Manipulating Immersion: The Impact of Perceptual Incongruence on Perceived Plausibility in VR

Larissa Brübach[*]
HCI[†] and PIIS[‡] Group
University of Würzburg
Germany

Mona Röhm
University of Würzburg
Germany

Franziska Westermeier
HCI and PIIS Group
University of Würzburg
Germany

Marc Erich Latoschik
HCI Group
University of Würzburg
Germany

Carolin Wienrich
PIIS Group
University of Würzburg
Germany

Figure 1: Virtual supermarket environment used in the user study.

Abstract

This work presents a study where we used incongruencies on the cognitive and the perceptual layer to investigate their effects on perceived plausibility and, thereby, presence and spatial presence. We used a 2x3 within-subject design with the factors familiar size (cognitive manipulation) and immersion (perceptual manipulation). For the different levels of immersion, we implemented three different tracking qualities: rotation-and-translation tracking, rotation-only tracking, and stereoscopic-view-only tracking. Participants scanned products in a virtual supermarket where the familiar size of these objects was manipulated. Simultaneously, they could either move their head normally or need to use the thumbsticks to navigate their view of the environment. Results show that both manipulations had a negative effect on perceived plausibility and, thereby, presence. In addition, the tracking manipulation also had a negative effect on spatial presence. These results are especially interesting in light of the ongoing discussion about the role of plausibility and congruence in evaluating XR environments. The results can hardly be explained by traditional presence models, where immersion should not be an influencing factor for perceived plausibility. However, they are in agreement with the recently introduced Congruence and Plausibility (CaP) model and provide empirical evidence for the model's predicted pathways.

Index Terms: plausibility, immersion, presence, VR, congruence

[*]e-mail: larissa.bruebach@uni-wuerzburg.de
[†]Human-Computer Interaction Group
[‡]Psychology of Intelligent Interactive Systems Group

1 Introduction

The effects of different levels of immersion on one's XR experience have been studied extensively in the past [6, 16, 17]. It is quite an important concept in extended reality (XR, short for virtual reality VR, augmented reality AR, and mixed reality MR) research as "Immersion provides the boundaries within which PI (Place Illusion) can occur" [24]. Immersion can be defined as a set of sensorimotor and effective valid actions supported by the system (i.e., as a system characteristic). Its objective nature allows us to systematically influence it and study its effects on other qualia in XR. Qualia (plural for quale) are subjective and internal feelings that are caused by sensory perceptions [22]. This could be the feeling of presence, placeness, embodiment, social presence, and many more. The influences on presence, i.e., specifically on the PI, could be confirmed repeatedly [6].

PI has been defined to be orthogonal to the plausibility illusion (Psi) [24], i.e., according to Slater's definition, both should be independent. Given immersion's often validated and confirmed effects on PI, immersion should, therefore, not influence Psi, neither directly nor through the PI. However, the recent Congruence and Plausibility model (CaP) [14] challenges this assumption. The CaP model dissolves the dichotomy and instead describes the interplay of congruence cues on the sensation, perception, and cognition layer. PI would, therefore, still be significantly influenced by manipulations on the sensation or perception layer. Plausibility, which is focused on the model, is created as a weighted activation from all three layers. Accordingly, in addition to cognitive cues, sensational and perceptional manipulations caused by the system (i.e., immersion) can also influence the feeling of plausibility.

The aim of this paper is to empirically clarify this theoretical contradiction. To determine whether different levels of immersion ultimately influence perceived plausibility, we present a study using congruence manipulations on different layers of the CaP model. We chose the perceptual layer, which is manipulated through three different levels of immersion using different VR headset tracking types. Additionally, we caused incongruencies on the cognitive layer by manipulating the familiar size of objects within the scene. We conducted a user study to investigate how these incongruencies affected the perceived plausibility and demonstrated that the manipulations on both layers affect plausibility, providing empirical evidence for the CaP model.

2 Related Work

A meta-review from 2016 by Cummings and Bailenson [6] found 83 studies examining immersion's effects on presence. They found that technical immersion has a medium-sized effect on presence. However, since 2016, a lot has changed. XR hardware is changing and improving rapidly. Inevitably, more studies concerning immersion and its influence have been published.

In 2018, Born et al. [3] used a so-called exergame (where the player controls the game with body movements) to investigate the influence of two different immersion types: room-scale VR and standard screen. They found that a higher immersion positively affected presence, embodiment, motivation, and performance. Similarly, Seibert and Shafer [20] investigated how using a VR headset compared to a standard screen would influence spatial presence. Their results show that the higher immersion through the VR headset positively affected the user's feeling of spatial presence. Ahn et al. [1] tested two different display modes to manipulate immersion. Results show that the higher immersion display positively affected social presence. However, a lower immersion can also lead to adverse effects. In 2023, Wenk et al. [29] published a study examining immersion's influence on cognitive load, motivation, usability, and embodiment. They tested a VR headset, an AR headset, and a standard computer screen against each other. The AR headset and the computer screen, which were categorized as lower immersion, led to lower usability. Likewise, Tang et al. [27] found that a higher immersion led to a significantly lower cognitive load when testing a VR headset against an iPad. Jicol et al. [12] propose that while technical factors do not directly influence users' feeling of presence, human factors might mediate this relationship. Their study showed that a higher FoV seems to increase presence only when the users feel agency. What all these works have in common is that the immersion was manipulated by different display types. However, the tasks of the experiments were often different. Sometimes, the participants only had to observe something [1]. Sometimes, they had to interact with the environment, and the method of interaction may have been the same [3, 20, 29] or different between conditions [27]. This makes comparing the studies difficult, as there may have been unwanted confounds from the different presentation and interaction methods.

As we can see from Cummings review and the discussed papers, a higher immersion can have a positive effect on the user's feeling of presence and other factors, like motivation or embodiment. It is, therefore, an important criterion for XR experiences and plays a central role in one of the most widespread models to explain XR experiences and the emergence of presence by Slater [24]. Presence is defined as the realistic response from users to a virtual environment [24]. This has replaced the previous opinion where presence was defined as "the sense of being in a virtual environment" and was, therefore, closer to the definition of PI. In Slater's model, immersion is seen as the influencing factor for place illusion (PI) and as an objective system characteristic. It gives the frame within which PI can arise. PI is defined as *"the strong illusion of being in a place in spite of the sure knowledge that you are not there."* [24, p. 3]. This PI then influences the user's feeling of presence. Slater also introduced the plausibility illusion (Psi) as a cognitive construct influencing presence. Psi is, in turn, influenced by coherence [21]. Psi is the illusion that what is apparently happening is really happening, even though the user knows for sure that it is not. PI and Psi both contribute to presence and are seen as orthogonal factors, so they should not influence each other. This model can be seen in Fig. 2. Recently, Slater et al. [25] looked at this model again. Here, they reinforce their view that plausibility is one of the most important components in XR. It is, therefore, important to investigate its influence on the VR experience. They propose a combination of psychophysiological and qualitative measurement methods.

Objectively measuring Psi has been a challenge in the past. Slater et al. [26] employ a variant of the color matching theory, which posits that although individuals may perceive the same color differently, they can still agree on its label, such as "red." In their study, participants in a VR environment adjusted settings related to illumination, display size, navigation, and avatar until they matched their perceived levels of presence (PI) and plausibility (Psi) to a maximized reference experience. The findings indicate that participants' configurations varied based on their focus on either PI or Psi. Skarbez et al. [23] also examined how different coherence characteristics are prioritized by the participants using a similar approach. In their study, participants could control the settings for virtual human behavior coherence, virtual body coherence, physical interaction coherence, and scenario coherence. Both experiments were able to hint at which aspects of an XR environment are most important for participants. However, it is not possible to tell *how* plausible the environments were perceived by the participants. This makes it difficult to judge which aspects influence the plausibility evaluation and compare the results to those of other experiments.

According to Finnegan et al., [7], the correct distance perception is a critical factor in how realistic users perceive an XR application to be. Their work has shown that an intentional misalignment of audio and visual can lead to a more precise distance estimation in users [7, 8]. This more realistic perception may increase the sense of presence and spatial presence.

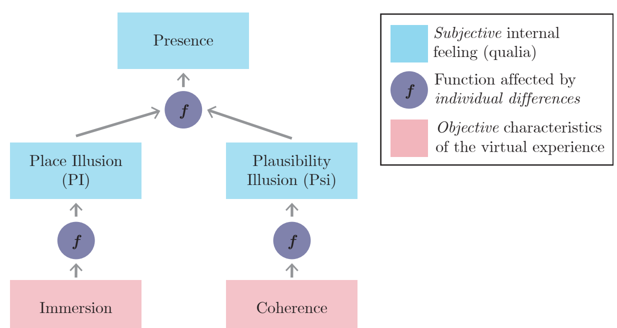

Figure 2: Slater's model, which describes the emergence of presence through place illusion and plausibility illusion [24] extended by Skarbez [21]. (layout redesigned and simplified by the authors)

Wirth et al. [32] proposed a model on the formation of spatial presence in various media applications. With new technologies emerging, they saw a lack of incorporating other media psychology aspects, like attention and involvement. Over time, many subtypes of presence were discussed, with spatial presence being closest to Minsky's original definition [15] of "the feeling of being there". In their view, the conviction of users to really be located in an environment is a crucial factor that can intensify existing media effects, like involvement or enjoyment. An important factor in their model is that the attention towards the environment must not be interrupted by higher cognitive processes [32] to not influence spatial presence. Incongruencies on a cognitive level in an XR environment might interrupt this attention and, therefore, spatial presence.

A more recent model, called the Congruence and Plausibility model (CaP), by Latoschik and Wienrich [14] changed this view and the role of immersion. They see plausibility as the main influencing factor on one's XR experience. In contrast to Slater's understanding, plausibility is seen as a holistic construct rather than just a cognitive one. Qualia like placeness (in replacement for the place illusion), embodiment, presence, and others are seen to be influenced by plausibility. The plausibility, in turn, results from a weighted function of different (in)congruencies on the three-layer manipulation space. These three layers are the cognitive (top-down), perceptual (bottom-up), and sensation (bottom-up) layer. This model allows for individual manipulation of congruence on these layers. While all resulting incongruencies influence plausibility, it is possible that specific manipulations target specific qualia. Perceptual incongruencies, for example, might influence spatial qualia, like spatial presence, more than other qualia. In this model, immersion is seen to cause congruencies or incongruencies at the perceptual layer, which impact plausibility via the weighted activation function. There are first studies that investigated the direction of causality within the model [4, 30, 5]. Brübach et al. [4] used a perceptual manipulation (missing gravity) to cause incongruencies and influence plausibility. Additionally, they tried to counteract these incongruencies with a cognitive manipulation (contextual framing). Their results show that the perceptual manipulation, which is seen to be on a lower level than the cognitive manipulation, was able to influence plausibility. However, the cognitive manipulation was insufficient to counteract the perceptual incongruencies. After that, Brübach et al. [5] looked at ways to systematically influence the plausibility of a VR environment. They tested four different incongruencies on different CaP model layers: familiar size, object placement (both on the cognitive layer), audio, and light (both on the perceptual layer). The familiar size manipulation was shown to be the only manipulation that caused reduced plausibility. So far, the results look promising and support the implications of the new model. If the causality direction is correct, immersion and plausibility are not orthogonal factors, but immersion would influence plausibility through congruence. This new model and where immersion is located can be seen in Fig. 3.

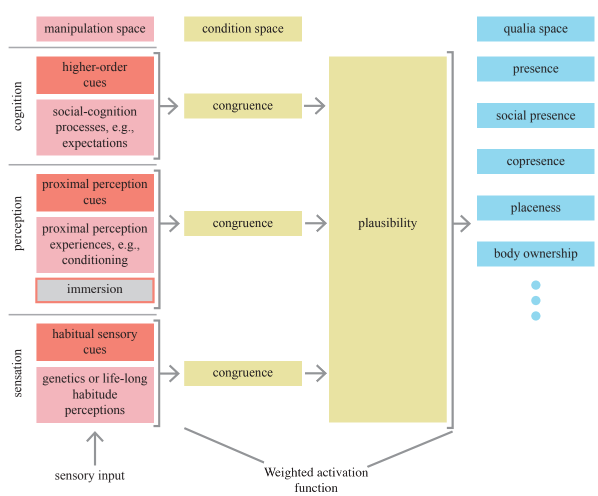

Figure 3: The new Congruence and Plausibility model by Latoschik and Wienrich [14] and the immersion location. (layout redesigned by the authors)

In summary, the different models result in different predictions that motivate the research question of the current paper: Does immersion influence plausibility?

2.1 Summary and Contribution

As we have seen, the influence of immersion has been studied to a great extent. The concept has been around for almost 40 years [6]. However, the influence of immersion on plausibility and vice versa is understudied. There are contradictions in models that either postulate an orthogonality [24, 21] or a direct influence [14] of the two factors: immersion (and thus, PI) and plausibility. The recent CaP sees immersion as an essential factor influencing the user's perceived plausibility. So, we must look at the relationship between these two concepts. This can also help to test the causality direction within the CaP model. In this way, it is possible to evaluate and, if necessary, improve the new theoretical model. This can lead to a better understanding of the user experience and the resulting qualia in XR.

We propose a 2x3 factorial study that uses incongruencies on the cognitive and the perceptual layer to narrow this research gap. The cognitive incongruencies are implemented through a familiar size manipulation as proposed by Brübach et al. [5]. The perceptual incongruencies are implemented using three levels of immersion that affect the VR headset tracking. Following the CaP model, we assume that both factors influence perceived plausibility.

2.2 Reasoning for the choice of the manipulation on the cognitive layer

We use different depth cues, like motion parallax or occlusion, in our everyday lives to assess distances. Another depth cue is the so-called familiar size. We can estimate distances by comparing the perceived size of an object with the size of the object that we know. Familiar size, therefore, relies on our previous knowledge to work. Or the other way around, if we know how big an object is, then we also know how the size must change relative to the distance to the object. If the object is further away, it should be smaller than if it is closer [10]. This well-known fact from the psychology of perception offers a simple way of manipulation. If objects do not become smaller even though they are further away, this creates an incongruence on a cognitive level.

In previous work, the familiar size was categorized as a cognitive manipulation and was shown to be able to influence the perceived plausibility [5]. For this reason, we used this manipulation to influence the cognitive level. We implemented a simple supermarket scene where participants were asked to scan different objects. The familiar size of these objects to be scanned was manipulated as they moved along the conveyor belt.

2.3 Reasoning for the choice of the manipulation on the perceptional layer

In previous studies, the immersion was often manipulated through different types of media (i.e., VR glasses vs. desktop applications) [3, 20, 1, 27, 29]. However, this has led to other unintended differences, such as the control of the application or the resolution of the screens. For our study, we wanted to keep as many aspects similar as possible. That is why we manipulated the tracking qualities within a VR application. For the perceptual manipulation, we implemented three different VR headset tracking qualities, which can be seen as three different levels of immersion. We used normal rotation and translation tracking, one where the translation is disabled and one where both the rotation and translation are disabled. In the manipulated tracking conditions, participants can use the controller thumbsticks to move their VR view to fulfill the task. This manipulation is used to manipulate the perceptual layer, as immersion is generally processed on the perception layer.

2.4 Hypotheses

Previous research has shown that incongruencies on both the cognitive and the perceptual layer can have an influence on perceived plausibility [4, 11, 30, 5]. To further validate the CaP model we focused on the influence of the perceptual level on the perceived plausibility. Our main goal is to find out whether manipulation on the perceptual level through immersion can influence the perceived plausibility. Following the predictions of the CaP model, we have the following first two hypotheses:

- **H1** Incongruencies caused by the familiar size manipulation will result in a significantly lower perceived plausibility.
- **H2** Incongruencies caused by the tracking manipulation will result in a significantly lower perceived plausibility.

In line with previous research, we expect the immersion manipulation to result in lower presence scores [6]. We also expect the familiar size manipulation to affect presence. We formulate the following two hypotheses:

- **H3** Incongruencies caused by the familiar size manipulation will result in a significantly lower presence.
- **H4** Incongruencies caused by the tracking manipulation will result in a significantly lower presence.

The tracking manipulation is on a lower level than the familiar size manipulation. The CaP model hypothesizes that higher-level manipulations have no downward effect. As spatial presence is a lower-level quale, we do not expect the higher-level manipulation to have an influence [4]. However, the lower-level manipulation should influence spatial presence. Therefore, our last hypothesis is as follows:

- **H5** Incongruencies caused by the tracking manipulation will result in a significantly lower spatial presence, while the familiar size manipulation will have no effect.

3 Methods

3.1 Study Design

The study had a 2x3 factorial within-subject design with the variables familiar size and VR headset tracking types. The resulting conditions can be seen in Tab. 1.

Table 1: The different conditions in the two experiments.

Tracking Type	Familiar Size Manipulated	Not Manipulated
Rotation and Translation Tracking (RT)	A	B
Rotation Only Tracking (RO)	C	D
Stereoscopic View Only (SVO)	E	F

The first independent variable, familiar size, had two conditions: manipulated and not manipulated. In the manipulated condition, we affected the familiar size of the objects that participants had to scan. Specifically, this meant that the objects retained their size regardless of the distance to the participant. In the not manipulated condition, the objects' size changed as expected and known from reality when they moved along the conveyor belt. The second independent variable, VR headset tracking types, had three conditions: Rotation and Translation Tracking (RT Tracking), Rotation Only Tracking (RO Tracking), and Stereoscopic View Only (SVO Tracking). In the RT Tracking condition, the movement of the headset was as expected. In the RO Tracking condition, the participants could rotate their head, and the camera view in VR would follow. However, the translation of the headset was not transferred to the VR game view. To change their view, participants had to use the thumbstick of one controller. In the last condition, SVO Tracking, neither the rotation nor the translation was transferred from the headset to the VR game view. This meant that no movement of the participant affected the camera view in VR. Here, participants only had a stereoscopic view of the scene. Again, participants had to use the controllers to change their VR view. One thumbstick was responsible for changing the rotation, while the other was responsible for the translation.

Table 2: Latin square used in our study to avoid sequence effects. Note that the last two columns are not part of the Latin square and were just appended.

A	D	B	C	E	F
B	A	C	D	E	F
C	B	D	A	F	E
D	C	A	B	F	E

We considered the condition with normal familiar size and RT Tracking as the control condition.

As fully counterbalancing the conditions would require a very large number of participants, we used a Latin square to control for sequence effects. However, we expected the SVO tracking to influence the VR sickness of the participants significantly, so it was decided that these two conditions should always be last. Therefore, they were excluded from the Latin square and appended at the end with two different orders. The used Latin square can be seen in Tab. 2.

3.2 Application

We used a high-end computer with an Nvidia Geforce RTX 3080 GPU with 64 GB of RAM and an Intel i9-11900K CPU. The application was developed in the Unity Engine (v2020.3.21f1)[1] using the Steam VR Plugin (v2.7.3) [2] and the XR Interaction Toolkit (v2.3.0) [3]. The HP Reverb headset was used for development and user study.

A supermarket asset was used to build a virtual supermarket environment [4]. We tried to keep aspects like the objects' sizing, the conveyor belt's length, and the lighting as realistic as possible. The supermarket was made of three tills and several shelves. The participant was placed behind one of the tills as (s)he was supposed to act as a cashier. They had no virtual body, and only the controllers were shown in VR. The environment can be seen in Fig. 1.

A tutorial was implemented to help people get familiar with the controllers. This was especially important for the rotation-only and the stereoscopic-view-only conditions, as participants had to use the thumbsticks to move their view. It consisted of a simple blue tiled room [5] with a conveyor belt that moved simple cubes from left to right. Participants were instructed to move their view around and grab cubes. They were told to do this until they felt comfortable with all controls. The tutorial can be seen in Fig. 4.

[1] https://unity.com/releases/editor/whats-new/2020.3.21
[2] https://github.com/ValveSoftware/steamvr_unity_plugin/releases
[3] https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@2.3/manual/index.html
[4] https://assetstore.unity.com/packages/3d/environments/modern-supermarket-186122
[5] https://assetstore.unity.com/packages/3d/environments/speedtutor-test-scene-free-159460

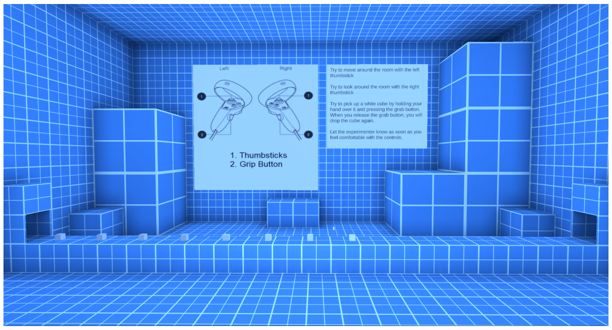

Figure 4: Tutorial scene where participants could familiarize themselves with the controls. The asset *SpeedTutor Test Scene* from the Unity Asset Store was used to create this environment.[5]

3.2.1 Familiar Size Manipulation Implementation

We used the familiar size to manipulate the objects. To manipulate the familiar size, we kept the object's size seemingly the same no matter the distance to the participant. So when the objects moved along the conveyor belt, they seemed to remain the same size. We used everyday objects like juice cartons and washing powder cartons that people could easily recognize. Normally, the relationship between the apparent size of an object and its distance from the observer is indirectly proportional. This can be expressed with the formula

$$apparentsize = initialsize * (initialdistance/currentdistance) \tag{1}$$

Depending on the distance, we must scale the object up or down to keep the apparent size the same. The scaling factor is calculated as follows

$$objectscalingfactor = currentdistance/initialdistance \tag{2}$$

The normal size change can be seen in Fig. 5a, and the manipulated familiar size can be seen in Fig. 5b.

3.2.2 Tracking Manipulation Implementation

In the RT tracking condition, everything worked normally. Participants could rotate their head and move it around, and the VR view would change accordingly. This serves as our control condition.

In the RO tracking condition, the VR view of the participants would only follow the head rotation. However, if they moved the headset in any direction, the VR view would not follow. If participants wanted to change the location of their VR view, they had to use the thumbstick on the controller. To avoid the controllers moving away from the headset when the participant moved their head, we moved them inversely to the head movement. This way, they would stay in the same position in relation to the headset.

In the SVO tracking condition, neither the rotation nor the translation of the headset was transferred onto the VR view. This meant that the VR view would stay the same no matter how the participant would rotate or move the headset. The headset acted as a stereoscopic display in this scenario. To look around, participants had to use one thumbstick for their rotation and one for the translation.

To avoid excessive VR sickness, we asked participants in the RO and SVO tracking conditions to use the thumbstick controls only when necessary.

3.3 Measures

A variation of the questions from Brübach et al. [4] was used to measure the perceived plausibility. The word "object" was changed to "scenario" to ensure that participants focused on the whole scene and what was happening, similar to Brübach et al. [5]. The thirteen items are on a 7-point Likert scale ranging from *I do not agree at all (1)* to *I fully agree (7)*. The questionnaire can be seen in Tab. 3.

Table 3: Perceived plausibility questionnaire proposed by Brübach et al. [4]. The questions were adapted for this study by replacing *object behavior* with *scenario* as in Brübach et al. [5].

no	question
1	I am used to scenarios behaving this way.
2	In everyday life, I expect scenarios to behave this way.
3	I have seen scenarios behave this way in real life.
4	The behavior of the scenario is unusual for me. [1]
5	I do not know the behavior of this scenario from real life. [1]
6	I had a prior expectation of how the scenario would behave.
7	I expected the behavior of the scenario.
8	I have seen this scenario behavior in movies, games etc. before.
9	I was surprised by the behavior of the scenario. [1]
10	I had no idea that the scenario would behave this way. [1]
11	The behavior of cause and effect matched the scenario.
12	The behavior of the scenario made sense.
13	I think this behavior of the scenario is impossible. [1]

[1] Question is inverted.

We used the Igroup Presence Questionnaire (IPQ) by Schubert et al. [19] to measure presence, with the three subscales, *spatial presence*, *involvement*, and *experienced realism*, as well as one item that does not belong to a subscale. The fourteen items of the questionnaire are on a scale from 0 to 6. There are five questions on the *spatial presence* subscale, four on the *involvement* subscale, and four on the *experienced realism* subscale. The wording of the endpoints varies between the questions.

We expect using the thumbsticks to control the VR view will cause a higher workload. It is quite unintuitive to need external controls to change our view. To control for possible differences between conditions in the workload of participants during the experiment, the NASA-TLX by Hart et al. [9] was used. It has six subscales: *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustrations*. The items are measured on a scale from 0 to 100.

The Virtual Reality Sickness Questionnaire (VRSQ) by Kim et al. [13] was used to control for VR sickness. It measures sickness caused by virtual reality with the two dimensions *oculomotor* and *disorientation*. It has five for *disorientation* and four items for *oculomotor*. Each item can have a score from 0 (not at all) to 3 (strong).

Additionally, to control for individual differences between participants regarding the tendency to feel immersed, we used the Immersive Tendency Questionnaire (ITQ) by Witmer et al. [33]. It has three subscales, *focus*, *involvement*, and *games*, with seven, seven, and two questions, respectively, as well as two questions that do not belong to a subscale. The items are on a scale from 1 to 7 and have different wordings for the endpoints and anchors.

3.4 Procedure

The procedure can be seen in Fig. 6. The experiment took approximately one hour to complete. Participants started by signing the consent forms. They answered some demographical questions, the ITQ, and questions about their previous VR usage. Afterward, they started with the tutorial. The experimenter briefly explained the controls and asked only to use the thumbstick controls when necessary as they had a risk of increasing VR sickness. (S)he also reminded the participants that they could stop the experiment at any point without any consequences for them. When the participants felt comfortable with the controls and had no more questions, they

(a) Normal Familiar Size.

(b) Manipulated Familiar Size.

Figure 5: Familiar Size Manipulations.

filled out the first VRSQ as a baseline, and the experiment started. They began with one of the six conditions. The items appeared on the conveyor belt one after another with 3 seconds in between. Their task was to pick up the items on the conveyor belt, scan them, and put them to their left. They were free to use their dominant hand. A typical beep sound would play every time an item was scanned. They had to scan a total of seventeen items and re-scan the items until the beep could be heard. After scanning the participants placed the items on the other side of the register where they remained until the condition was over. After each condition, they answered the plausibility questionnaire, the IPQ, the NASA-TLX, and the VRSQ. After they had completed all six conditions, they were told about the intentions of the experiment.

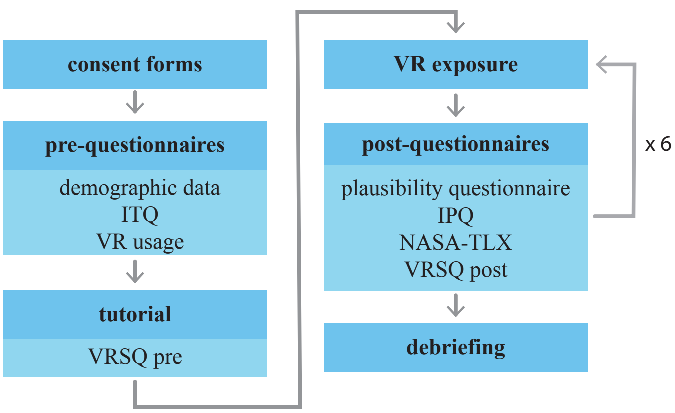

Figure 6: Experiment procedure.

3.5 Participants

A power analysis showed that a minimum of 28 participants was necessary for the experiment (effect size of 0.2, estimated power of 0.8). However, to enhance the robustness of our findings, we decided to include a slightly larger sample size. Given the experimental design, which comprises four condition orders, it was important to ensure the number of participants was divisible by 4. To meet this criterion, we recruited a total of 34 participants. This number allowed us to account for the exclusion of two participants while still maintaining the required sample size for our experimental conditions.

Thirty-four participants took part in the experiment. They received compensation equivalent to 12 $ in the currency of the country where the experiment was conducted for their participation. Two participants had to be excluded due to technical difficulties during the experiment, leaving thirty-two participants for the data analysis. The pool was divided into twenty-two female and ten male participants. The age ranged from 19 to 61, with a mean age of $M = 28.56$ ($sd = 11.02$). Twenty-one were students, eight were employees, one had no current occupation, one was a retiree, and one was a pupil. Seven had none to one hour of VR experience, five had one to three hours of experience, four had five to ten hours of experience, five participants had ten to twenty hours of experience, and one had more than twenty hours of VR experience. Twenty-two participants played video or smartphone games for less than one hour a day, eleven between one and three hours a day and only one participant between three and five hours daily. The ITQ showed a mean value of 4.15 (SD = .74) and a normal distribution within the sample.

3.6 Analysing methods

We used both ANOVA and ANCOVA for the data analysis with a significance level of $\alpha < .05$. The assumption of normal distribution was violated for most of the dependent variables. However, an ANOVA is generally robust against these violations, especially with a larger sample size [2, 18]. The assumption of sphericity was also violated in some cases. Wherever this was the case, the Greenhaus-Geisser correction for $\varepsilon < .75$ or the Huynh-Feldt correction for $\varepsilon > .75$ was applied. This follows the recommendations of Verma [28].

4 Results

All means, and standard deviations for each condition and questionnaire can be seen in Tab. 4.

4.1 Control Variables

The sphericity assumption was violated for the NASA-TLX, and, therefore, the corresponding correction was applied. We found a significant main effect for the NASA-TLX between all tracking conditions ($F(2,62) = 83.82, p < .001 \eta^2 = .638$). We found no significant main effect between the familiar size manipulation conditions. We found no interaction effect between the familiar size and tracking manipulation.

Post-hoc tests using the Holm correction revealed a significant main effect with a higher workload between the SVO tracking and both the RT tracking ($p_{holm} < .001$) and the RO tracking ($p_{holm} < .001$).

The sphericity assumption was violated for the VRSQ. Therefore, the Greenhouse-Geisser correction was applied. The VRSQ scores were calculated as the difference between before and after the exposure. We found a significant main effect for the headset tracking condition ($F(1.34, 41.56) = 23.12, p < .001, \eta^2 = .184$),

Table 4: Means and standard deviation for each condition and questionnaire.

Tracking		RT		RT		RO		RO		SVO		SVO	
Familiar size manipulation		yes		no		yes		no		yes		no	
Questionnaire	Subscale	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
VRSQ		- 2.87	10.81	-0.91	3.34	1.52	4.28	0.48	3.21	6.85	12.64	8.65	13.45
Perceived Plausibility		4.53	1.91	5.64	0.93	4.13	1.82	5.09	1.37	2.46	1.43	2.75	1.16
IPQ	overall	3.34	1.03	3.51	0.96	3.17	1.03	3.43	1.07	2.48	1.12	2.43	1.35
IPQ	spatial presence	3.78	1.24	3.81	1.17	3.52	1.21	3.66	1.21	2.66	1.44	2.61	1.59
NASA-TLX	overall	11.67	12.29	12.22	13.89	18.93	17.40	20.3	17.30	48.42	20.27	47.53	22.93
NASA-TLX	mental	10.92	15.89	9.38	12.18	17.86	19.23	16.03	17.76	52.13	25.49	52.33	29.25
NASA-TLX	physical	10.36	13.23	11.34	18.77	18.47	20.79	17.03	17.60	42.73	26.45	41.78	30.03
NASA-TLX	temporal	11.31	12.24	15.58	17.02	19.84	21.45	22.02	22.78	32.97	26.60	35.47	27.19
NASA-TLX	performance	16.50	21.84	17.22	21.68	24.02	25.19	28.06	26.20	53.81	26.28	51.53	26.09
NASA-TLX	effort	10.67	15.41	11.28	14.01	17.98	17.33	17.31	16.75	53.86	26.91	53.52	28.22
NASA-TLX	frustration	10.22	13.99	8.48	16.33	15.36	20.01	19.69	24.36	54.94	32.60	50.48	28.56

but no significant effect for the familiar size manipulation condition. We found no interaction effect between the familiar size and tracking manipulation.

Post-hoc tests using the Holm correction showed a significant main effect with higher VR sickness for the SVO tracking than the RT tracking ($p < .001$) and the RO tracking ($p < .001$).

4.2 Plausibility

As the perceived plausibility questionnaire has not been validated before, we calculated Cronbach's α to check the internal consistency. We combined the answers of the 32 participants over the 6 conditions, resulting in 192 answers. For the 13 items, Cronbach's α is .954, which indicates a very high internal consistency.

The sphericity assumption was not violated for the plausibility questionnaire, and we did not apply any corrections. We found a significant main effect for both the headset tracking ($F(2,62) = 62.92, p < .001, \eta^2 = .012$) and the familiar size manipulation ($F(1,31) = 14.09, p < .001, \eta^2 = .06$), with lower plausibility ratings in the manipulated condition. We found no interaction effect between the familiar size and tracking manipulation.

Post-hoc tests using the Holm correction revealed that the SVO tracking had a significant main effect with lower plausibility ratings than both the RT tracking ($p_{holm} < .001$) and the RO tracking ($p_{holm} < .001$). The RO tracking also had a significant main effect with lower plausibility ratings than the RT tracking ($p_{holm} = .046$).

As we have found a significant main effect for both control variables for the tracking manipulation, we wanted to check whether they influenced the perceived plausibility. We calculated an ANCOVA using the NASA-TLX and the VRSQ as covariates. We still found a significant main effect between headset tracking ($F(2,184) = 21.86, p < .001, \eta^2 = .144$) conditions and the familiar size conditions ($F(1,184) = 14.01, p < .001, \eta^2 = .047$. However, we did not find a significant main effect for the two covariates. We, therefore, assume that the higher workload and higher VR sickness in the tracking conditions did not affect the plausibility ratings.

4.3 IPQ

The sphericity assumption was violated for the IPQ. Therefore, the Greenhouse-Geisser correction was applied. The ANOVA showed a significant main effect for both the headset tracking ($F(1.36,42.15) = 30.64, p < .001, \eta^2 = .396$) and the manipulated familiar size ($F(1,31) = 4.5, p = .042, \eta^2 = .009$) with lower presence scores in the manipulated condition.

Post-hoc tests using the Holm correction showed a significant main effect with lower IPQ scores for SVO tracking compared to both the RT tracking ($p_{holm} < .001$) and the RO tracking ($p_{holm} < .001$). There was no significant main effect between the RT and the RO tracking condition.

We found a significant main effect in the subscale *spatial presence* for the tracking manipulation ($F(1.38,42.68) = 28.37, p < .001, \eta^2 = .361$). However, there was no significant main effect for the familiar size manipulation.

Post-hoc tests using the Holm correction showed a significant main effect with lower spatial presence ratings between the SVO tracking and both the RT tracking ($p_{holm} < .001$) and the RO tracking ($p_{holm} < .001$).

We found no interaction effect between the familiar size and tracking manipulation for both the overall presence score and the spatial presence subscale.

We also wanted to check if the workload and VR sickness affected the IPQ scores between the tracking conditions. Again, we calculated an ANCOVA using the NASA-TLX and the VRSQ as covariates. We still found a significant main effect for the headset tracking ($F(2,184) = 11.61, p < .001, \eta^2 = .102$). However, we did not find a significant main effect for the familiar size conditions and for the two covariates. We, therefore, assume that the higher workload and higher VR sickness in the tracking conditions did not affect presence.

5 Discussion

The first hypothesis **H1** *Incongruencies caused by the familiar size manipulation will result in a significantly lower perceived plausibility.* can be accepted. We found a significant main effect between the control and manipulated familiar size conditions. Similarly, we can accept **H2** *Incongruencies caused by the tracking manipulation will result in a significantly lower perceived plausibility*. Both the RO and the SVO tracking lead to significantly lower plausibility scores than the control condition, RT tracking. As we have seen, higher workload and higher VR sickness did not have an influence here. We did not find a significant interaction effect between the two manipulations. This means that the two variables do not seem to affect each other and can independently affect the perceived plausibility. These results are not surprising for the familiar size manipulation. Both the current presence model [24, 21] and the CaP [14] model predict that a cognitive incongruence will influence the perceived plausibility. However, it is different in terms of the influence of the perceptual manipulation. According to the Slater model, Place Illusion, which arises within the immersion frame, and Plausibility Illusion are orthogonal factors. Immersion should, therefore, not influence perceived plausibility directly or indirectly through Place Illusion. In contrast, our results show that tracking manipulation does affect the perceived plausibility. This relationship, on the other hand, can be explained by the CaP model, where different levels of immersion can lead to incongruence on the perceptual layer and, therefore, influence plausibility.

We can also accept both hypothesis **H3** *Incongruencies caused by the familiar size manipulation will result in a significantly lower presence* and **H4** *Incongruencies caused by the tracking manipulation will result in a significantly lower presence*. We found significant main effects for the IPQ for both manipulations compared to the control condition. An ANCOVA showed that the significantly higher workload and VR sickness did not affect the presence ratings in the tracking conditions. We expected these results following both the Slater and the CaP model. Previous studies have shown that a higher immersion can lead to a higher fielding of presence, so the reverse effect is not surprising [3, 20, 1].

Lastly, we can accept the hypothesis **H5** *Incongruencies caused by the tracking manipulation will result in a significantly lower spatial presence while the familiar size manipulation will have no effect*. We found significant effects on the spatial presence subscale of the IPQ for the tracking manipulation. However, no significance was found for the familiar size manipulation. The tracking manipulation is on a lower level than the familiar size manipulation in the CaP model, which might explain this effect. As spatial presence is seen as a low-level quale, it is likely only influenced by lower-level manipulations. The downward impact of the cognitive manipulation is not strong enough to have an effect on the participant's spatial presence. This is in line with the findings of Brübach et al. [4]. They also found that their lower-level perceptual manipulation (missing gravity) could not be counteracted by the higher-level cognitive manipulation (framing). The Slater / Skarbez model can also explain the influence of the immersion manipulation on spatial presence. In this model, immersion directly influences the PI. However, the familiar size manipulation would be a congruence manipulation and should, therefore, not influence the PI. This aligns with our findings in the spatial presence subscale of the IPQ.

5.1 Limitations and Future Work

The general flaws of a within-subjects experiment also apply to this study. We tried to keep sequence effects as minimal as possible using a Latin square. However, as the SVO tracking condition was excluded here, we cannot guarantee that there were no sequence effects. None of the participants had to stop the experiment early due to VR sickness, which can partly be attributed to the sequence design of putting the SVO condition last.

We acknowledge that there is a lack of a valid measurement instrument for plausibility and that the Perceived Plausibility Questionnaire has not yet undergone formal validation. Therefore, we used the operationalizations that have been commonly used in recent studies on the subject [11, 4, 5, 30, 31] where this questionnaire was also used. This approach was undertaken to facilitate comparability with existing literature and ensure methodological coherence within the field.

Another issue with the Perceived Plausibility Questionnaire is that the instruments used to measure perceived plausibility might not be entirely appropriate for measuring PSI, as Slater defines it. While there are approaches to compare plausibility through behavioral methods, there is, to our knowledge, no dedicated, objective questionnaire regarding plausibility or Psi. Certain compromises were made in the pursuit of comparing these models.

Many participants were confused by the SVO tracking at first. They asked whether this was intentional or a problem with the system. They could continue the experiment after assuring them this was intentional and a short reiteration of the controls. Interestingly, some commented that they felt like their movements and rotations were inversed. This might be because when we move our head forward, the objects in front of us come closer and seem bigger. However, because there was no translation in the VR view, participants felt like the objects moved away from them as they could feel their head moving forward.

Even though we controlled for a high workload and VR sickness in our results, we cannot entirely disregard the significant differences concerning the task complexity between the conditions. The SVO tracking, in particular, caused an increased workload and intense VR sickness. The unintuitive use of the thumbsticks to control the VR view was an added barrier. This might also be due to the low gaming experience of the participants, as the controls in video games are sometimes similar, and this experience could have helped with understanding the controls. Future work should try and find alternative ways to manipulate the immersion further. Maybe a third-person perspective or a standard screen in combination with a controller would be better suited. A manipulation of the field of view, for example, could be interesting.

Lastly, the supermarket environment was quite abstract. There were no sounds, like other customers or background music, except the beeping. Also, the objects appeared out of nowhere on the conveyor belt and had no weight. These details made the experience quite different from a real supermarket. This could have caused a priori influences on the perceived plausibility. However, we still found significant main effects between the manipulations and the control conditions. We, therefore, believe that the lower perceived plausibility and lower presence were not affected by the faults in the environment.

6 Conclusion

Recent developments in understanding XR experiences led to a new CaP model. While previous research supports this model, more work is needed to fully understand the emergence of different qualia and the direction of causality within this model. It is also essential to understand the differences and, therefore, the advantages over previous models.

We present a study that uses incongruencies on different layers to research their influence on perceived plausibility, presence, and spatial presence. Our conditions included perceptual incongruencies through immersion with different VR headset tracking types (RT, RO, and SVO) and cognitive incongruencies through the familiar size of objects with the factors manipulated and not manipulated.

Our results show that cognitive incongruence affects perceived plausibility. This influence, in turn, impacts presence but not spatial presence. Both the Slater and the CaP models explain these results. However, our results also show that different types of immersion cause incongruencies that affect perceived plausibility and, thereby, presence and spatial presence. Following previous presence models, immersion should not influence plausibility, neither directly nor indirectly, through the place illusion. In the CaP model, immersion is located on the perceptual layer in the manipulation space and can affect plausibility by causing incongruencies on this layer. The results can be explained with the CaP model but not all of them with previous presence models. This is another indication of the validity of this model. Future research should continue to test the different models. Only then will it be possible to understand XR experiences better and design them in the best possible way for users.

Acknowledgments

This research has been funded by the Bavarian State Ministry For Digital Affairs in the project XR Hub (project number A5-3822-2-16).

References

[1] D. Ahn, Y. Seo, M. Kim, J. H. Kwon, Y. Jung, J. Ahn, and D. Lee. The effects of actual human size display and stereoscopic presentation on users' sense of being together with and of psychological immersion in a virtual character. *Cyberpsychology, Behavior, and Social Networking*, 17(7):483–487, 2014. doi: 10.1089/cyber.2013.0455 2, 3, 8

[2] M. J. Blanca Mena, R. Alarcón Postigo, J. Arnau Gras, R. Bono Cabré, and R. Bendayan. Non-normal data: Is anova still a valid option? *Psicothema, 2017, vol. 29, num. 4, p. 552-557*, 2017. doi: 10.7334/psicothema2016.383 6

[3] F. Born, S. Abramowski, and M. Masuch. Exergaming in vr: The impact of immersive embodiment on motivation, performance, and perceived exertion. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8. IEEE, 2019. doi: 10.1109/VS-Games.2019.8864579 2, 3, 8

[4] L. Brübach, F. Westermeier, C. Wienrich, and M. E. Latoschik. Breaking Plausibility Without Breaking Presence - Evidence For The Multi-Layer Nature Of Plausibility. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2267–2276, May 2022. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2022.3150496 3, 4, 5, 8

[5] L. Brübach, F. Westermeier, C. Wienrich, and M. E. Latoschik. A systematic evaluation of incongruencies and their influence on plausibility in virtual reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ISMAR '23, pp. 894–901. IEEE Computer Society, 2023. doi: 10.1109/ISMAR59233.2023.00105 3, 4, 5, 8

[6] J. J. Cummings and J. N. Bailenson. How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media psychology*, 19(2):272–309, 2016. doi: 10.1080/15213269.2015.1015740 1, 2, 3, 4

[7] D. J. Finnegan, E. O'Neill, and M. J. Proulx. Compensating for distance compression in audiovisual virtual environments using incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 200–212, 2016. doi: 10.1145/2858036.2858065 2

[8] D. J. Finnegan, E. O'Neill, and M. J. Proulx. An approach to reducing distance compression in audiovisual virtual environments. In *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–6. IEEE, 2017. doi: 10.1109/SIVE.2017.7901607 2

[9] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, pp. 904–908. Sage publications Sage CA: Los Angeles, CA, 2006. doi: 10.1177/154193120605000909 5

[10] C. B. Hochberg and J. E. Hochberg. Familiar size and the perception of depth. *The Journal of Psychology*, 34(1):107–114, 1952. 3

[11] M. Hofer, T. Hartmann, A. Eden, R. Ratan, and L. Hahn. The Role of Plausibility in the Experience of Spatial Presence in Virtual Environments. *Frontiers in Virtual Reality*, 1, 2020. doi: 10.3389/frvir.2020.00002 4, 8

[12] C. Jicol, C. Clarke, E. Tor, R. M. Dakin, T. C. Lancaster, S. T. Chang, K. Petrini, E. O'Neill, M. J. Proulx, and C. Lutteroth. Realism and field of view affect presence in vr but not the way you think. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2023. doi: 10.1145/3544548.3581448 2

[13] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics*, 69:66–73, 2018. doi: 10.1016/j.apergo.2017.12.016 5

[14] M. E. Latoschik and C. Wienrich. Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. *Frontiers in Virtual Reality*, 3:694433, June 2022. doi: 10.3389/frvir.2022.694433 1, 3, 7

[15] M. Minsky. Telepresence. *OMNI magazine*, pp. 45–52, 1980. 2

[16] S. Morélot, A. Garrigou, J. Dedieu, and B. N'Kaoua. Virtual reality for fire safety training: Influence of immersion and sense of presence on conceptual and procedural acquisition. *Computers & Education*, 166:104145, 2021. doi: 10.1016/j.compedu.2021.104145 1

[17] T. Rose, C. S. Nam, and K. B. Chen. Immersion of virtual reality for rehabilitation - review. *Applied Ergonomics*, 69:153–161, 2018. doi: 10.1016/j.apergo.2018.01.009 1

[18] E. Schmider, M. Ziegler, E. Danay, L. Beyer, and M. Bühner. Is it really robust? *Methodology*, 2010. doi: 10.1027/1614-2241/a000016 6

[19] T. Schubert, F. Friedmann, and H. Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281, June 2001. doi: 10.1162/105474601300343603 5

[20] J. Seibert and D. M. Shafer. Control mapping in virtual reality: effects on spatial presence and controller naturalness. *Virtual Reality*, 22(1):79–88, 2018. doi: 10.1007/s10055-017-0316-1 2, 3, 8

[21] R. Skarbez, F. P. Brooks, and M. C. Whitton. Immersion and coherence in a stressful virtual environment. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11. ACM, Tokyo Japan, Nov. 2018. doi: 10.1145/3281505.3281530 2, 3, 7

[22] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton. A Survey of Presence and Related Concepts. *ACM Computing Surveys*, 50(6):1–39, Nov. 2018. doi: 10.1145/3134301 1

[23] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton. A Psychophysical Experiment Regarding Components of the Plausibility Illusion. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1369–1378, Apr. 2017. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2017.2657158 2

[24] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, Dec. 2009. doi: 10.1098/rstb.2009.0138 1, 2, 3, 7

[25] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality. *Frontiers in Virtual Reality*, 3:914392, June 2022. doi: 10.3389/frvir.2022.914392 2

[26] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics*, 29(4):1–9, July 2010. doi: 10.1145/1778765.1778829 2

[27] Q. Tang, Y. Wang, H. Liu, Q. Liu, and S. Jiang. Experiencing an art education program through immersive virtual reality or ipad: Examining the mediating effects of sense of presence and extraneous cognitive load on enjoyment, attention, and retention. *Frontiers in Psychology*, 13:957037, 2022. doi: 10.3389/fpsyg.2022.957037 2, 3

[28] J. Verma. *Repeated measures design for empirical researchers*. John Wiley & Sons, 2015. 6

[29] N. Wenk, J. Penalver-Andres, K. Buetler, T. Nef, R. M. Müri, and L. Marchal-Crespo. Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Reality*, 27(1):307–331, 2023. doi: 10.1007/s10055-021-00565-8 2, 3

[30] F. Westermeier, L. Brübach, M. E. Latoschik, and C. Wienrich. Exploring Plausibility and Presence in Mixed Reality Experiences. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–10, 2023. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2023.3247046 3, 4, 8

[31] F. Westermeier, L. Brübach, C. Wienrich, and M. E. Latoschik. Assessing depth perception in vr and video see-through ar: A comparison on distance judgment, performance, and preference. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2140 – 2150, 2024. doi: 10.1109/TVCG.2024.3372061 8

[32] W. Wirth, T. Hartmann, S. Böcking, P. Vorderer, C. Klimmt, H. Schramm, T. Saari, J. Laarni, N. Ravaja, F. R. Gouveia, et al. A process model of the formation of spatial presence experiences. *Media psychology*, 9(3):493–525, 2007. doi: 10.1080/15213260701283079 2

[33] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998. doi: 10.1162/105474698565686 5