showed the effectiveness of the hybrid combination, i. e., task completion time is reduced with consistent accuracy. The authors conclude that "users could seamlessly switch between methods and use the most efficient action to correctly transform the object while keeping high time performance" [11].

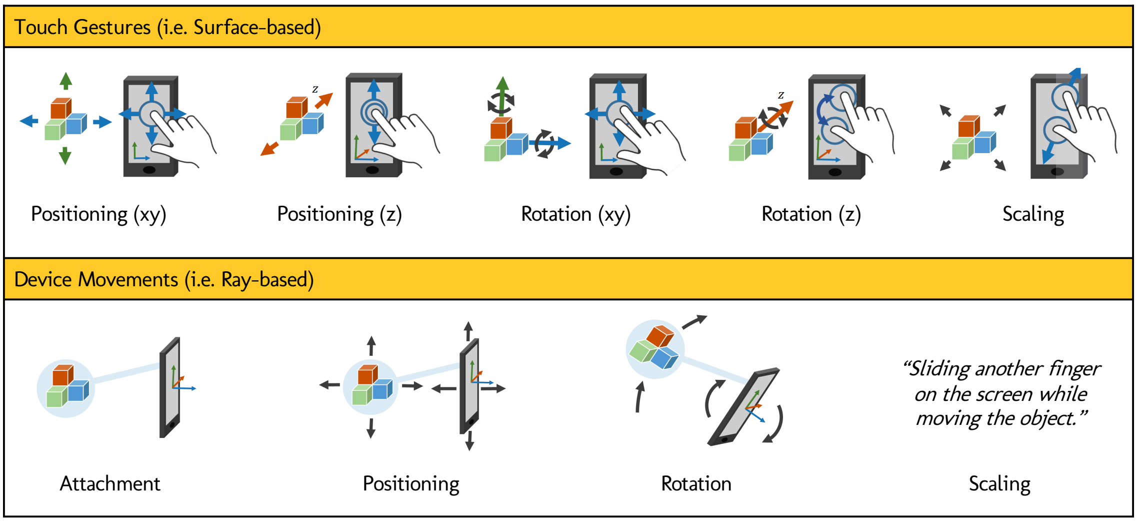


Table 3.3: The modified version of the prototype by Grandi et al. again implements a *Hybrid Approach* but alters the interaction techniques. In this version, manipulation is either performed with touch gestures (i. e., surface-based) or with device movements (i. e., ray-based). *Redrawn from [11] and slightly adapted.*

In a second study, they focused "on the evaluation of collaborative aspects when two users are manipulating virtual objects in the same scene" [11]. They aimed "at observing and classifying the strategies that emerge when users are free to make their task organization" [11]. This time, not individual participants but dyads had to solve the docking task using the hybrid method. To stimulate collaboration, they created three conditions that varied the level of occlusion, as shown in Figure 3.14. The results of their study revealed two main strategies. 66% of their teams choose to work together, which means that they collaboratively manipulated the same object (*shared interaction*). The members of the other pairs manipulated the objects individually (*independent interaction*). Interestingly, groups did not change their strategy with the level of occlusion.

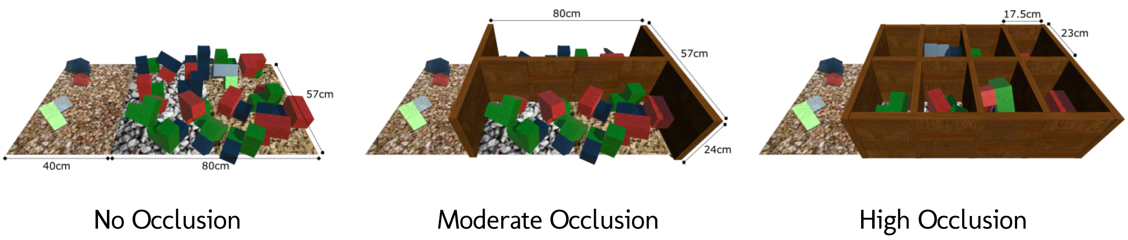


Figure 3.14: In their second study, Grandi et al. [11] created three conditions that varied the occlusion during the docking task. During each condition (no occlusion, moderate occlusion, high occlusion), the static objects in the left had to be docked with the right movable objects from the right. *Taken from [11]*

This indicates that "the strategies are less related to the environmental factors and more related to the users and pairs profile." Whereas the results showed no differences according to task completion times, they revealed that "pairs that adopted the shared interaction felt more involved in the task than pairs that adopted the independent strategy." Unfortunately, they do not report on how the two groups differed concerning accuracy. Also, one would like to get more insights into how the pairs of the group with shared interaction manipulated the objects. How many groups really manipulated the same DOFs simultaneously i. e., relied on a *Composition of Users' Actions*? Or were their groups that tended towards a *Separation of DOFs* like Grandi et al. reported it in their other work [5]?

3.4 Comparison of Different Approaches

The last three sections presented related work that investigated three different approaches for the collaborative 3D object manipulation: A *Separation of DOFs*, a *Composition of Users' Actions*, and the combination as a *Hybrid Approach*. But which of those approaches one should choose? Unfortunately, during the literature research for works that compare the approaches, only the publication of Aguerreche et al. [8] could be found, which compares at least a *Separation of DOFs* with a *Composition of Users' Actions*. During their study, participants had the task to assemble a car hood presented in a projection-based VE. Twenty-four participants performed the task in pairs with three conditions. In the first condition, they had to use the tangible device shown in Figure 3.15 on the left side, which could be reconfigured to match the shape of the virtual object to be manipulated. Participants grabbed it at different ends as if they would carry a physical object. Movements and rotations of this tangible device were directly coupled to the virtual car hood. For the other two conditions, participants used spatially tracked controllers that they held in one hand. The second condition implemented a *Composition of Users' Actions*. Participants' actions were composed using the merge policy that computes the mean of their inputs. Inputs were made by moving or rotating the controller (cf. Figure 3.16, middle). The third condition used a *Separation of DOFs* with one participant performing the positioning and the other one the rotation of the object (cf. Figure 3.16, right).

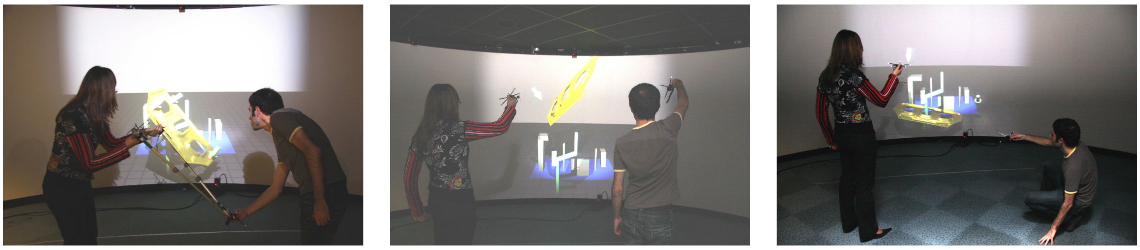


Figure 3.15: Aguerreche et al. [8] developed a reconfigurable tangible device for 3D object manipulation *(left)*. They compared it with a *Composition of DOFs (middle)* and a *Separation of DOFs (right)*. *Taken from [8].*

As dependent variables, Aguerreche et al. measured task completion times and accuracy operationalized by the number of collisions. Their results speak clearly in favor of the condition implementing a *Composition of Users' Actions*. Taking both measures together, participants performed significantly better with this condition compared to the one that used a *Separation of DOFs* (cf. Figure 3.16).

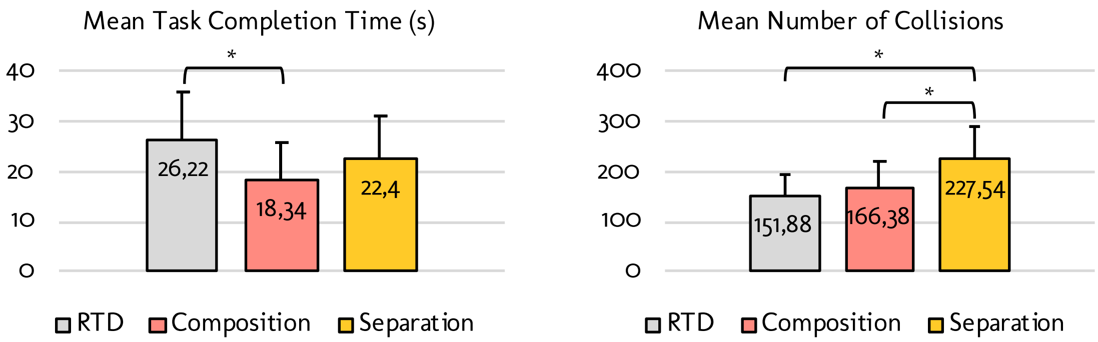


Figure 3.16: During their study, Aguerreche et al. [8] measured task completion times *(left)* and the number of collisions *(right)*. Disregarding the reconfigurable tangible device (RTD), both speak in favor of the *Composition of DOFs*. *Redrawn from [8].*

3.5 Conclusion

A view on related work for the approaches *Separation of DOFs*, *Composition of Users' Actions*, and their combination as a *Hybrid Approach* suggests that a collaborative manipulation can be more effective and can increase the usability of a system – with any of the presented approaches.

However, it is difficult to answer the question on which approach should be used for the collaborative 3D object manipulation with hand-held AR displays. This has different reasons: At first, only one work could be found, that investigated an AR scenario. All other publications focus on virtual environments presented by HMDs, projectors, or computer monitors. Secondly, many of the summarized publications only report the results of preliminary studies that evaluate a specific implementation of one approach. Of those who conducted formal evaluations, only two works provide insights on differences between the approaches. However, their results are quite contradictory: Aguerreche et al. [8] compared a *Separation of DOFs* with a *Composition of Users' Actions*. In terms of accuracy, their results are in favor of a *Composition of Users' Actions*. In contrast, Grandi et al. [5], who investigated a *Hybrid Approach*, concluded that the accuracy of the collaborators increased with their specialization on one transformation, which is in favor of a *Separation of DOFs*.

Consequently, to be able to decide on which of the three approaches one should use for a collaborative 3D object manipulation on hand-held AR displays, further research is required. The study prototype described in the next chapter and the experimental comparison described in Chapter 5 are a first step towards this direction.

4 Study Prototype

For conducting the experimental comparison, it was required to develop a study prototype that implements the three approaches for collaborative 3D object manipulation: *A Separation of DOFs*, a *Composition of Users' Actions*, and a *Hybrid Approach*. Since the domain of interior design motivates this thesis, the prototype was implemented as a hand-held AR application for furnishing a home. The chapter starts with a list of requirements for such an application, which derive from the foundations in Chapter 2 and the related work discussed in Chapter 3. Afterward, the specific interaction concepts for the three approaches are explained as they are investigated in this thesis. Finally, the concrete implementation of the study prototype is described in detail. The whole chapter is grounded on the Master-Project Report [31] that preceded this thesis.

4.1 Requirements

The crucial factor of network communications translates into the first rather technical requirement. Reliable, low-latency networking is essential for enabling collaborative 3D object manipulation.

> **R01 Networking:** The devices of the users need to communicate with each other using reliable, low-latency networking to exchange the state of the virtual furniture inside a room.

Further, users need to be able to add virtual furniture to the room by instantiating it, which is addressed by the second requirement. Every manipulation technique has to implement the canonical task selection/release. Therefore, this task directly translates into the third requirement. The canonical tasks positioning and rotation form the next two requirements. Since scaling is not relevant to the scenario of interior design, it is the only canonical task that is left out. Finally, users also need to be able to delete furniture again, which adds another requirement to the following list.

> **R02 Instantiation:** The user needs to be able to instantiate pieces of furniture.
>
> **R03 Selection and Release:** The user needs to be able to select and release instantiated pieces of furniture.
>
> **R04 Positioning:** The user needs to be able to change the position of a selected piece of furniture by moving it along the floor of the room.
>
> **R05 Rotation:** The user needs to be able to rotate a selected piece of furniture around its yaw axis.
>
> **R06 Deletion:** The user needs to be able to delete a selected piece of furniture.

Another three requirements address the crucial factor of action integration and derive from the two different approaches and their combination as a *Hybrid Approach*. The study prototype needs to implement all three to enable their comparison.

R07 **Separation:** The study prototype needs to implement a mode that uses a *Separation of DOFs*.

R08 **Composition:** The study prototype needs to implement a mode that uses a *Composition of Users' Actions*.

R09 **Hybrid:** The study prototype needs to implement a mode that uses a combination of the other two approaches as a *Hybrid Approach*.

The last requirement covers the crucial factor of feedback, which is needed to support the collaborators in being aware of the actions of each other.

R10 **Awareness Cues:** The app must provide awareness cues that visualize the actions of the collaborators to each other.

4.2 Interaction Concepts for the Three Approaches

The scenario of furnishing a home with virtual pieces of furniture using a hand-held AR device required to develop specific concepts for the three approaches of collaborative 3D object manipulation. They were implemented as the modes *Separation*, *Composition*, and *Hybrid* into the study prototype. These interaction concepts are designed for two users. Due to the scenario, only the manipulation tasks positioning and rotation needed to be considered. Scaling is not required since the size of a virtual piece of furniture needs to match with the real-world counterpart. Besides, objects can only be moved along the floor. Consequently, the lifting or lowering of an object is not needed, which means only the two other DOFs of the task positioning are investigated. Out of the same reason, rotations are limited to the yaw axis. Hence, in total, 3 DOFs needed to be considered.

The interaction concepts for *Separation* divide the DOFs of the canonical tasks rotation and positioning between the users. While one user controls the position of the object, the other one is responsible for its rotation. Consequently, users can only perform different manipulation tasks simultaneously. Figure 4.1 illustrates this with an example: While the user colored in blue moves the object by two meters, the user in red rotates it simultaneously by 45 degrees.

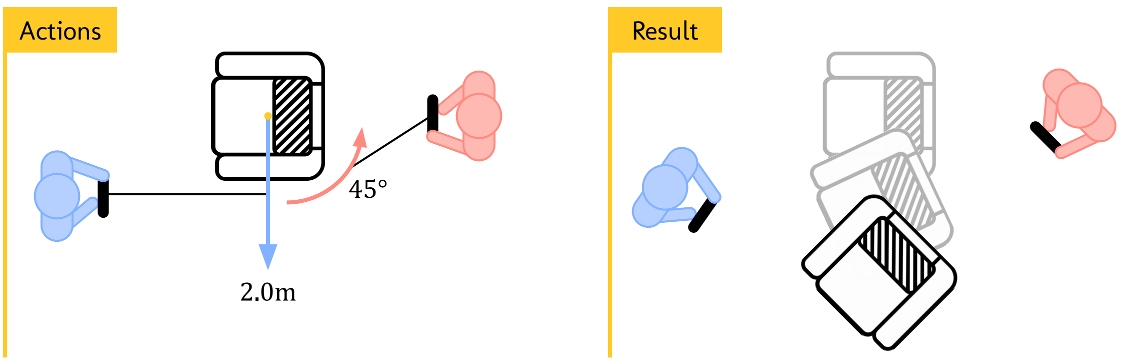


Figure 4.1: The concepts for the mode *Separation* of the study prototype divide the DOFs of the canonical tasks rotation and positioning between the users. That is, if one user rotates the object, the other one can only position it. *Taken from [31].*

For the mode *Composition*, the interaction concepts allow the users to perform only the same manipulation tasks simultaneously. Further, the computation of the sum is used as the merge policy. Figure 4.2 illustrates this with examples for simultaneous translations of an object. The examples cover three different cases: Actions towards the same direction reinforce each other. On the contrary, actions directed towards opposite directions equalize each other. The decision for the computation of the sum as the merge policy is reasoned with the third case. If users stand at the sides of the same corner of a virtual piece of furniture, they can focus on the alignment of the side at which they stand.

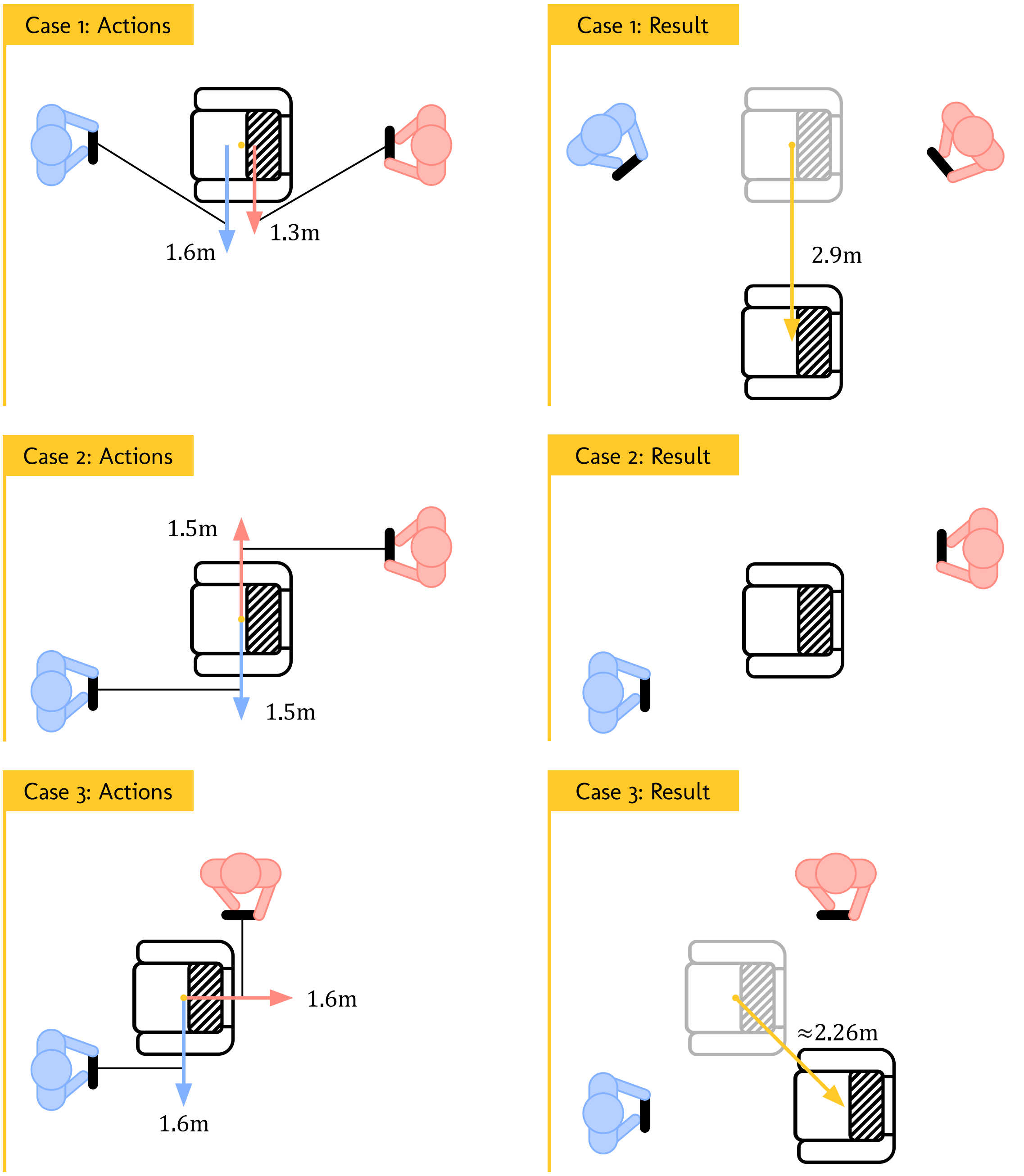


Figure 4.2: The concepts for the mode *Composition* of the study prototype use the computation of the sum as a merge policy. Therefore, actions towards the same direction reinforce each other *(Case 1)*, and actions towards opposite directions equalize each other *(Case 2)*. If users stand at the sides of the same corner, it allows them to focus only on the side at which they stand *(Case 3)*. *Taken from [31].*

Since rotations are only allowed around the yaw-axis of a virtual piece of furniture, only two cases can occur: Rotations in the same direction reinforce each other, and those towards opposite directions equalize each other. Figure 4.3 illustrates each case with an example: If one user rotates the object counterclockwise by 20 degrees and the other one in the same direction by 45 degrees, the resulting rotation adds up to 65 degrees. On the contrary, if one user rotates the object clockwise by 20 degrees and the other one counterclockwise by 45 degrees, the object is rotated counterclockwise by 25 degrees.

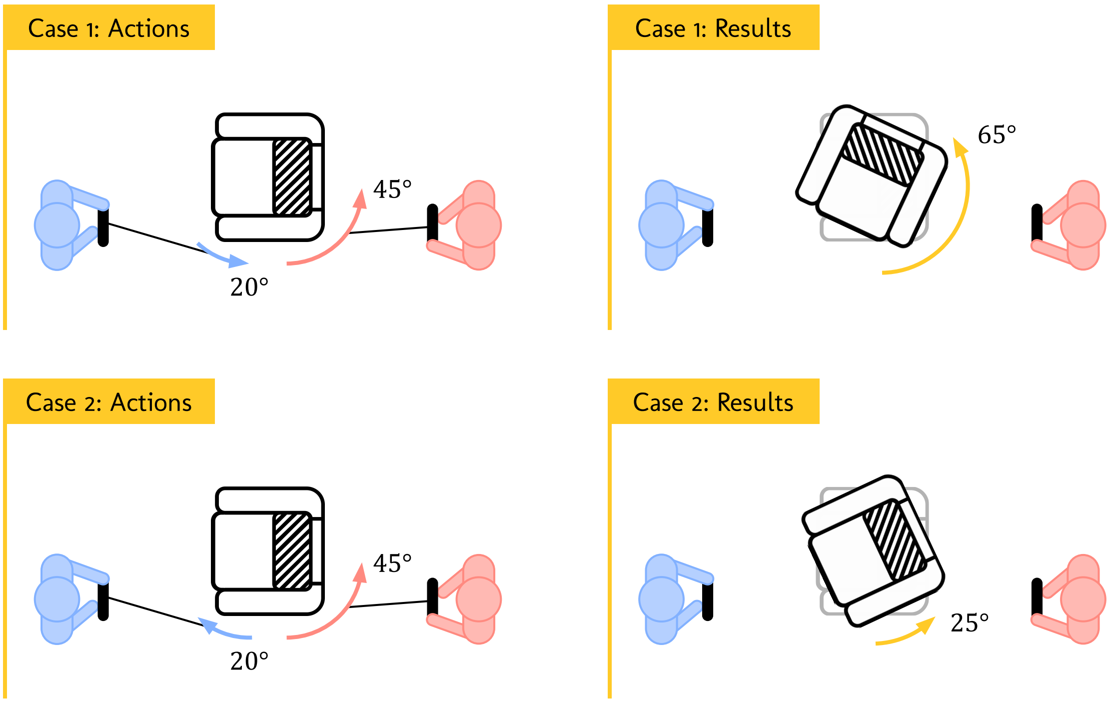


Figure 4.3: In terms of the canonical task rotation, only two cases can occur during the mode Composition: Rotations into the same direct reinforce each other *(Case 1)* while those towards opposite directions equalize each other *(Case 2)*.

For the mode *Hybrid*, the concepts for the other two approaches are combined, which allows the users to perform any manipulation action at any time. Therefore, they can decide on their own whether and when they split the canonical task between each other. Depending on the users' inputs, the study prototype dynamically switches between the modes *Composition* and *Separation*. Therefore, concurrent manipulations with the same canonical task are added up, whereas the manipulations with different canonical tasks are directly applied.

4.3 Implementation

The order of the requirements structures the description of the study prototype. It was built with the game engine Unity [32]. Since the application utilizes Apple's ARKit [33, 34] to realize the AR experience, it runs exclusively on iOS devices. The Master-Project Report [31] provides further technical details on the implementation.

4.3.1 Networking (R01)

As illustrated in Figure 4.4, the study prototype implements a classical client-server architecture to enable communication between the devices of the users. As soon as one user performs an action on his or her device that needs to be synchronized with the other user's devices, the application sends a message to the server. This message contains the necessary information to reproduce the action. The server processes the received message and forwards it to the other user's device, which then also executes the action. Besides, this client-server architecture allows controlling the devices from the server-side. For example, the server assigns a unique color (i. e., red or blue) to each device at the start of the application.

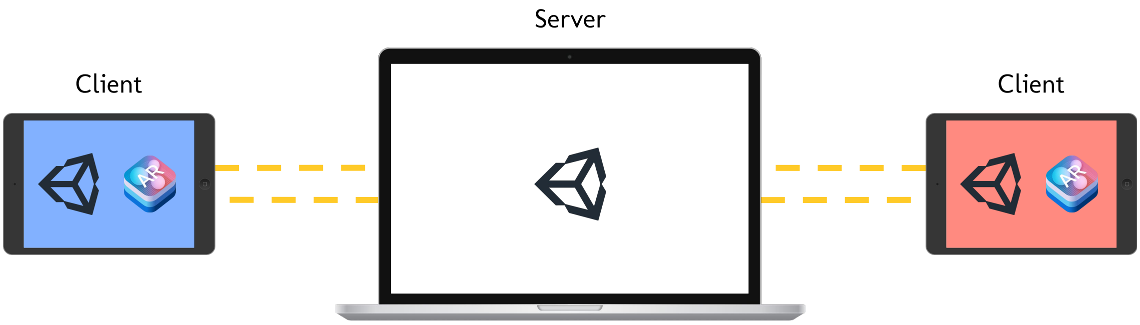


Figure 4.4: The study prototype uses a classic client-server architecture to enable communication and synchronization between the users' devices. *Taken from [31].*

4.3.2 Instantiation (R02)

The app provides a catalog that lists all available pieces of furniture. As shown in Figure 4.5 on the left side, the user needs to press the "+"-button to open it. The catalog then slides in from the bottom edge of the screen. It is vertically scrollable and divided into different categories of chairs, armchairs, tables, and sofas. Each category contains one or more tiles that show the name, a preview image, and the number of obtainable instances of each available piece of furniture. If there are more than three pieces of furniture inside a category, the tiles are horizontally scrollable. To instantiate a piece of furniture, the user needs to press on the respective tile, as shown on the right side of Figure 4.5. In succession, the catalog closes itself, and the piece of furniture appears on the floor two meters in front of the user. To close the catalog without instantiating a piece of furniture, a user needs to press on the "x"-button on the right side of the screen.

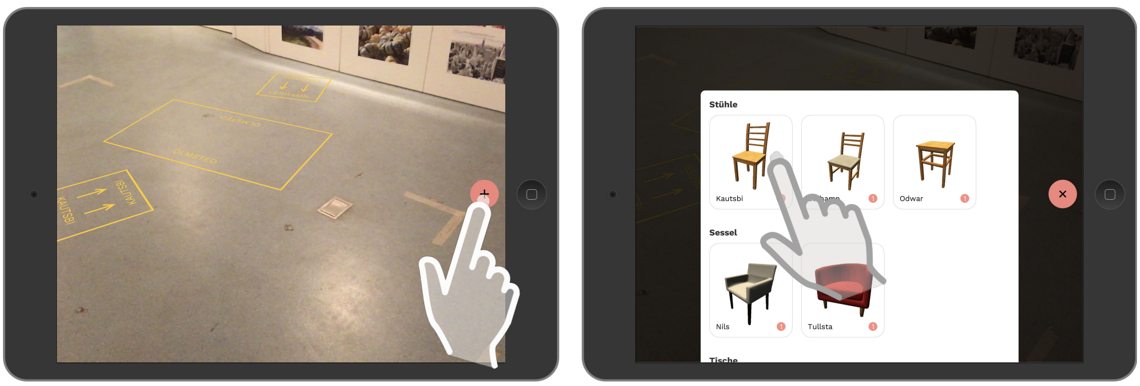


Figure 4.5: To instantiate a specific piece of furniture, a user needs to open the catalog *(left)* and tap on its tile *(right). The white hand outlines indicate where the user touches the screen. Taken from [31].*

4.3.3 Selection and Release (R03)

After its instantiation, a piece of furniture is automatically selected. To select an already instantiated piece of furniture manually, the user needs to tap on it (cf. Figure 4.6, left). The furniture then gets highlighted by a white plate that appears below it. When a user then taps on a different piece of furniture, the app deselects the current one since only one object can be selected at a time. To release a selected piece of furniture manually, the user can also tap elsewhere on the screen. The white plate then disappears again.

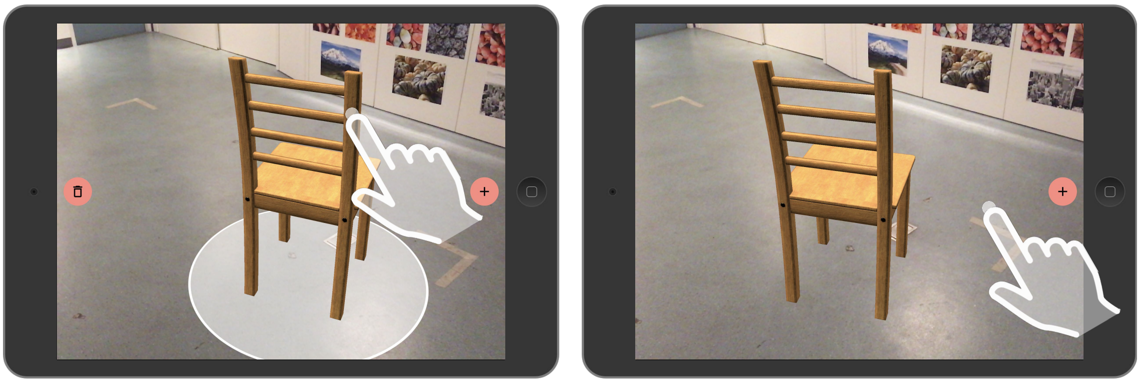


Figure 4.6: A user selects an instantiated piece of furniture by tapping on it *(left)*. Its release requires a tap elsewhere on the screen *(right)*. *The white hand outlines indicate where the user touches the screen. Taken from [31].*

4.3.4 Positioning (R04)

To move a piece of furniture along the floor, the user needs to touch and drag the white plate below the selected piece of furniture, as illustrated in Figure 4.7. During the action, the associated plate is tinted in the user's color. Further, an icon for positioning appears on the plate to give the user feedback about this action. Technically, ray-casting is used: An invisible ray is shot into the scene through the point where the user touches the screen. While the user moves his or her finger alongside the screen, the intersection point between this ray and the room's floor is calculated. To this intersection point, the piece of furniture is then moved every frame. As already explained, users can not raise or lower a piece of furniture.

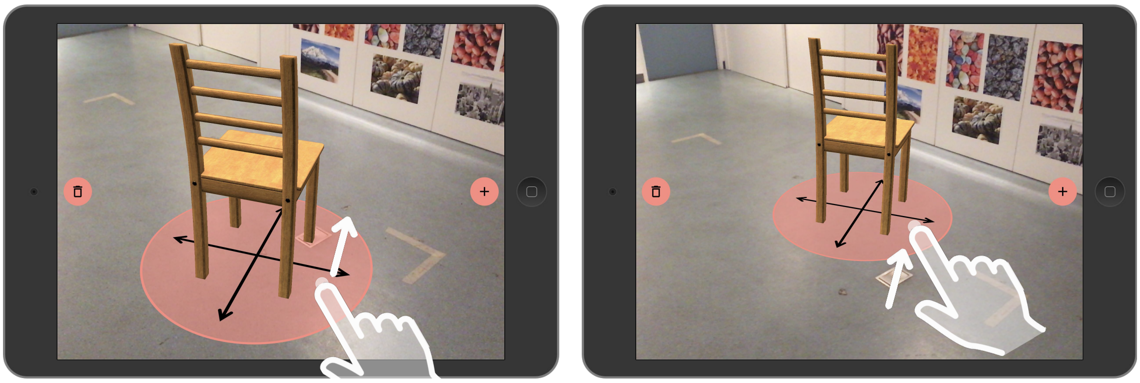


Figure 4.7: For positioning, users need to drag the plate below the selected piece of furniture. During the action, the plate is tinted with the user's color and equipped with an icon for positioning to provide feedback. *The white hand outlines were added to indicate where the user touches the screen. Taken from [31].*

4.3.5 Rotation (R05)

The rotation of a piece of furniture requires the user to perform the established gesture with two fingers, as shown in Figure 4.8 on the left side. Thereby, the user needs to touch the screen with both fingers and to rotate one finger around the other. The rotation angle of the user's fingers is then directly mapped to a change in the rotation of the selected piece of furniture. During rotation, the app again tints the plate below the piece of furniture in the user's color and displays an icon for rotation on it to provide the user with feedback. In contrast to the gesture for positioning this gesture can be performed anywhere on the screen (cf. Figure 4.8, right).

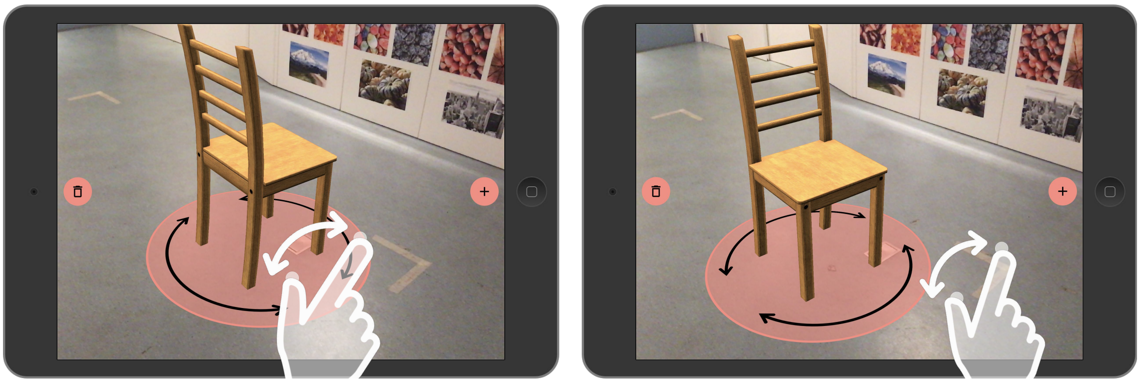


Figure 4.8: To rotate the selected piece of furniture, the user needs to perform the familiar gesture for rotation *(left)*. It can be executed anywhere on the screen *(right). The white hand outlines indicate where the user touches the screen. Taken from [31].*

4.3.6 Deletion (R06)

A user deletes the selected piece of furniture by tapping on the "delete"-button on the left side of the screen (cf. Figure 4.9, left). The button is only visible if an object is selected. Hence, a deletion always suspends the current selection and hides the "delete"-button (cf. Figure 4.9, right). Besides, a deleted piece of furniture is added back to the catalog.

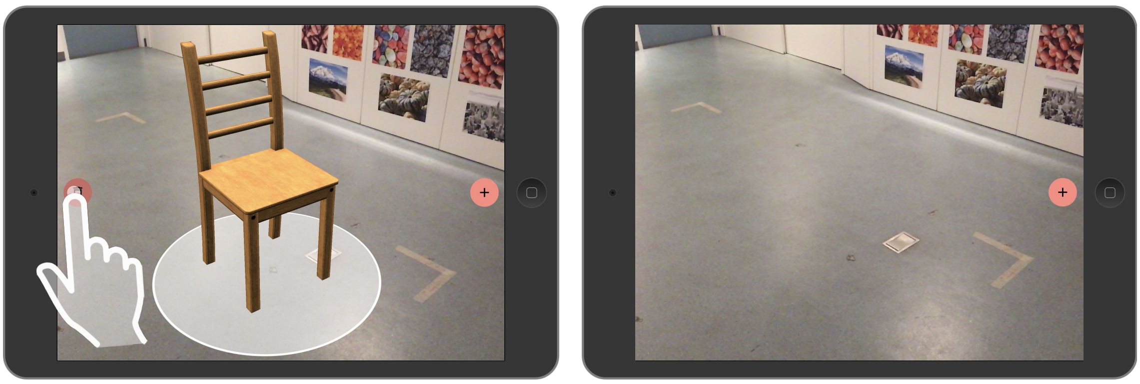


Figure 4.9: A tap on the "delete"-button deletes the selected piece of furniture *(left)*. The button is only visible if a piece of furniture is currently selected *(right). The white hand outlines indicate where the user touches the screen. Taken from [31].*

4.3.7 Separation (R07)

The mode *Separation* of the app allows users to perform mutually exclusive manipulations, e.g., while one user is rotating the object, the other one can only move it and vice versa (cf. Figure 4.10). Thereby, the first-come, first-served (FCFS) principle takes effect. The user who starts manipulating the object first determines what the other user is allowed to do. To minimize confusion, manipulation of the first user provokes the appearance of the icon for the respective other manipulation on the plate of the second user. That is, for example, if one user is rotating the object, the other user will see the icon for positioning on his or her plate. Additionally, arrows around the plate indicate the direction towards which the actions of the first user are directed. As soon as the second user starts manipulating the object, those arrows also appear around the first user's plate.

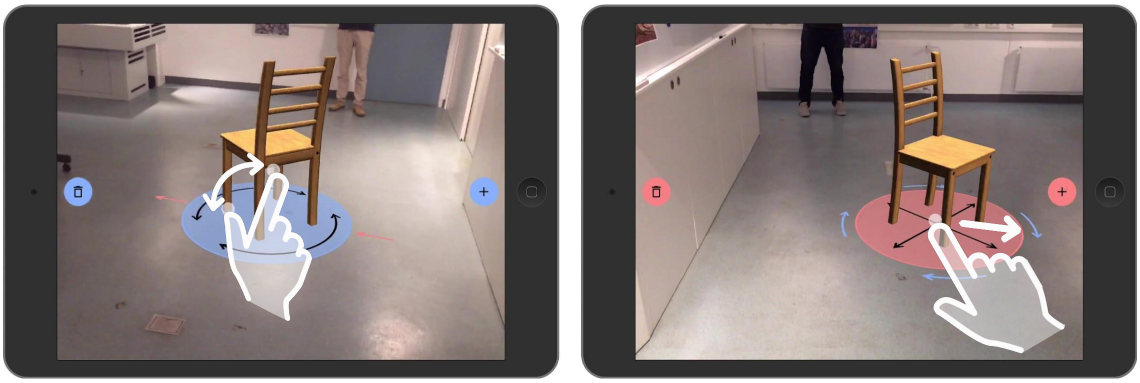


Figure 4.10: During the mode *Separation*, users can only perform different actions at the same time. That is, if one user rotates the object *(left)*, the other one can only move it *(right)* and vice versa. *The white hand outlines indicate where the users touch the screen. Taken from [31].*

4.3.8 Composition (R08)

During *Composition*, both users are required to do the same action simultaneously. Thereby, the FCFS principle is employed again. The user who starts manipulating the object determines what the other user is allowed to do. If the first user starts rotating the piece of furniture, the second user can also only rotate it. The same applies to positioning. Since the mode *Composition* sums up the inputs of the users, actions towards the same direction enforce each other. An object that is, for example, rotated by both users in the same direction, rotates accordingly faster (cf. Figure 4.11).

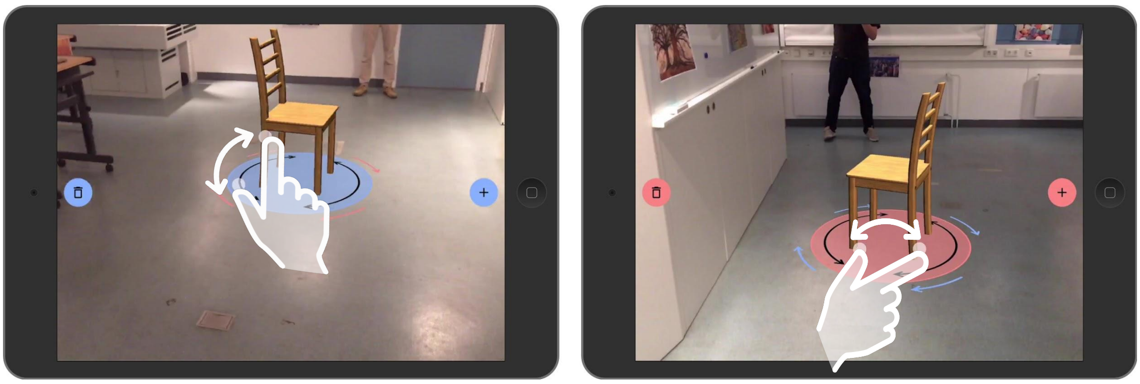


Figure 4.11: When the app is in the mode *Composition*, rotations in the same direction reinforce each other, e. g., if both users are rotating the object in the same direction, it rotates faster. *Taken from [31].*

If the translations of the users are directed towards different or opposite directions, a problem occurs: One user can move the plate below the piece of furniture away from the other user's finger. Consequently, the one-to-one mapping between finger and plate is suspended. As a solution, the concept introduces a proxy. A smaller plate appears, that always stays underneath the user's finger (cf. Figure 4.12). Inspired by the *Bent Pick Ray* of Riege et al. [27] (cf. Chapter 3, p. 20) it is linked to the bigger plate by a line that dynamically changes its length. When the user ends moving the object by lifting his or her finger, the small plate disappears again. Apart from that, the same awareness cues as during *Separation* are used to provide the users with feedback. Own manipulations are indicated by the icon on the plate that is tinted in the user's color during manipulation. The arrows around the plate visualize the other user's actions.

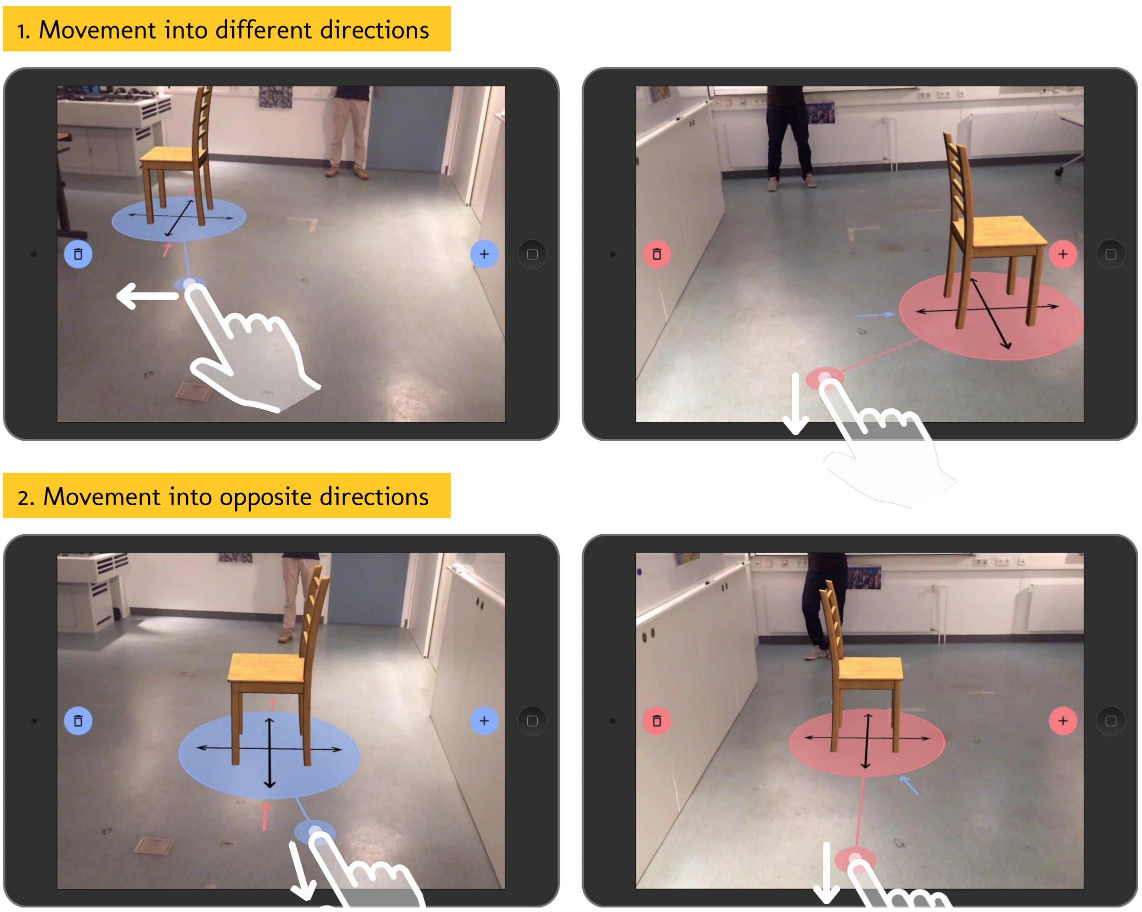


Figure 4.12: Two examples for positioning during *Composition*: Users performing actions towards different directions *(1)*, and towards opposite directions *(2). The white hand outlines indicate where the users touch the screen. Taken from [31].*

4.3.9 Hybrid (R09)

The mode *Hybrid* of the study prototype enables both users to rotate and move the selected piece of furniture at any time. From a technical point of view, the app dynamically switches between *Separation* and *Composition* depending on the users' inputs. For example, if both users are moving the selected piece of furniture, *Composition* is used. If one of the users then starts rotating it, the app seamlessly switches to *Separation*.

4.3.10 Awareness Cues (R10)

The explanations of the implementation for the other requirements already touched upon the different awareness cues that the app provides. All of them appear either directly on or next to the plate below the selected piece of furniture. The three different cues are the color of the plate, the icon it shows, and the arrows around it. Figure 4.13 provides an overview of all possible combinations of the different awareness cues for the modes *Composition* and *Separation*. A plain white plate indicates that the piece of furniture is only selected and that none of the users is manipulating it. If user A starts manipulating the piece of furniture, the plate is tinted in his or her color. On the plate of user B, either the icon for rotation or positioning appears – dependent on the approach. The icon signifies the action user B currently can perform. Besides, arrows around user B's plate indicate the direction towards which the action of user A is directed. When user B also starts manipulating the piece of furniture, his manipulations are reflected with the same awareness cues. His or her plate is colored, and the arrows appear around user A's plate. During the mode *Hybrid*, all variants illustrated in Figure 4.13 can occur – with one difference. Since users are unrestricted in their actions, the white plate never shows an icon in this mode.

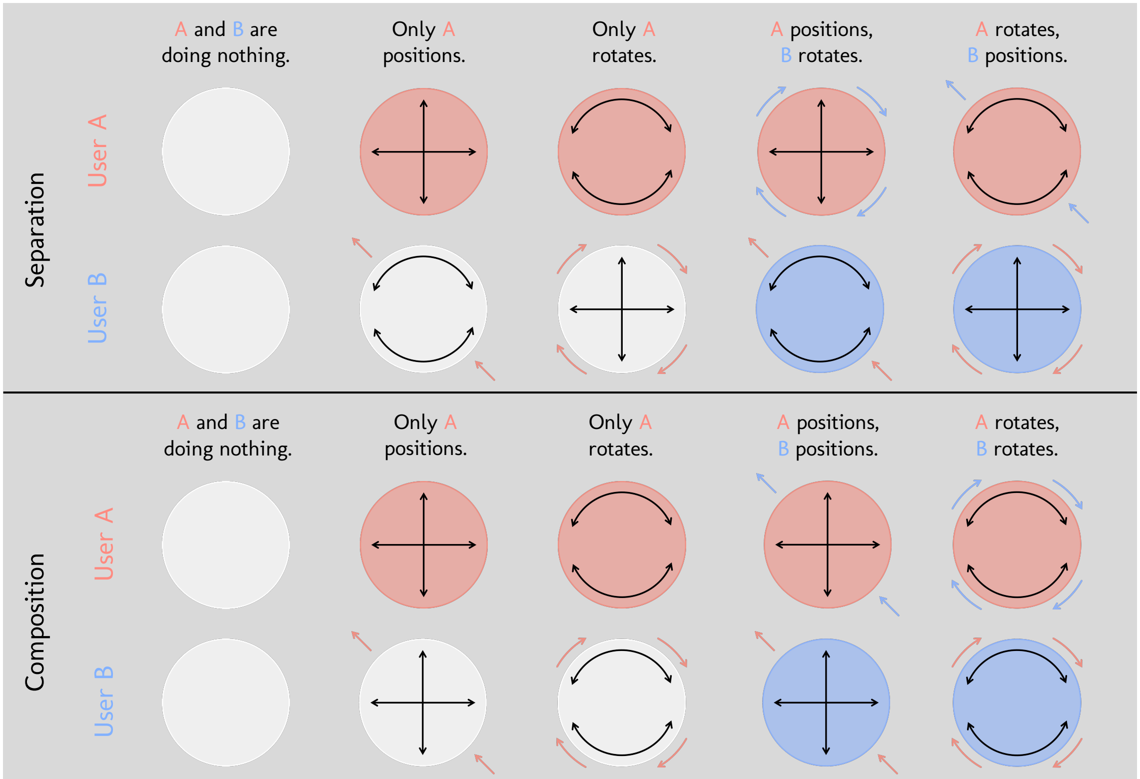


Figure 4.13: The different awareness cues the study prototype provides concentrate on the plate below a selected piece of furniture. The color of the plate indicates if one is performing an action. The icon on the plate shows either the action one does or one can perform. Arrows around the plate signify the directions of the other user's actions. *Taken from [31].*

5

Experimental Comparison

The study prototype described in the previous chapter was used in a controlled lab study to compare the three implemented approaches for action integration with each other (i. e., *Separation*, *Composition*, and *Hybrid*). This chapter describes this experimental comparison first. Afterward, it reports the results and their analysis.

5.1 Study Design

The study was conducted as a controlled lab experiment. Action integration represented the counterbalanced within-subjects factor with the three approaches *Separation*, *Composition*, and *Hybrid* being the conditions. To assess the question from the introduction of which approach should be used for collaborative furnishing on hand-held devices, the study was designed to answer the following three more specific research questions:

RQ1 Performance: How do the approaches differ in accuracy and task completion time?

RQ2 Workload: How do the approaches differ in subjectively perceived workload?

RQ3 User Experience: Which approach do users prefer?

5.1.1 Participants

Forty-eight participants (28 female, 20 male) between 18 and 40 years ($M = 22.96$, $SD = 4.15$) took part in the study. Forty-five of them were students (teacher trainee ($n = 9$), biology ($n = 6$), political economy ($n = 6$), computer science ($n = 4$), economics ($n = 3$), psychology ($n = 3$), chemistry ($n = 2$), history ($n = 2$), literature-art-media ($n = 2$), linguistics ($n = 2$), mathematics ($n = 2$), law ($n = 1$), literature ($n = 1$), philosophy ($n = 1$), not specified ($n = 1$)) and three employed persons (biological technical assistant, industrial engineer, nursery school teacher). The participants formed 24 dyads, of which 9 consisted of two females, 5 of two males, and 10 of both one male and one female participant. All of them already knew their partner before the experiment. Only 4 of them had made experiences with AR in advance. Twenty-six participants previously furnished a real home together with another person. Further, participants were asked to rate their tablet experience ($M = 3.98$) and their gaming experience ($M = 2.40$) on a scale from 1 (very inexperienced) to 5 (very experienced).

5.1.2 Task

For all three conditions, each dyad had to collaboratively furnish the same room with the seven virtual pieces of furniture shown on the right side of Figure 5.1. Thereby, virtual floor markings predefined their positions and rotations (cf. Figure 5.1, left) of the furniture. Inside each floor marking, participants found the name of the piece of furniture they had to instantiate from the catalog first and to place there afterward. For the pieces of furniture with a perceptible front side, arrows inside the floor marking determined to which side participants had to orient it. For those without a perceptible front side, different orientations were possible: Due to their symmetric characteristics, the table "Olmsted" could be placed with two different orientations inside the marker (i. e., turned by 180 degrees), and the stool "Odwar" even with four (i. e., turned by 90 degrees). During the task, collaboration was guided by allowing the participants to select and manipulate only the same piece of furniture at a time.

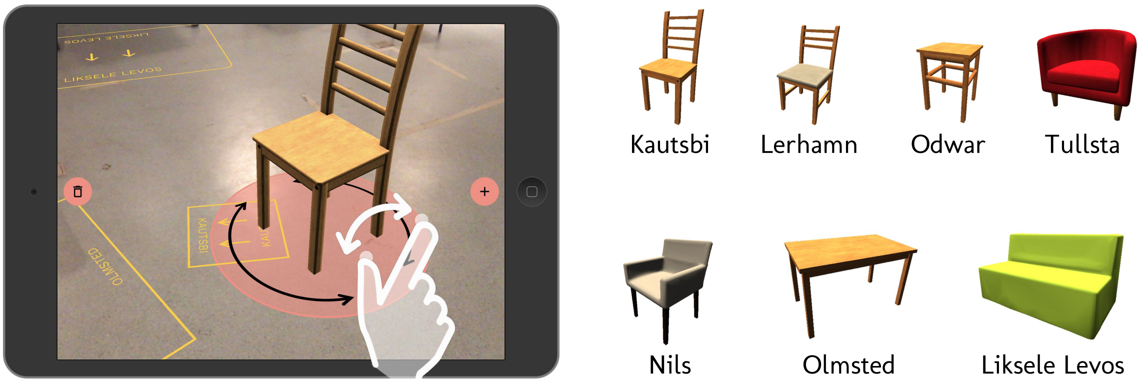


Figure 5.1: Dyads had to furnish a room with seven virtual pieces of furniture collaboratively *(right)*. Yellow floor markings predefined their positions and rotations *(left)*. For pieces of furniture with a perceptible front side, arrows inside the marking indicated the orientation. *Taken from [31].*

During a training phase before each task, dyads had time to get used to the individual approach. Thereby, the catalog listed only two virtual card boxes of different sizes, which dyads had to instantiate and align to fitted yellow floor markings (cf. Figure 5.2). For the bigger card box, its required orientation within the marker was designated by arrows on top of it. Those arrows had to match with those inside the floor marking. The other, smaller box required no specific orientation, which means participants had two possibilities to orient it inside the marking (i. e., turned by 180 degrees).

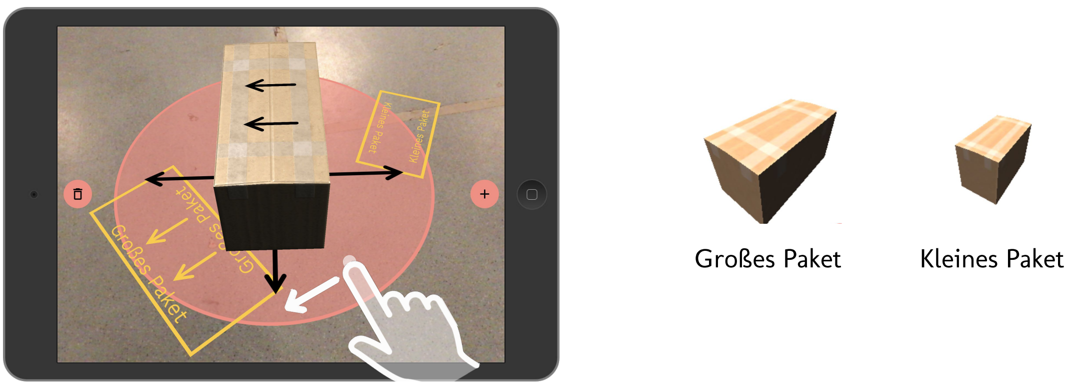


Figure 5.2: A training phase preceded each task during which dyads had to align two cardboard boxes to fitted yellow floor markings. *Taken from [31].*

5.1.3 Dependent Variables and Operationalization

Accuracy and task completion time were measured to operationalize the *Performance (RQ1)* of the three approaches. For accuracy, the error between every piece of furniture and its associated floor marking was measured separately for position and rotation. For the position errors, the euclidean distances between the actual positions and the target positions of the pieces of furniture were calculated. The smallest angle difference between the rotation of marking and piece of furniture determined the rotation error. The raw, unweighted NASA Task Load Index (NASA TLX) [35] was used to measure subject *Workload (RQ2)*. *User Experience (RQ3)* was assessed via the User Experience Questionnaire (UEQ) [36]. Additionally, a semi-structured interview was conducted with each dyad to gain further insights. Participants were asked which condition they preferred during the task and which they would want to use for furnishing their place without the predefined positions of the furniture. Also, they were requested to mention the advantages and disadvantages of each approach.

5.1.4 Procedure

The study followed the procedure shown in Figure 5.3, which was revised by a pilot study with two dyads. After been welcomed (cf. Appendix B, participants filled out a demographic questionnaire first (cf. Appendix D). Then the investigator explained the study task and the study prototype with a short presentation. This presentation consisted of step-by-step instructions using images and videos showing the cardboard boxes participants later used in the training phase (cf. Appendix E). Further, participants were asked to solve the task as quickly and as accurately as possible. Then, the approach of the first condition was explained to the participants in a second presentation (cf. Appendix F). Afterward, they started with the training task for this condition. During the training, they could familiarize themselves with the control concept of the application and the particular approach of the condition. Then, two passes of the actual task followed during which accuracy and task completion time were measured. To rate the condition, participants filled out the NASA TLX (cf. Appendix G) and the UEQ (cf. Appendix H) afterward. The whole procedure was then repeated for the two other conditions, starting with the presentation of the approach. The complete session took approximately one hour and ended with the concluding, semi-structured interview (cf. Appendix I). Participants received compensation for their time.

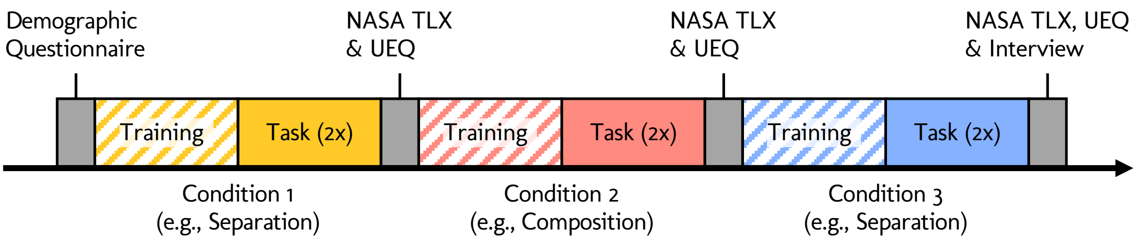


Figure 5.3: A session started with a demographic questionnaire. For each condition, dyads completed a training phase first. Then, they performed the task twice. Participants assessed each condition by filling out the NASA TLX and the UEQ. Each session ended with a concluding interview. Conditions were counterbalanced. *Taken from [31].*

5.1.5 Apparatus

Each participant of a dyad was provided with an iPad Pro that has a mass of 437 grams. The device's screen has a size of 9.7 inches and a resolution of 2048 × 1536 pixels. Consequently, it can display 264 pixels per inch [37]. For creating the AR experience, the iPad's rear-facing camera with a resolution of 12 megapixels was utilized. The server ran on a high-end desktop PC with Windows 10 installed. For network communications, the aim was to create optimal conditions: A dedicated router was used, which was not connected to the internet during the study. The router created a wireless network using the 5 GHz bands to which the iPads were connected. The connection between the server and the router relied on an ethernet cable. No other devices were connected to the router during the study.

As shown in Figure 5.4, participants used the iPads not only to fulfill the tasks but also to answer the questionnaires. Therefore, all three questionnaires were designed and implemented on the tablets as part of the study prototype, which had various advantages: First, control mechanisms ensured that participants answered all questions. Secondly, problems of illegible handwriting could not occur. And thirdly, it expedited the evaluation of the questionnaires since they did not have to be digitized. After answering a questionnaire, participants hit the sent button, and their replies were directly sent to the server, which stored them inside a file. Therefore, the investigator was not even required to transfer them manually from the iPads to a PC. The same holds for the interaction logs. Since the iPads communicated their positions, rotations, and all the users' interactions with and over the server, interaction logging could be performed on it. For each dyad, the server automatically created a log file and wrote all received and sent messages to this file.

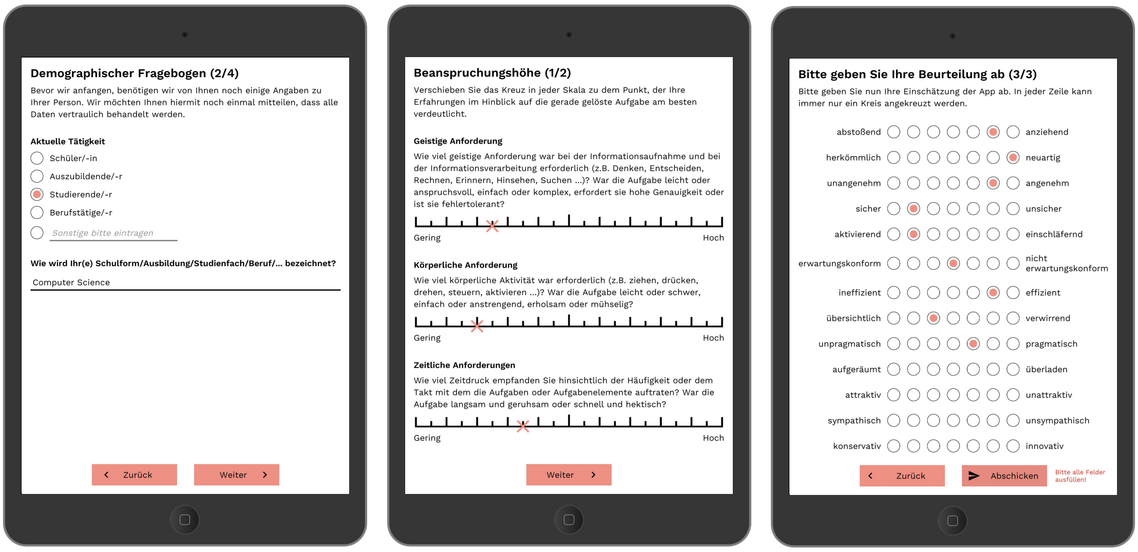


Figure 5.4: Participants answered the different questionnaires directly on the iPads. Therefore, they were specially built and directly integrated into the study prototype. Parts of the demographic questionnaire *(left)*, the NASA TLX *(middle)* and the UEQ *(right)* are shown. *Taken from [31].*

For the study, the server was equipped with the visual interface shown in Figure 5.5. It supported the investigator in conducting the study. The area bordered in blue in Figure 5.5 provided instruments to monitor the study. The "Clients"-panel in the upper left corner listed all clients logged-in to the server, showed their IP address, and their battery level. The "Statistics"-panel below showed the count of all messages that the server received and sent so far. The "Live View"-panel in the middle provided a top-down view on the study setting that visualized the current position of the participants and the already instantiated pieces of furniture. A red frame borders the panels in Figure 5.5 that allowed to set up and control the study. The "Settings"-panel in the upper right corner allowed to disable logging and enable different features for debugging. The "World Map Setup"-below was used to align the floor markings inside the physical room. The sequence view in the bottom enabled the investigator to control the procedure of the experiment. At the beginning of each session, the investigator was required to enter the group number of the current dyad. The sequence of tasks and questionnaires then automatically rearranged itself respecting counterbalancing. During the study, the investigator only needed to click through this sequence to switch between the conditions in the right order and issue the various questionnaires to the participants' iPads.

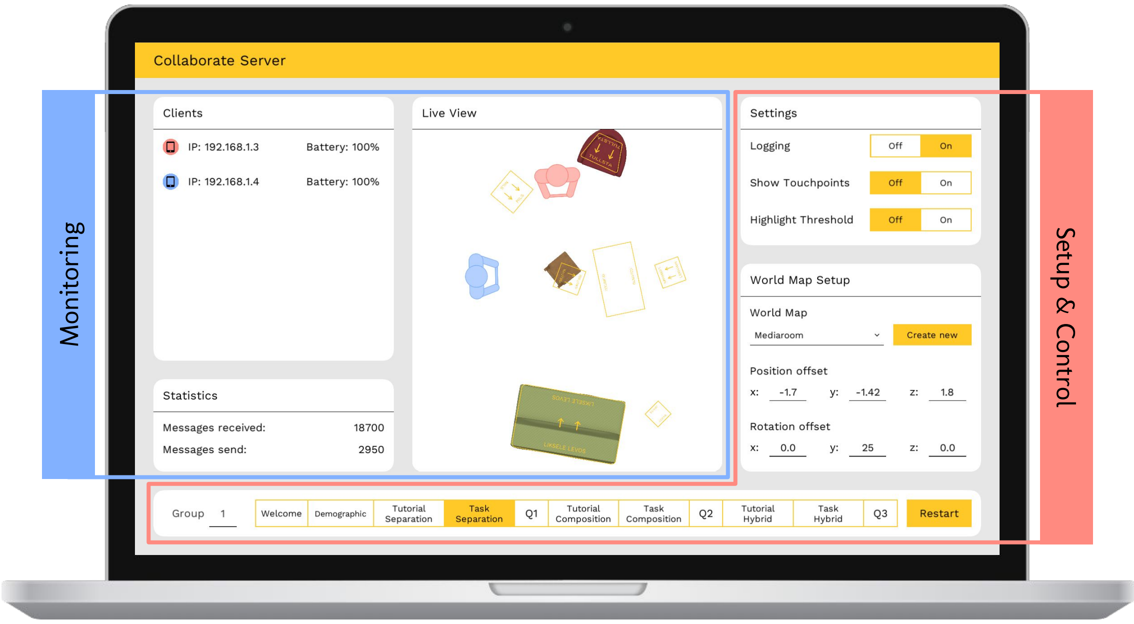


Figure 5.5: The server was equipped with a visual interface to ease controlling, setting up and monitoring the study. *Taken from [31].*

The room in which the study took place is shown in Figure 5.6. Its dimensions were $5.70 \times 7.06 \times 2.93$ meters with a walkable area of approximately 4.0×7.0 meters. Pretests revealed that ARKit required more visual features for stable tracking than the room itself provided. Consequently, the room was equipped with printed photographs to solve this issue (cf. Figure 5.6, right). The study investigator used the large display for the explanatory presentations before each condition (cf. Figure 5.6, left). At the table in front of this display, participants filled out the questionnaires. They also sat there during the concluding interview. During the tasks and the interview, the investigator sat on the smaller table in the left.

Figure 5.6: The room in which the study took place. At the table in front of the large display participants filled out the questionnaires. The investigator of the study sat on the smaller table to their left *(left)*. One wall was equipped with various pictures to improve tracking quality *(right)*.

5.2 Results

Following the order of the research questions, this section presents the results of the study. To indicate the approach for action integration, subscript $_S$ is used for *Separation*, subscript $_C$ for *Composition*, and subscript $_H$ for *Hybrid*. The Shapiro-Wilk test was used to test if the results for the different measures were normally distributed. As the data violated the assumption of a normal distribution, a non-parametric approach was used to analyze it. Further, medians (*Mdn*) are reported. To detect differences between the three conditions for each pass, Friedman's ANOVA was used. If this test showed differences, post hoc analysis was performed with Dunn-Bonferroni-Tests. The Wilcoxon signed-rank test was used for comparing the passes within each condition (i.e., pass 1 versus pass 2). Generally, an alpha level of 0.05 was assumed for all statistical tests. For the pairwise comparisons, the Bonferroni correction required to adjust the alpha level to $(.05/3) \approx .016$.

5.2.1 Performance (RQ1)

To assess the performance of the three approaches, task completion times and accuracy were measured. During *Hybrid*, participants could seamlessly switch between *Separation* and *Composition*. Therefore, also details about the percentage use of these two conditions during *Hybrid* are provided.

Task Completion Times

The task completion times for all three conditions and both passes are shown in Figure 5.7. For the first pass, there was a significant difference between the three conditions ($\chi^2(2) = 12.583$, $p < .05$). Task completion times for the conditions *Separation* ($Mdn_S = 153.22s$, $z = -3.320$, $p < .016$) and *Hybrid* ($Mdn_H = 184.96s$, $z = 2.742$, $p < .016$) were significantly lower than for the condition *Composition* ($Mdn_C = 220.38s$). For the second pass, task completion times were also significantly different ($\chi^2(2) = 8.583$, $p =< .05$). Here, participants were significantly faster with the condition *Separation* ($Mdn_S = 141.45s$, $z = -2.887$, $p < .016$) compared to the condition *Composition* ($Mdn_C = 174.02s$). Comparing the two passes of each condition with each other, it showed that for the conditions *Separation* ($z = -2.829$, $p < .05$) and *Composition* ($z = -2.800$, $p < .05$) participants were significantly faster in the second pass.

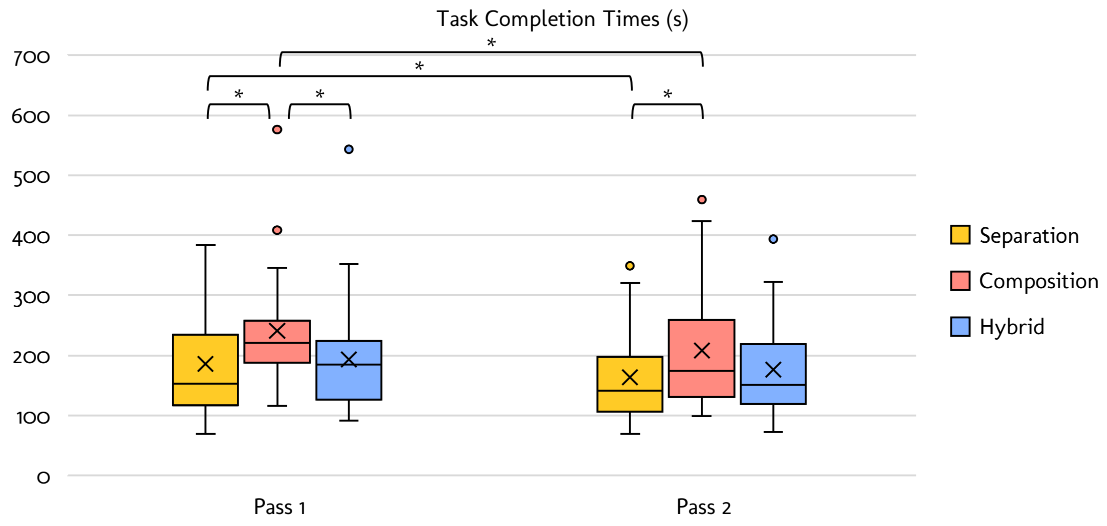


Figure 5.7: Box plots of the task completion times for all three conditions and both passes. Within each box, the line indicates the median value and the cross the mean value.

Use of Separation and Composition during Hybrid

As participants could dynamically switch between *Separation* and *Composition* during the condition *Hybrid*, Figure 5.8 presents the quantification of the relative duration of each approach for action integration. The mean percentage values relate to the simultaneous manipulation times during task completion for each pass. This simultaneous manipulation time excludes individual interactions and periods during which participants were not manipulating the objects (e. g., coordination, selection of task objects, general communication). In pass 1 of *Hybrid*, participants averagely used *Separation* 48.63% and *Composition* 51.37% of the simultaneous manipulation time. During pass 2 of *Hybrid*, they averagely used *Separation* 53.46% and *Composition* 46.54% of the simultaneous manipulation time.

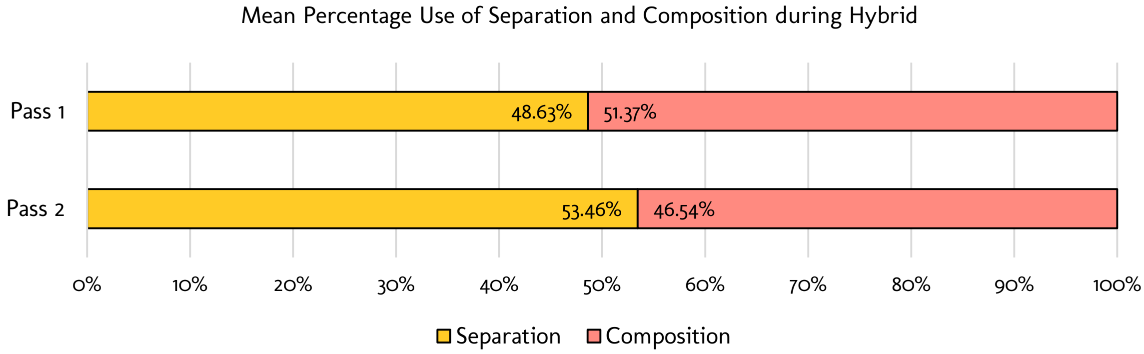


Figure 5.8: The mean percentage usages of *Separation* and *Composition* during *Hybrid* for both passes. Percentage values refer to the simultaneous manipulation time which excludes individual interactions and periods during which participants were not manipulating an object.

Figure 5.9 additionally shows the percentage increase in the use of either *Separation* or *Composition* during *Hybrid* between passes 1 and 2. Again, the percentage values refer to the simultaneous manipulation time, which excludes individual interactions and periods during which participants were not manipulating objects. Fourteen groups increased their percentage use of *Separation* and the other ten groups their percentage use of *Composition* in pass 2 compared to pass 1.

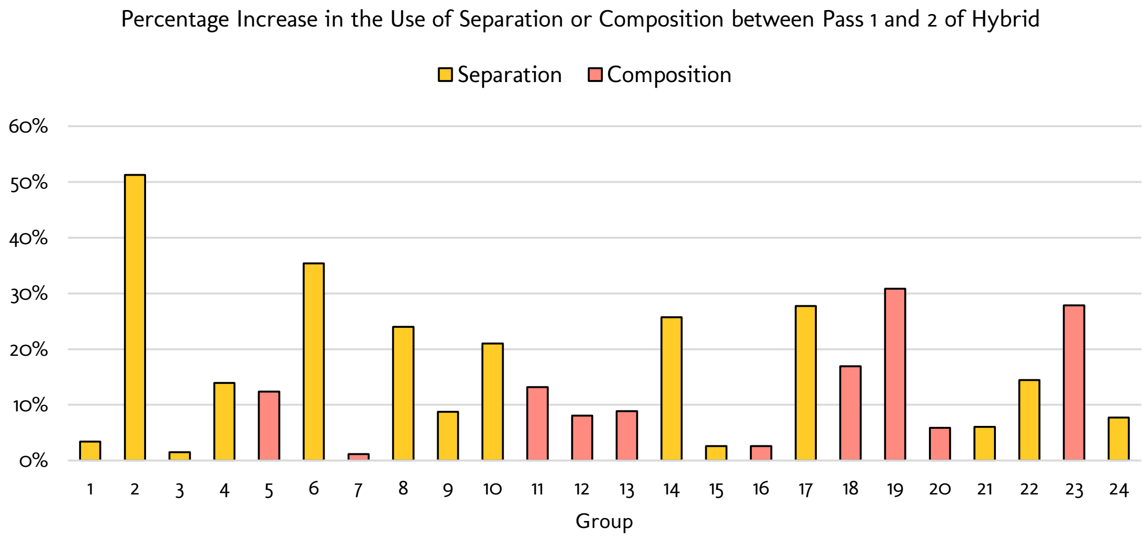


Figure 5.9: The percentage increase in the use of either *Separation* or *Composition* between pass 1 and pass 2 of *Hybrid* for each group. Percentage values refer to the simultaneous manipulation time which excludes individual interactions and periods during which participants were not manipulating an object.

Accuracy

Figure 5.10 shows the position errors in millimeters for all three conditions and both passes. The analysis showed that the differences for both the first pass ($\chi^2(2) = 3.083$, $p > .05$) and the second pass ($\chi^2(2) = 5.083$, $p > .05$) were not statistically significant. However, a trend for the benefit of the condition *Composition* is recognizable which has the lowest median values for both passes (Pass 1: $Mdn_S = 14.8mm$, $Mdn_C = 10.97mm$, $Mdn_H = 14.90mm$; Pass 2: $Mdn_S = 13.90mm$, $Mdn_C = 12.93mm$, $Mdn_H = 13.78mm$). The comparison of the passes within each condition also showed no statistical differences.

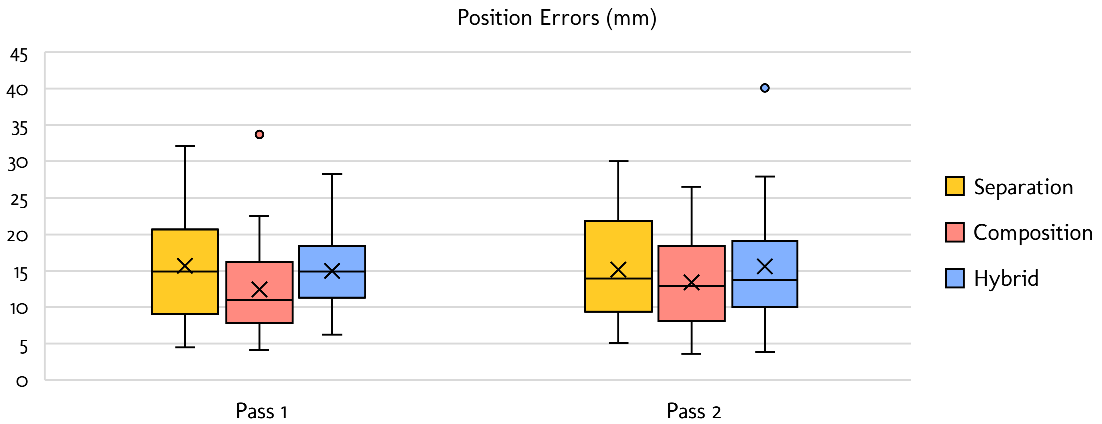


Figure 5.10: Box plot of the position error in millimeters for all three conditions and both passes. Within each box, the line indicates the median value and the cross the mean value.

The rotation errors in degrees for all three conditions and both passes are depicted in Figure 5.11. As for the position errors, no significant differences could be revealed – neither for the first pass ($\chi^2(2) = 1.750$, $p > .05$) nor for the second pass ($\chi^2(2) = 0.250$, $p > .05$). Also, the comparison of the passes within each condition showed no statistical differences. But again, the median values of the condition *Composition* are the lowest for both passes (Pass 1: $Mdn_S = 0.67°$, $Mdn_C = 0.57°$, $Mdn_H = 0.63°$; Pass 2: $Mdn_S = 0.64°$, $Mdn_C = 0.64°$, $Mdn_H = 0.69°$), which reasserts the trend that was also visible for the positioning errors.

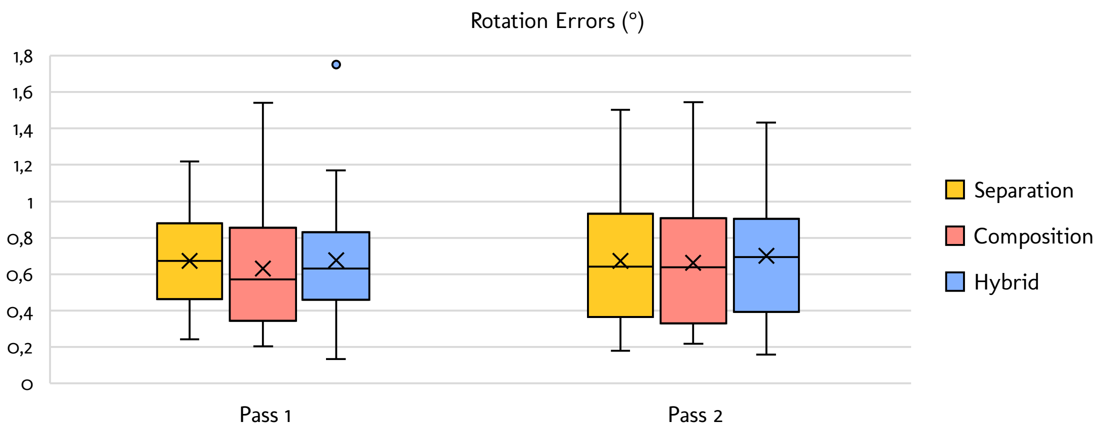


Figure 5.11: Box plots of the rotation errors in degrees for all three conditions and both passes. Within each box, the line indicates the median value and the cross the mean value.

5.2.2 Work Load (RQ2)

Figure 5.12 shows the box plots for the mean overall scores of the NASA TLX questionnaire for each condition. Analysis revealed significant differences between the three conditions ($\chi^2(2) = 9.937$, $p < .05$). Overall scores for the condition *Separation* were significantly better than those for the condition *Composition* ($Mdn_S = 37.08$, $Mdn_C = 45.42$, $z = -3.113$, $p < .016$).

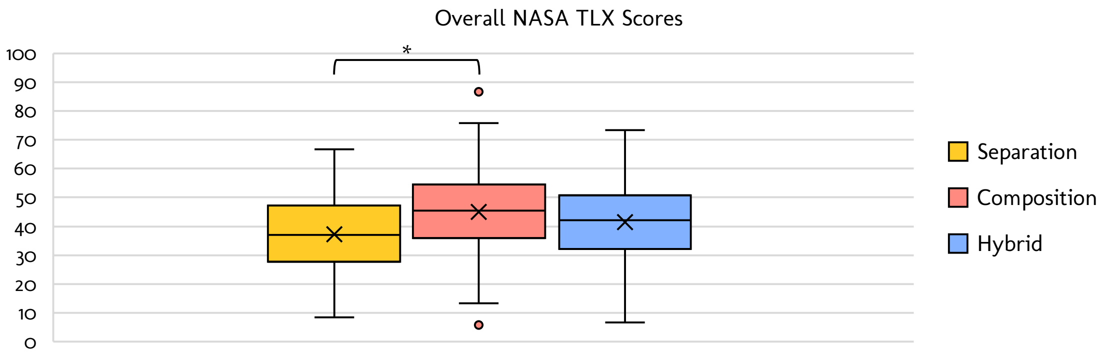


Figure 5.12: Box plots of the the overall NASA TLX scores for all three conditions and both passes. Within each box, the line indicates the median value and the cross the mean value.

Figure 5.13 shows the breakdown of the overall scores into the different dimensions of the NASA TLX questionnaire. Analysis revealed significant effects for the dimensions *Physical Demand* ($\chi^2(2) = 9.337$, $p < .05$), *Performance* ($\chi^2(2) = 10.683$, $p < .05$), *Effort* ($\chi^2(2) = 6.356$, $p < .05$), and *Frustration* ($\chi^2(2) = 8.503$, $p < .05$). The dimensions *Mental Demand* ($\chi^2(2) = 2.958$, $p > .05$) and *Temporal Demand* ($\chi^2(2) = 5.035$, $p > .05$) showed no significant differences. Post hoc analysis revealed, that the condition *Separation* was significantly ranked better than the condition *Composition* for the dimensions *Physical Demand* ($Mdn_S = 35$, $Mdn_C = 42.5$, $z = -2.909$, $p < .016$), *Performance* ($Mdn_S = 20$, $Mdn_C = 25$, $z = -2.858$, $p < .016$), *Effort* ($Mdn_S = 40$, $Mdn_C = 52.5$, $z = -2.398$, $p < .016$) and *Frustration* ($Mdn_S = 22.5$, $Mdn_C = 37.5$, $z = -2.501$, $p < .016$). The other pairwise comparisons showed no significant results.

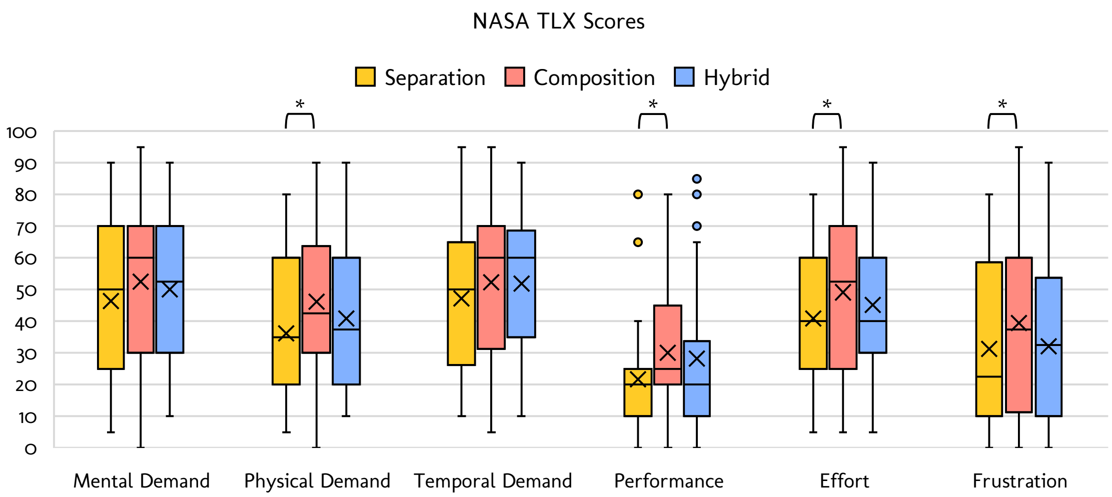


Figure 5.13: Box plots of the the NASA TLX scores broken down into the six dimensions for all three conditions and both passes. Within each box, the line indicates the median value and the cross the mean value.

5.2.3 User Experience (RQ3)

User experience was assessed via the User Experience Questionnaire (UEQ) and a semi-structured interview. At the beginning of the interview participants were requested to rank the three approaches according to their preference and the amount of coordination they required. Afterward, they were asked to reason their ranking and name advantages and disadvantages of the three approaches.

User Experience Questionnaire (UEQ)

The responses to the UEQ are visualized in Figure 5.14. Analysis revealed differences for the dimensions *Attractiveness* ($\chi^2(2) = 22.613$, $p < .05$), *Perspicuity* ($\chi^2(2) = 16.926$, $p < .05$), *Efficiency* ($\chi^2(2) = 14.619$, $p < .05$), *Dependability* ($\chi^2(2) = 17.190$, $p < .05$), and *Stimulation* ($\chi^2(2) = 7.560$, $p < .05$). No difference could be revealed for the dimension *Novelty* ($\chi^2(2) = 3.574$, $p > .05$).

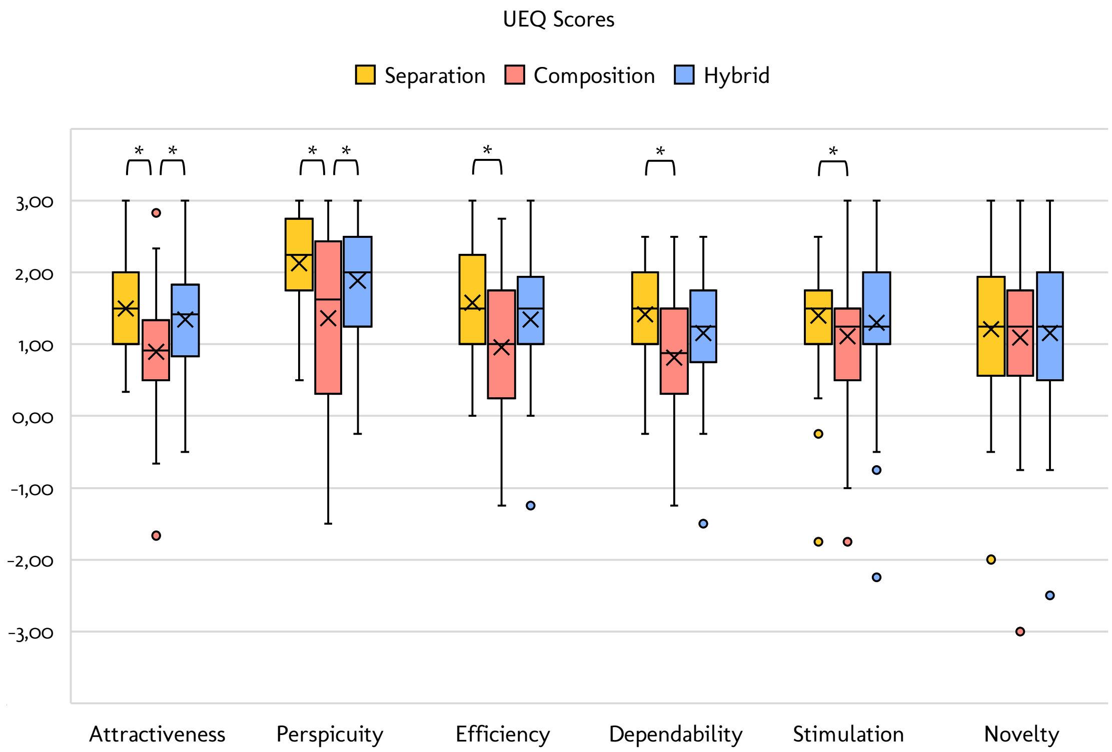


Figure 5.14: Box plots of the UEQ scores for all three conditions and both passes. Within each box, the line indicates the median value and the cross the mean value.

Dimension	Separation - Composition	Hybrid - Composition
Attractiveness	Mdn_S: 1.50, Mdn_C: 0.92, z = 4.287, p < .016	Mdn_H: 1.42, Mdn_C: 0.92, z = -3.368, p < .016
Perspicuity	Mdn_S: 2.25, Mdn_C: 1.63, z = 3.674, p < .016	Mdn_H: 2.00, Mdn_C: 1.63, z = -2.603, p < .016
Efficiency	Mdn_S: 1.50, Mdn_C: 1.00, z = 3.572, p < .016	
Dependability	Mdn_S: 1.50, Mdn_C: 0.88, z = 3.878, p < .016	
Stimulation	Mdn_S: 1.25, Mdn_C: 1.25, z = 2.501, p < .016	

Table 5.1: Detailed results of the post hoc analysis for the UEQ scores. Only significant results are listed.

For the dimension *Attractiveness*, post hoc analysis showed that participants scored the conditions *Separation* and *Hybrid* significantly better than the condition *Composition*. The same holds for the dimension *Perspicuity*, for which participants again scored the conditions *Separation* and *Hybrid* significantly better than the condition *Composition*. In terms of *Efficiency*, *Dependability*, and *Stimulation*, only the condition *Separation* was scored significantly better than the condition *Composition*. Table 5.1 lists the detailed results of the post hoc tests that showed significant differences.

Concluding Interview

During the concluding interview, participants were asked to rank the three conditions according to two criteria. The first criterion was their general preference (*"Which of the three variants did you like most?"*). The second criterion required participants to assess the amount of coordination that was required (*"Which of the three variants required the least coordination?"*). For analyzing the results of the participants, a fractional ranking was used. Therefore, scores were assigned to the conditions based on their ranked position (i.e., 1 for the condition ranked best, 2 for the condition ranked second best, and 3 for the condition ranked worst). If participants ranked two conditions equally, the mean of two scores was assigned. For example, if a participant ranked the conditions *Separation* and *Hybrid* equal but better than *Composition*, a score of 1.5 was assigned to *Separation* and *Hybrid*, and *Composition* received a score of 3. The results of both rankings are depicted in Figure 5.15.

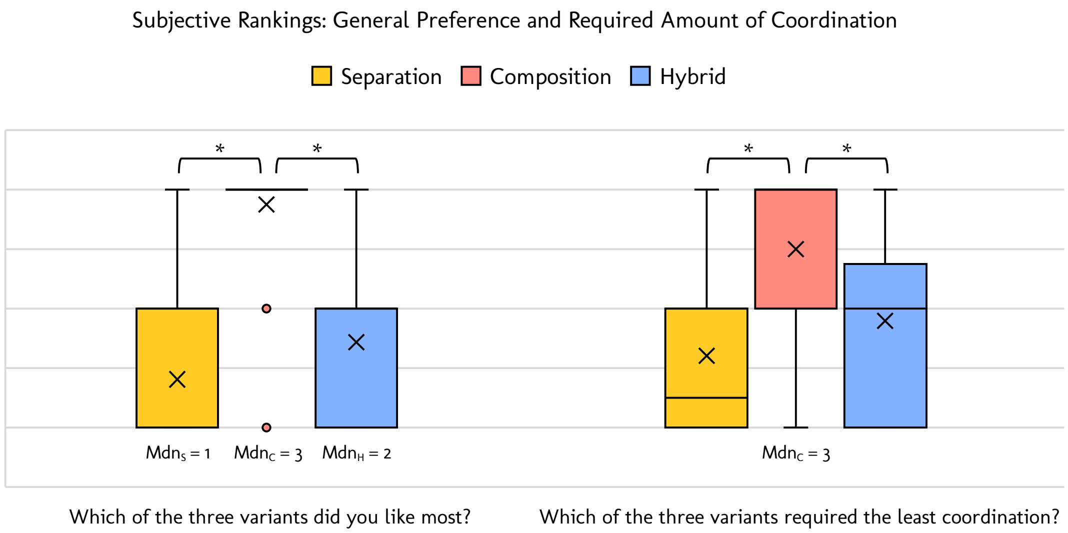


Figure 5.15: Box plots visualizing the results of the participant's subjective ranking of the three conditions according to their general preference and the required amount of coordination. Within each box, the line indicates the median value and the cross the mean value. For the boxes, in which the median value is not recognizable, it is written down below the box.

Statistical analysis showed significant differences in the ranking of participant's general preferences ($\chi^2(2) = 57.770$, $p < .05$). Post hoc analysis revealed that participants ranked the conditions *Separation* ($Mdn_S = 1$, $z = -7.195$, $p < .016$) and *Hybrid* ($Mdn_H = 2$, $z = 5.664$, $p < .016$) significantly better than the condition *Composition* ($Mdn_C = 3$). Also, for the ranking according to the required amount of coordination, the analysis revealed statistical differences ($\chi^2(2) = 21.618$, $p < .05$). The conditions *Separation* ($Mdn_S = 1.25$, $z = -4.389$, $p < .016$) and *Hybrid* ($Mdn_H = 2$, $z = 2.960$, $p < .016$) were ranked to require significantly less coordination compared to the condition *Composition* ($Mdn_C = 3$).

Participants were also asked to reason their ranking and list the advantages and disadvantages of each condition. Table 5.2 provides an overview of their statements.

Separation	Composition	Hybrid
+ Clear division of work (37) + Specialization on one task (16) + Prevents from getting in each other's way (11) + Requires only an initial communication (5) + Perform different actions simultaneously (4) + General positive sensations (16), e.g., *"organized"*, *"structured"*, *"fluent"*	+ Both focus on the same task (6) + Increased communication (4) + Two perspectives for the same action (2) + Fast for large-scale actions (2)	+ Freedom of action (25) + Quick to correct each other (13) + Higher individual share of each collaborator (12) + Less coordination required (9) + Different and same actions simultaneously (4) + Increased accuracy since everyone feels responsible for both actions (2) + General positive sensations (8), e.g., *"unbent"*, *"easier"*, *"flexible"*
- Restricted freedom of action (14) - Harder to correct each other (5) - Less accurate due to one perspective on each task (4)	- Unpredictable addition lead to overshooting (21) - Feeling of working against each other (16) - Lead to alternating work (8) - No different actions simultaneously (8) - General negative sensations (25), e.g., *"frustrating"*, *"unpredictable"*, *"chaotic"*	- Possibility of hindering each other (14) - High amount of coordination required (6) - General negative sensations (7), e.g., *"chaotic"*, *"complicated"*, *"unclear"*, *"uncoordinated"*

Table 5.2: A summary of the advantages and disadvantages participants mentioned during the concluding interview regarding each condition.

Concerning positive aspects of the condition *Separation*, at least one participant of every group ($n = 37$) valued the clear division of work. For example, they said, that *"one had clearly assigned roles or tasks"* (#8a) or that *"it was clear who does what"* (#2a). According to many participants ($n = 11$), this had the positive effect that *"you don't get in each other's way"* (#15b), e. g., they said that *"you can't meddle into the actions of each other"* (#23a), or that *"you don't put one's oar in"* (#24a). Five participants noted that this division of work required only an *"initial communication"* (#19a), but after this *"short coordination [in the beginning] ... everyone knew what he had to do"* (#21a). Additionally, participants considered as beneficial that this condition allowed the specialization on one task ($n = 16$). For example, they said that *"one can focus on one [task]"* (#3a), or *"everyone has one task for which he is responsible"* (#7a), and *"you do not have to think about that many things"* (#6a). One participant said that this allows for *"handing over control over the part you can't do so well [to your collaborator] and rely on him"* (#17a). Further, some participants ($n = 4$) highlighted that the division of work also allowed to do different tasks simultaneous, which decreased the time required to complete the task. For example, one said that "*if you are able to perform different actions at the same time, you reach the goal faster*" (#5b). Finally, participants ($n = 16$) mentioned general positive sensations regarding the condition *Separation*. For example, they *"found it easier [to solve the task]"* (#4b) and *"more intuitive"* (#5b). Also, the perceived the condition as *"fast and accurate"* (#17a) and *"efficient"* (#3a). They also said that it was *"organized and structured"* (#9a/b) and *"better for working productively"* (#9a). One participant even described it as the *"clearest, tidiest, and easiest"* (#8a) condition, another perceived it as *"fluent"* (#19b), and a third said that *"it was less error-prone"* (#17a).

When asked about negative aspects of the condition *Separation*, participants remarked that the freedom of action was restricted ($n = 14$). For example, one participant said that the condition is *"relatively inflexible if you want to do [the task of your collaborator]"* (#13a). Another declared that *"one restricts oneself because you can't directly [perform the task of the other person]"* (#19a). Participants further criticized that the missing freedom of action made it harder to correct each other ($n = 5$), e. g., *"one has to wait until [the other] is unlocked again"* (#19a), or *"if you see something you could correct ... you can't react that fast"* (#7b). Finally, some participants ($n = 4$) mentioned that the accuracy of the condition *Separation* could suffer from the fact that you *"consider only one perspective"* (#2b). For example, one participant said that *"one looked more at his own [task] and less at whether the other really aligned it completely correctly"* (#4b). To compensate for this problem, *"you had to walk more around the object"* (#11b).

Asked about positive aspects of the condition *Composition*, some participants ($n = 6$) mentioned that because *"both participants focused on the same task, [they] were more accurate"* (#2b). Others ($n = 4$) valued that it increased communication, e. g., *"people talk more to each other"* (#12a). Some participants ($n = 2$) reasoned that the condition would be faster for large-scale actions.

But there were also participants ($n = 3$), who said that the condition *"has practically no advantages"* (#24b). Especially, many participants ($n = 16$) reported that they had the feeling of working against each other. For example, one participant said that it was *"like a tug war"* (#6b). Participants further found it hard to assess the addition of their actions, which lead to overshooting ($n = 21$). For example, one said that *"you quickly turned or moved too far when both did something"* (#11b), another complained that they were *"constantly overshooting the floor marking, constantly swinging back and forth"* (#22b). Therefore, many participants ($n = 8$) said that they decided to work alternating to avoid the overshooting, e. g., "'not both of us at the same time' was our strategy" (#14b). Furthermore, participants ($n = 8$) faulted that they could not do different tasks at the same time, which decreased the time to complete the task. Consequently, the condition provoked a lot of negative sensations in the participants ($n = 25$), e. g., *"it is so frustrating"* (#1a), *"[produces] nervousness and anxiety"* (#1b), *"it was really hard"* (#2a), *"unpredictable"* (#2b), *"exhausting, makes depressive"* (#6a), *"one becomes aggressive"* (#6b), *"chaotic"* (#7b), *"stupid"* (#10a), *"pesky"* (#11a), *"confusing"* (#10b), *"complicated"* (#12a), *"overcharging"* (#16a), *"inefficient"* (#17a), *"unintuitive"* (#19a). One participant even stated that *"if you do it long enough, you wouldn't want to see your partner again"* (#1a).

Asked about the condition *Hybrid*, participants ($n = 25$) mainly valued the freedom they had with this approach of action integration. For example, they said, *"you are free to do whatever you want to"* (#1b) or *"you have an incredible amount of room for maneuver"* (#4b) and *"you can develop your own strategy"* (#4b). Closely related to that, participants ($n = 12$) also positively highlighted the higher individual share of each collaborator, e. g., *"you could always do something"* (#5a). Further, many ($n = 13$) liked that they could quickly correct each other (#17b). Besides, a lot of participants ($n = 9$) stated that this approach requires less coordination. Some ($n = 4$) said that they were faster with the approach since they *"could perform both different and same actions simultaneously"* (#2a). Finally, two participants said that they were more accurate *"because* *everyone felt responsible for both actions"* (#4b). Regarding general positive sensations, participants ($n = 8$) perceived the approach as *"efficient"* (#2a), *"unbent"* (#13a), *"easier"* (#1a), *"more individual"* (#7a), *"flexible"* (#13a), *"more intuitive"* (#19a), *"fast"* (#19a), *"fluent"* (#19a), and *"being fun"* (#9b).

Regarding negative aspects, participants ($n = 14$) mainly complained about the possibility of hindering each other. Some ($n = 6$) also noted that it required a high amount of coordination. As general negative sensations they ($n = 7$) mentioned that it was *"chaotic"*, *"complicated"*, *"unclear"* (#14b), and *"uncoordinated"* (#24a).

At the end of the concluding interview, participants were also asked, if the collaborative manipulation of the same object would be relevant for them if they would use the app for planing the furnishing of a room together with a flatmate. As Figure 5.16 shows, 27 out of 48 participants ($\approx 56.25\%$) would find a collaborative manipulation not relevant in such a scenario. As reasons, they mentioned that in such a scenario you would have *"no time pressure"* (#11b), that they would *"do it with one tablet"* (#10a) or that they would prefer *"to do it on their own"* (#9a) and *"compare their settings afterward"* (#15b). However, for 21 out of 48 participants ($\approx 43.75\%$) it would be relevant. They mainly reasoned their decision with the importance of agreeing on a furnishing and the different points of view they have during the collaborative manipulation.

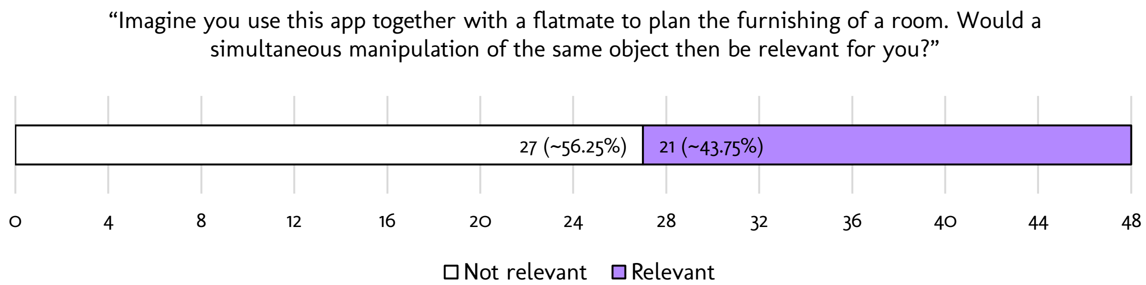


Figure 5.16: Participants were asked to assess if a collaborative object manipulation would be relevant for planning the furnishing their homes together with a flatmate.

The 21 participants who would find a collaborative manipulation relevant were further asked, which of the three conditions they then would like to use in the described scenario. Fifteen participants ($\approx 71.43\%$) would prefer the condition *Hybrid*, only four ($\approx 19.05\%$) the condition *Separation*, and two ($\approx 9.52\%$) were undecided between *Hybrid* or *Separation*. No participant mentioned *Composition* as the preferred approach (cf. Figure 5.17).

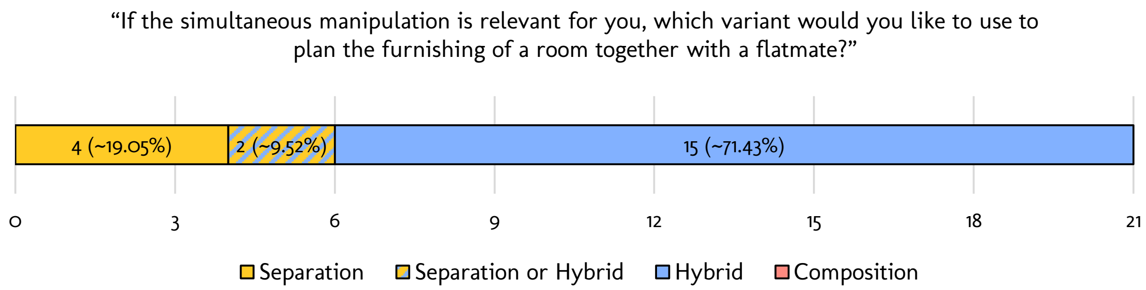


Figure 5.17: The participants that assessed the collaborative manipulation as relevant, were asked which variant the would like to use for furnishing their homes together with a flatmate.

6 Discussion

This chapter discusses the results and draws implications for future applications. The structure thereby follows the research questions. Besides, this chapter discusses the limitations of this work and highlights topics for future work.

6.1 Performance (RQ1)

> **RQ1 Performance:** How do the approaches differ in accuracy and task completion time?

According to accuracy, results revealed no significant differences between the three approaches – neither for positioning errors nor for rotation errors. Consequently, participants were able to achieve the same level of accuracy with all three implemented approaches of action integration. Another look at the qualitative statements gives reasons for the lack of significant differences. Fourteen participants of 7 groups explained that – contrary to the instructions – they primarily focused on accuracy and neglected task completion time. For example, one participant said: *"Accuracy does not depend on the variant. Either you are an awfully accurate person or a sloppy person"* (#16a). This implies that the differences concerning accuracy are manifested in the task completion times, which revealed significant differences (cf. Figure 5.7, p. 42).

During the first pass, participants were significantly faster with the condition *Separation* than with the condition *Composition*. The same applies to the condition *Hybrid*. With it, participants were also significantly faster than with the condition *Composition*. The analysis revealed no significant difference between *Separation* and *Hybrid*. Two factors can be identified for the fact that participants were faster with the conditions *Separation* and *Hybrid*. First, during the condition *Composition*, participants could only perform one manipulation task at a time, i. e., either rotation or positioning. During the other two conditions, however, they could perform both manipulation tasks simultaneously – which means participants could theoretically perform twice as many manipulations in the same period. Admittedly, during the approach *Composition*, participants could generally move the furniture faster due to the addition of their forces. However, this advantage was not relevant during vernier adjustment. The reason for that lies in the second factor, which is the disadvantage of *Composition* that participants mentioned most often ($n = 21$) during the concluding interview: Since they could not estimate the addition of the forces, they often overshot the target.

The second pass generated different results. Here, only the significant difference between *Separation* and *Composition* could be revealed again. The reason for that can be found in the comparison of the two passes within each condition. Only with the condition *Hybrid* participants were not significantly faster in the second pass. These results imply that only the condition hybrid did not involve a learning effect. One explanation for the missing learning effect during *Hybrid* could be that the condition allowed participants to switch between *Separation* and *Composition* seamlessly. As confirmed by participants' statements during the concluding interview, the condition *Hybrid* thus encouraged participants to try out different strategies during the task, which may have hindered the learning effect. For example, they said that *"you can work out your own strategy with this condition"* (#4a), or *"we wanted to try out several [strategies]"* (#17b), and *"we had to talk about how we proceed"* (#14a). On the contrary, participants noted that the other two conditions already preset strategies, e. g., *"[during Separation] the strategy was clear, [since] we had no other option"* (#14a), or *"similar to [Separation] the roles are clear [during Composition] – if one moves the other moves too."* (#7b). The quantitative results also support these qualitative statements: During both passes of *Hybrid*, participants averagely used approximately half of the simultaneous manipulation time *Composition* and the other half *Separation* (cf. Figure 5.8, p. 43). However, fourteen groups increased their use of *Separation* in pass 2 and the other ten their use of *Composition* (cf. Figure 5.9, p. 43). Additionally, it is notable that for many groups, relatively big changes in their use of *Separation* and *Composition* are recognizable between the passes, which further affirms that participants were trying out different strategies during *Hybrid*.

Qualitative statements from the concluding interview further imply that with the training required for finding the perfect strategy, participants could even perform best with the condition *Hybrid*. For example, one participant said: *"If you know each other well and practice often this condition [Hybrid] has the highest potential for speed and accuracy. But this requires more training"* (#17a). Another said: *"If you coordinate it well, you'd be the quickest [with the condition Hybrid]"* (#11b). However, to be able to make reliable statements, future work needs to have a more in-depth look into the different strategies participants developed during the condition *Hybrid*.

Finally, it is notable that the results imply that the learning effect was higher for *Composition* than for *Separation*. During the first pass, medians of the task completion times of *Separation* and *Composition* differed by 67.2 seconds. For the second pass, this difference reduces to 32.6 seconds. With more training, this difference could possibly decrease further. However, future work is needed to assess the maximal decrease.

6.2 Subjective Workload (RQ2)

> **RQ2 Workload:** How do the approaches differ in subjectively perceived workload?

The results of the NASA TLX draw a similar picture as the task completion times (cf. Figure 5.12, p. 45) of the second pass. Overall, the condition *Separation* yields a significantly lower subjective workload than the condition *Composition*. Between *Hybrid* and the other two conditions, no significant differences could be revealed. A look at the scores broken down to the different dimensions of the NASA TLX shows that participants rated *Separation* significantly better than *Composition* for the dimensions *Physical Demand*, *Performance*, *Effort*, and *Frustration*. Between the condition *Hybrid* and the other two conditions, post hoc tests revealed no significant differences (cf. Figure 5.13, p. 45).

During all three conditions, participants needed to perform the same physical activities: They needed to physically locomote through the room, hold the tablet, and execute the gestures for rotation and positioning on it with their fingers. The ratings of the dimension *Physical Demand* imply that during the condition *Composition*, participants were required to perform those tasks more often or more intensive than during the condition *Separation*. Different reasons for this rating can be identified. One could argue that since collaborators felt responsible for both tasks (i.e., rotation and positioning), the condition *Composition* required more physical locomotion. However, participants also had two perspectives on the object for both tasks, which should reduce the required amount of walking – assuming sufficient communication and coordination. Besides, participants also reasoned for the condition *Separation* a higher amount of physical locomotion, which makes this explanation even more unlikely: *"Since you do only one task and the other person concentrates on the other task, you need to walk more around the object to check if it matches with the marking"* (#11b). Another explanation could be that participants mistakenly also included virtual pushing and rotating into their ratings – guided by the associated description of the NASA TLX of the dimension Physical Demand, which explicitly mentions those terms: *"How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.). Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?"* [35]. This would be in line with the participants complaining about overshooting the floor marking often during *Composition*. Finally, the significantly higher task completion times could also be the reason: Participants needed longer to perform the task, and therefore they were also required to be longer physically active.

With the dimension *Performance* of the NASA TLX, participants rated how successful they thought they were in accomplishing the goals of the task [35]. Therefore, the scores of this dimension agree with the quantitative measures for performance (RQ1). Compared with the condition *Composition*, participants could fulfill the task with the condition *Separation* not only significantly faster but also with the same level of accuracy.

By rating the dimension *Effort*, participants had to assess *"how hard [they had] to work (mentally and physically) to accomplish [their] level of performance."* Once more, they rated the condition *Separation* significantly better than the condition *Composition*. Together with their rating of the dimension *Performance*, this means that according to their subjective opinion, they needed less effort to perform better with the condition *Separation* compared to *Composition*. The task completion times confirm the better performance of *Separation*. The reason for the higher effort *Composition* required can be identified in the addition that the approach performs for action integration. Participants reported that they could not assess the effect of this addition, which often led to overshooting. Correcting this overshooting required more effort than in the condition *Separation* were the effect did not occur.

Participants expressed a significant higher subjective frustration induced by *Composition* compared to *Separation*. They reported this frustration not only with their rating of the dimension *Frustration* in the NASA TLX questionnaire but also during the concluding interview – primarily through the general negative sensations they mentioned ($n = 25$). For example, they called the condition *"frustrating," "unpredictable,"* or *"chaotic."* The reason for this frustration can again be identified in the unpredictable addition. Due to the resulting overshooting, participants had the feeling of working against each other ($n = 16$), which can be a very frustrating experience when you try to solve a task together.

6.3 User Experience (RQ3)

RQ3 User Experience: Which approach do users prefer?

During the concluding interview, participants had to rank the three conditions according to their general preference and the required amount of coordination (cf. Figure 5.15, p. 47). Statistical analysis showed that participants ranked the conditions *Separation* and *Hybrid* significantly better than the condition *Composition*. Between the rankings of *Separation* and *Hybrid* no statistically significant differences could be revealed. These results indicate that participants' preferences were polarized between the conditions *Separation* and *Hybrid*. The explanation for this polarization can be found in the qualitative statements participants made during the concluding interview: On the one hand, they valued the clear division of work the condition *Separation* promotes ($n = 37$). On the other hand, they complained about restricted freedom of action during this condition ($n = 14$). And the condition *Hybrid* exactly provided this freedom of action, which was highly valued by the participants ($n = 25$). Consequently, participants were polarized between the *"organized and structured"* (#9a/b) experience of *Separation* and the *"unbent and flexible"* (#13a) one of *Hybrid*.

However, this polarization applies only to the specific task participants had to solve during the study. When asked which approach they would like to use for furnishing their own home, 15 out of the 21 participants who found the collaborative manipulation relevant, reported that they would prefer *Hybrid* over the other conditions (cf. Figure 5.17, p. 50). Their qualitative statements suggest that in this scenario, the freedom of *Hybrid* outweighs the structuredness of *Separation*.

The polarization of participants' preferences between *Separation* and *Hybrid* regarding the study task also becomes apparent in the results of the UEQ. Especially, the scores for the dimension *Attractiveness* confirm the polarization. Participants rated the conditions *Separation* and *Hybrid* significantly more attractive than the condition *Composition*. In addition to *Attractiveness*, also the dimensions *Perspicuity*, *Efficiency*, *Dependability*, and *Stimulation* of the UEQ showed statistical differences.

With the dimension *Perspicuity* participants rated how easy it was to learn the condition and to get familiar with it. Again, they gave significantly better scores for the conditions *Separation* and *Hybrid* compared to *Composition*. For the significantly lower scores of *Composition*, the addition of the actions of the users' actions is standing to reason. Since participants found this addition hard to assess, they perceived the condition as *"complicated"* (#12a). This effect could not appear during the condition *"Separation"* since participants were not able to influence the actions of each other. Also, it allowed them to focus only on a single task (i. e., either rotation or positioning) to reduced the learning effort. Admittedly, the hard to assess addition could also happen during the condition *Hybrid*. However, participants were not required to agree upon who does what, which could be the reason for the significantly higher scores in terms of perspicuity compared to *Composition*.

The significant higher subjective *Efficiency* of *Separation* compared to *Composition* agrees with the quantitative results. With *Separation*, participants could reach their level of accuracy not only significantly faster but also with a significantly lower subjective workload.

Also, in terms of the dimension *Dependability*, participants rated *Separation* significantly better than *Composition*. Again, the problems participants had with estimating the addition and the resulting overshooting during *Composition* may provide an explanation. During *Hybrid*, this overshooting could also occur, but participants could also avoid it entirely by dividing the tasks between each other. This explains why no significant differences could be revealed between *Hybrid* and the other two dimensions for the dimension of *Dependability*.

With the dimension *Stimulation*, participants scored how exciting and motivating the use of the condition was [36]. The analysis of the scores for this dimension also revealed that participants rated the condition *Separation* significantly better than the condition *Composition*. Surprisingly, the polarized preferences of the participants could not be revealed for this condition, since only the condition *Separation* was rated significantly better as the condition *Composition*. A reason for the not revealable difference between *Hybrid* and *Composition* could be again the overshooting. It was possible only during these approaches and could demotivate the participants. Also, participants were only during these both conditions able to quickly correct each other. Especially during the condition *Hybrid*, many participants complained about this possibility of hindering each other.

The dimension *Novelty* of the UEQ prompted participants to rate how creative the design of each approach was and if it caught their interest [36]. Since all three conditions provided the same functionality and differed only in the approach for action integration is not surprising that their scores regarding this dimension were similarly high. Statistical analysis showed no differences between them.

6.4 Implications

> Which of the three approaches (*Separation of DOFs, Composition of Users' Actions, Hybrid Approach*) should a hand-held AR application implement to enable users to furnish a room with digital furniture collaboratively?

The quantitative results of the experimental comparison suggest using *Separation* for such an application: Participants achieved their level of accuracy not only faster but also with less subjective workload. However, during the concluding interview, it turned out that participants' preferences were polarized between the conditions *Hybrid* and *Separation*. On the one hand, they highly valued the clear division of work during *Separation*, and on the other hand, they liked the freedom of action during *Hybrid*. This polarized preferences of the participants imply to choose the approach depending on scenario and task. Applications that require high accuracy and speed possibly should rely on the clear division of work *Separation* provides. In contrast, for more exploratory tasks like furnishing a place, the freedom of action that *Hybrid* allows could be more suitable. Future work should, therefore, have a more in-depth look at different tasks to further investigate this trade-off between a clear division of work and freedom of action.

Both the quantitative and qualitative results speak against the usage of *Composition*. Only in terms of accuracy, a trend in favor of this condition was recognizable. During both passes, participants were, on average, slightly more accurate than with the other two conditions. Hence, future work could investigate other implementations of this approach, which could increase its accuracy further. Especially using the mean instead of the sum seems to be a promising alternative for performing action integration. Also, one could think about an approach that dynamically switches between mean and sum for action integration depending on the participants' view on the object to be manipulated. If the trend towards a higher accuracy of the approach can be confirmed by future work, one probably should use this approach for tasks that are not time-critical but require a high accuracy.

6.5 Limitations

The study has some limitations that should be addressed in future work. At first, the experimental comparison of the three approaches considered the real-world task of furnishing a place that required only 3 DOFs for manipulation. Participants could rotate the virtual object around the yaw axis and move it along the floor. Rotations around the other two axes were not possible. Besides, participants could not lift, lower, or scale the objects. Future work could investigate different scenarios and tasks that require the manipulation of all 9 DOFs. A second limitation concerns the number of people that collaborate. The study reported by this work investigated dyads. But one could also consider groups with three or even more members. Finally, this work investigated three specific variants of the three different approaches for action integration related work revealed. However, also other implementations of the three approaches are possible. During *Composition*, one could also compute the mean of the users' inputs instead of their sum. Also, if more than 3 DOFs are investigated, *Separation* would allow splitting the DOFs of one canonical task between the users. For example, in terms of the canonical task rotation, one user could then rotate the object around the yaw axis, another one around the roll axis, and a third one around the pitch axis.

6.6 Future Work

The study and its limitations give rise to further research questions that could be addressed by future work: First, one should have a more in-depth look at the participants' strategies during the three conditions. Is there a general strategy for each approach that works best? Or does the performance depend on the users' skills and the current constellation? During *Separation*, for example, it could be better to keep the same roles throughout the whole task, i. e., one collaborator rotates the object while the other one moves it. On the other hand, it could also be beneficial to change the roles depending on the perspective on the object to be manipulated. For *Hybrid*, it could be interesting to find out when and why participants used either *Composition* or *Separation*. Especially during *Composition*, it could also be crucial how participants positioned themselves towards the object. For example, did participants perform better if they stood on different sides of the same corner of the virtual piece of furniture? And which constellation (i.e. *f-formation* [38]) did they mostly use, i.e., face-to-face, side-to-side, or corner-to-corner (cf. Figure 6.1)? The investigation of strategies could also include their talking to each other and their movements through the room. How much did they talk to each other? Did they walk around a lot, or were they rather stationary?

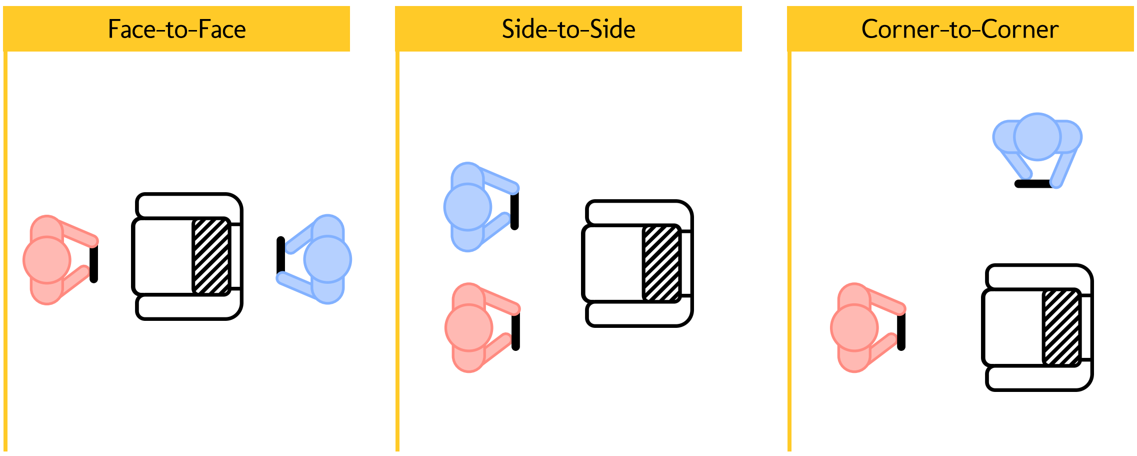


Figure 6.1: Future work could investigate, which *f-formations* [38] users prefer and with which of them they perform best.

Further, one could think about how to prevent the frustration provoked mainly during *Composition* but also during *Hybrid*. The probably most obvious approach to solve this issue is the examination of different merge policies. What are their benefits and drawbacks? Are there tasks for which a certain merge policy works best? Which merge policy do the users like most? For example, it could be that the users can better estimate the resulting manipulations of a merge policy that computes the mean of the users' inputs. Especially in terms of positioning, it could also be beneficial to choose the merge policy either based on the current f-formation or even more fine-grained based on the angle between the vectors of the users' actions.

Practically, such an adaptive merge policy could interpolate between the computation of the sum and the mean dependent on the current constellation (cf. Figure 6.2). For example, if the angle between the vectors towards the users' actions are directed accounts for 90 degrees, the merge policy could compute the sum. For angles of 0 and 180 degrees, the mean could be used. For the angles in between, the merge policy could interpolate between sum and average.

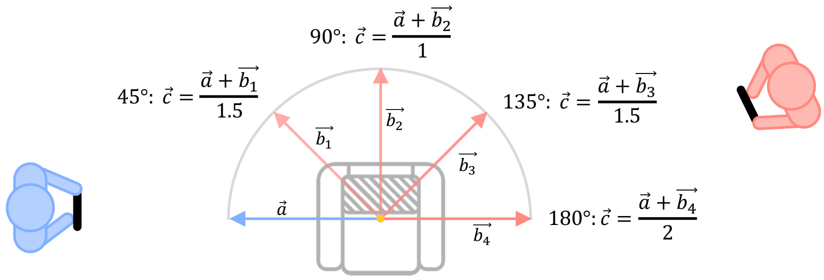


Figure 6.2: An adaptive merge policy could interpolate between the computation of the sum and the mean based on the angle between the vectors towards the users' actions are directed.

Besides, one could address how the different approaches could be implemented for more than two collaborators. Dependent on scenario and task also the approach *Separation of DOFs* could support more than two collaborators. A third collaborator could, for example, be added if one also includes the canonical task scaling. For more than three collaborators, the DOFs of one canonical task needed to be separated between the users. In terms of positioning, one could, for example, take up the "Sliding Feature" together with the ray-based technique as introduced by Pinho et al. [7] to separate the DOFs of this canonical task between two users (cf. Figure 6.3, left). A possible modification of the "Sliding Feature" could be, that both users are equipped with rays and that the object is always positioned where they cross (cf. Figure 6.3, right).

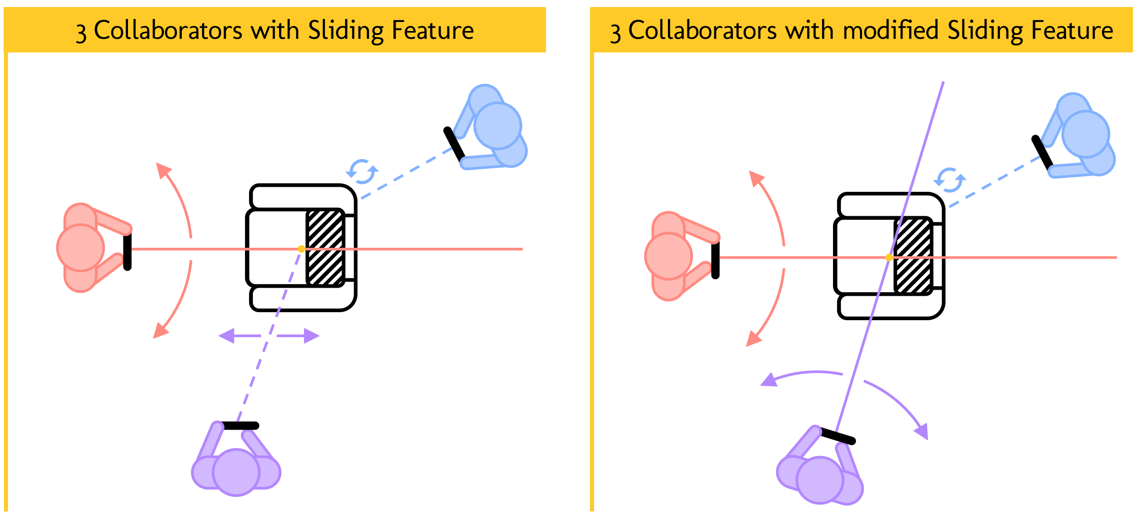


Figure 6.3: Using a ray-based technique together with the "Sliding Feature" as introduced by Pinho et al. [7] one could easily support three collaborators *(left)*. A modified version of could equip two users with a ray and position the object always at the position where they cross *(right)*.

If one takes that to the extreme, one could theoretically support up to nine collaborators, one for each of the nine possible DOFs to be manipulated. However, it is questionable if a real-world scenario can be found that could profit from this extreme form of *Separation*, with nine collaborators manipulating the same object simultaneously. The implementation of the approaches *Composition of Users' Actions* and the investigated *Hybrid Approach* presented in this work already support a theoretically unlimited number of collaborators, since simultaneous manipulations of the same DOFs are combined by adding them up. However, it is questionable how many collaborators could practicably work together with a system using this merge policy. Grandi et al. [5] showed it for four collaborators with the computation of the sum as the merge policy. But is this merge policy also practicable for more collaborators? And how does it work for other merge policies? Is there a real-world task that could profit from more than four collaborators manipulating the same object?

For a system that supports more than two collaborators, the awareness cues become even more important. Consequently, future work could also investigate them more in-depth. Which cues work best? Are there different cues needed for the different approaches? And how can they be designed to support different numbers of collaborators? While this work focuses on visual awareness cues, one could also think of haptic cues in different stages – from simple vibrations to electric muscle stimulation as Lopes et al. [39] proposed it (cf. Figure 6.4). Such haptic feedback would allow for making conflicting manipulations of the collaborators sensible or even more difficult. Consequently, also Ruddle et al. [26] stress its potential for increased coordination. Further, it could also help both collaborators to perceive the constraints of the physical room. As demonstrated by Lopes et al. [39], with electric muscle stimulation, users could even be hindered from pushing a virtual object into a physical wall. While vibrations could work for hand-held AR displays, electric muscle stimulation seems more suitable for HMDs, which provide a hands-free experience.

Figure 6.4: Lopes et al. [39] suggest electric muscle stimulation for haptic feedback in AR. *Taken from [39].*

Moreover, the comparison of specific implementations of the three approaches is still missing on AR HMDs, which could be a further scope of future work. In contrast to hand-held AR displays, those devices are capable of producing stereoscopic 3D, which allows assessing depth more easily. Also, they come with different input options like gesture recognition. Such a touch-less manipulation would no longer be limited to the two-dimensional touchscreen but could mimic the real-world grabbing and handling of an object. Especially the way Microsoft presents it for the HoloLens 2 [16] looks very promising (cf. Figure 6.5, left). But with Project Soli [40] also Google already introduced an interesting sensor that they advertise for the recognition of precise gestures (cf. Figure 6.5, right).

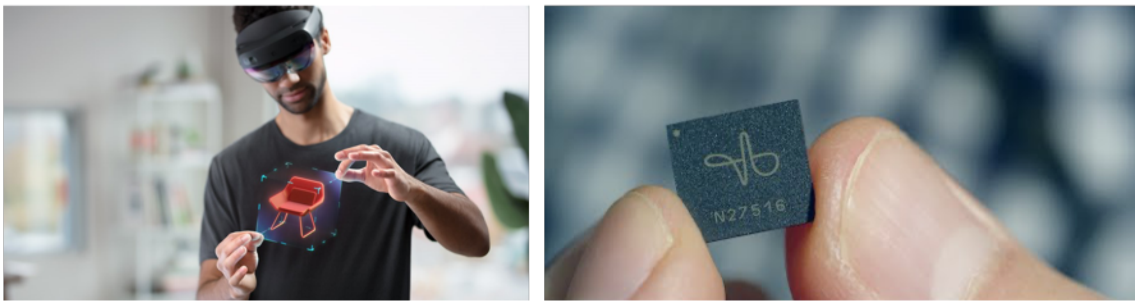


Figure 6.5: Both Microsoft *(left)* and Google *(right)* integrated a gesture recognition into their AR devices. *Taken from [16] and [41].*

7

Conclusion

Motivated by the scenario of interior design with hand-held AR displays, this work raised the question of how one could implement a collaborative 3D object manipulation for those devices. A literature review revealed three different approaches. The first approach is called *Separation of DOFs* and separates the manipulation of the different DOFs of an object between the users. The second approach is called *Composition of Users' Actions* and implements a merge policy to combine the users' inputs to one resulting manipulation. As a third, *Hybrid Approach* combinations of both variants are used. Independently from the approach, the summary of related work revealed, that a collaborative manipulation of virtual 3D objects can be more efficient and increase the usability of a system. Unfortunately, based on related work, it was difficult to decide which approach one should for the collaborative 3D object manipulation on hand-held AR displays: First, only one publication could be found, that investigates an AR scenario. Most of the works focus on virtual environments presented by HMDs, projectors, or computer monitors. Secondly, only two publications provide insights into the differences between the approaches. However, their conclusions are quite contradictory. One work speaks in favor of a *Composition of Users' Actions* and the other one for a *Separation of DOFs*.

To be able to answer the research question anyhow, three specific implementations of the three approaches were developed and integrated into a study prototype. The mode *Separation* of the prototype separates the DOFs of the canonical tasks rotation and positioning between the users. That is, while one user positions the object, the other one can only rotate it. *Composition* uses the sum as the merge policy and allows users to manipulate only the DOFs of either rotation or positioning simultaneously. Finally, the mode *Hybrid* switches dynamically between both approaches depending on the users' inputs. All three approaches rely on the same touch-based interaction. To investigate these specific implementations of the three approaches in terms of *performance*, *workload*, and *user experience*, a controlled lab study was conducted, in which 24 dyads took part.

Regarding performance and workload, the results of the study suggest using *Separation* for such an application since participants performed better while sensing less subjective workload. However, participants' preferences were polarized between *Separation* and *Hybrid*: On the one hand, they valued the clear division of work during *Separation*, and on the other hand, the liked the freedom of action provided by *Hybrid* and would, therefore, prefer this approach for furnishing a home. The mode *Composition* only showed a trend towards a higher accuracy but performed worst for all other measures. These polarized preferences imply to choose the approach depending on scenario and task. If a high performance in terms of accuracy and speed is required, one should possibly rely on *Separation*. For more exploratory tasks like furnishing a place, *Hybrid* could be more suitable due to the freedom it provides. If future work can confirm the trend towards a higher accuracy of *Composition*, one should probably use this approach for tasks that are not time-critical but require high precision.

With the experimental comparison of the three approaches on hand-held devices, this work provides first insights into their strengths and weaknesses. Besides, it raised a wide variety of further research questions with different scopes that should be addressed by future work.

8

List of Literature, Figures and Tables

Literature

[1] Mark Billinghurst and Hirokazu Kato. "Collaborative Augmented Reality". In: *Commun. ACM* 45.7 (July 2002), pages 64–70. ISSN: 0001-0782. DOI: 10.1145/514236.514265. URL: http://doi.acm.org/10.1145/514236.514265 (cited on page 1).

[2] *Envisioning the Future with Windows Mixed Reality*. URL: https://youtu.be/2MqGrF6JaOM (visited on 03/20/2018) (cited on page 1).

[3] *Neue AR-App IKEA Place*. URL: https://ikea-unternehmensblog.de/article/2019/ikea-place-app (visited on 11/14/2019) (cited on pages 1, 2).

[4] Marcio S. Pinho, Doug A. Bowman, and Carla M. Dal Sasso Freitas. "Cooperative object manipulation in collaborative virtual environments". In: *Journal of the Brazilian Computer Society* 14.2 (June 2008), pages 53–67. ISSN: 1678-4804. DOI: 10.1007/BF03192559 (cited on pages 2, 10, 13, 14, 16–18).

[5] J. G. Grandi et al. "Collaborative manipulation of 3D virtual objects in augmented reality scenarios using mobile devices". In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 2017, pages 264–265. DOI: 10.1109/3DUI.2017.7893373 (cited on pages 2, 10, 13, 22–24, 26, 57).

[6] *Ikea Place*. URL: https://youtu.be/-3vjDF74fok (visited on 03/21/2018) (cited on pages 2, 6).

[7] Márcio S. Pinho, Doug A. Bowman, and Carla M.D.S. Freitas. "Cooperative Object Manipulation in Immersive Virtual Environments: Framework and Techniques". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '02. Hong Kong, China: ACM, 2002, pages 171–178. ISBN: 1-58113-530-0. DOI: 10.1145/585740.585769 (cited on pages 2, 10, 14–16, 57).

[8] Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. "Evaluation of a Reconfigurable Tangible Device for Collaborative Manipulation of Objects in Virtual Reality". In: *Proceedings of the Theory and Practice of Computer Graphics Conference (TP.CG)*. Sept. 2011 (cited on pages 2, 18, 25, 26).

[9] T. Duval, A. Lecuyer, and S. Thomas. "SkeweR: a 3D Interaction Technique for 2-User Collaborative Manipulation of Objects in Virtual Environments". In: *3D User Interfaces (3DUI'06)*. Mar. 2006, pages 69–72. DOI: 10.1109/VR.2006.119 (cited on pages 2, 10, 19, 20).

[10] M. L. Chenechal et al. "When the giant meets the ant an asymmetric approach for collaborative and concurrent object manipulation in a multi-scale environment". In: *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*. Mar. 2016, pages 18–22. DOI: 10.1109/3DCVE.2016.7563562 (cited on pages 2, 10, 13, 18, 21, 22).

[11] J. G. Grandi et al. "Design and Assessment of a Collaborative 3D Interaction Technique for Handheld Augmented Reality". In: *2018 IEEE Virtual Reality (VR)*. Mar. 2018 (cited on pages 2, 10, 13, 23, 24).

[12] Ronald T. Azuma. "A Survey of Augmented Reality". In: *Presence: Teleoperators and Virtual Environments* 6.4 (1997), pages 355–385. DOI: 10.1162/pres.1997.6.4.355 (cited on page 5).

[13] Mark Billinghurst, Adrian Clark, and Gun Lee. "A Survey of Augmented Reality". In: *Foundations and Trends® in Human–Computer Interaction* 8.2-3 (2015), pages 73–272. ISSN: 1551-3955. DOI: 10.1561/1100000049. URL: http://dx.doi.org/10.1561/1100000049 (cited on page 5).

[14] Paul Milgram and Fumio Kishino. "A Taxonomy of Mixed Reality Visual Displays". In: vol. E77-D, no. 12 (Dec. 1994), pages 1321–1329. URL: http://etclab.mie.utoronto.ca/people/paul_dir/IEICE94/ieice.html (cited on pages 5, 6).

[15] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. Natick, MA, USA: A. K. Peters, Ltd., 2005. ISBN: 1568812302 (cited on page 6).

[16] *Microsoft HoloLens*. URL: https://www.microsoft.com/en-us/hololens (visited on 11/07/2019) (cited on pages 6, 58).

[17] Dieter Schmalstieg and Tobias Höllerer. *Augmented Reality: Principles and Practice*. Addison–Wesley Professional, June 2016. ISBN: 9780133153217 (cited on pages 6, 7, 9).

[18] Robert Johansen. *GroupWare Computer Support for Business Teams*. New York, NY, USA: The Free Press, 1988. ISBN: 0029164915 (cited on page 7).

[19] *Computer-supported cooperative work*. URL: https://en.wikipedia.org/wiki/Computer-supported_cooperative_work (visited on 02/21/2018) (cited on page 7).

[20] David Margery, Bruno Arnaldi, and Noël Plouzeau. "A General Framework for Cooperative Manipulation in Virtual Environments". In: *Virtual Environments '99*. Edited by Michael Gervautz, Dieter Schmalstieg, and Axel Hildebrand. Vienna: Springer Vienna, 1999, pages 169–178. ISBN: 978-3-7091-6805-9 (cited on pages 7, 8).

[21] Joseph J. LaViola et al. *3D User Interfaces: Theory and Practice*. Addison–Wesley Professional, Mar. 2017. ISBN: 9780134034478 (cited on pages 8, 17).

[22] *Ikea Ekerö*. URL: https://www.ikea.com/PIAimages/0602695_PE680474_S5.JPG (visited on 01/11/2019) (cited on page 8).

[23] D. Mendes et al. "A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments". In: *Computer Graphics Forum* 38.1 (2019), pages 21–45. DOI: 10.1111/cgf.13390 (cited on page 9).

[24] *ARCore Google Developers*. URL: https://developers.google.com/ar/ (visited on 02/15/2018) (cited on page 9).

[25] E. S. Goh, M. S. Sunar, and A. W. Ismail. "3D Object Manipulation Techniques in Handheld Mobile Augmented Reality Interface: A Review". In: *IEEE Access* 7 (2019), pages 40581–40601. DOI: 10.1109/ACCESS.2019.2906394 (cited on page 9).

[26] Roy A. Ruddle, Justin C. D. Savage, and Dylan M. Jones. “Symmetric and Asymmetric Action Integration During Cooperative Object Manipulation in Virtual Environments”. In: *ACM Trans. Comput.-Hum. Interact.* 9.4 (Dec. 2002), pages 285–308. ISSN: 1073-0516. DOI: 10.1145/586081.586084. URL: http://doi.acm.org/10.1145/586081.586084 (cited on pages 10, 18, 19, 58).

[27] K. Riege et al. “The Bent Pick Ray: An Extended Pointing Technique for Multi-User Interaction”. In: *3D User Interfaces (3DUI'06)*. Mar. 2006, pages 62–65. DOI: 10.1109/VR.2006.127 (cited on pages 10, 20, 35).

[28] Jonathan Wieland. “Supporting Collaborative Construction Tasks in Co-Located Mixed Reality”. Unpublished term paper. July 2018 (cited on page 13).

[29] W. Broll. “Interacting in distributed collaborative virtual environments”. In: *Proceedings Virtual Reality Annual International Symposium '95*. Mar. 1995, pages 148–155. DOI: 10.1109/VRAIS.1995.512490 (cited on page 13).

[30] Doug A. Bowman and Larry F. Hodges. “An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments”. In: *Proceedings of the 1997 Symposium on Interactive 3D Graphics*. I3D '97. Providence, Rhode Island, USA: ACM, 1997, 35–ff. ISBN: 0-89791-884-3. DOI: 10.1145/253284.253301. URL: http://doi.acm.org/10.1145/253284.253301 (cited on pages 13, 16).

[31] Jonathan Wieland. “Collaborate! – Collaborative Interior Design in Augmented Reality”. Unpublished term paper. Mar. 2019 (cited on pages 27–36, 38–41).

[32] *Unity*. URL: https://unity.com/ (visited on 03/18/2019) (cited on page 30).

[33] *Apple Developer: Augmented Reality*. URL: https://developer.apple.com/augmented-reality/ (visited on 08/28/2019) (cited on page 30).

[34] *Unity ARKIT Plugin*. URL: https://bitbucket.org/Unity-Technologies/unity-arkit-plugin (visited on 03/18/2019) (cited on page 30).

[35] *NASA Task Load Index*. URL: https://humansystems.arc.nasa.gov/groups/TLX/ (visited on 09/25/2019) (cited on pages 39, 53).

[36] *User Experience Questionnaire*. URL: https://www.ueq-online.org/ (visited on 09/25/2019) (cited on pages 39, 54, 55).

[37] *iPad Pro*. URL: https://en.wikipedia.org/wiki/IPad_Pro (visited on 08/31/2019) (cited on page 39).

[38] Adam Kendon. *Conducting interaction: Patterns of behavior in focused encounters*. Volume 7. CUP Archive, 1990 (cited on page 56).

[39] Pedro Lopes et al. “Adding Force Feedback to Mixed Reality Experiences and Games Using Electrical Muscle Stimulation”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: ACM, 2018, 446:1–446:13. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174020. URL: http://doi.acm.org/10.1145/3173574.3174020 (cited on page 58).

[40] *Soli*. URL: https://atap.google.com/soli/ (visited on 11/12/2019) (cited on page 58).

[41] *Welcome to Project Soli*. URL: https://www.youtube.com/watch?v=0QNiZfSsPc0&t=9s (visited on 11/12/2019) (cited on page 58).

Tables

Figures

9

Appendix

A. Declaration of Independent Work

Ich versichere hiermit, dass ich die anliegende Masterarbeit mit dem Thema

"Comparing Different Approaches for the Collaborative Manipulation of Virtual 3D Objects in Augmented Reality"

(Deutsch: "Vergleich verschiedener Ansätze für die kollaborative Manipulation von virtuellen 3D Objekten in Augmented Reality")

selbstständig verfasst und keine anderen Hilfsmittel und Quellen als die angegebenen benutzt habe.

Die Stellen, die anderen Werken (einschließlich des Internets und anderer elektronischer Text- und Datensammlungen) dem Wortlaut oder dem Sinn nach entnommen sind, habe ich in jedem einzelnen Falle durch Angabe der Quelle bzw. der Sekundärliteratur als Entlehnung kenntlich gemacht.

Weiterhin versichere ich hiermit, dass die o.g. Arbeit noch nicht anderweitig als Abschlussarbeit einer Bachelor- bzw. Masterprüfung eingereicht wurde. Mir ist ferner bekannt, dass ich bis zum Abschluss des Prüfungsverfahrens die Materialien verfügbar zu halten habe, welche die eigenständige Abfassung der Arbeit belegen können.

Darüber hinaus reiche ich die Arbeit zusätzlich auch in elektronischer Form, als Datei, beim Dozenten ein.

Konstanz, den 20. November 2019

Jonathan Wieland

B. Welcome Letter

Konstanz, den 21.12.2018

Herzlich willkommen!

Vielen Dank, dass Sie sich dazu bereit erklärt haben, an unserer Studie teilzunehmen. Sie unterstützen damit unsere Forschung maßgeblich! Bevor es losgeht, möchten wir Ihnen kurz vermitteln, um was es bei der Untersuchung geht und welche Rolle Sie dabei spielen.

Ziele und Ablauf der Studie

Im Rahmen der Studie untersuchen wir, inwiefern sich neue Interaktionstechnologien für die computergestützte Zusammenarbeit eignen. Dazu werden Sie gemeinsam mit Ihrem Teampartner oder Ihrer Teampartnerin drei Varianten einer App testen. Bei jeder der drei Varianten erhalten Sie als Team die Aufgabe, den Raum, in dem Sie sich befinden, gemeinsam mit virtuellen Möbelstücken auszustatten. Es geht dabei um Genauigkeit und Schnelligkeit. Im Anschluss an jede Variante bitten wir Sie, uns von Ihren Erfahrungen und Eindrücken zu berichten. Vor jeder Variante gibt es einen Übungsteil. Fragen zum generellen Ablauf oder zum System können Sie während dieses Übungsteils stellen. Bitte haben Sie jedoch Verständnis dafür, dass wir während der eigentlichen Aufgabe keine Fragen beantworten können, um eine Verzerrung der Daten zu verhindern.

Um möglichst umfassende Erkenntnisse zu erhalten, zeichnen wir die Studie zusätzlich in Bild und Ton auf. Für diese Aufzeichnungen ist Ihr Einverständnis erforderlich. Im Gegenzug verpflichten wir uns dazu, das Material pseudonymisiert und lediglich zu Auswertungszwecken zu verwenden. In diesem Zusammenhang haben wir eine Einverständniserklärung vorbereitet, die diesem Schreiben beiliegt. An dieser Stelle möchten wir darauf hinweisen, dass wir nicht Sie oder Ihre Leistung bewerten, sondern ausschließlich an der Tauglichkeit der Varianten der App interessierst sind.

Zeitrahmen und Entlohnung

Die Dauer der Studie beträgt insgesamt ca. 1 Stunde. Falls Sie sich zu irgendeinem Zeitpunkt unwohl fühlen und Ihre Teilnahme beenden möchten, ist das selbstverständlich auch ohne Angabe von Gründen möglich. Bitte wenden Sie sich dann an den Versuchsleiter.

Nach der Durchführung der Studie werden Sie für Ihre Hilfe mit 10 Euro entlohnt. Wir wünschen Ihnen jetzt gutes Gelingen und bedanken uns noch einmal recht herzlich für Ihre Unterstützung!

Jonathan Wieland

Arbeitsgruppe Mensch-Computer-Interaktion,
Fachbereich Informatik und Informationswissenschaft,
Universität Konstanz

C. Consent Form

Einverständniserklärung

ID: ______

Informationen zur Studienleitung

Studienleiter: Jonathan Wieland

Institution: Arbeitsgruppe Mensch-Computer Interaktion, Fachbereich Informatik und Informationswissenschaft, Universität Konstanz

Erklärung

Über das Ziel, den Inhalt und die Dauer der Studie wurde ich informiert. Im Rahmen dieser Studie werden in Fragebögen personenbezogene Daten erhoben. Zusätzlich wird die Studie auf Video aufgezeichnet, es werde Audioaufnahmen gemacht und Bewegungsdaten erfasst.

Hiermit bin ich darüber aufgeklärt, dass die personenbezogenen Daten vertraulich behandelt werden und nicht an Dritte weitergegeben werden. Video-, Audio- und Bewegungsdaten werden pseudonymisiert analysiert. Die Ergebnisse der Analyse werden eventuell in späteren Publikationen veröffentlicht. Wir garantieren dabei absolute Diskretion. Es wird zu keinem Zeitpunkt Rückschluss auf Sie als Person möglich sein.

Optionale Punkte *(Bei Zustimmung bitte ankreuzen)*

- [ ] Ich bin damit einverstanden, dass meine Video- und Bewegungsdaten zusätzlich zu internen Präsentationszwecken genutzt werden können.
- [ ] Ich bin an den Ergebnissen der Studie interessiert und bitte deshalb um Übersendung entsprechender Informationen.

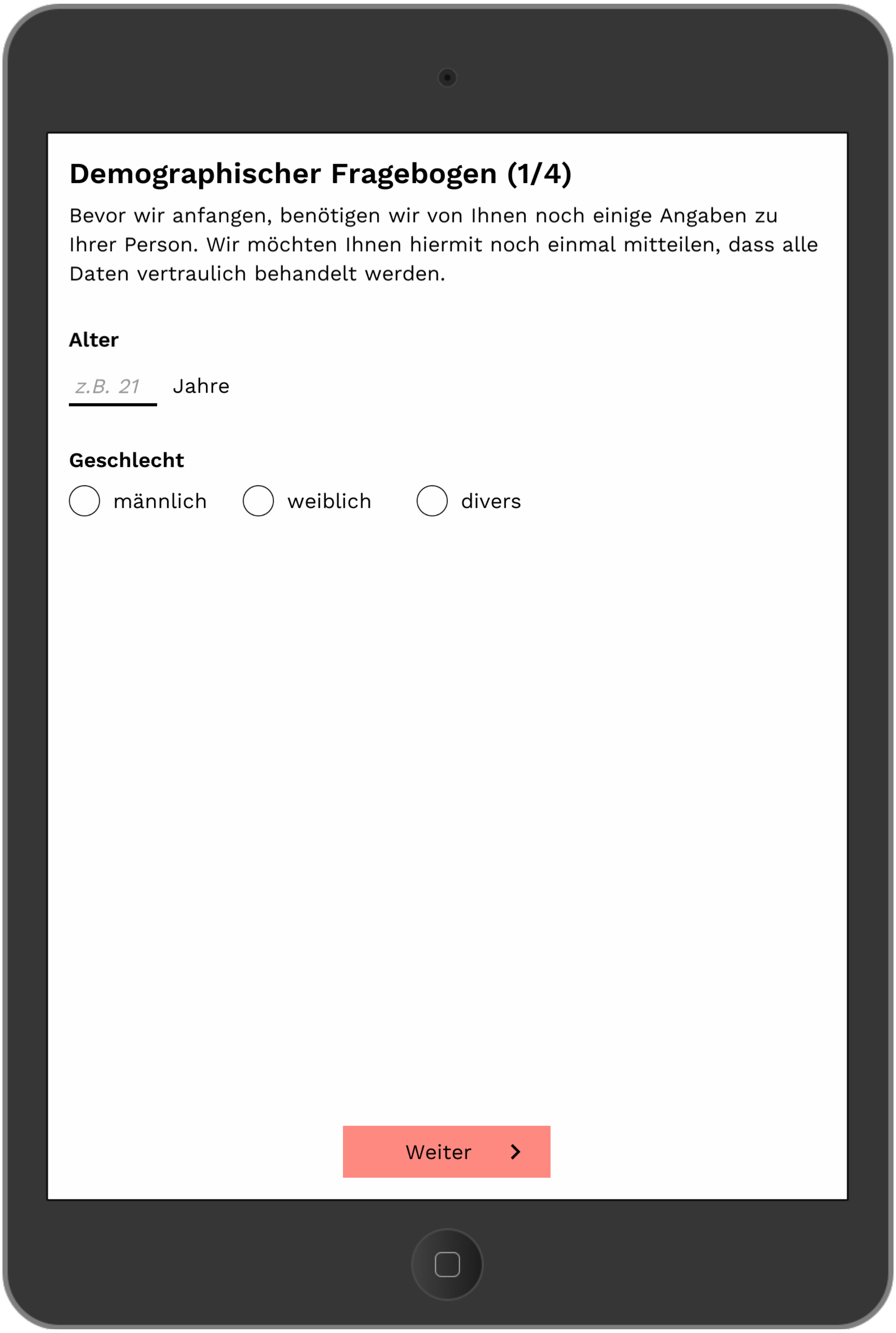
Hiermit erkläre ich mich mit den unter „Erklärung" genannten Punkten und den angekreuzten optionalen Punkten einverstanden:

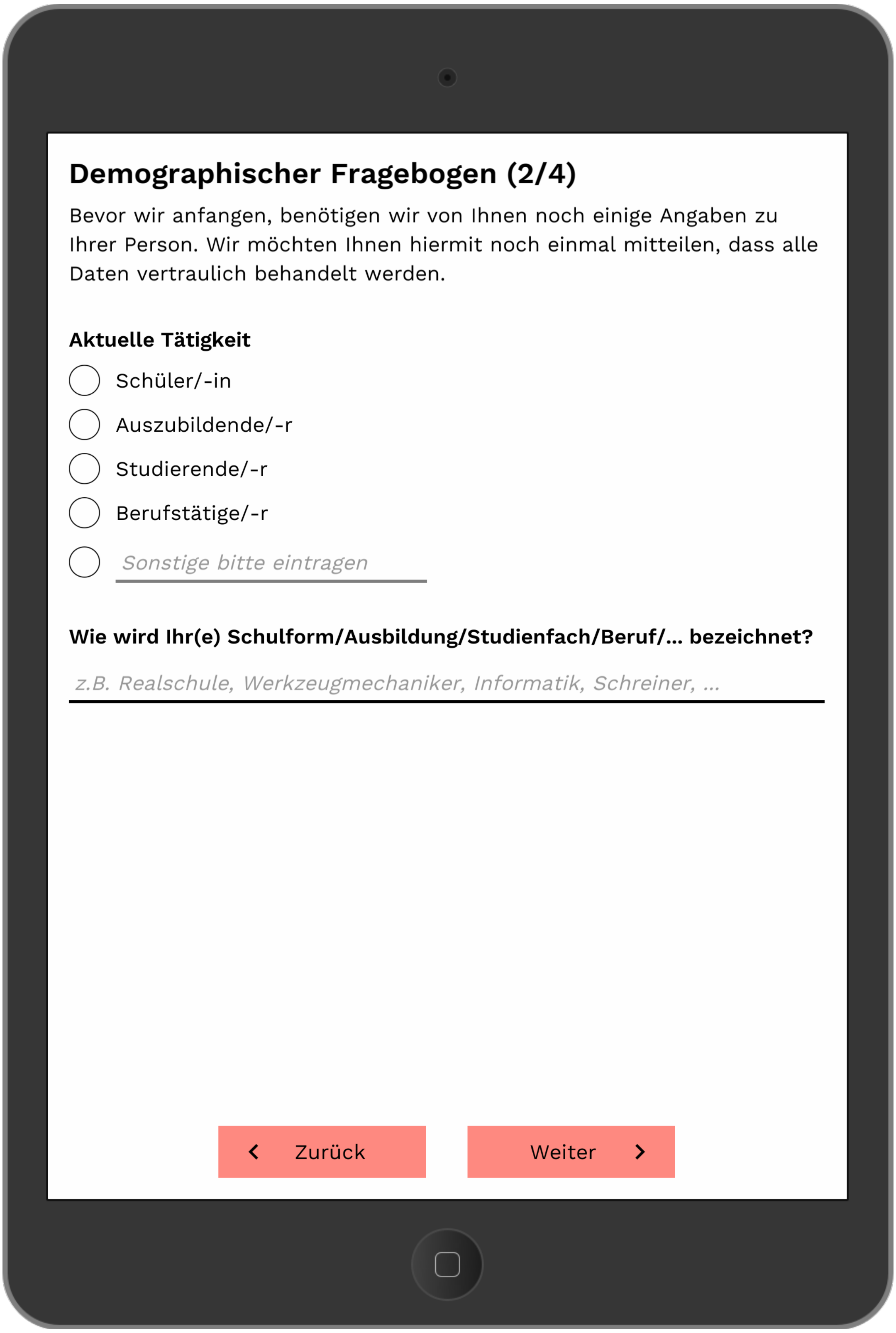
	Konstanz,	
(Name)	(Ort, Datum)	(Unterschrift)

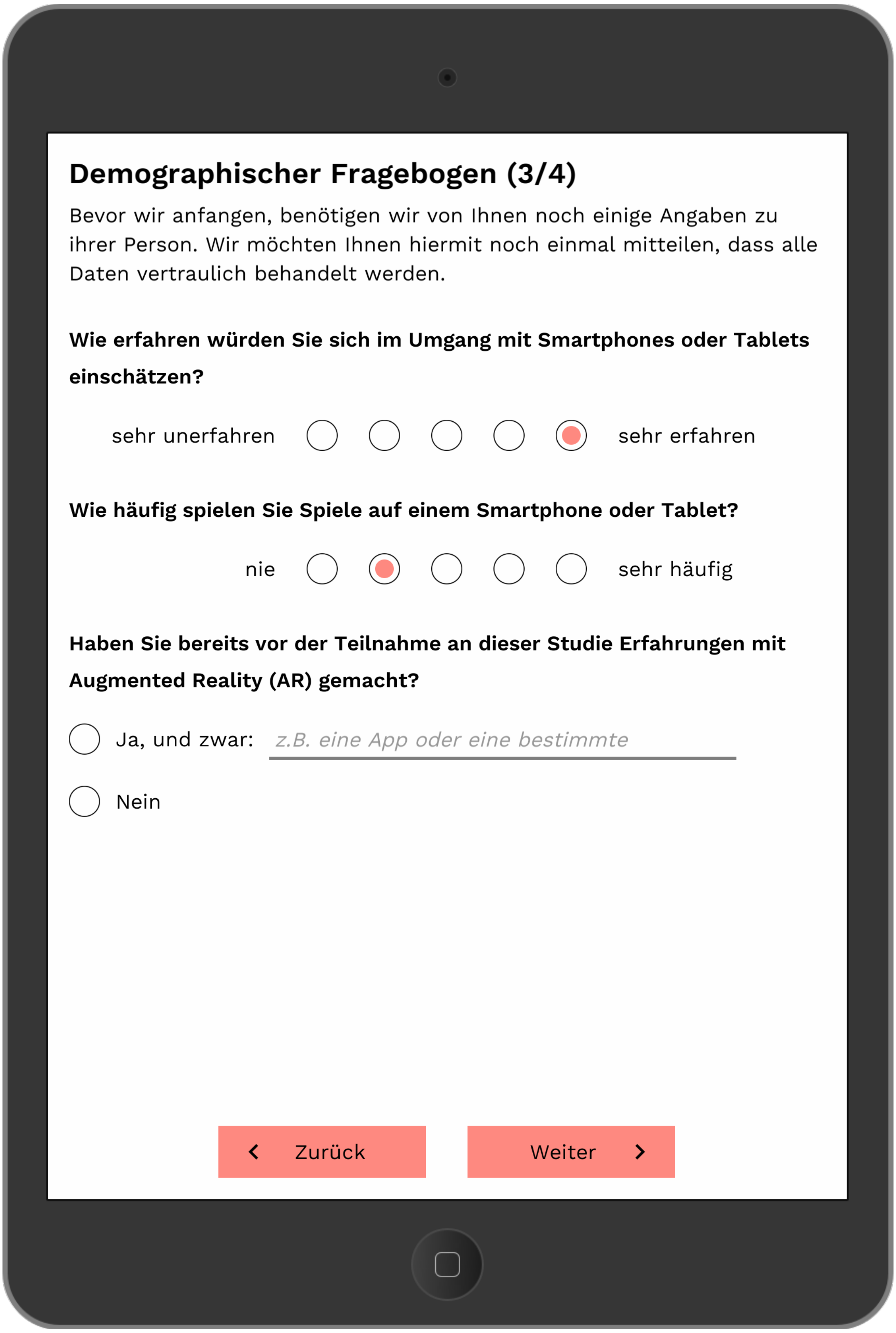
Hiermit verpflichtet sich die Studienleitung, die Video- und Audioaufzeichnung sowie sämtliche sonstigen gewonnenen Daten lediglich zu Auswertungszwecken im Rahmen dieser Untersuchung zu verwenden:

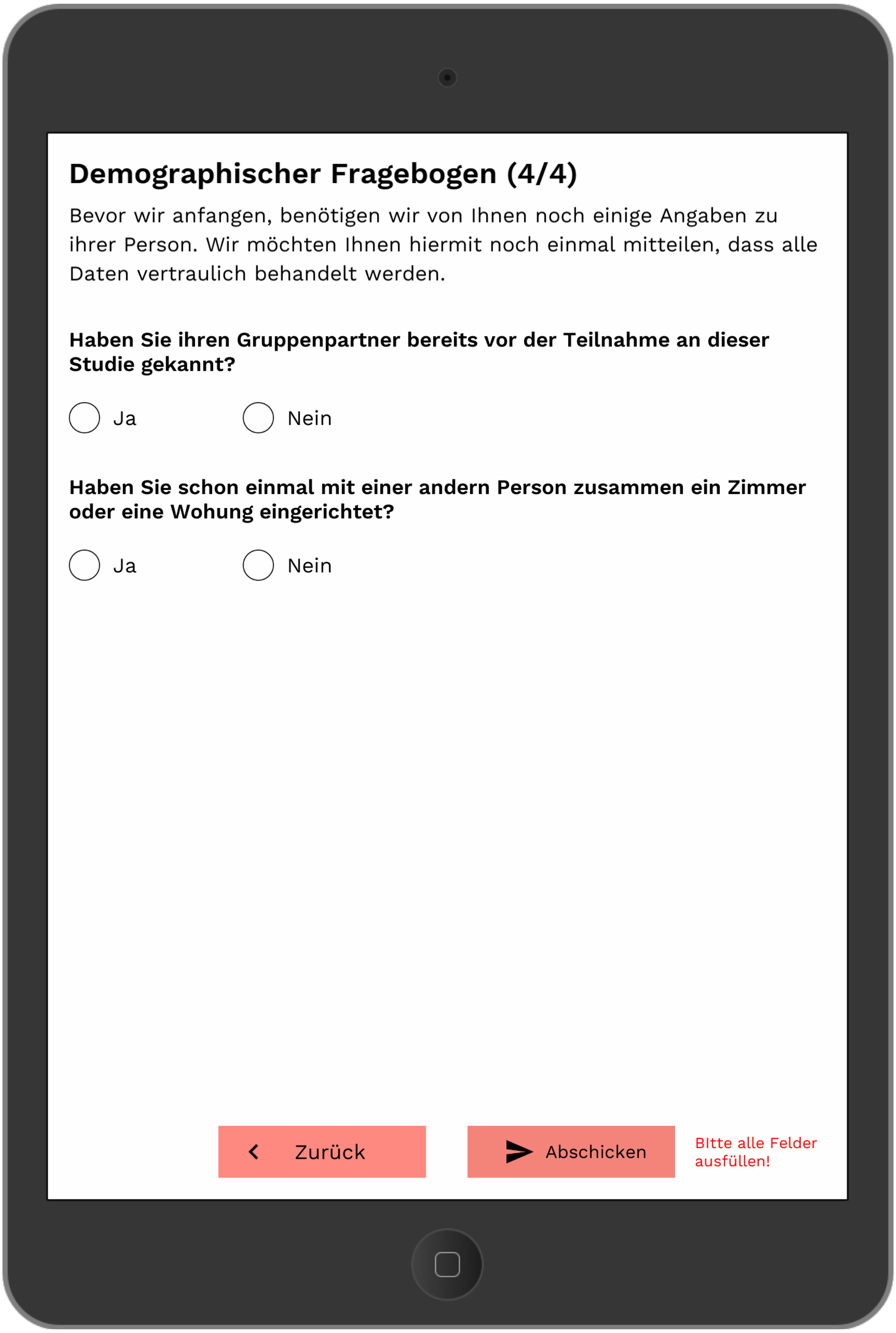
Jonathan Wieland	Konstanz,	
(Name)	(Ort, Datum)	(Unterschrift)

D. Demographic Questionnaire




Demographischer Fragebogen (2/4)
Bevor wir anfangen, benötigen wir von Ihnen noch einige Angaben zu Ihrer Person. Wir möchten Ihnen hiermit noch einmal mitteilen, dass alle Daten vertraulich behandelt werden.
Aktuelle Tätigkeit
Schüler/-in
Auszubildende/-r
Studierende/-r
Berufstätige/-r
Sonstige bitte eintragen
Wie wird Ihr(e) Schulform/Ausbildung/Studienfach/Beruf/... bezeichnet?
z.B. Realschule, Werkzeugmechaniker, Informatik, Schreiner, ...
Zurück
Weiter


Demographischer Fragebogen (3/4)
Bevor wir anfangen, benötigen wir von Ihnen noch einige Angaben zu ihrer Person. Wir möchten Ihnen hiermit noch einmal mitteilen, dass alle Daten vertraulich behandelt werden.
Wie erfahren würden Sie sich im Umgang mit Smartphones oder Tablets einschätzen?
sehr unerfahren
sehr erfahren
Wie häufig spielen Sie Spiele auf einem Smartphone oder Tablet?
nie
sehr häufig
Haben Sie bereits vor der Teilnahme an dieser Studie Erfahrungen mit Augmented Reality (AR) gemacht?
Ja, und zwar:
z.B. eine App oder eine bestimmte
Nein
Zurück
Weiter


Demographischer Fragebogen (4/4)
Bevor wir anfangen, benötigen wir von Ihnen noch einige Angaben zu ihrer Person. Wir möchten Ihnen hiermit noch einmal mitteilen, dass alle Daten vertraulich behandelt werden.
Haben Sie ihren Gruppenpartner bereits vor der Teilnahme an dieser Studie gekannt?
Ja
Nein
Haben Sie schon einmal mit einer andern Person zusammen ein Zimmer oder eine Wohung eingerichtet?
Ja
Nein
Zurück
Abschicken
BItte alle Felder ausfüllen!

E. Explanation of Study Prototype and Task

Augmented Reality Studie „Collaborate“

Aufgabenstellung

1

Sucht einen Marker am Boden.

2

EINER von euch „zieht“ das passende Objekt aus dem Katalog.

3

Platziert das Objekt so, dass es mit dem Marker übereinstimmt.

4

Tippt neben die Plattform, um die Auswahl aufzuheben.

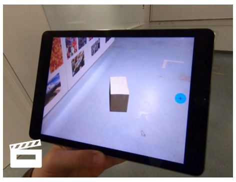
Z

Zusammenfassung.

Erledigt die Aufgabe so schnell wie möglich und so genau wie möglich!

Augmented Reality Studie „Collaborate“

Verschieben und Drehen

1

Tippt das Objekt an, um es auszuwählen.

2

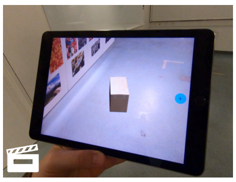
Verschieben: Bewegt mit einem Finger die Plattform.

3

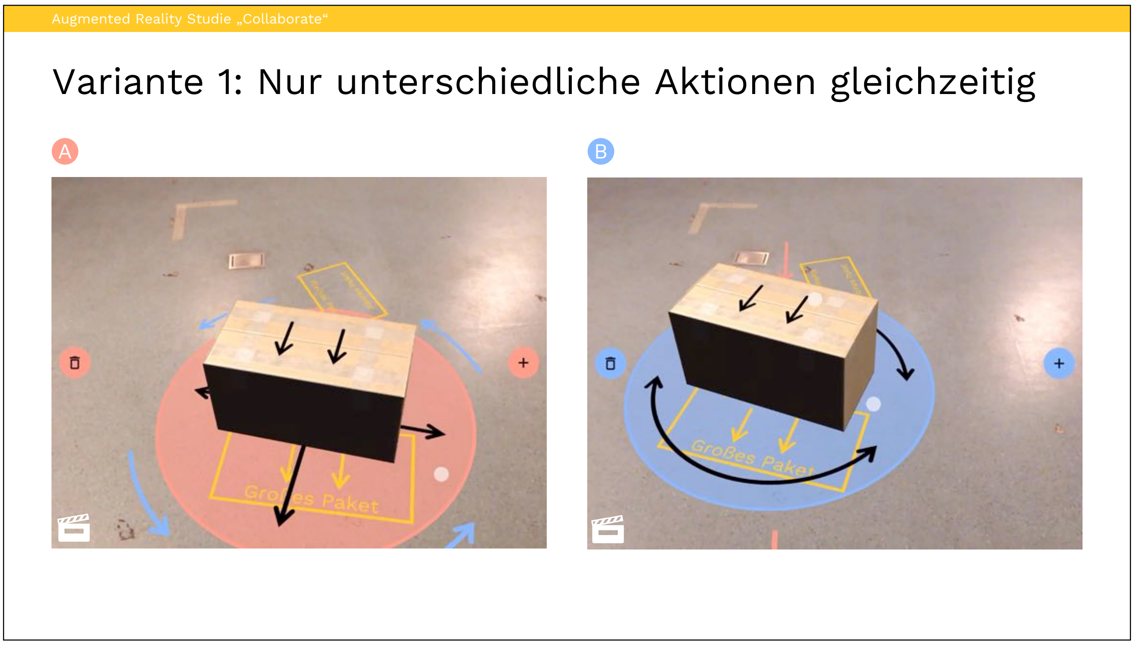
Drehen: Führt mit zwei Fingern die Drehgeste durch.

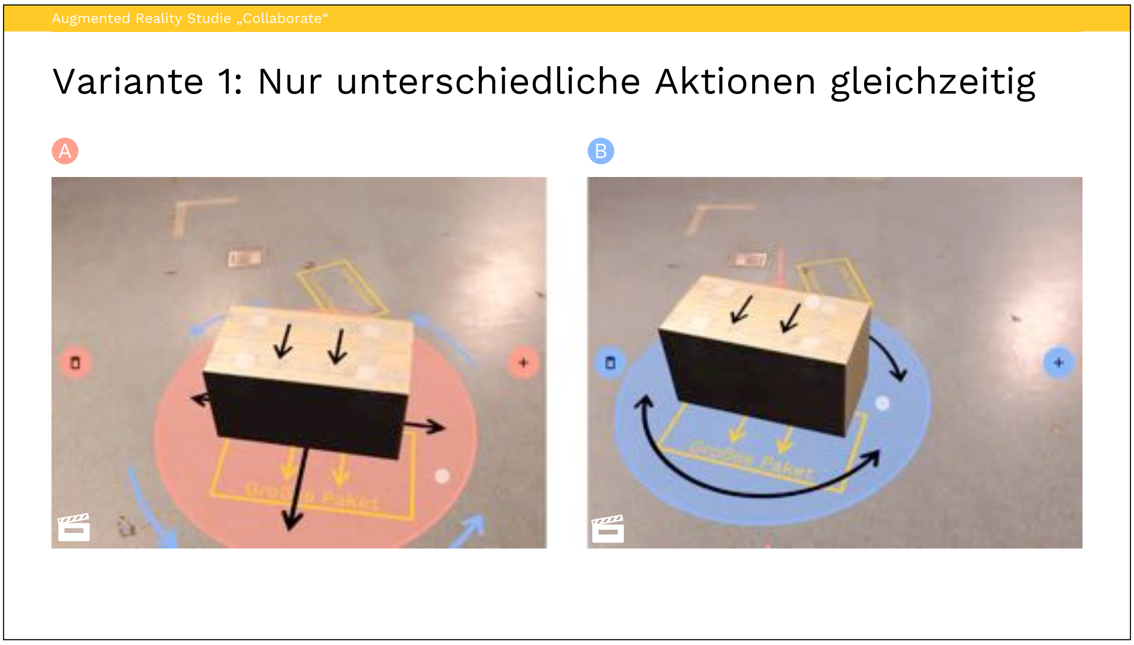
4

Tippt neben die Plattform, um die Auswahl aufzuheben.

Z

Zusammenfassung.

F. Presentation of the Approaches

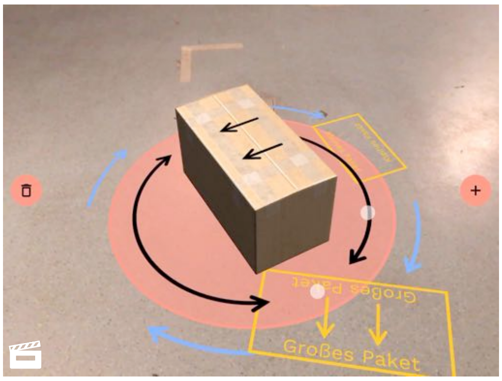




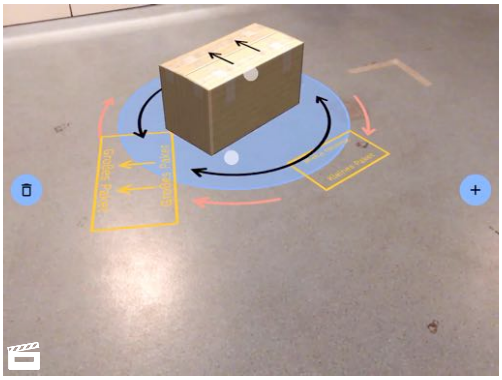
Augmented Reality Studie „Collaborate“

Variante 2: Nur dieselbe Aktion gleichzeitig

A



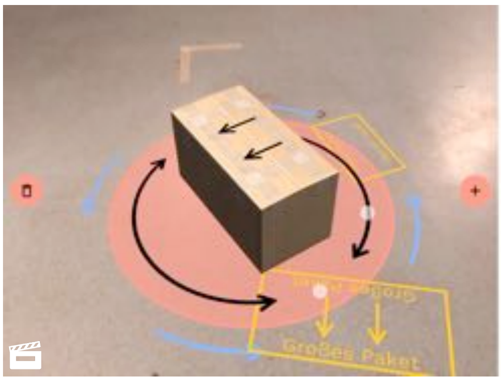
B



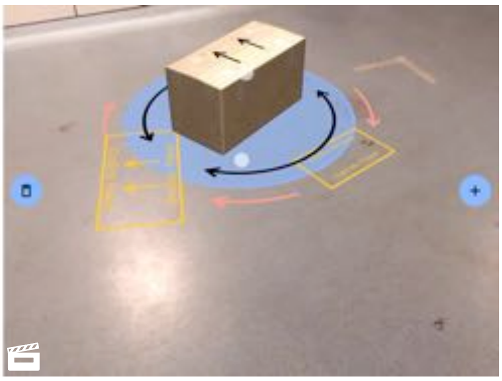
Augmented Reality Studie „Collaborate“

Variante 2: Nur dieselbe Aktion gleichzeitig

A



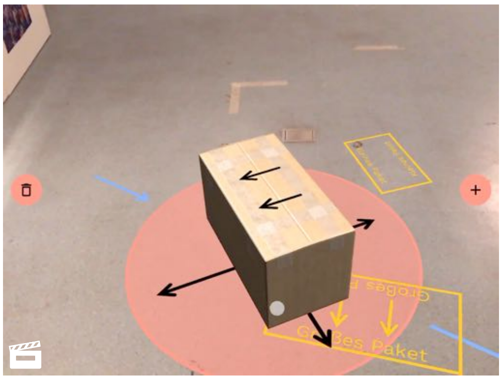
B



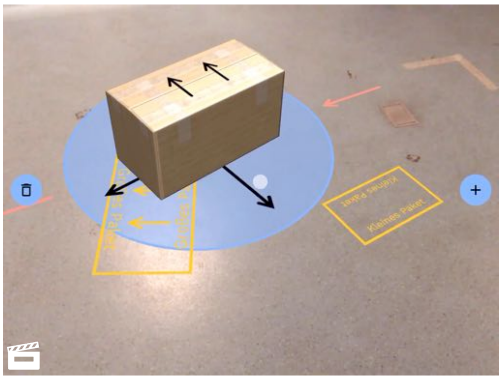
Augmented Reality Studie „Collaborate“

Variante 3: Beliebige Aktionen gleichzeitig

A



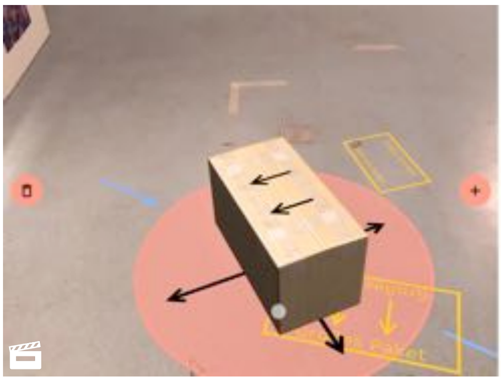
B



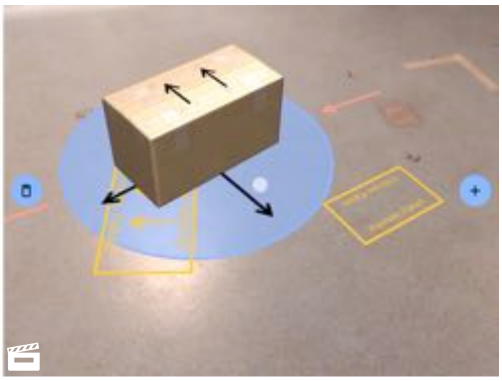
Augmented Reality Studie „Collaborate“

Variante 3: Beliebige Aktionen gleichzeitig

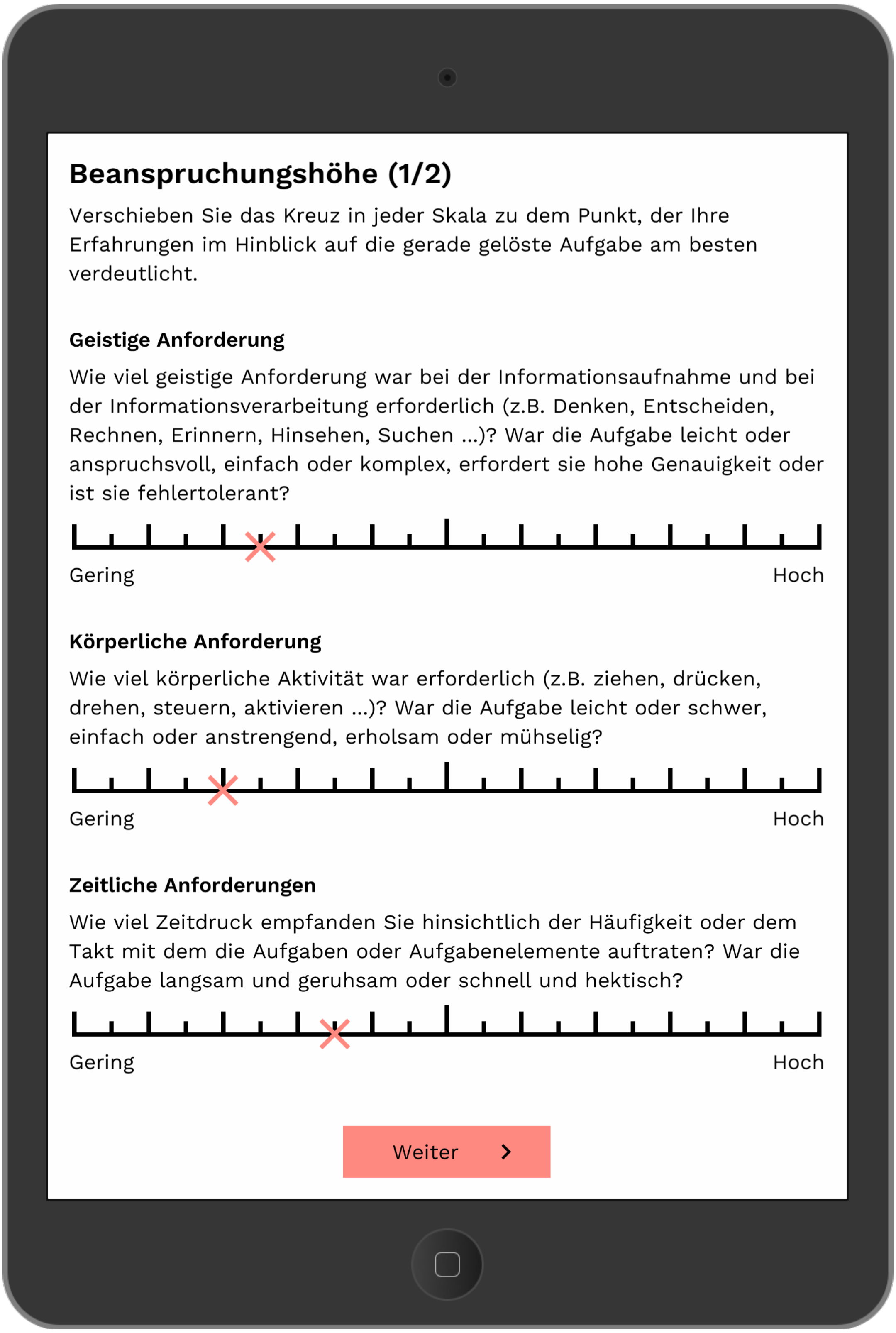
A

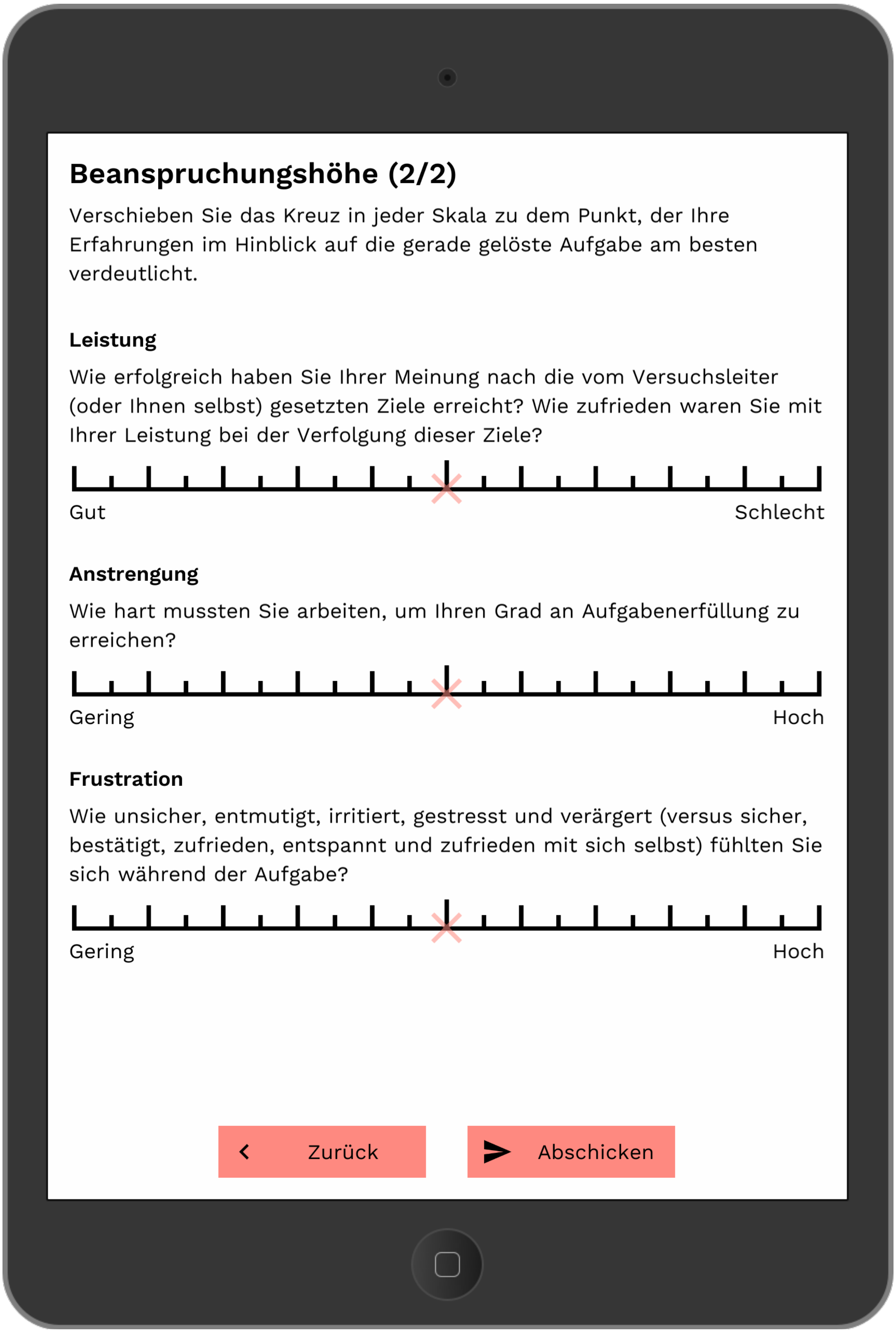


B

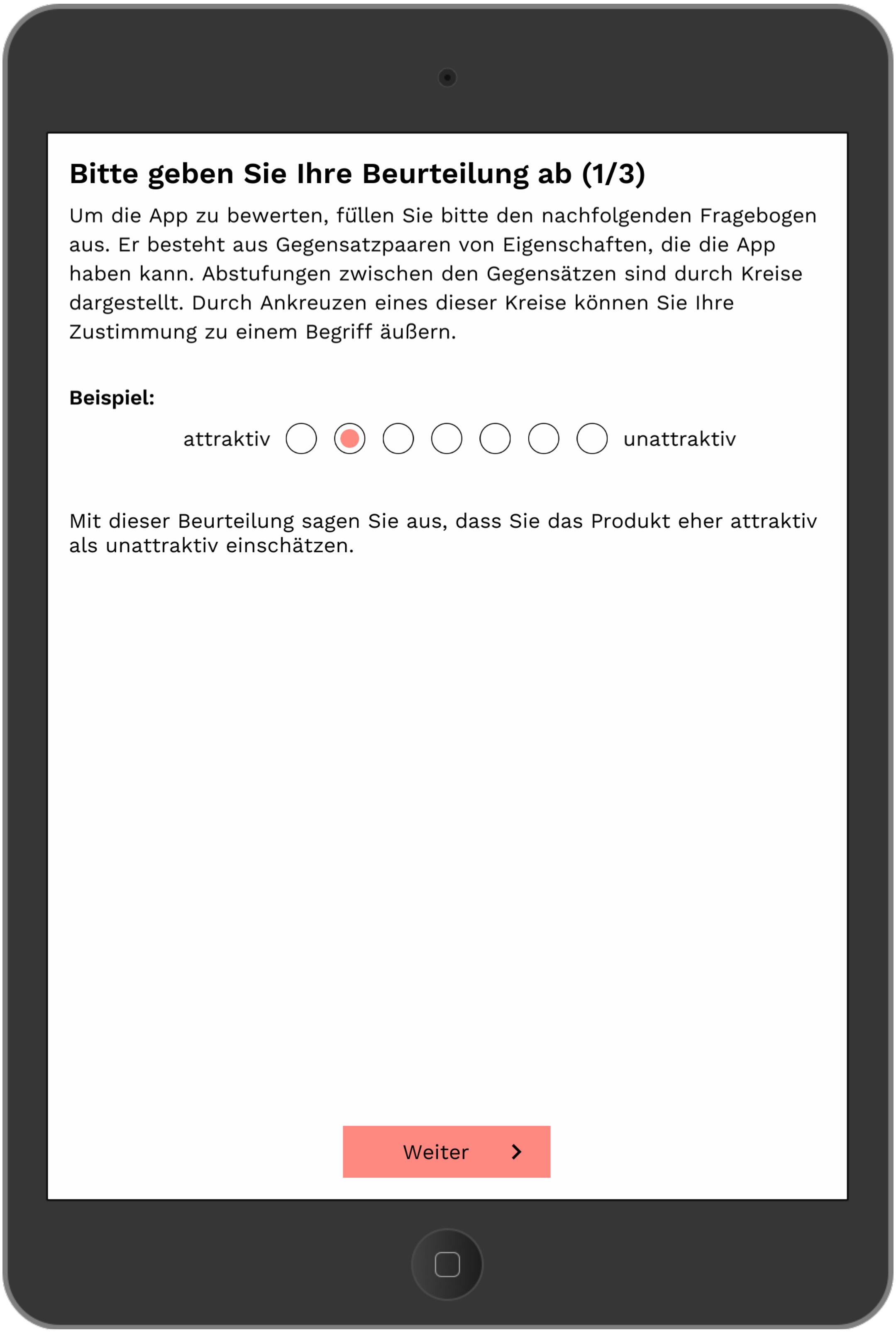


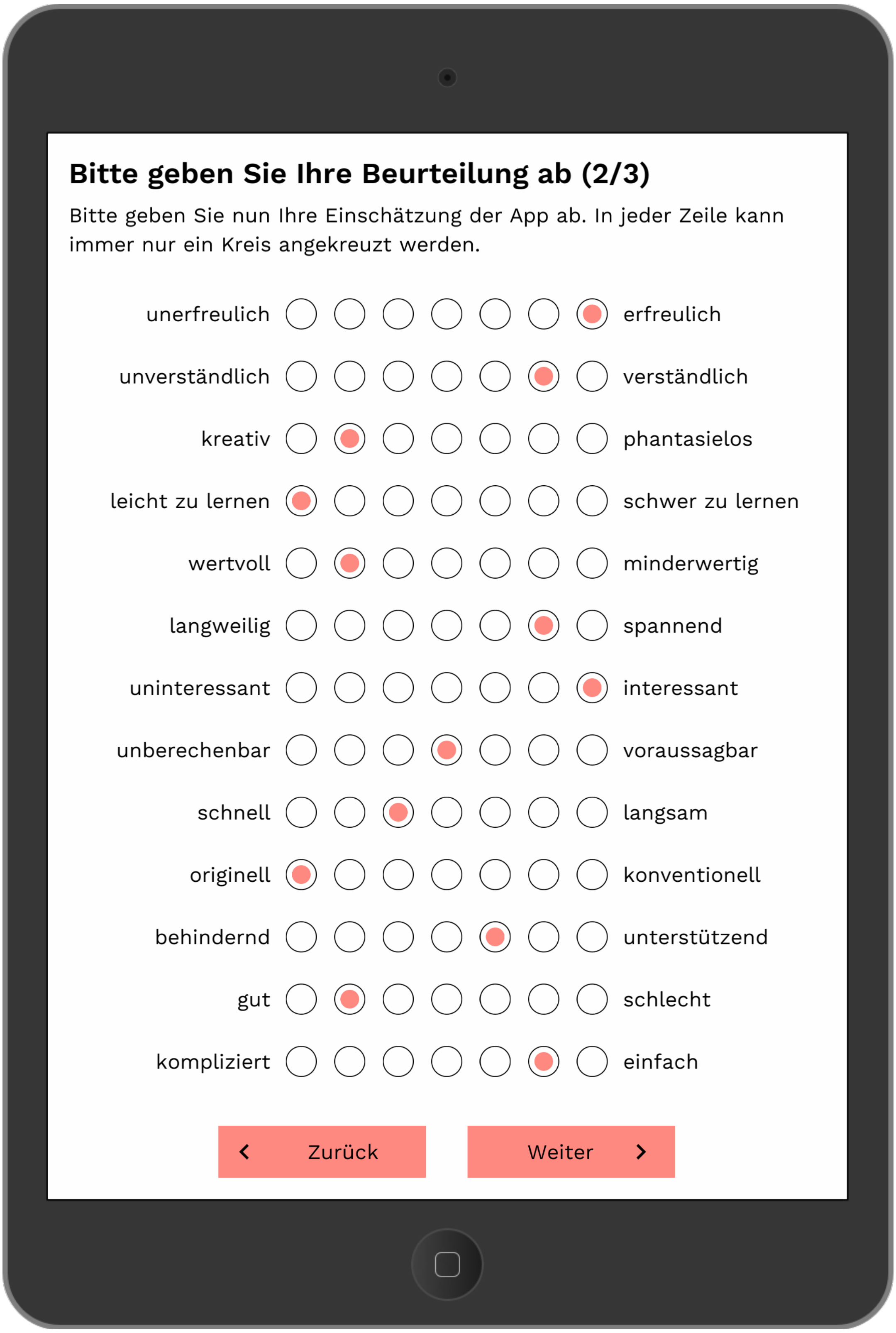
G. NASA Task Load Index (NASA TLX)

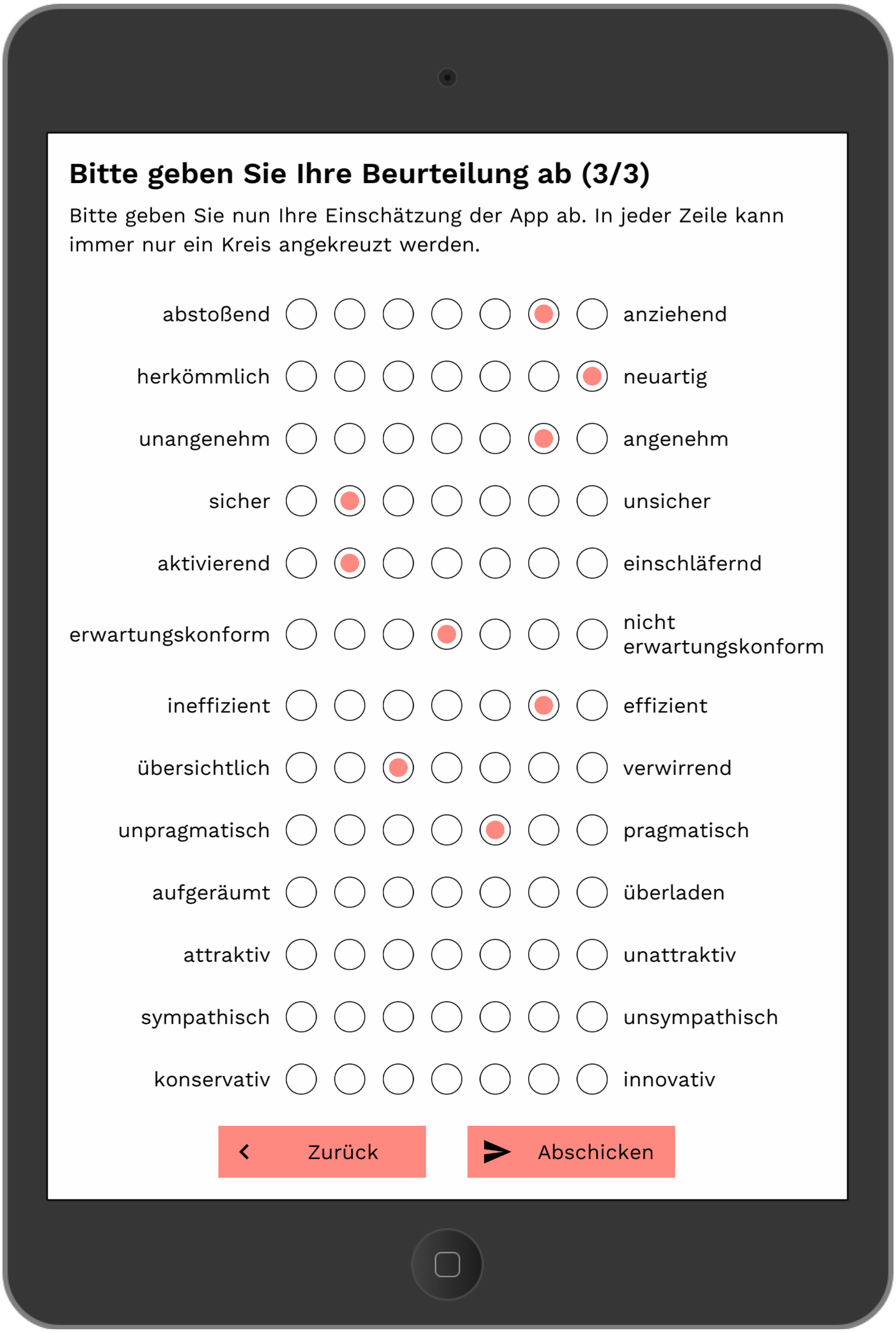



Beanspruchungshöhe (2/2)
Verschieben Sie das Kreuz in jeder Skala zu dem Punkt, der Ihre Erfahrungen im Hinblick auf die gerade gelöste Aufgabe am besten verdeutlicht.
Leistung
Wie erfolgreich haben Sie Ihrer Meinung nach die vom Versuchsleiter (oder Ihnen selbst) gesetzten Ziele erreicht? Wie zufrieden waren Sie mit Ihrer Leistung bei der Verfolgung dieser Ziele?
Gut
Schlecht
Anstrengung
Wie hart mussten Sie arbeiten, um Ihren Grad an Aufgabenerfüllung zu erreichen?
Gering
Hoch
Frustration
Wie unsicher, entmutigt, irritiert, gestresst und verärgert (versus sicher, bestätigt, zufrieden, entspannt und zufrieden mit sich selbst) fühlten Sie sich während der Aufgabe?
Gering
Hoch
Zurück
Abschicken

H. User Experience Questionnaire (UEQ)




Bitte geben Sie Ihre Beurteilung ab (2/3)
Bitte geben Sie nun Ihre Einschätzung der App ab. In jeder Zeile kann immer nur ein Kreis angekreuzt werden.
unerfreulich erfreulich
unverständlich verständlich
kreativ phantasielos
leicht zu lernen schwer zu lernen
wertvoll minderwertig
langweilig spannend
uninteressant interessant
unberechenbar voraussagbar
schnell langsam
originell konventionell
behindernd unterstützend
gut schlecht
kompliziert einfach
Zurück
Weiter


Bitte geben Sie Ihre Beurteilung ab (3/3)
Bitte geben Sie nun Ihre Einschätzung der App ab. In jeder Zeile kann immer nur ein Kreis angekreuzt werden.
abstoßend
anziehend
herkömmlich
neuartig
unangenehm
angenehm
sicher
unsicher
aktivierend
einschläfernd
erwartungskonform
nicht erwartungskonform
ineffizient
effizient
übersichtlich
verwirrend
unpragmatisch
pragmatisch
aufgeräumt
überladen
attraktiv
unattraktiv
sympathisch
unsympathisch
konservativ
innovativ
Zurück
Abschicken

I. Questions of the Concluding Interview

ID: ______

Beobachtungen Studie „Collaborate“

Rot	Blau
Separation	Separation

Rot	Blau
Composition	Composition

Rot	Blau
Hybrid	Hybrid

Abschlussgespräch Studie „Collaborate“

Darauf hinweisen, dass die Teilnehmer individuell antworten können.

Ihr habt mit der App drei Mal die gleiche Aufgabe durchgeführt. Der Unterschied bestand in der Art wie die Zusammenarbeit umgesetzt war. In der ersten Variante wurden Rotation und Positionierung auf euch aufgeteilt, d.h. ihr konntet nur unterschiedliche Manipulationen gleichzeitig ausführen. In der zweiten Varianten wurden eure Eingaben kombiniert, d.h. ihr konntet nur dieselbe Manipulation gleichzeitig durchführen. In der dritten Variante konnte jeder von euch zu jeder Zeit Positionieren und Rotieren. Ich möchte euch als erstes bitten, die drei Varianten nach verschiedenen Kriterien zu ranken. Dafür vergebt ihr für jede Variante von 0-10 Punkten. 0 ist schlecht, 10 ist sehr gut.

1. **Favorit.** (a) Welche der drei Varianten hat euch am besten gefallen? (b) Stellt euch vor, ihr müsstet es nochmal machen. Welche Variante würdet ihr dann am liebsten verwenden?

Rot	Blau
(a)	(a)
(b)	(b)

Begründung für das Ranking?

Rot	Blau

2. **Absprache.** (a) Bei welcher der drei Varianten musstet ihr euch am wenigsten Absprechen? (b) Stellt euch vor, ihr müsstet es nochmal machen. Mit welcher Variante wäre dann am wenigsten Absprache nötig?

Rot	Blau
(a)	(a)
(b)	(b)

Begründung für das Ranking?

Rot	Blau

3. **Schnelligkeit.** Mit welcher der drei Varianten wart ihr am schnellsten? (b) Stellt euch vor, ihr müsstet es nochmal machen. Mit welcher Variante wärt ihr dann am schnellsten?

Rot	Blau
(a)	(a)
(b)	(b)

Begründung für das Ranking?

Rot	Blau

4. **Genauigkeit.** (a) Mit welcher der drei Varianten wart ihr am genausten? (b) Stellt euch vor, ihr müsstet es nochmal machen. Mit welcher Variante wärt ihr dann am genausten?

Rot	Blau
(a)	(b)
(a)	(b)

Begründung für das Ranking?

Rot	Blau

5. **Vor- und Nachteile.** Was sind Vor- und Nachteile der drei Varianten? Bzw. was hat die Aufgabe vereinfacht oder vielleicht erschwert?

Rot	Blau
Separation	Separation
Composition	Composition
Hybrid	Hybrid

6. **Strategien.** (a) Mit welcher Strategie seid ihr bei den drei Varianten vorgegangen? War das immer dieselbe, oder habt ihr eure Strategie an die Variante angepasst? (b) Mit welcher Strategie würdet ihr vorgehen, wenn ihre es nochmal machen müsstet?

Rot	Blau
Separation (a) (b)	Separation (a) (b)
Composition (a) (b)	Composition (a) (b)
Hybrid (a) (b)	Hybrid (a) (b)

7. **Pfeile.** Die Aktionen eures Teampartners wurden mit verschiedenen Pfeilen visualisiert. Wie hat euch das gefallen? War es eher nützlich oder eher hinderlich? Habt ihr auf die Pfeile geachtet, oder sie eher ignoriert? Habt ihr Verbesserungsvorschläge?

Rot	Blau

8. **Verbesserungsvorschläge.** Habt ihr sonstige Verbesserungsvorschläge oder Ideen für weitere Features? Habt ihre eine Idee für eine vierte Variante?

Rot	Blau

9. **Probleme.** Gab es Probleme bei der Verwendung der App, allgemeine oder in einer bestimmten Bedingung? Habt Ihre Lösungsvorschläge für diese Probleme?

Rot	Blau

10. **Gleichzeitiger Zugriff.** Stellt euch vor, ihr seid in eine WG gezogen und richtet mit der App eure WG ein. Könntet ihr euch dann vorstellen, eine der drei Varianten zu verwenden? Oder wäre der gleichzeitige Zugriff auf die Möbel für euch nicht relevant?

Rot	Blau

11. **Sonstiges.** Habt ihr sonstige Anmerkungen oder Fragen?

Rot	Blau

J. SD Card with a Digital Copy of this Work

This thesis encloses an SD card that stores a digital copy of this work.