Check for updates

A Qualitative Investigation of User Transitions and Frictions in Cross-Reality Applications

Julius von Willich
Telecooperation Lab
Technical University of Darmstadt
Darmstadt, Hesse, Germany
willich@tk.tu-darmstadt.de

Frank Nelles
PEASEC
Technical University of Darmstadt
Darmstadt, Germany
nelles@peasec.tu-darmstadt.de

Wen-Jie Tseng
Technical University of Darmstadt
Darmstadt, Hesse, Germany
wen-jie.tseng@tu-darmstadt.de

Jan Gugenheimer
Technical University of Darmstadt
Darmstadt, Germany
Télécom Paris - LTCI
Institut Polytechnique de Paris
Paris, France
jan.gugenheimer@tu-darmstadt.de

Sebastian Günther
Telecooperation Lab
Technical University of Darmstadt
Darmstadt, Germany
guenther@tk.tu-darmstadt.de

Max Mühlhäuser
Telecooperation Lab
Technical University of Darmstadt
Darmstadt, Germany
max@informatik.tu-darmstadt.de

Abstract

Research in Augmented Reality (AR) and Virtual Reality (VR) has mostly viewed them in isolation. Yet, when used together in practical settings, AR and VR each offer unique strengths, necessitating multiple transitions to harness their advantages. This paper investigates potential challenges in Cross-Reality (CR) transitions to inform future application design. We implemented a CR system featuring a 3D modeling task that requires users to switch between PC, AR, and VR. Using a talk-aloud study (n=12) and thematic analysis, we revealed that frictions primarily arose when transitions conflicted with users' Spatial Mental Model (SMM). Furthermore, we found five transition archetypes employed to enhance productivity once an SMM was established. Our findings uncover that transitions have to focus on establishing and upholding the SMM of users across realities, by communicating differences between them.

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**; **Virtual reality**; **Laboratory experiments**.

Keywords

Cross-Reality Transitions, Augmented Reality, Virtual Reality, Cross-Device Interaction, Transitional Interfaces

ACM Reference Format:
Julius von Willich, Frank Nelles, Wen-Jie Tseng, Jan Gugenheimer, Sebastian Günther, and Max Mühlhäuser. 2025. A Qualitative Investigation of User Transitions and Frictions in Cross-Reality Applications. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3713921

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713921

1 Introduction

In modern society, computers play a vital role in many productive settings. Novel technologies, such as Augmented Reality (AR) and Virtual Reality (VR) demonstrated potential in training for a wide range of use cases [9, 45, 55, 66]. With previous research demonstrating that users can transfer information seamlessly between different realities [62], incorporating different technologies into one application, leveraging their respective benefits, seems sensible. However, it is still unclear how to design ideal applications across multiple realities, and even when systems attempt to merge these technologies to capitalize on their individual strengths [63], the transitions between them are often overlooked.

While research has focused on developing and evaluating transition mechanisms between realities [36, 60], a comprehensive understanding of the underlying frictions and issues remains a critical yet relatively unexplored area, particularly in the context of AR [2]. To address this gap and gain deeper insights into these underlying frictions, we pose three research questions aimed at exploring how users perceive and interact with Cross-Reality (CR) transitions.

RQ1: Which frictions do the users perceive during transitions between PC, AR, and VR?
RQ2: What are contributing factors for these frictions?
RQ3: How do users use transitions in their workflow?

To answer these questions, we conducted a qualitative study (n=12) to explore and understand transitions between different realities. We designed a 3D modeling task that required frequent transitions between PC, AR, and VR. During the experiment, we used a *talk-aloud* approach, asking participants to verbalize their plans and impressions, reporting frictions they experienced.

We conducted a semi-structured interview after the 3D modeling task to ask about problems with the participants' workflow in the context of technology transitions and possible mitigation mechanisms. All the recorded audio from the modeling task and interview was transcribed, and the task recording was coded using a mixed inductive and deductive approach [27].

Our thematic analysis revealed 1) friction points, such as forgetting one's position in a specific reality, 2) that participants formed consistent Spatial Mental Models (SMMs) across realities, and 3) distinct transition Archetypes used by participants, such as the Gradual Transition, which uses AR as a stepping stone between PC and VR. These findings are consistent with related work finding specific transitions to be avoided, and highlight the importance of expressive CR transition mechanisms, explaining differences between realities to the user.

With our findings, we could also contextualize SMMs of transitional interfaces across realities. Based on this model, we could formulate clear design goals for CR transitions, such as minimizing or explaining changes in position, rotation, and scale between realities to the user. Furthermore, we provide multiple examples of CR applications tackling problems most commonly encountered in our user study, providing suggestions for future, transitional interface design.

Our main contributions are 1) findings in the form of themes and their connections from the qualitative study (n=12), 2) insights resulting from the thematic analysis of these findings, highlighting the importance of preserving SMMs in CR applications, and 3) an explanation on how SMMs exist in the context of CR applications.

2 Related Work

This chapter provides an overview of CR, necessary to understand where our research will take place. We explain key terminology used in this paper, highlight how past work explored Multi-Reality and Cross-Reality workflows, and give an introduction to (spatial) mental models, explaining how users understand complex applications.

2.1 Key Terminology

Since the definition of terminology surrounding Mixed Reality (MR) by Milgram and Kishino [53], it has been adapted and refined throughout the years. As such, we will clarify which terminology will be used throughout this paper to avoid confusion: We will use terminology based on the recent survey paper by Auda et al. [2]: With Mixed Reality (MR) as an umbrella term for everything on the continuum defined by Milgram and Kishino [53], including Augmented Reality (AR) and Virtual Reality (VR). VR describing an entirely synthetic environment, which can be experienced using a Head-Mounted Display (HMD). AR denoting applications overlaying virtual content onto the real world, or vice versa, merging both VR and reality. We chose to omit Augmented Virtuality, as the distinction between Augmented Reality and Augmented Virtuality is subjectively ambiguous and often related to the technology employed. Finally, we use Cross-Reality (CR) to denote interfaces and applications crossing the borders between reality, AR, and VR.

2.2 Multi-Reality Applications

Several works have researched the user experience of applications using more than one reality. We divided the existing research into multiple segments: 1) Mixed Reality featuring scenarios where the user does not transition along the reality-virtuality continuum but is presented content from different realities, 2) CR interfaces with the user acting in multiple realities, and 3) transition mechanisms, aiding the user when they move along the reality-virtuality continuum.

2.2.1 Mixing Realities. MR approaches can work in different directions on the reality-virtuality continuum. Either by including real-world content in VR or by overlaying virtual content onto the real world. Automated approaches for specific objects, such as coffee mugs [18], writing pens and hands [3], or passers-by [52, 74] exist. But also approaches giving control to the user on where they want to see content from a different reality, such as "RealityLenses" [76] or "MagicLens" [16], aim to soften the border between realities. Mixing a real environment, AR, and VR, Ayyanchira et al. [4] built an application to navigate through a real-world building. Using a small scale AR/VR model, the user can overlook the whole building, which can also be used in 1:1 scale using a smartphone or HMD.

Another group of approaches tackles this disconnect between the virtual and the real world from the other direction, overlaying virtual content onto the real world. They either use the real world as input [67] or to provide context [6]. Using the real world as input allows for more intuitive interactions [51] while using it to provide context allows for more intuitive understanding [11, 21, 35, 71] of data. Cheng et al. developed an application that automatically arranges AR interfaces in the real world, drawing from the benefits of this intuitive, spatial understanding [22]. "ModularHMD" allows the user to add or remove modules on an HMD to select between a fully virtual and a real-world view [32]. The authors used one HMD with three removable modules around the user's field of view, each allowing four states: occluding (to get a narrower field of view), a wider field of view in VR, video see-through of the real world, and removed.

From these approaches, it is clear that mixing interaction paradigms in order to draw from their respective strengths is a promising approach. However, as Chioss et al. have shown, AR interfaces can also be detrimental to cognitive load. Thus, the decision to use them must be well-considered [23]. Overall, it is clear that there are still open challenges in the field of MR, including "more mature and better-tested HCI concepts" [24].

2.2.2 Cross Reality and Transitional Interfaces. Multi-device applications are a good entry point for understanding how users think across multiple realities.

Brudy et al. [17] provide a valuable taxonomy of this research's extensive body, establishing multiple categories for classification, as well as showing that users use different devices for specific subtasks. This separation might also exist for different realities in CR tasks. Jokela et al. studied the use of multiple information devices, like smartphones and laptops, in everyday tasks [42]. They found that if multiple devices were used in parallel, they were usually used for multiple perspectives on the same task, dedicating each device to a specific view or application. Cools et al. [26] explored extending interfaces beyond a normal screen using a desktop-AR prototyping framework. Within their prototype, objects created and manipulated on PC could be moved seamlessly into AR to get a 3D perspective of the previously flat representation. They describe the ability to do more detailed work on the PC and to have a more natural view of the objects in AR, which supports the spatial model of the user, a concept influencing our study design.

While AR is usually employed to display information across realities, CR and transitional interfaces [20, 29] allow users to also act across realities. The corresponding design space, as defined by Wang et al. [79], shows the wide applicability of these approaches. Employing them to effectively transfer information and act along the reality-virtuality continuum [62] enables the design of interfaces drawing from their individual strengths.

An early example of CR transitional interfaces is the "MagicBook" by Billinghurst et al. [8], showcasing the low barrier of entry these interfaces have to offer. Roo et al. created "One Reality" [63], leveraging interactive surfaces, AR and VR for their respective strengths in different stages of a continuous workflow. However, they did not look into transitions between realities. Cools et al. [25] present five different interaction methods to transition objects between AR and VR. Their "Auto Blended Space" method, rated best by users, blends the source and target realities when the user grabs an object in the source reality, allowing the user to seamlessly move into the target reality along with the object. Extending CR with tangible user interfaces, Kaimoto et al. [44] implemented an application using tangibles across realities. Their application allows users to draw digital sketches in AR influencing tiny robots that act as tangibles and vice versa. Jung et al. [43] implemented "In Limbo" to transition users from reality to VR to establish a mental link between the user's own body and their virtual body in VR, suggesting a connection between different realities. Wang et al. showcased "Slice of Light" [77], allowing users to move between different VR instances, featuring VR to VR transitions in a multi-user environment.

CR approaches can be especially powerful if they employ interaction paradigms users already learn in their daily lives. Approaches such as interactive, modular architecture models [65] reporting their state to an analyzing software. Or "DualCAD" by Millette et al. [54], a CR CAD application, using AR and motion controls, and "BISHARE" [82] which uses a smartphone as input for a CAD application, similar to "DualCAD". Using a smartphone or tablet is a common use-case, showcasing how applications can benefit from cross-device interaction [5, 12, 30, 31, 48]. These applications were generally well received for their ease of use and learnability.

The success of these applications in the design space of CR shows the potential for future transitional interfaces. However, we still need a suitable understanding of how to design them in an engaging and usable manner.

2.2.3 Exploring Transitions. If applications leverage different realities' strengths, users need to switch reality often, depending on their current task. For transitional and CR interfaces to support the user efficiently, these transitions must work without hindrances, keeping the user in their current workflow. Thus, exploring and understanding the frictions of transitions between realities is key. From their elicitation study, Piumsomboon et al. [59] gathered a wide variety of MR disengagement mechanisms suitable for transitioning users out of MR. Knibbe et al. [47] evaluated the moment of exiting VR, seeking to understand what exactly happens from the user's point of view at that moment. They found that participants mentally prepared their transitions and that spatial disorientation played a vital role, a result we could also find in our study.

In contrast to transitions back to reality, transitions across different realities offer a richer design space. These transitional interfaces are still an active research area with many interesting approaches [7, 40]. George et al. conceptualized different transition methods, evaluating two portal-based metaphors in a search task across realities [36]. They concluded that a hand-held portal metaphor can support transitions between realities, by providing a link between realities, a theme we also found in our study. Pointecker et al. evaluated the influence of different transition mechanisms on participants' performance in an information-gathering task across multiple realities [60]. They found a simple fade to work best for productivity applications, while a portal metaphor is best suited for hedonic applications, such as games.

Wang et al. [78] explored the impact of state synchronization on transition perception and behavior. They concluded that users in general try to avoid transitions as much as possible, with synchronization of states between realities lowering this friction, a statement consistent with our findings. Schröder et al. have thoroughly analyzed the interplay between two users moving along the reality-virtuality continuum, solving a common task [64]. They found that users avoided specific transitions rather than realities. For example, transitions from VR to PC were avoided, while transitions from AR to PC and VR to AR were used readily. This also matches our findings.

However, the topic of how to transition a single user between realities is not yet conclusively examined, as researchers and designers cannot draw on their previous experience with traditional interfaces [80], as CR frictions are not well understood.

2.3 Mental Models

In order to understand how users interact with a specific application, we need to find a suitable representation of how they form this understanding. One suitable theory is mental models, explaining the interplay between the user's idea of how an application works and how the application is designed to be used.

Johnson et al. describe a *Mental Model* as a "small-scale model" containing a user's internal representation of an external reality [41]. Users form these models based on their assumptions of how different parts of, e.g., a computer application relate and influence each other [19, 46, 58].

In our paper, we will focus on Spatial Mental Models (SMMs), describing spatial relations between different systems, as the participants' understanding of the system's assumed inner workings was not our main concern.

2.3.1 Spatial Mental Models. Besides user interfaces, a *Mental Model* can also represent other concepts. One prominent example is the *Spatial Mental Model (SMM)* that preserves coarse spatial relations of an environment coherently [73]. Prior research has found spatial relations can be easily comprehended from language [34, 70] as well as from direct experience. Recent research also found humans can form SMMs from virtual environments [37, 72, 75] and that suitable design can strengthen these SMMs [1].

2.3.2 Research Gap. Both the *Mental Model* and SMM are relevant in the context of transitioning between different realities. The *Mental Model* describes how the interface changes and the SMM describes how the space around the user changes. Although existing interface design theories can be employed to optimize the user's *Mental Model*, the role SMMs play in CR transitions is still not understood.

3 Method: Task, Apparatus, and Procedure

To gain insights into what happens before, during, and after a reality switch, we created a CR application and accompanying tasks aimed to induce transitions between PC, AR, and VR. For example, participants would be tasked to use their body as a scale reference in VR, then switch to AR to adjust accordingly.

When participants switched between realities, we asked them to *talk-aloud* and report why they did it. By recording their responses in combination with a semi-structured interview, we formed a body of transition data to answer our research questions.

3.1 Spatial Interaction Task

The task we gave the participants was to create a statue in the city center of Darmstadt. The task was divided into ten sub-tasks, in combination with separating capabilities between different realities, this induced additional transitions. The subtasks included building the statue using PC, placing the statue in the city using AR, and observing it in the city context, using VR. We designed the subtasks with the goal of eliciting transitions by requiring the use of different capabilities, spread across the different realities. The tasks were handed out on flash cards and the participants were verbally reminded of the task when they asked. The exact sub-task descriptions in English and German are listed in Appendix A.

3.2 Spatial Interaction Apparatus

The objective of this apparatus was to enable the spatial interaction task across three realities. While previous work suggested not to make capabilities unique to one reality [54], eliciting transitions was our primary concern. This artifact encompasses a traditional keyboard-and-mouse interface (PC), an AR, and a VR component. For AR and VR we used a Varjo-XR3 HMD, enabling seamless switching between AR and VR. The headset was connected to the PC under the desk by a cable. The interaction was facilitated using two HTC Vive controllers. In addition, we utilized a 1:1000 scale model of Darmstadt, placed on a table, so it was within reach.

Users were tasked to create an abstract statue in the city model. The city model also served as a reference in AR. All three modes, PC, AR, and VR had different capabilities to encourage more transitions between realities, similar to the approach by Schröder et al. [64]. However, we decided to replace the tablet-based AR interface with an HMD-based interface to provide both very simple and very complex processes for transitioning hardware-wise. The application featured a unified workspace, allowing users to modify the same statue across all modes.

3.2.1 Interplay of the Different Realities. The apparatus spans three different realities, PC, AR, and VR, which are explained below. Users could freely transition between realities whenever they wanted. When transitioning away from PC, users had to put on the HMD,

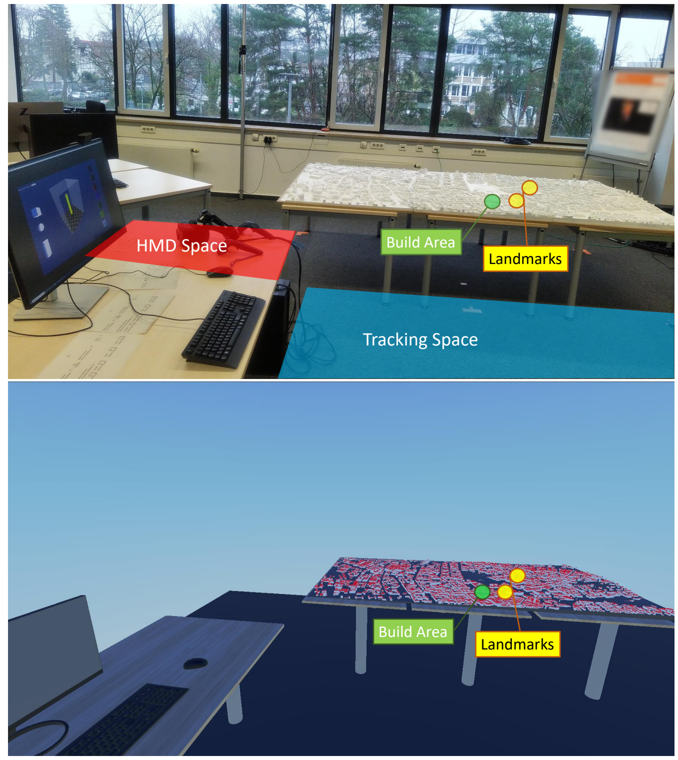

Figure 1: Pictures of the study setup. Representing the real world on the left and the virtual copy on the right. Participants could place down the HMD and controllers on the table (red) and could freely move in front of the city model (blue). The two landmarks were marked with small stickers, highlighted in yellow in this picture. The area the statue would be built in is shown in green.

which automatically detected if it was worn and changed to the appropriate reality. The HMD saved the last used reality, returning to it when put back on. Participants could change this state by clicking a button in the PC interface or pressing a physical button on the side of the HMD when worn. Returning to PC was achieved by simply taking off the HMD.

The whole working area in which the statue was supposed to be built was contained in a see-through cube, which we call EditZone. This EditZone was faintly visible in all three realities, could only be moved in AR, and the PC camera always pointed at it. Note that the EditZone was chosen to be unobtrusive. As such, it is barely visible in still frames but could be observed more easily in moving 3D. Its position and rotation were updated in all realities when it was moved in AR.

When switching to VR, the participant returned to the position where they left VR. This was done because it ensures that actions in other realities do not affect the state of the VR environment apart from the statue. Furthermore, adding a different behavior would mitigate issues this setup is supposed to uncover. At the start of the study, the city model in VR was correctly aligned with its real-world counterpart and thus AR.

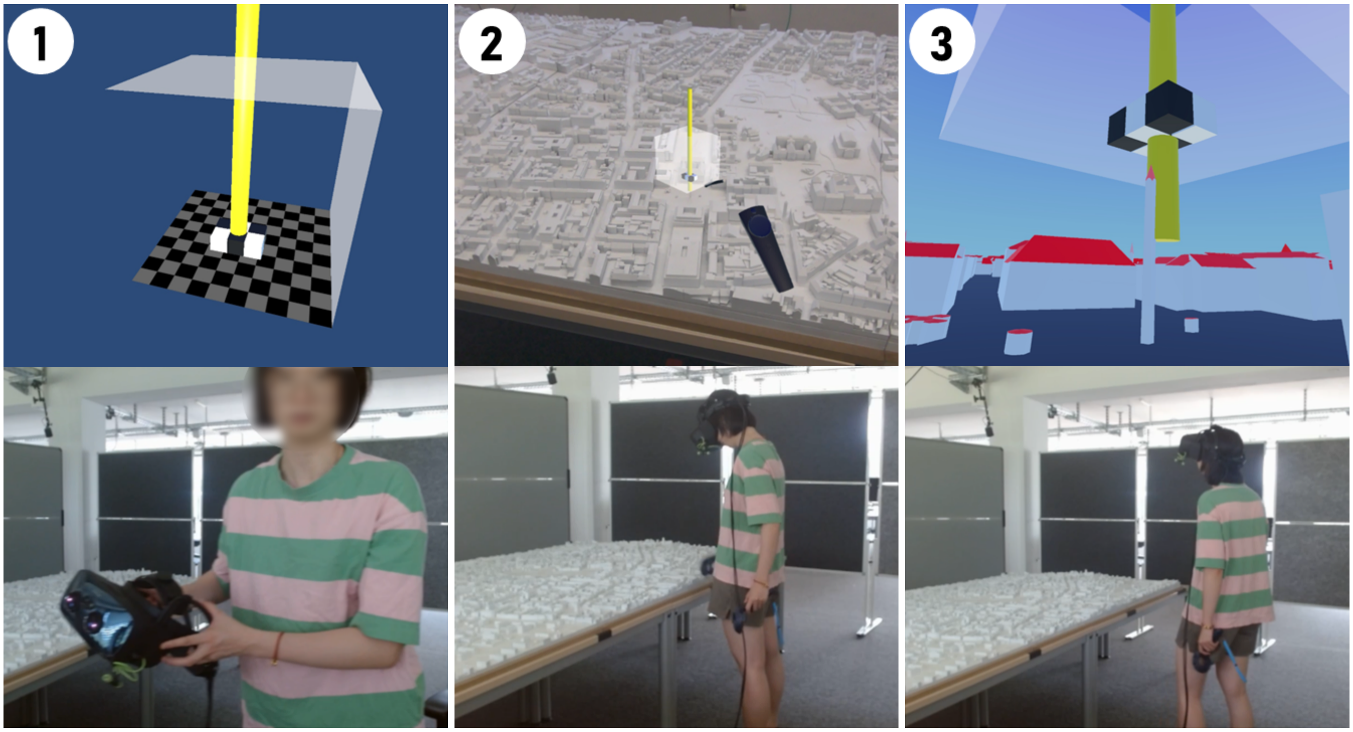

Figure 2: Example of the three realities used in our apparatus. 1) Shows the participant preparing to move from PC to AR, 2) shows the participant in front of the city model, looking at their statue in the overall city context, and 3) shows the participant looking at their model within the city at a 1:1 scale. The opacity of the box indicating the EditZone was increased for better visibility in the screenshots.

3.2.2 PC. The PC part of the application features a simple, CAD software-inspired interface. Here, users could translate, rotate, scale, color, place new, and delete existing primitives. The viewpoint could be changed by moving a virtual camera with the mouse. In this mode, users could not see the digital city model; as such, they needed to use either AR or VR to see the statue in the city context.

The users controlled most operations via keyboard shortcuts (single key). Only the operations for placing new primitives and selecting a color were implemented as clickable items within the GUI. For the translate, rotate, and scale operations, gizmos were shown attached to the object. They enabled interaction with the primitive in the respective mode, comparable to the object manipulation gizmos available in Unity. The PC interface is depicted in Figure 3.

New primitives spawn in the middle of the EditZone on a checkerboard-patterned floor. The checkerboard pattern is only visible in the PC view, serving as a frame of reference. The rest of the walls are only slightly visible to not inhibit the user.

We chose these capabilities, as we found PC interfaces to be well-suited for quick and accurate positioning tasks. We chose to only allow coloring in this mode in order to encourage additional transitions, even though AR and VR are also suitable. The desktop was configured for seated interaction, and an office chair was provided.

3.2.3 AR. In this mode, participants could translate and rotate primitives with the controllers, but scaling was only possible for the EditZone and, thus, the whole statue. Furthermore, users could move the EditZone in the physical city model, enabling them to view it in the greater context of the city.

Controlling the interface was done via the HTC Vive controllers' buttons, the same applies to VR.

At the beginning of the study, the EditZone was placed on the city model in a spot unrelated to the task. The users could not see the PC UI in this mode, as we turned the screen black while the HMD was worn. This was done to clearly track which reality participants were in at any given time.

3.2.4 VR. In this mode, participants could scale and translate the city model around them, effectively allowing them to move through the city at any scale they desire. A simple point-and-teleport locomotion was also implemented. As such, this mode enabled them to correctly grasp the scale of buildings and their statue and judge obstructions from different viewing angles. The statue as a whole could not be scaled or otherwise modified. However, single primitives could be translated and rotated.

The colors of the city model were slightly changed in comparison to the real model to make the buildings more discernible: the roofs were colored red, and the floor was shaded by in-engine lighting (Figure 2.3).

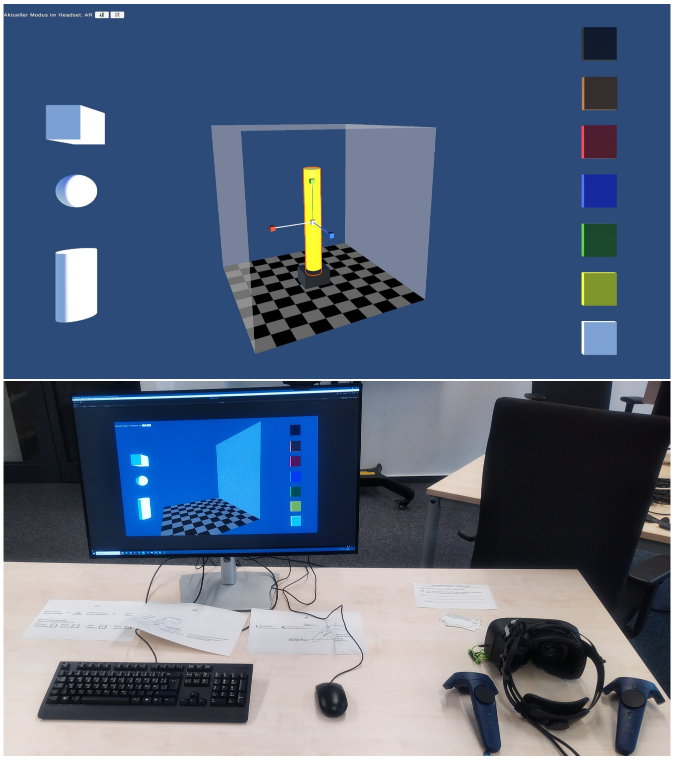

Figure 3: PC setup of the apparatus. Left the view participants had when using PC input, with primitive selection on the left, EditZone with manipulation gizmo (visibility of the surrounding cube exaggerated for better visibility), and color selection on the right. The picture on the right shows the setup in the real world, including input cheat sheets and task flashcards.

To make the participants more confident not to walk into obstacles, we added the desk with the computer and the table holding the city model in VR. This was necessary, as once the city model was moved, there would have been no indication of where the tables were. The participants could disable these tables by pressing a button. The desk and tables were re-enabled with every transition to VR to minimize the risk of participants walking into them.

3.3 Procedure

When the participants arrived in the lab, we provided a consent form, explained the experiment, informed them that they were being recorded, and could quit the experiment at any time without consequences. We introduced them to the concept of the *talk-aloud* approach, walked them through the different user interfaces and provided cheat sheets for the controls should they need to refresh their memory. Furthermore, we explicitly mentioned that the study was designed to elicit more transitions by splitting up the application's functionality across realities and by providing the task piece-wise so that they could not bundle work packages for each reality. We allowed the participants to get used to the application before giving them their first sub-task.

During and directly after transitions, participants were asked to verbalize their plans and impressions, reporting the friction they experienced. We refrained from interrupting participants mid-work step so as not to distract them from the task at hand. A negative side effect of *think-aloud* approaches, also highlighted by Nielsen et al. [56, 57]. Thus, we used a *talk-aloud* process as opposed to a *think-aloud* process to improve data quality.

After they finished the task, we conducted a semi-structured interview, asking about transitions in detail, problems encountered, and suggested solutions or missing features. At the end, the participants filled out a questionnaire about their person and previous experiences in 3D Modeling, AR and VR. The procedure lasted one to two hours, depending on the participants' progress during the *talk-aloud* study.

The study was conducted in compliance with our university's ethics guidelines.

3.4 Recording

The *talk-aloud* study and the semi-structured interview were recorded. We used a webcam facing the user on the PC and the area used in AR and VR. The PC and HMD interfaces were recorded as screen captures. Additionally, we used four microphones across the working area: One in the webcam, one attached to the HMD, one above the city model in the middle of the room, and one in front of the interviewer. The application recorded transitions, including their directions and timestamps.

The individual audio recordings of the *talk-aloud* studies and the semi-structured interviews were transcribed using whisper[1]. They were manually merged and corrected when the automated approach did not yield satisfying results. For the *talk-aloud* transcripts, transitions were added at the correct time it was performed, including the information from which to which reality participants transitioned. These transcripts formed the basis for our qualitative analysis.

3.4.1 A Comment on Translation. The study was performed in either German or English, depending on the participant's preference. We decided to leave all transcripts in their respective languages since the authors involved in coding speak both languages. While translation is possible and has been done before [13], we decided to avoid it, as even partial translations [68] cost time and may introduce inaccuracies [38]. We created the codebook in English to avoid inaccurate translations after the coding process, as suggested by Esfehani and Walters [33].

3.5 Participants

We recruited 12 participants via word of mouth from different university institutes until no new insights were gained, thus reaching saturation [39]. Eight participants identified as male, three as female, and one as agender. They were between 22 and 41 years old ($M = 28.8y, SD = 6.7$). Five of the participants were researchers, the others being students. Participants were asked for their experience with AR, VR, CAD/3D modeling, and their geographical knowledge of Darmstadt on a 5-point Likert-scale (1=no experience, 5=expert). The responses were diverse, averaging in medium

[1] https://github.com/openai/whisper

knowledge (AR: M=2.50, SD=1.19; VR: M=3.00, SD=1.29; CAD/3D: M=3.16, SD=1.14; geography: M=3.08, SD=1.11). Six participants wore glasses, three wore them during the study, the others did not. Two of the participants reported having a red-green color weakness. Three participants performed the study in English and nine in German. Each participant's run took between 60 and 90 minutes, depending on the level of detail with which they wanted to follow the task instructions and how well they coped with the different realities and transitions. No compensation was offered in accordance with local regulations.

3.6 Qualitative Data Analysis and Coding

With the prepared transcripts, we performed a descriptive, mixed deductive and inductive coding approach [13]. While these labels help roughly describe the approach used, we will still explain how exactly we approached the analysis to avoid confusion, as qualitative research is often custom-tailored to the question at hand and researchers employing it, as explained in detail by Braun and Clarke [14].

After three authors got a first impression of all transcripts, we held an initial code-finding session. In this session, we created a first, thematically structured codebook focused on describing transitions. This initial codebook included the concept of participants forming a *Mental Model* across realities, as well as different transition Archetypes.

With this codebook, one author pre-labeled the first three *talk-aloud* transcripts, adding new codes as they appeared in the text. These three transcripts and new codes were discussed among the same three authors to create a unified understanding of the subject. Afterward, two authors coded five transcripts each, with one transcript as overlap to compute the Intercoder Coefficient, which we computed as 0.64 using Brennan and Prediger [15]. Disagreements in this transcript were then resolved together, resulting in only one coded document per transcript.

In every case, unclear transitions were marked and discussed with the other authors. The video recordings were used throughout the coding in situations where the transcripts did not suffice, and the actions of the participants were important.

The final codebook, including criteria, can be found in Appendix B.

4 Results and Observations

In this section, we present common themes identified in our thematic analysis about how participants perform and perceive transitions across VR, AR, and PC. We will start by presenting the frictions we identified, discuss which factors contribute to them, and which strategies participants employed to solve their task in a CR workflow. Quotes in this section are verbatim if the participant spoke English or translated analogously from German if the participant spoke German. Participants are labeled with P01 to P12, and timestamps are given at 30-second intervals. Codes from our code book, see Appendix B, will be written in SMALL CAPITALS.

4.1 Frictions (RQ1)

Overall, we could identify three main frictions during our experiment: 1) disorientation, 2) Fear of the Unknown, and 3) (physical and mental) effort. The first two frictions are inherent to CR transitions, while effort can mainly be attributed to the technology employed and our apparatus.

- **Disorientation** Participants often experienced DISORIENTATION immediately after transitioning between different realities, not understanding where they were in the new environment. This was especially prominent between AR and VR.
- **Fear of the Unknown** Not knowing what to expect can be very deterring [61], an effect we could also observe in our experiment. Participants reported negatively on being uncertain where they would appear after a transition to VR. Some participants further reported avoiding transitions if they were not sure what they would see in the other reality.
- **Effort** Finally, participants also talked about how discomfort and effort made them avoid transitions. They reported delaying taking off or putting on the HMD as long as possible. They even completely omitted transitions, using more strenuous alternatives, such as reading text in the AR video pass-through instead of taking off the HMD.

While identifying these frictions is a vital step toward understanding how users perceive CR transitions, it is also important to understand which factors contribute to them. Understanding how to minimize these frictions and how different factors contribute to them is crucial.

4.2 Contributing Factors (RQ2)

This section will discuss the individual contributing factors we identified and which frictions they contribute to. In Table 1 we provide an overview of which codes commonly occurred together and which frictions the associated problems influence.

4.2.1 Uncertainty. Participants talked about how they did not know what to expect in the beginning of the experiment and how this lack of EXPECTATION influenced their transition behavior. "*[After one of the first transitions from AR to VR] I still don't know where I will end up*" (P05, Minute 24:30). They also reported not knowing where they would influence their transition behavior, which contributed to their Fear of the Unknown. "*[Transitioning from PC to VR via AR] I don't want to go to VR again. Because I don't know what I will see. And I will lose the sense of direction*" (P04, Minute 53:00), or "*[Before transition to VR] I'm still learning where I will appear [after a transition to VR]*" (P05, Minute 73:30).

In general, participants usually had fewer problems accurately voicing EXPECTATIONS regarding their POSITION the longer they used the application. For example, one participant stated, "*[Before transitioning from PC to VR] I expect that I will be standing on [landmark] when putting on the headset. [Transition to VR] Good, that worked*" (P01, Minute 09:00). We observed that forming accurate expectations counteracted both disorientation after a transition and Fear of the Unknown before a transition. Especially towards the end of their trial, participants started formulating their EXPECTATIONS when transitioning. They usually commented on how everything worked as EXPECTED and how this made transitions more pleasant, e.g. "*[Transition to VR] This transition now was pleasant. I roughly knew where I am standing, and I saw what I expected.*" (P08, Minute 23:30). This ability to plan suggests the existence of some kind of

Code Group	Code (count)	Occuring with (count)	Frictions		
			D	FoU	E
Acting Across Realities	Anchor Point (44)	Alignment (6), Position (5)	×	×	
	Real World Adjustment (9)	-			×
	Task Preparation (9)	Anchor Point (3)		×	
Discomfort	Cognitive Effort (6)	-			×
	Cognitive Strain (13)	-			×
	Physical Effort (15)	-			×
	Physical Strain (14)	-			×
Occurrences	Disorientation (20)	Position (9), Expectation (7), Consistency (5), Forgetting (5)	×	×	
	Forgetting (29)	Position (6), Disorientation (5)	×	×	
Requirements	Consistency (48)	Expectation (24), Position (21), Rotation (10)	×	×	
	Control (5)	-		×	
	Expectation (56)	Position (26), Consistency (24), Rotation (10)	×	×	
Spatial Mental Model	Alignment (21)	Anchor Point (6), Consistency (6), Position (5)	×		
	Different Perspective (20)	Alignment (4), Position (3), Consistency (2), Scale (2)	×	×	
	Position (78)	Expectation (26), Consistency (21), Disorientation (9), Rotation (7)	×		
	Rotation (21)	Position (13), Expectation (10), Consistency (7)	×		
	Scale (14)	Expectation (5)	×		

Table 1: Representation of relevant codes, grouped by the code groups used in the code book. The number in brackets denotes the number of occurrences and joint occurrences. The last column (Frictions) denotes which frictions these codes are relevant to. D = Disorientation, FoU = Fear of the Unknown, and E = Effort

SMM the participants use to understand how the different realities are connected.

4.2.2 Forgetting. Participants reported FORGETTING details from realities they were not currently in. A common theme was participants remarking on FORGETTING where they left VR, and losing orientation when returning to VR, disorienting them. "*[In VR] I simply forgot where I was standing*" (P09, Minute 27:00), or "*[After switching to VR] Where is my cube?*" (P01, Minute 05:00). In some cases, participants avoided transitioning to VR as FORGETTING their position in VR increased their Fear of the Unknown.

We could also observe participants returning to a reality they just left because they needed a detail they FORGOT about, or reverting because they lost their bearings "*[Interviewer] Why did you transition back [from VR to AR]? [Participant] I wanted to see about where [landmark] is*" (P05, Minute 62:30).

4.2.3 Alignment. When participants mentioned ALIGNMENT, they often noted slight offsets between the city model in AR and the city model in VR or between the real and virtual tables. These offsets were caused by tracking errors, resulting in a drift between reality and MR over time. However, they also talked about ALIGNMENT owed to their interaction and not technical issues of the apparatus.

Participants reported that it was easier for them to transition from AR to VR when the city models in both realities were in ALIGNMENT. This usually happened at the beginning of the study when the real and the virtual city model in AR and VR were still in their initial, aligned configuration. We could also observe participants having less problems switching between AR and PC. These realities were inherently aligned in our apparatus since the video see-through shows the real world as is. "*[After transitioning from AR to PC] Here, I always appear where I think I will appear. I find the transition from AR to PC the easiest*" (P05, Minute 40:30).

Participants reported getting lost more easily when the city model in AR and VR did not align. This would occur after they moved around in VR, or when they ended up in a different location or scale than they EXPECTED, becoming disoriented.

4.2.4 Consistency. Participants commented positively on CONSISTENCY helping them with orientation. They commented either positively on how the position stayed CONSISTENT and as EXPECTED between AR and VR, or how they got disoriented if it did not. When participants talked about CONSISTENCY without explicitly mentioning EXPECTATIONS, they remarked on past transitions being good or pleasant because "everything" stayed "the same".

Interestingly, the participants' notion of CONSISTENCY changed over the experiment. Especially at the beginning, CONSISTENCY was strongly connected to ALIGNMENT, as participants expected the city model in VR to always re-align with the physical city model in AR. They expected the VR model to stay CONSISTENT with the physical model. Later during the experiment, they used CONSISTENCY to refer to their position in VR staying the same after switches. "*[In VR] That is what I expected before, if I am working at a specific spot, and want to look there again and again, this [appearing in VR in the same position as you left] is very sensible*" (P09, Minute 29:00).

Both notions of CONSISTENCY helped participants predict their positions across realities, reducing their Fear of the Unknown. Multiple participants also remarked that AR and VR are DIFFERENT PERSPECTIVES of the same environment. Especially towards the end of the experiment, this allowed them to avoid the effort of transitions, as they could simply imagine a perspective they would experience in VR by looking at the model in AR. "*[Interviewer] You did not switch to VR now to look at it? [Participant] I could see in AR which direction it is. [...] I am looking at it from below [from the imaginary position on the model]*" (P04, Minute 42:30). In general, participants had fewer problems predicting POSITIONS the longer they used the application.

We could also observe a consistent state to be perceived as inconsistent because of missing cues. For example, when participants rotated the EditZone in AR, this was also reflected in the PC interface, however as the city model was not visible on PC, participants did not seem to connect those two interfaces.

Participants commented negatively on the inconsistency of their statue's initial POSITION, which was neither linked to the PC screen nor to the area they were supposed to work on in the city. Another negative experience with CONSISTENCY was a small offset between the statue's POSITION between AR and VR which happened due to tracking inaccuracies for some participants, e.g. "*[In AR] I've put it on there now, but of course it's not exactly the same. The positioning is not accurate.*" (P10, Minute 50:30).

4.2.5 *Strain.*

Participants reported STRAIN being introduced when taking off or putting on the HMD. Especially participants with glasses and long hair reported having problems while taking off or putting on the HMD as they brushed the glasses off their face or their hair got tangled in the HMD strap. Another more long-term STRAIN participants reported on was the HMD getting warm and heavy over time. And needing to STRAIN their eyes to read the flashcards using the video pass-through.

From a mental standpoint, participants reported that they had to actively think about consolidating the different realities. Participants reported doing this either proactively before a transition or reactively after switching realities. The ability to proactively consolidate different realities strongly suggests the existence of an SMM, holding both realities and relating them to one another.

These additional STRAINS contributed to the overall, perceived effort for switching realities.

4.3 Transitions as part of the Workflow (RQ3)

When it comes to how users employ transitions in their workflows, we identified three main motifs directly linked to working across multiple realities. These motifs can be characterized as 1) connecting realities to make reality switches easier, 2) fetching information from a different reality, and 3) preparing actions in different realities. These different reasons also align with the different frictions we could observe and show how participants tried to improve the connection between different realities. We could also identify unique types of transitions in those groups, which we call Archetypes, as they also appear as codes in our code book (see Appendix B). They will be highlighted with SMALL CAPITALS. They are defined by their intention and by how many realities and switches are involved. While the reasons for transitioning are biased by our study design, we can reflect on higher-level motifs participants had for using these Archetypes.

4.3.1 *Common Transitions.*

In addition to the three main motifs, participants transitioned without an ulterior motif. We called this type of transition DEFAULT, as the transition itself is the aim of the transition, without any additional motifs. This transition appeared equally between all realities, contained only one reality switch, and comprised most of the labeled transitions.

We could also observe participants transitioning on ACCIDENT when they did not intend to perform a transition or ended up in the "wrong" reality. This Archetype appeared the least between VR and PC. Between AR and VR, this Archetype happened when participants pressed the transition button by ACCIDENT and, in some cases, by habit. Between PC and the other realities, this Archetype happened when participants meant to click the button in the PC UI to change which reality they would transition into but forgot to actually click it.

4.3.2 *Connecting Realities.*

A Common behavior we could observe, providing a connection between realities, was using AR as an intermediate step when moving from PC to VR or vice versa. During these transitions, participants usually used the video pass-through to pick up or put down the controllers. They also mentioned that this intermediate step made it easier to switch between realities and participants used this transition type less the longer the experiment was running. We called transitions with this pattern of using AR to interact with the peripherals and the working environment as GRADUAL TRANSITION if participants did not use AR as part of their modeling workflow.

Apart from simply transitioning through AR, participants also made quick adjustments or refreshed their memory during these IN-BETWEEN transitions. We called transitions IN-BETWEEN PEEK if participants took a quick look at their statue or the model, and IN-BETWEEN TOUCH if they made small adjustments in AR before finally switching to VR or PC. The two sub-Archetypes appeared mainly from PC to VR, and not as much from VR to PC.

Typical use cases included participants putting on the HMD at the table, moving towards the physical city model, quickly looking at the statue to regain their bearing, and then transitioning into VR. Or participants making small adjustments to the statue when arriving at the physical city model, then immediately entering VR.

During these transitions, participants often used what we call ANCHOR POINTS to strengthen the connection between realities further. While most participants used the controllers as connecting elements between the different realities, some also used the tables mirrored in VR. They commented on how it helps to have something to "take along", usually during intermediate transitions. The tables especially also served as a scale reference, consistent between the real world and VR. Participants could accurately judge the real table and, assuming a consistent size, used the virtual tables to bring real-world measurements to VR. Although participants had to get used to it first, some participants put down their controllers on the virtual tables towards the end of the study instead of first switching to AR, commenting positively on it. "*[Interviewer] How did this work for you without AR as an intermediate step? [Participant] Good, I did not expect that I would like the table. But it was astonishingly easy.*" (P08, Minute 49:30).

4.3.3 *Preparing Actions.*

We could furthermore observe participants deliberately preparing actions before switching realities. For example, participants either pre-placed primitives on PC to use in a different reality or moved real-world objects to not collide with their surroundings when in VR, such as the chair or HMD cable. Some even used primitives to temporarily mark one side of the statue, which they deleted later. These preparations happened on a bigger scale, thus not directly related to a single transition.

Participants also exhibited pre-planned actions on a per-transition level, showcasing quick transitions back and forth between two realities. These transitions came in two variants, the simple QUICK TOUCH and the iterative, more complex QUICK WORKFLOW. We called transitions QUICK TOUCH if participants switched to a different reality, made a small adjustment, and then returned to where they started. Participants exhibited this behavior, for example, when they saw a small misalignment in VR and switched to AR to move the statue a bit to the side, returning to VR to continue with their original task. Their time in the second reality was usually short, and the adjustment they wanted to make was seemingly already planned before they initiated the transition. This was evident from their comments and also from how deliberately they worked in the different realities.

A similar behavior, which was observable, especially after participants had some time to get acquainted with the software, involved multiple switches back and forth. This usually happened when they needed VR to confirm the outcome of their actions and AR to make adjustments. In these cases, CONSISTENT states between realities were especially important and participants mentioned when their EXPECTATIONS were confirmed or violated. We called these transitions QUICK WORKFLOW, one instance of which could contain many reality switches, with the longest encompassing 19.

4.3.4 Fetching Information. Another dominant group of transitions with discernible motifs are those employed to access information from a different reality. This group comprises three distinct Archetypes: QUICK PEEK, REALITY QUICK PEEK, and LAZY REALITY.

QUICK PEEK describes situations when participants switch to a reality to look at something and switch back without performing any other actions. With 84 transitions assigned, it is the biggest Archetype in our data set. Participants used this type of transition when they needed to refresh their memory about where their statue was situated in the city or when they needed to look up small details. This transition occurred almost exclusively between AR and VR.

REALITY QUICK PEEK is similar in concept to QUICK PEEK but was used by participants to refresh their memory of their real-world surroundings rather than the statue. Typical use cases were orienting themselves in the real-world room or untangling from the HMD cable. Both REALITY QUICK PEEK and QUICK PEEK encompass two reality switches, into one reality and back again. They remarked that for quickly reading the flashcards or addressing the experimenter, this worked fine, but they would take off the HMD for longer work steps.

If the transition only encompassed one switch, without a return to the original reality, we called them LAZY REALITY. In these cases, participants often remarked that they did not intend to take off the HMD as it was cumbersome, and they did not know yet which reality they needed next.

Overall, these Archetypes and behaviors demonstrate participants finding methods to mitigate some of the problems they are presented with by our apparatus. As we mentioned above, some problems originate from the technology, and others from the deliberate apparatus design, prompting frequent reality switches. However, while the problems are amplified by our design, their root causes are still relevant for CR transitions, giving a clear indication of which problems need to be addressed in future CR applications.

In the following chapter, we will discuss how the different frictions, contributing factors, and mitigation attempts suggest the existence of (multiple) SMMs and breaking thereof. We will also discuss how future transition mechanisms can account for these breaks, conserving the users' SMMs. We will highlight how existing approaches have already addressed these issues.

5 Discussion

Through our study, we identified three main motifs in transitions between VR, AR, and PC environments: connecting realities, preparing actions, and fetching information. This classification allowed us to assess underlying problems and understand participants' coping mechanisms in CR applications. Notably, fetching information emerged as the most common reason for CR transitions, highlighting the importance of making information readily available across realities as a key design goal. Additionally, we observed that ANCHOR POINTS play a crucial role in connecting different realities, helping users form a cohesive Application SMM.

We use the concept of Spatial Mental Model (SMM) to interpret how participants coordinate transitions between realities. Our results indicate two SMMs. The Application SMM consists of the virtual environment across VR, AR, and PC, while the Real-world SMM contains the physical room. Based on our results and related work [10, 49, 81], the Application SMM is malleable over time in the study. Participants might continuously gain or lose information in their SMMs. In the following, we outline the individual problems, how they relate to SMMs, and provide suitable solutions based on both existing work and our results.

5.1 Understanding User Needs in Cross-Reality Transitions

One core motivation for transitions was to fetch information from a different reality. Participants were aware of multiple realities and structures but could not remember all the details. Therefore, they used different transitions to perform quick actions that they had prepared beforehand. Archetypes associated with this behavior are IN-BETWEEN TOUCH, QUICK TOUCH, and QUICK WORKFLOW, with participants only remaining in the reality they acted in for a short time. They used QUICK PEEK, IN-BETWEEN PEEK, and to some degree QUICK WORKFLOW to get the information they forgot or were uncertain of or how their actions had affected a different reality.

We frequently observed this behavior later in the task once participants had developed a good understanding of how the realities were connected and had become familiar with the system. For example, they would adjust the statue using the PC interface and then briefly hold the HMD to their face to view it in VR without standing up. These quick updates allowed participants to obtain necessary information from another reality and continue their task in the current one. Additionally, participants often employed Lazy Reality to avoid taking off the HMD when getting new tasks from the flashcards since it would be too strenuous. Instead, they used the video pass-through to view new tasks, planning their next actions before deciding whether to remove the HMD. This observation highlights how current VR hardware can cause physical strain, compromising users' willingness to remove the HMD and switch back to PC. Furthermore, the results demonstrate that participants could effectively plan actions across different realities, anticipating how outcomes in one reality would influence another.

5.2 Supporting Cross-Reality Transitions

To effectively support and partially replace information fetching transitions like Quick Peek, we propose the concept of implementing windows between realities. The goal is to enable users to fetch information from different realities. Thus allowing users to prepare or even omit fetching transitions. These windows can be implemented by software, for instance, presenting a portal or phone view showing another reality [36, 77], the World in Miniature (WIM) approach [28, 69], or attaching information to real-world objects using AR [6, 51]. By allowing users to see a different reality and act in it through the window, they can replace transitions like In-Between Touch. Alternatively, a cross-reality system can capture necessary information, like showing different realities or people from outside VR in the virtual environment [18, 74, 76]. Finally, hardware solutions, like flipping up the HMD instead of taking it off, can reduce the physical strain, thus potentially supporting update-type transitions to reality.

While windows between realities help avoid fetching transitions, Anchor Points can help users understand and remember how different realities are connected. Using Anchor Points may prevent forgetting and disorientation, help users form expectations, and improve perceived consistency between realities. Thus, they address and mitigate disorientation and fear of the unknown. One prominent example of using Anchor Points to design comprehensible transitions is the "MagicBook" [8], using a physical book as an Anchor Point helped users connect real and virtual content. Some participants suggested creating visible AR indications, such as small figures, to show where they left VR. This type of Anchor Point can, for example, be realized using a miniature that indicates user position in AR [4].

These two solutions, windows between realities and Anchor Points, can greatly decrease the frictions users experience when using CR interfaces. By reducing the need for transitions and helping users understand how realities are connected, should they actually need to transition. While these solutions help mitigate frictions, we use the concept of SMMs to interpret how users perceive different realities and the transitions between them.

5.3 Understanding Cross-Reality Transitions Through Spatial Mental Models

The results of our study show how people process spatial information in different realities. In the following, we examine our findings through the lens of Spatial Mental Models (SMMs) [34, 73], which capture the coarse spatial relations among elements in small or well-learned environments coherently. The setup in the study consists of the AR-view of the city plan, the VR-view in the city, and the PC in the lab with office chair and desks. We suggest the simultaneous existence of two SMMs, the Application SMM describing the virtual worlds of the application and how they are connected, and the Real-world SMM describing the real-world room users are in.

5.3.1 Real-world Spatial Mental Model. Even with the virtual tables disabled in VR, participants avoided colliding with the real-world tables. This observation indicates that they knew where the real tables were, suggesting the existence of an SMM for the real world. Previous research has identified people leveraging Real-world SMM in VR while exiting an immersive experience [47] or interacting with safety boundaries [72]. This model either exists parallel to or as part of the SMM describing the application.

5.3.2 Application Spatial Mental Model. The Application SMM contains spatial information for the CR application spanning VR, AR, and PC as well as their connection, although they present different viewpoints on the same world. The connection between VR and PC is the weakest, most likely because the PC interface only presents the statue without additional visualization in the virtual environment. This connection would be stronger if the PC interface presented the visual of the city to the user. Jung et al. [43], and Roo et al. [62] demonstrated a connection between realities on a conceptual level, which further supports the concept of an Application SMM.

We observed that participants used the Application SMM to plan actions in the other realities and perform them while considering their outcome. They could formulate Expectations regarding relative positions and rotations across different realities, e.g., looking towards where the statue would be in VR before transitioning to VR. They also used their Application SMM to understand how transitions work and connect the different realities. For instance, participants reported how different realities relate to each other as different Perspectives of the same city. This ability to place one reality in another and mentally switch between them indicates the existence of SMMs [34]. Breaking these Expectations might cause Confusion. In the experiment of Knibbe et al. [47], participants were confused when they did not stand at their expected location in the real world during a transition from VR to reality.

Participants started forming and using this Application SMM very early during the experiment, based only on the short description of the CR interface. They expressed confusion when a statue was not aligned with its assumed position "in" the monitor at the beginning of the study. This assumed consistency of objects between realities, termed Anchor Points, could also be beneficial for the formation of the Application SMM. A similar concept was also well-received in the study by Cools et al. [25], with their "Blend Spaces" using objects as triggers and anchors for CR transitions. These "Blend Spaces" showed objects from different realities simultaneously, allowing users to transition to the reality they reside in. This concept is a good example on how Anchor Points could help users understand how different realities align. Participants used tables or controllers as a connecting element between realities, allowing them to anchor themselves during transitions by focusing on an unchanging element existing in the origin and target reality. If participants perceived something as an Anchor Point, perceived mis-alignment between realities lead to confusion and Disorientation.

During the early stages of the experiment, participants experienced confusion when the scale between VR, AR, and PC were different. They reported disorientation when switching into VR while zoomed in. "*[After transition to VR] This was unexpected because I badly lost orientation, and I was zoomed in further than expected.*" (P08, Minute 03:30) In contrast, they also noted that matching scale improved transitions: "*[Interviewer] How was this transition? [Participant] I find transitions from PC to AR always more pleasant because the world [city model] is not so big.*" (P05, Minute 56:30).

Early in the experiment, participants were observed splitting up the transition from PC to VR by transitioning to AR first, so-called Gradual Transitions. This suggests a weaker or more complex connection between PC and VR in the Application SMM. This finding also aligns with the study of Schröder et al. [64], who noted that participants avoid specific transitions rather than realities, further indicating that some transitions are more taxing.

Over time, participants demonstrated increasing confidence and ease when working across and changing between realities, indicating continuous improvement of their Application SMM. The perception of Anchor Points also evolved throughout the experiment. While participants became confused if the real-world city model and the city model in VR were not aligned when they moved or zoomed in VR at the start of the experiment; toward the end of their trials, participants often remarked that the Alignment between the virtual city model and the real city model was no longer necessary. They noted that now it made sense for them to appear in VR exactly where they left off instead of next to and above the virtual city.

Participants also started using previously avoided transitions, such as moving directly between PC and VR. This suggests they formed a robust Application SMM explaining how AR and VR are or are not connected, using it to plan and execute transitions more efficiently. Some participants even developed a robust enough Application SMM to visualize different realities and their points of view within them without the need for actual transitions. Finally, while participants improved their Application SMM over time, they also reported forgetting parts of it, indicating that these models are volatile and require constant maintenance.

5.3.3 Implications for CR Design. Based on these findings, applications should guide the user in the formation of a reliable Application SMM, reducing the risk of breaking expectations. This can be achieved by designing for and designating dedicated Anchor Points and by designing CR transitions so they explain to users what changes during a transitions[2]. Mechanisms, like windows between worlds [36], can help users avoid transitions, thus reducing the need to process them. These approaches can help users navigate and understand the relationships between different realities more effectively, potentially reducing confusion and improving the overall user experience in Cross-Reality interfaces.

[2]From a purely Application SMM focused standpoint, this would make a slow fly in an ideal transition. However, this would be time-consuming and severely increase discomfort by inducing cybersickness.

6 Limitations and Future Work

Since our CR application was deliberately designed to distribute capabilities across different realities, it is important to note that the observed workflows may not represent an ideal or typical CR application. The interfaces for different realities were specifically chosen to provide both easy (AR to VR) and more complex (PC to VR) transitions, encompassing a wide variety of transition types. Furthermore, we intentionally avoided certain sensible design decisions regarding transitions and interfaces, such as allowing participants to work on PC while wearing the HMD. This approach was taken to avoid mitigating or obscuring the frictions we aimed to uncover through our experiment.

Recruiting participants via word of mouth at a STEM-focused university did yield a rather homogeneous participant pool regarding their education. This may have skewed the results, but we argue that the underlying issues are generalizable.

An intriguing avenue for future research would be to study the actual workflows of users interacting with a well-designed CR interface. This could involve:

- Exploring which transition archetypes are used and their frequency when capabilities are not artificially distributed across realities.
- Analyzing user behavior and preferences in well-designed CR environments and interfaces.
- Investigating how users adapt to and utilize different realities when given more intuitive and seamless transition options.

Such studies could provide valuable insights for further design recommendations for CR interfaces, potentially leading to more efficient and user-friendly systems.

Another promising area for future work is a closer examination of the ideal design and application of Anchor Points. This research could focus on:

- Identifying the most effective types of Anchor Points for different CR scenarios.
- Exploring how Anchor Points can be dynamically adapted or personalized for individual users.
- Investigating the potential of Anchor Points to enhance spatial awareness and navigation in complex CR environments.

Examining the long-term effects of well-designed Anchor Points on user performance and comfort in CR applications. Delving deeper into these aspects, can unlock the full potential of Anchor Points and significantly improve the user experience in Cross-Reality interfaces.

Furthermore, based on established models, we provided a possible explanation of why and how users transition, which is still untested. Future work may test our findings by designing transitions under the theory provided to see if they hold up or need extensions. One possible method for quantifying how well different transition designs alleviate the problem can be psychophysiological measurements. These have been shown to detect breaks in presence and vary with the severity of the breaks by Liebold et al. [50]

7 Conclusion

In conclusion, we performed a controlled, *talk-aloud* lab study, evaluating the user perspective on transitions behavior and experience when using a Cross-Reality (CR) application. We found distinct frictions, factors that contribute to them, behaviors used by the participants to mitigate them, and the concept of Anchor Points strengthening connections between realities. We found that participants form multiple Spatial Mental Models (SMMs), one for the CR application they are currently working in, and one for the real world they reside in, even when in Virtual Reality (VR). We called them Application SMM and Real-world SMM, respectively; the latter has also been observed in related work [47].

We propose that these frictions can be reduced by designing CR applications as well as suitable transition mechanisms. The design goal in both cases is to strengthen the connection between different realities. This can happen either by giving the user an explanation of the nature of the connection, such as breadcrumbs, or by visualizing their position in the target reality before transitioning. Transition mechanisms should always help the user understand how different realities are connected. This can be achieved by using slow, gradual transitions, giving the users time to understand how the connection works. Overall, being aware of the existence of the Application SMM and influencing how users form it, as well as supporting them in maintaining it, is a key factor in designing future CR applications.

Acknowledgments

This work has been funded by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE and by the LOEWE initiative (Hesse, Germany) within the emergenCITY centre (funding code: LOEWE/1/12/519/03/05.001(0016)/72).

References

[1] Ali Almutawa. 2021. Effect of Smooth Transition and Hybrid Reality on Virtual Realism : A Case of Virtual Art Gallery. *Advances in Science, Technology and Engineering Systems Journal* 6, 3 (2021), 231–240. https://doi.org/10.25046/aj060325

[2] Jonas Auda, Uwe Gruenefeld, Sarah Faltaous, Sven Mayer, and Stefan Schneegass. 2023. A Scoping Survey on Cross-Reality Systems. *Comput. Surveys* (2023).

[3] Jonas Auda, Roman Heger, Uwe Gruenefeld, and Stefan Schneegass. 2021. VRSketch: Investigating 2D Sketching in Virtual Reality with Different Levels of Hand and Pen Transparency. In *Human-Computer Interaction – INTERACT 2021 (Lecture Notes in Computer Science)*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 195–211. https://doi.org/10.1007/978-3-030-85607-6_14

[4] Akshay Ayyanchira, Elias Mahfoud, Weichao Wang, and Aidong Lu. 2022. Toward cross-platform immersive visualization for indoor navigation and collaboration with augmented reality. *Journal of Visualization* 25, 6 (Dec 2022), 1249–1266. https://doi.org/10.1007/s12650-022-00852-9

[5] Huidong Bai, Li Zhang, Jing Yang, and Mark Billinghurst. 2021. Bringing full-featured mobile phone interaction into virtual reality. *Computers & Graphics* 97 (Jun 2021), 42–53. https://doi.org/10.1016/j.cag.2021.04.004

[6] Daniel Bambušek, Zdeněk Materna, Michal Kapinus, Vítězslav Beran, and Pavel Smrž. 2022. Handheld Augmented Reality: Overcoming Reachability Limitations by Enabling Temporal Switching to Virtual Reality. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 698–702. https://doi.org/10.1109/HRI53351.2022.9889434

[7] João Belo, Tiare Feuchtner, Chiwoong Hwang, Rasmus Lunding, Mathias Lystbæk, Ken Pfeuffer, and Troels Rasmussen. 2021. Challenges of XR Transitional Interfaces in Industry 4.0. (2021).

[8] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: A Transitional AR Interface. *Computers & Graphics* 25, 5 (Oct. 2001), 745–753. https://doi.org/10.1016/S0097-8493(01)00117-0

[9] James P. Bliss, Philip D. Tidwell, and Michael A. Guest. 1997. The Effectiveness of Virtual Reality for Administering Spatial Navigation Training to Firefighters. *Presence: Teleoperators and Virtual Environments* 6, 1 (Feb. 1997), 73–86. https://doi.org/10.1162/pres.1997.6.1.73

[10] Jeremy E. Block and Eric D. Ragan. 2020. Micro-Entries: Encouraging Deeper Evaluation of Mental Models Over Time for Interactive Data Systems. In *2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV)*. 38–47. https://doi.org/10.1109/BELIV51497.2020.00012

[11] Daniel Boland and Mark McGill. 2015. Lost in the rift: engaging with mixed reality. *XRDS: Crossroads, The ACM Magazine for Students* 22, 1 (Nov 2015), 40–45. https://doi.org/10.1145/2810046

[12] Sabah Boustila, Thomas Guégan, Kazuki Takashima, and Yoshifumi Kitamura. 2019. Text Typing in VR Using Smartphones Touchscreen and HMD. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 860–861. https://doi.org/10.1109/VR.2019.8798238

[13] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806 arXiv:https://doi.org/10.1080/2159676X.2019.1628806

[15] Robert L Brennan and Dale J Prediger. 1981. Coefficient kappa: Some uses, misuses, and alternatives. *Educational and psychological measurement* 41, 3 (1981), 687–699.

[16] Leonard D. Brown and Hong Hua. 2006. Magic Lenses for Augmented Virtual Environments. *IEEE Computer Graphics and Applications* 26, 4 (July 2006), 64–73. https://doi.org/10.1109/MCG.2006.84

[17] Frederik Brudy, Christian Holz, Roman Rädle, Chi-Jui Wu, Steven Houben, Clemens Nylandsted Klokmose, and Nicolai Marquardt. 2019. Cross-Device Taxonomy: Survey, Opportunities and Challenges of Interactions Spanning Across Multiple Devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–28. https://doi.org/10.1145/3290605.3300792

[18] Pulkit Budhiraja, Rajinder Sodhi, Brett Jones, Kevin Karsch, Brian Bailey, and David Forsyth. 2015. Where's My Drink? Enabling Peripheral Real World Interactions While Using HMDs. *arXiv preprint arXiv:1502.04744* (2015). arXiv:1502.04744

[19] John M. Carroll and Judith Reitman Olson. 1988. Mental Models in Human-Computer Interaction / This Chapter Appeared in Its Entirety and Is Reprinted from Mental Models in Human Computer Interaction: Research Issues about What the User of Software Knows, J.M. Carroll and J.R. Olson, Editors,-The Report of the Workshop on Software Human Factors: Users Mental Models, Nancy Anderson, Chair, Sponsored by the Committee on Human Factors, Commission on Behavioral and Social Sciences and Education, National Research Council, Published by the National Academy Press, 1987. In *Handbook of Human-Computer Interaction*. Elsevier, 45–65. https://doi.org/10.1016/B978-0-444-70536-5.50007-5

[20] Felipe G. Carvalho, Daniela G. Trevisan, and Alberto Raposo. 2012. Toward the Design of Transitional Interfaces: An Exploratory Study on a Semi-Immersive Hybrid User Interface. *Virtual Reality* 16, 4 (Nov. 2012), 271–288. https://doi.org/10.1007/s10055-011-0205-y

[21] Marco Cavallo, Mishal Dholakia, Matous Havlena, Kenneth Ocheltree, and Mark Podlaseck. 2019. Dataspace: A Reconfigurable Hybrid Reality Environment for Collaborative Information Analysis. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 145–153. https://doi.org/10.1109/VR.2019.8797733

[22] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. 2021. SemanticAdapt: Optimization-based Adaptation of Mixed Reality Layouts Leveraging Virtual-Physical Semantic Connections. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 282–297. https://doi.org/10.1145/3472749.3474750

[23] Francesco Chiossi, Uwe Gruenefeld, Baosheng Hou, Joshua Newn, Changkun Ou, Rulu Liao, Robin Welsch, and Sven Mayer. 2024. Understanding the Impact of the Reality-Virtuality Continuum on Visual Search Using Fixation-Related Potentials and Eye Tracking Features. *Proceedings of the ACM on Human-Computer Interaction* 8 (July 2024), 33. https://doi.org/10.1145/3676528

[24] Arzu Çöltekin, Ian Lochhead, Marguerite Madden, Sidonie Christophe, Alexandre Devaux, Christopher Pettit, Oliver Lock, Shashwat Shukla, Lukáš Herman, Zdeněk Stachoň, Petr Kubíček, Dajana Snopková, Sergio Bernardes, and Nicholas Hedley. 2020. Extended Reality in Spatial Sciences: A Review of Research Challenges and Future Directions. *ISPRS International Journal of Geo-Information* 9, 7 (July 2020), 439. https://doi.org/10.3390/ijgi9070439

[25] Robbe Cools, Augusto Esteves, and Adalberto L. Simeone. 2022. Blending Spaces: Cross-Reality Interaction Techniques for Object Transitions Between Distinct Virtual and Augmented Realities. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 528–537. https://doi.org/10.1109/ISMAR55827.2022.00069

[26] Robbe Cools, Matt Gottsacker, Adalberto Simeone, Gerd Bruder, Greg Welch, and Steven Feiner. 2022. Towards a Desktop-AR Prototyping Framework: Prototyping Cross-Reality Between Desktops and Augmented Reality. *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)* (Oct 2022), 175–182. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00040

[27] Juliet Corbin and Anselm Strauss. 2008. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory.* SAGE Publications, Inc. https://doi.org/10.4135/9781452230153

[28] Kurtis Danyluk, Barrett Ens, Bernhard Jenny, and Wesley Willett. 2021. A Design Space Exploration of Worlds in Miniature. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 122, 15 pages. https://doi.org/10.1145/3411764.3445098

[29] Nicolas J. Dedual, Ohan Oda, and Steven K. Feiner. 2011. Creating Hybrid User Interfaces with a 2D Multi-Touch Tabletop and a 3D See-through Head-Worn Display. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality.* 231–232. https://doi.org/10.1109/ISMAR.2011.6092391

[30] Amit P. Desai, Lourdes Peña-Castillo, and Oscar Meruvia-Pastor. 2017. A Window to Your Smartphone: Exploring Interaction and Communication in Immersive VR with Augmented Virtuality. In *2017 14th Conference on Computer and Robot Vision (CRV).* 217–224. https://doi.org/10.1109/CRV.2017.16

[31] Seyed Amir Ahmad Didehkhorshid, Siju Philip, Elaheh Samimi, and Robert J. Teather. 2020. Text Input in Virtual Reality Using a Tracked Drawing Tablet. In *HCI International 2020 – Late Breaking Papers: Virtual and Augmented Reality (Lecture Notes in Computer Science)*, Constantine Stephanidis, Jessie Y. C. Chen, and Gino Fragomeni (Eds.). Springer International Publishing, Cham, 314–329. https://doi.org/10.1007/978-3-030-59990-4_24

[32] Isamu Endo, Kazuki Takashima, Maakito Inoue, Kazuyuki Fujita, Kiyoshi Kiyokawa, and Yoshifumi Kitamura. 2021. ModularHMD: A Reconfigurable Mobile Head-Mounted Display Enabling Ad-hoc Peripheral Interactions with the Real World. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 100–117. https://doi.org/10.1145/3472749.3474738

[33] Minoo H. Esfehani and Trudie Walters. 2018. Lost in Translation? Cross-language Thematic Analysis in Tourism and Hospitality Research. *International Journal of Contemporary Hospitality Management* 30, 11 (Jan. 2018), 3158–3174. https://doi.org/10.1108/IJCHM-10-2017-0701

[34] Nancy Franklin, Barbara Tversky, and Vicky Coon. 1992. Switching Points of View in Spatial Mental Models. *Memory & Cognition* 20, 5 (Sept. 1992), 507–518. https://doi.org/10.3758/BF03199583

[35] Steffen Gauglitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14)*. Association for Computing Machinery, New York, NY, USA, 449–459. https://doi.org/10.1145/2642918.2647372

[36] Ceenu George, An Ngo Tien, and Heinrich Hussmann. 2020. Seamless, Bi-directional Transitions along the Reality-Virtuality Continuum: A Conceptualization and Prototype Exploration. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 412–424. https://doi.org/10.1109/ISMAR50242.2020.00067

[37] Valérie Gyselinck, Chiara Meneghetti, Monica Bormetti, Eric Orriols, Pascale Piolino, and Rossana De Beni. 2013. Considering Spatial Ability in Virtual Route Learning in Early Aging. *Cognitive Processing* 14, 3 (Aug. 2013), 309–316. https://doi.org/10.1007/s10339-013-0557-1

[38] Nelofer Halai. 2007. Making Use of Bilingual Interview Data: Some Experiences from the Field. *The Qualitative Report* 12, 3 (Sept. 2007), 344–355. https://doi.org/10.46743/2160-3715/2007.1621

[39] Monique Hennink and Bonnie N Kaiser. 2022. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social science & medicine* 292 (2022), 114523.

[40] Hans-Christian Jetter, Jan-Henrik Schröder, Jan Gugenheimer, Mark Billinghurst, Christoph Anthes, Mohamed Khamis, and Tiare Feuchtner. 2021. Transitional Interfaces in Mixed and Cross-Reality: A New Frontier?. In *Companion Proceedings of the 2021 Conference on Interactive Surfaces and Spaces (ISS '21)*. Association for Computing Machinery, New York, NY, USA, 46–49. https://doi.org/10.1145/3447932.3487940

[41] Philip N. Johnson-Laird. 1983. *Mental Models: Towards A Cognitive Science of Language, Inferences, and Consciousness.* Harvard University Press, Cambridge, Massachusetts.

[42] Tero Jokela, Jarno Ojala, and Thomas Olsson. 2015. A Diary Study on Combining Multiple Information Devices in Everyday Activities and Tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3903–3912. https://doi.org/10.1145/2702123.2702211

[43] Sungchul Jung, Pamela J. Wisniewski, and Charles E. Hughes. 2018. In Limbo: The Effect of Gradual Visual Transition Between Real and Virtual on Virtual Body Ownership Illusion and Presence. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 267–272. https://doi.org/10.1109/VR.2018.8447562

[44] Hiroki Kaimoto, Kyzyl Monteiro, Mehrad Faridan, Jiatong Li, Samin Farajian, Yasuaki Kakehi, Ken Nakagaki, and Ryo Suzuki. 2022. Sketched Reality: Sketching Bi-Directional Interactions Between Virtual and Physical Worlds with AR and Actuated Tangible UI. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3526113.3545626

[45] Maged N. Kamel Boulos, Zhihan Lu, Paul Guerrero, Charlene Jennett, and Anthony Steed. 2017. From Urban Planning and Emergency Training to Pokémon Go: Applications of Virtual Reality GIS (VRGIS) and Augmented Reality GIS (ARGIS) in Personal, Public and Environmental Health. *International Journal of Health Geographics* 16, 1 (Dec. 2017), 7, s12942–017–0081–0. https://doi.org/10.1186/s12942-017-0081-0

[46] David E. Kieras and Susan Bovair. [n. d.]. The Role of a Mental Model in Learning to Operate a Device. *Cognitive Science* 8 ([n. d.]), 255–273.

[47] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornbæk. 2018. The Dream Is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174057

[48] Pascal Knierim, Dimitri Hein, Albrecht Schmidt, and Thomas Kosch. 2021. The SmARtphone Controller. *i-com* 20, 1 (2021), 49–61. https://doi.org/doi:10.1515/icom-2021-0003

[49] D. Kristen Gilbert, Wendy A. Rogers, and Mary E. Samuelson. 2004. Long-Term Retention of a Spatial Mental Model for Younger and Older Adults. *Experimental Aging Research* 30, 2 (April 2004), 217–224. https://doi.org/10.1080/03610730490274266

[50] Benny Liebold, Michael Brill, Daniel Pietschmann, Frank Schwab, and Peter Ohler. 2017. Continuous Measurement of Breaks in Presence: Psychophysiology and Orienting Responses. *Media Psychology* 20, 3 (July 2017), 477–501. https://doi.org/10.1080/15213269.2016.1206829

[51] Chuan-en Lin, Ta Ying Cheng, and Xiaojuan Ma. 2020. ARchitect: Building Interactive Virtual Experiences from Physical Affordances by Bringing Human-in-the-Loop. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376614

[52] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 2143–2152. https://doi.org/10.1145/2702123.2702382

[53] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[54] Alexandre Millette and Michael J. McGuffin. 2016. DualCAD: Integrating Augmented Reality with a Desktop GUI and Smartphone Interaction. In *2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. 21–26. https://doi.org/10.1109/ISMAR-Adjunct.2016.0030

[55] María Murcia López. 2018. The Effectiveness of Training in Virtual Environments. *Doctoral thesis, UCL (University College London).* (Dec. 2018).

[56] Janni Nielsen, Torkil Clemmensen, and Carsten Yssing. 2002. Getting Access to What Goes on in People's Heads?: Reflections on the Think-Aloud Technique. In *Proceedings of the Second Nordic Conference on Human-computer Interaction.* ACM, Aarhus Denmark, 101–110. https://doi.org/10.1145/572020.572033

[57] Janni Nielsen, Carsten Yssing, et al. 2004. *The disruptive effect of Think Aloud.* Technical Report.

[58] Donald A. Norman. 1986. *Cognitive Engineering* (0 ed.). CRC Press, Boca Raton, 31–62. https://doi.org/10.1201/b15703-3

[59] Thammathip Piumsomboon, Gavin Ong, Cameron Urban, Barrett Ens, Xiaoliang Bai, and Simon Hoermann. 2022. Ex-Cit XR: Expert-elicitation of XR Techniques for Disengaging from IVEs. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 710–711. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00148

[60] Fabian Pointecker, Judith Friedl, Daniel Schwajda, Hans-Christian Jetter, and Christoph Anthes. 2022. Bridging the Gap Across Realities: Visual Transitions Between Virtual and Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 827–836. https://doi.org/10.1109/ISMAR55827.2022.00101

[61] Steve Prentice. 2022. The Fear of the Unknown. , 65–79 pages. https://doi.org/10.1007/978-1-4842-8101-7_5

[62] Joan Sol Roo, Jean Basset, Pierre-Antoine Cinquin, and Martin Hachet. 2018. Understanding Users' Capability to Transfer Information between Mixed and Virtual Reality: Position Estimation across Modalities and Perspectives. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3173937

[63] Joan Sol Roo and Martin Hachet. 2017. One Reality: Augmenting How the Physical World Is Experienced by Combining Multiple Mixed Reality Modalities. In *Proceedings of the 30th ACM User Interface Software and Technology Symposium.* ACM, New York, NY, USA, 787–795. https://doi.org/10.1145/3126594.3126638

[64] Jan-Henrik Schröder, Daniel Schacht, Niklas Peper, Anita Marie Hamurculu, and Hans-Christian Jetter. 2023. Collaborating Across Realities: Analytical Lenses for Understanding Dyadic Collaboration in Transitional Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3544548.3580879

[65] Gerhard Schubert, Christoph Anthes, Dieter Kranzlmüller, and Frank Petzold. 2012. *From Physical to Virtual: Real-time Immersive Visualizations from an Architect's Working Model.*

[66] Songjia Shen, Hsiang-Ting Chen, William Raffe, and Tuck Wah Leong. 2021. Effects of Level of Immersion on Virtual Training Transfer of Bimanual Assembly Tasks. *Frontiers in Virtual Reality* 2 (May 2021). https://doi.org/10.3389/frvir.2021.597487

[67] Keng Hua Sing. 2016. Garden : A Mixed Reality Experience Combining Virtual Reality and 3D Reconstruction. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. ACM, New York, NY, USA, 180–183. https://doi.org/10.1145/2851581.2890370

[68] Allison Squires. 2009. Methodological Challenges in Cross-Language Qualitative Research: A Research Review. *International Journal of Nursing Studies* 46, 2 (Feb. 2009), 277–287. https://doi.org/10.1016/j.ijnurstu.2008.08.006

[69] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. ACM Press/Addison-Wesley Publishing Co., USA, 265–272. https://doi.org/10.1145/223904.223938

[70] Holly A Taylor and Barbara Tversky. 1992. Spatial Mental Models Derived from Survey and Route Descriptions. *Journal of Memory and Language* 31, 2 (April 1992), 261–292. https://doi.org/10.1016/0749-596X(92)90014-O

[71] Y. P. Tsang, T. Yang, Z. S. Chen, C. H. Wu, and K. H. Tan. 2022. How Is Extended Reality Bridging Human and Cyber-Physical Systems in the IoT-empowered Logistics and Supply Chain Management? *Internet of Things* 20 (Nov. 2022), 100623. https://doi.org/10.1016/j.iot.2022.100623

[72] Wen-Jie Tseng, Petros Dimitrios Kontrazis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2024. Understanding Interaction and Breakouts of Safety Boundaries in Virtual Reality Through Mixed-Method Studies. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 482–492. https://doi.org/10.1109/VR58804.2024.00069

[73] Barbara Tversky. 1993. Cognitive Maps, Cognitive Collages, and Spatial Mental Models. In *Spatial Information Theory A Theoretical Basis for GIS (Lecture Notes in Computer Science)*, Andrew U. Frank and Irene Campari (Eds.). Springer, Berlin, Heidelberg, 14–24. https://doi.org/10.1007/3-540-57207-4_2

[74] Julius Von Willich, Markus Funk, Florian Müller, Karola Marky, Jan Riemann, and Max Mühlhäuser. 2019. You Invaded My Tracking Space! Using Augmented Virtuality for Spotting Passersby in Room-Scale Virtual Reality. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 487–496. https://doi.org/10.1145/3322276.3322334

[75] David Waller. 2000. Individual Differences in Spatial Learning from Computer-Simulated Environments. *Journal of Experimental Psychology: Applied* 6, 4 (2000), 307–321. https://doi.org/10.1037/1076-898X.6.4.307

[76] Chiu-Hsuan Wang, Bing-Yu Chen, and Liwei Chan. 2022. RealityLens: A User Interface for Blending Customized Physical World View into Virtual Reality. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3526113.3545686

[77] Chiu-Hsuan Wang, Chia-En Tsai, Seraphina Yong, and Liwei Chan. 2020. Slice of light: Transparent and integrative transition among realities in a multi-HMD-user environment. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 805–817.

[78] Jia Wang and Robert Lindeman. 2014. Coordinated 3D Interaction in Tablet- and HMD-based Hybrid Virtual Environments. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction (SUI '14)*. Association for Computing Machinery, New York, NY, USA, 70–79. https://doi.org/10.1145/2659766.2659777

[79] Nanjia Wang and Frank Maurer. 2022. A Design Space for Single-User Cross-Reality Applications. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3531073.3531116

[80] Nanjia Wang and Frank Maurer. 2022. User-Centered Prototyping for Single-User Cross-Reality Virtual Object Transitions. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00039

[81] Julia Wertheim and Terrence C. Stewart. 2018. Explaining Reasoning Effects: A Neural Cognitive Model of Spatial Reasoning. *Proceedings of the Annual Meeting of the Cognitive Science Society* 40, 0 (2018).

[82] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376233

A Tasks

During the study, we gave the participants 10 sub-tasks to work on. As we had German and English participants, we provided the task descriptions in their preferred language, the content being the same. The tasks are provided in both languages.

The phrase "Markierung auf dem Modell"/"marking on the model" indicates that we put a colored sticker on the task sheet and the point of interest on the physical city model in order to make them easily recognizable. In VR, there was no such marker.

In the paper, we used the substitution "[landmark]" for Schloss and Kantplatz to avoid confusion while reading. The choice of these specific landmarks should not matter in the context of our analysis.

(1) Der Luisenplatz soll eine neue Statue bekommen, der Stadtrat hätte gerne etwas Abstraktes. Die Skulptur soll aus einer rechteckigen Plattform mit einem Seitenverhältnis von 2:3 bestehen. Diese soll ein Schachbrettmuster haben. Also mindestens 2x3 Bodenelemente.
(2) Die Skulptur soll ganz oben an der Spitze des Luisenplatz im Modell platziert werden.
(3) Die kurze Seite soll 4 Armspannweiten (Controller zu Controller) lang sein.
(4) An den 2 Ecken einer kurzen Seite sollen zylindrische Säulen stehen, welche ca. 1 Mensch hoch sind, du selbst bist die Referenz.
(5) Vom Schloss aus (Markierung auf dem Modell) sollen beide Säulen sichtbar sein. Die komplette Skulptur soll nicht von den Häusern abgeschnitten sein. Evtl. muss diese also umplatziert werden.
(6) In der Mitte der Plattform soll eine hohe Säule stehen, mit ca. einer Armlänge Durchmesser. Diese Säule soll vom Kantplatz (Markierung auf dem Modell) aus sichtbar sein.
(7) Drehe die Skulptur um 45°.
(8) Vom Schloss aus sollen nur blaue Steine sichtbar sein.
(9) Wie sieht das aus, wenn wir die Skulptur jetzt auf den Kopf stellen?
(10) Mach alles blaue jetzt grün, und wieder alles vom Schloss aus erkennbare blau. Alles von Häusern verdeckte rot.

(1) The Luisenplatz will get a new sculpture. The major [sic][3] wants something abstract. The sculpture should have a rectangular base, the sides should have a ratio of 2:3. The base should have a checker board pattern, so you need at least 2x3 primitives.
(2) The sculpture should be placed on top of the Luisenplatz in the city model.
(3) The short side should be 4 arm spans (controller to controller) wide.
(4) On the 2 corners of one short side, there should be cylindrical pillars, which are 1 human high. You yourself shall be the scale.
(5) From Schloss (marking on the model) both pillars should be visible. The whole sculpture should not be obstructed from the surrounding buildings. You may have to reposition the sculpture.

[3] mix up during translation, participants did not question/complain

(6) In the middle of the platform, there should be a high pillar, with a diameter of about 1 arm length. This pillar should be visible from Kantplatz (marking on the model).
(7) Turn the sculpture 45°.
(8) All primitives visible from Schloss should be blue.
(9) Turn the sculpture on its head.
(10) Color every blue primitive green, and again everything blue that is visible from Schloss. Color everything red that is obstructed by buildings.

B Codebook

After coding we structured and renamed the codes to aid comprehensibility for persons not involved in the coding process. We will now present the codes from our codebook, explaining their concepts and application rules. While these transitions may work in scenarios with more than three distinct realities, they were conceived in a setup with only three realities. The code occurrences are listed in Table 2.

B.1 Mental Model

Codes in this group were used to mark something indicative of a mental model.

B.1.1 UI Model. Codes in this group were used to indicate something connected specifically to the UI of our apparatus, rather than the apparatus as a whole.

Lack of Expected Feature Used when participants talked about a feature they wanted to have.

UI Consistency Used when participants talked about desiring consistent input over all three realities.

B.1.2 Requirements. Codes in this group were used when participants talked about something they may require in an ideal prototype.

Control This code came up during coding and turned out to only be applicable for three participants. It was used when participants talked about feeling more in control after a transition, usually when transitioning from VR to Augmented Reality (AR).

Consistency Used when participants talked about the consistency between realities, for example, the statue staying at the same place or them returning to the same place in VR.

Expectation Used, when participants talked about expecting something after a transition, mainly them or something being at a specific position.

B.1.3 Spatial Mental Model. This group of codes denotes different aspects, indicative of the participants building a SMM of the apparatus or their surroundings.

Different Perspective Used, when participants indicated understanding that PC, AR, and VR represent the same environment, but from different viewing angles.

Alignment Used when participants talked about the alignment of objects in two different modalities. A typical situation was participants noticing that the real and virtual city model were aligned in the beginning or that they missed this alignment later during the task.

Scale Used when participants talked about the scale of them or an object. Typical situations were participants talking about scale discrepancies between the real and the virtual city model or when they became aware of their statues scale when switching to VR.

Position Used when participants talked about their position or that of an object. Typical situations were participants predicting their or an object's position in VR or them being confused when they appeared in an unexpected location after transitioning to VR.

Rotation Used when participants talked about a rotational aspect of themselves or an object. A typical situation was them realizing that their viewing angle was the only thing being conserved when transitioning to VR, with the rest being as it was when they left.

Code with Structure	Count
Archetypes > Accident	14
Archetypes > Back-and-Forth > Quick Peek	55
Archetypes > Back-and-Forth > Quick Touch	23
Archetypes > Back-and-Forth > Quick Workflow	19
Archetypes > Back-and-Forth > Reality Quick Peek	40
Archetypes > Default	251
Archetypes > In-Between > Gradual Transition	39
Archetypes > In-Between > In-Between Peek	14
Archetypes > In-Between > In-Between Touch	7
Archetypes > Lazy Reality	43
Comfort/Discomfort > Cognitive Effort	6
Comfort/Discomfort > Cognitive Strain	13
Comfort/Discomfort > Physical Effort	15
Comfort/Discomfort > Physical Strain	14
Connotation > Negative	41
Connotation > Positive	46
Mental Model > Occurences > Disorientation	20
Mental Model > Occurences > Forgetting	29
Mental Model > Occurences > Learning	17
Mental Model > Requirements > Consistency	48
Mental Model > Requirements > Control	5
Mental Model > Requirements > Expectation	56
Mental Model > SMM > Alignment	21
Mental Model > SMM > Different Perspective	18
Mental Model > SMM > Position	78
Mental Model > SMM > Rotation	21
Mental Model > SMM > Scale	14
Mental Model > Tools > Anchor Point	44
Mental Model > Tools > Real World Adjustment	9
Mental Model > Tools > Task Preparation	9
Mental Model > UI Model > Lack of Expected Feature	3
Mental Model > UI Model > UI Consistency	6
Sum	**1038**

Table 2: Number of code occurrences in the user study.

B.1.4 *Acting Across Modalities.*
Sub-codes from this category were used when participants acted across modality transitions.

- **Task Preparation** Used when participants talked about preparing a task in a different modality, e.g. preparing elements in PC before accurately placing them in AR or VR.
- **Real World Adjustment** Used when participants used AR to adjust something in the real world, such as moving the chair or the Head-Mounted Display (HMD) cable.
- **Anchor Point** Used when participants used something as an anchor point, e.g. an auxiliary block to help them in orienting their statue, the tables visible in all modalities, or the controllers as something they can take with them from one modality to another.

B.1.5 *Occurrences.*
This group was used when something related to the mental model occurred.

- **Disorientation** Used when participants talked about being disoriented after a transition.
- **Forgetting** Used when participants forgot about a feature of the apparatus, influencing their decision-making, e.g. forgetting they can change which modality they transition to when putting on the HMD, resulting in presumably unneeded transitions.
- **Learning** Used when participants learned something about the apparatus, influencing their workflow. This code was mostly used at the beginning of each participant's study, as they were still getting used to the software. But also when participants realized they could solve something in a specific way, such as staying in the chair while transitioning to VR if they only wanted to take a quick look.

B.2 Discomfort

Codes in this category were used when the participants experienced comfort or discomfort with regard to cognitive or physical processes while transitioning.

B.2.1 *Physical Strain.*
Used when participants mentioned physical strain, stress, or pain, e.g. the eyes getting tired due to long-time working with the HMD on, or pulling their hair while removing the HMD.

B.2.2 *Physical Effort.*
Used when participants mentioned that they experienced physical effort during a transition, e.g., putting on the HMD or walking to a position.

B.2.3 *Cognitive Strain.*
Used when participants mentioned cognitive strain they experienced during or due to a transition. For example, the text for the sub-tasks is hard to read through the HMD display and therefore takes more effort to understand or re-orientate often while doing multiple transitions.

B.2.4 *Cognitive Effort.*
Used when participants mentioned cognitive effort they had during or due to a transition. For example, remembering the point of view in VR before doing the transition from AR to VR.

B.3 Connotations

The two connotation codes Positive and Negative are used to classify the statements and experiences of the participants.

B.4 Archetypes

We discovered multiple types of transitions, characterized by the realities involved and their integration into the participants' workflows. Furthermore, they are also defined by the modality they start and end in.

B.4.1 *Accident.*
A few transitions were not done on purpose and thus are not part of the intended workflow. These transitions occurred either by accidentally pressing the button on the HMD or when participants forgot to adjust the modality they load into when putting on the HMD.

B.4.2 *Lazy Reality.*
This transition is characterized by participants switching from VR to AR in order to see the real world, instead of taking off the HMD. Lazy Reality was used by participants to read new instructions or make eye contact with the examiner when talking.

B.4.3 *In-Between.*
In-Between describes a family of transitions, which are characterized by starting in one modality, moving into a second modality, and ending up in a third modality. Their sub-types are distinguished by the purpose of the transition.

- **Gradual Transition** The purpose of a gradual transition was, to insert a supporting transition, positioned "between" the start and end modality on the Milgram continuum. Participants reported that these gradual transitions helped them adjust to the modality they ended up in. In our application, these could only happen when transitioning from PC to VR or VR to PC, as there is no other constellation with a modality "between" the start and end.
- **In-Between Peek** This archetype describes transitions involving a support transition, which is used to quickly get information from the in-between modality. Such as transitioning from PC to VR in order to compare one's arm span to the statue, and then from VR to AR in order to adjust the scale of the statue.
- **In-Between Touch** Similar to In-Between Peek, In-Between Touch describes a transition involving all three modalities, where the participants made small adjustments in the intermediate modality. A typical scenario was participants transitioning from PC to VR via AR, adjusting the alignment of the statue in AR.

B.4.4 *Back-And-Forth.*
Back-And-Forth describes a family of transitions, which are characterized by starting in one modality, moving into a second modality, and ending up back in the first modality. Their sub-types are distinguished by the purpose of the transition.

- **Quick Peek** Similar to In-Between Peek, this transition is defined by the intention of quickly looking at something in a different modality, before returning to the original modality. This archetype appeared mostly between AR and VR but an interesting use was checking the correct coloring of the statue by putting on the HMD to take a short look and adjusting the statue at the PC accordingly.

Quick Touch Similar to In-Between Touch, this transition is defined by the intention of quickly adjusting something in a different modality, before returning to the original modality.

Quick Workflow While Quick Peek and Quick Touch feature one quick action in a different modality and are limited to two transitions, Quick Workflow describes chaining and interweaving of multiple Quick Peek and Quick Touch, with the participant ending up in one of the two involved modalities. This pattern was typically seen when participants tried to correctly scale the statue, making small adjustments in AR and checking the new size in VR.

Reality Quick Peek This Archetype is similar to Lazy Reality, however, the main difference is that participants intended to return to VR immediately. It was usually employed if participants wanted to push the chair aside or untangle themselves from the HMD's cable.

B.4.5 Default. This code was used to denote a transition that did not fit into the other Archetypes. These transitions usually had no use other than transitioning into a different reality and were not part of an overarching (sub-)task. They were mostly motivated by the design of our apparatus, accompanied by comments such as "I switch to AR because only there I can scale the whole statue".