2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)

2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3503-2838-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMAR59233.2023.00105

A Systematic Evaluation of Incongruencies and Their Influence on Plausibility in Virtual Reality

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

Franziska Westermeier
HCI and PIIS Group
University of Würzburg
Germany

Carolin Wienrich
PIIS Group
University of Würzburg
Germany

Marc Erich Latoschik
HCI Group
University of Würzburg
Germany

Figure 1: Virtual bowling environment where different incongruencies were integrated for a systematic evaluation.

ABSTRACT

Currently, there is an ongoing debate about the influencing factors of one's extended reality (XR) experience. Plausibility, congruence, and their role have recently gained more and more attention. One of the latest models to describe XR experiences, the Congruence and Plausibility model (CaP), puts plausibility and congruence right in the center. However, it is unclear what influence they have on the overall XR experience and what influences our perceived plausibility rating. In this paper, we implemented four different incongruencies within a virtual reality scene using breaks in plausibility as an analogy to breaks in presence. These manipulations were either located on the cognitive or perceptual layer of the CaP model. They were also either connected to the task at hand or not. We tested these manipulations in a virtual bowling environment to see which influence they had. Our results show that manipulations connected to the task caused a lower perceived plausibility. Additionally, cognitive manipulations seem to have a larger influence than perceptual manipulations. We were able to cause a break in plausibility with one of our incongruencies. These results show a first direction on how the influence of plausibility in XR can be systematically investigated in the future.

Keywords: plausibility, congruence, XR, evaluation, experience

Index Terms:
Human-centered computing—Empirical studies in HCI; Virtual reality; Human-centered computing—HCI theory, concepts and models; Virtual reality;

1 INTRODUCTION

A discussion about how different qualia in augmented, virtual, and mixed reality (AR, VR, and MR, short XR for extended reality) are influenced and how they influence each other is ongoing [8, 12, 16]. A quale (singular for qualia) is "defined as a subjective and internal feeling elicited by sense perceptions [12, p. 3]. Plausibility and congruence have gained more and more attention during these discussions and are now seen as important factors when it comes to the assessment of XR experiences. One definition of plausibility is from Slater et al. where it is the factor concerned with the sense-making or credibility of the scenario compared to expectations within the experience. This is connected to how well the scenario can produce events related to the participant that (s)he did not cause [15]. Congruence on the other hand is defined as *"the objective match between processed and expected information on the sensory, perceptual, and cognitive layers"* [8]. Previous approaches to catching the essentials of an XR experience are being revised [16]. More recently the role of plausibility was evaluated with a new model, called the Congruence and Plausibility model (Cap), by Latoschik and Wienrich [8] in mind [1, 19]. This model assumes a three-layer manipulation space with a cognition, a perception, and a sensation layer. The sensory input on each layer provides congruence which contributes to the plausibility of the scenario.

The effects of manipulated plausibility have been researched for quite some time [5, 13]. The strongest manipulation of plausibility would be a complete break of it. With breaks in plausibility, we address users' perception of discrepancies in the experience in analogy to breaks in presence. Previous presence research defined breaks in presence as countable events when the attention is shifted from the virtual to the real environment [18]. A break in plausibility would mean the congruence of the scenario is not in line with the user's expectations. This break could be a one-time event or a persistent state as a lot of different aspects can trigger it. Finding ways to systematically cause breaks in plausibility would provide central

---

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

DOI 10.1109/ISMAR59233.2023.00105

ways of manipulating plausibility. These methods could then be used to further investigate the overall formation and perception of plausibility.

For finding potential congruence manipulations that cause breaks in plausibility, we first analyze the concept theoretically. Based on this theoretical knowledge, we derive potential approaches that should break plausibility. In particular, we implemented four different congruence manipulations. Two of which were on the cognitive level of the CaP model and two of which were on the perceptual layer. Furthermore, two manipulations were in the context of the task, while two were not. We conducted a user study to evaluate the effect of our manipulations on the perceived plausibility as well as other qualia, i.e. presence and spatial presence. Results show that we were able to cause a break in plausibility. We did also find an effect on presence, however, there are mixed results for spatial presence.

2 RELATED WORK

Plausibility was first incorporated into the presence model by Slater [15] in 2009 as the plausibility illusion (Psi). In their latest work, Slater et al. [16] extend the influence of Psi and the factors on which Psi relies. They name three components that contribute to the emergence of Psi: (1) how well the environment reacts to users' actions, (2) the environment including events that relate to the participant, and (3) the XR environment should meet real-life expectations if its events could happen in reality. The Slater model was later used as a basis by Skarbez et al. [12] in an attempt to unify different presence models (see figure 2). They argue that there has to be a similar concept to immersion that influences Psi, which they specify as coherence. They define coherence as the level to which a "virtual scenario behaves in a reasonable and predictable way" [14]. Skarbez [14] also argues that the perceived plausibility depends on users' expectations of a virtual environment.

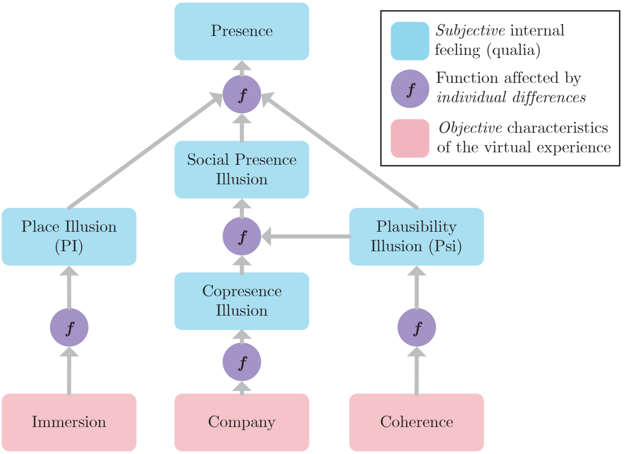

Figure 2: Model of relationships between presence concepts proposed by Skarbez et al. [11] (layout redesigned by the authors).

More recently, Hofer et al. [5] included the idea of a sensorimotor and cognitive distinction. They propose that presence emerges from the psychological dual-system for information processing. The plausibility judgment can be seen as a higher-order process that could be interrupted by secondary tasks, whereas spatial presence would be a fast and automatic process. They argue that these concepts influence each other, whereas Slater [15] sees these concepts as orthogonal. Furthermore, Hofer et al. categorize plausibility into two sub-components: external and internal plausibility. External plausibility refers to the degree to which the shown environment and events match our real-world knowledge. Internal plausibility is how much the environment and events make sense within the given scenario. This is in line with the view of Skarbez [14] that the prior expectations of participants play a role in the judgment of plausibility. Prior knowledge, in this case, could either be knowledge about the real world or knowledge about the setting of the virtual environment.

Latoschik and Wienrich proposed a new model called the *Congruence and Plausibility (CaP) model* that puts plausibility and congruence into focus as the main influencing factors for the XR experience [8] (see figure 3). Congruence replaces the term coherence. It is similar to the coherence definition by Skarbez but more restrictive. Congruence here arises from the three levels of the manipulation space: sensory, perceptual, and cognitive. These congruencies or incongruencies contribute to plausibility through a weighted activation function. The cognitive layer processes higher-order cues and is a top-down process. The perceptual layer processes proximal cues and the sensation layer habitual sensory cues. Both layers are bottom-up processes. They define plausibility as the result of congruencies of these different processing layers in the manipulation space, in contrast to Slater [15], who defines Psi as a pure cognitive construct. Additionally, they do not see the perceived plausibility as an illusion. It is merely a result of information processing on the three layers, and it leads to a subjective state or condition that the user actually feels, regardless if this feeling is triggered by real-world or artificial stimuli.

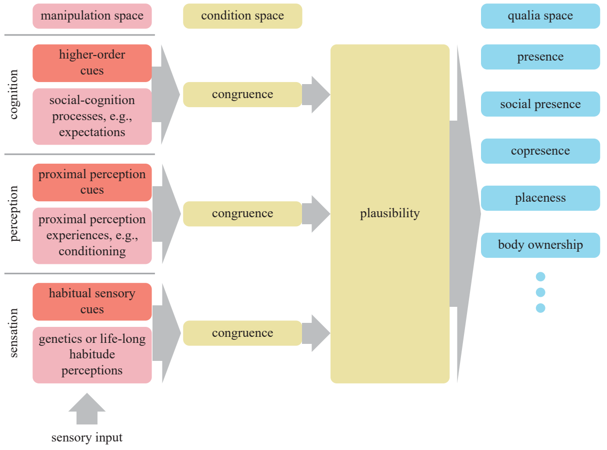

Figure 3: CaP model proposed by Latoschik and Wienrich [8] (layout redesigned by the authors).

As we have seen, plausibility plays a central role when talking about XR experiences. Consequently, it has been studied many times in the past. Some papers focused on what aspects of an XR environment affect the sense of plausibility.

In 2010, Slater et al. [17] tested different aspects of immersion (illumination, display size, navigation, and avatar). They were interested in which of these aspects were important for the Place Illusion (PI) and Plausibility Illusion (Psi). They created an environment where different levels for these aspects could be set by the participants. Participants were first presented with an environment where all aspects were set to their maximum. They were then placed in the same environment, however, this time the aspects were all set to their minimum. Participants were asked to adjust the aspects (one at a time) until they felt like their feeling of PI or Psi matched the feeling of the ultimate environment. Results show that participants choose different configurations depending on which factor, PI or Psi, they were told to focus on. For Psi, the most important factor was illumination. However, the avatar was important in both groups.

Skarbez et al. [13] also examined how different coherence characteristics are prioritized by the participants using a similar approach to Slater et al. [17]. Participants were first shown the ultimately coherent environment and were afterward placed in a downgraded version of the scenario. From there, they were able to upgrade different coherence factors until they felt like they experienced the same level of Psi as in the ultimately plausible environment. Four coherence factors were defined: virtual human behavior coherence, virtual body coherence, physical interaction coherence, and scenario coherence. The results show that the virtual body coherence was upgraded most often and should therefore be the most important factor influencing Psi. Scenario coherence seems to be the second most important factor. For the other characteristics, participants usually upgraded them by one level, however, not necessarily two levels.

Other work focused on breaks in plausibility as an analogy to breaks in presence. They were trying to find out what influence breaks in plausibility have on the perception of the XR environment.

Hofer et al. [5] researched which effect plausibility violations have on the perception of plausibility and spatial presence. They looked at two levels of immersion (display and HMD), two levels of cognitive load (none and high by remembering a nine-digit number), and two levels of plausibility (plausible and implausible). Participants had to walk through a virtual building while the (external) plausibility was manipulated by the objects' physical behavior and the environment's appearance. Objects, for example, would be upside down, rotated continuously, or changed their size. They utilized questions such as *The rooms I walked through are very similar to rooms in reals life* as a manipulation-check whether the perceived plausibility was influenced. They did find a significant effect in these questions, showing their plausibility manipulation was successful. However, they did not find significant effects from the plausibility manipulation on spatial presence.

Similarly, Brübach et al. [1] also manipulated objects within the environment. Two experiments were conducted by manipulating objects in a VR environment so that they would either adhere to gravity or not as the breaks in plausibility. Additionally, in the second experiment, the environment and a priori framing were either set to be on a container ship or on a spaceship. Therefore, the object's behavior would either match the environment or not. First, Participants had a short orientation phase where they could get familiar with the VR environment and their interaction possibilities. Afterward, an accident happened, and the objects within the environment needed to be put back into a dedicated space. This ensured that the participants interacted with the objects and notice their behavior and whether it matched the narrative or not. The authors define the object manipulation to be on the perceptual layer and the framing manipulation to be on the cognitive layer. However, we would argue that the object manipulation should be allocated to the cognitive layer as well since we expect a perceptual manipulation to affect the motion- or stereo-parallax. The layers in the CaP model [8] most likely have gradations, which means that each layer has higher or lower levels. Therefore, the object manipulation should be allocated to the cognitive layer as well, however, on a lower level than the framing manipulation. While the authors were able to detect breaks in plausibility caused by the lower-level manipulation, they could not find an effect on presence or spatial presence. The higher-level manipulation could not reduce the perceived plausibility, nor could it cancel the effect of the lower-level manipulation.

2.1 Summary and Contribution

As we have discussed, neither the concept of plausibility nor the question of how it is influenced is new. There has been a lot of research and discussion regarding this topic and the discussion is still ongoing. Breaks in plausibility, as an analogy to breaks in presence, seem to be a useful approach to research plausibility. However, a systematic investigation of different incongruencies, their ability to cause breaks in plausibility, and how these breaks influence other qualia is missing. It is essential that we understand the different ways we can manipulate congruence and cause breaks in plausibility. Only then it is possible to use these breaks in plausibility to investigate the directions of action and relationships between different XR qualia. To bridge this research gap, we propose a study investigating different congruence manipulations.

We implemented five different VR scenes, one for each manipulation and a control scene. We measured the perceived plausibility using a variation of the questions proposed by Brübach et al. [1]. Breaks in presence were previously defined as an attention shift from the virtual to the real environment. This attention shift is defined as a one-time event that participants can recover from [18]. Our incongruencies, however, are continuous during the experiment. Every time the participants interact with the environment or look around they can notice the discrepancy and a break in plausibility can occur. We, therefore, define that a break in plausibility occurred when there is a significant effect in the perceived plausibility questionnaire between a manipulation condition and the control condition. Additionally, we measured presence and spatial presence to see if the incongruencies had an effect on them.

The results of our study give new insights into how plausibility can be manipulated to systematically research its role in XR experiences.

3 PRESENT STUDY

3.1 Hypotheses

Manipulations that are not connected to the task merely cause changes in the environment. They might be more subtle because they are not the focus of attention. We expect that the effect that these manipulations have on the perceived plausibility is lower than the effect of the manipulations connected to the task. Therefore, our first hypothesis is as follows:

- **H1** Incongruencies in the context of the task have a stronger effect on the perceived plausibility than those not in the context of the task.

We want to induce incongruencies on both the perceptual and the cognitive layers. Based on the findings in Brübach et al. [1], we expect that the perceptual layer has a stronger influence on the perceived plausibility. We, therefore, assume that:

- **H2** Incongruencies on the perceptual layer cause a stronger effect on the perceived plausibility than incongruencies on the cognitive layer.

In order for these incongruencies to be used later to study the direction of action of different models on XR qualia, it is necessary that they also have an effect on other qualia in XR. Based on the direction of effects in the model by Latoschik and Wienrich [8] we assume that a break in the perceived plausibility has an effect on both presence and spatial presence. Our last two hypotheses are as follows:

- **H3** Incongruencies on the different layers have an effect on presence.

- **H4** Incongruencies on the different layers have an effect on spatial presence.

3.2 Study Design

We used a 1x5 within-subject design with four different congruence manipulations and one control condition. With the manipulations, we wanted to cover different levels of the new CaP model. As

	Connected to task	**Not connected to task**
Cognitive layer	Familiar Size	Object Placement
Perceptual layer	Audio	Light

Table 1: Assignment of the manipulations to the layers of the model by Latoschik and Wienrich [8] and their connection to the task.

the sensory layer can only be influenced by changing the sensory procession in the body we omit it from our experiment. Instead, we focused on the cognitive and perceptual layers, which are easier to manipulate.

Additionally, we wanted to see if the context of the manipulation made a difference. So we decided to use both, manipulations that were directly connected to the task in VR and ones that had nothing to do with it.

We decided on the four manipulations *familiar size*, *object placement*, *audio*, and *light*. Familiar size and object placement are located on the cognitive layer of the CaP model [8], while audio and light are located on the perceptual layer. The familiar size and the audio were directly connected to the task, whereas the object placement and the light were not. This assignment of the manipulations can also be seen in figure 1. The different manipulations are described in the next section.

3.3 Congruence Manipulations

3.3.1 Familiar Size

Familiar size is an important concept related to our depth perception. We are able to guess how far away an object is by comparing its size to the size of the object we are used to. If the object is smaller it is further away and if it is bigger it is closer to us. Familiar size is therefore based on past experience in contrast to other perceptual depth cues like overlapping [4]. Based on this need for prior knowledge we allocate this manipulation on the cognitive layer. For this manipulation, an object should be visualized at different distances in the scene, preferably with a movement so that participants can directly watch the change in size. We decided it would be best if the participants caused the movement of the object to enhance the focus on it. So the familiar size manipulation should be in the context of the task.

3.3.2 Object Placement

With the object placement manipulation, we want to make use of a phenomenon called ”change blindness”. It is difficult for humans to detect changes in visual stimuli even if they happen right in front of them. Our attention is required to recognize such changes [9]. Therefore, we allocate this manipulation on the cognitive layer as well. Disappearing objects in the context of the task might be frustrating for participants, leading to an unwanted negative effect. So it was decided that this manipulation should not be in the context of the task but rather affect the surrounding environment.

3.3.3 Audio

We use sound to locate objects and increase our spatial awareness. The ability to perceive and interpret sounds is an essential aspect of our overall perceptual experience. We locate the audio manipulation on the perceptual layer. Previous research has shown that we are sensible to incongruencies in sounds, especially when the mismatch is between the auditory and visual stimuli. [6]. To make sure participants pay attention to the sound we decided to manipulate the sound in the context of the task.

3.3.4 Light

Lastly, we allocated light manipulation on the perceptual layer as well. It is also part of our overall perceptual experience. It is quite difficult to find a task where light is an essential part. Usually, it is just a part of the environment with no active part. That is why we also decided that the light manipulation should not be in the context of the task.

4 METHODS

4.1 Application

There are some requirements for the environment and the task to be able to implement all manipulations. It needed to contain objects that would move away from the participant and make a sound after the interaction. It also required a surrounding environment that contained easily manipulable lights and objects. A bowling environment gave us the freedom to seamlessly integrate the different manipulations, as it contained active and passive parts. Participants were tasked with playing bowling in VR multiple times. Each time another aspect of the environment was manipulated. We tried to keep aspects like the sizing of the objects, length of the lane, and lighting as realistic as possible. The scene consists of a single lane, a ball return, windows with a neutral view, and some background decorations like plants and benches. Colorful spotlights were placed next to the lane on the right and illuminated the lane. The environment can be seen in figure 4.

(a) Bowling lane

(b) Back of the room

Figure 4: Virtual bowling environment used as a basis for the experiment.

Participants could play bowling in VR as they would in real life. They could pick up the bowling ball and throw it toward the pins at the end of the lane. Safety guards at each side of the lane ensured that participants would always hit pins. The number of pins that were hit is displayed above the end of the lane. After each throw, a barrier appeared at the end of the lane and all pins were set up again. The ball was dispensed at the left side of the lane. However, participants could use whichever hand they preferred to throw the ball.

We implemented a tutorial scene. Here, participants could get familiar with picking up and throwing the ball. This was especially important for participants with little experience with VR or the Valve Index controller. To avoid priming the tutorial scene was reduced to simple white spheres and cylinders. It can be seen in figure 5.

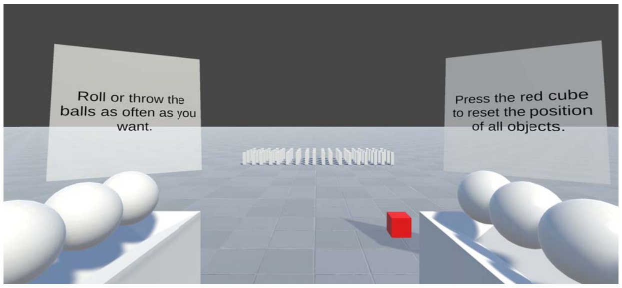

Figure 5: Tutorial Scene

The four different manipulations were implemented into four different scenes. One additional scene served as a control, where nothing was manipulated.

4.1.1 Familiar size

Normally, when objects move further away from us their size decreases. However, in this manipulation, the perceived size of the ball remained the same as it moved down the lane, away from the participant. To achieve this we increased the size of the ball object relative to its distance to the camera. The implementation of this manipulation can be seen in figure 6.

(a) (b)

Figure 6: Manipulation *Familiar Size*

4.1.2 Object Placement

For this manipulation objects (e.g. the benches and plants) changed their location when the participant looked away. As soon as one object was out of the field of view of the participant the object was moved to another location. They moved between a total of three different locations. The implementation of this manipulation can be seen in figure 7.

(a) (b)

Figure 7: Manipulation *Object Placement*

4.1.3 Audio

Similar to the familiar size manipulation audio in real life is louder the closer it is to us. In this manipulation, however, the audio of the rolling ball increased with the distance. The ball was therefore louder when it was at the end of the lane as it was in the beginning. The implementation of this manipulation can be seen in figure 8.

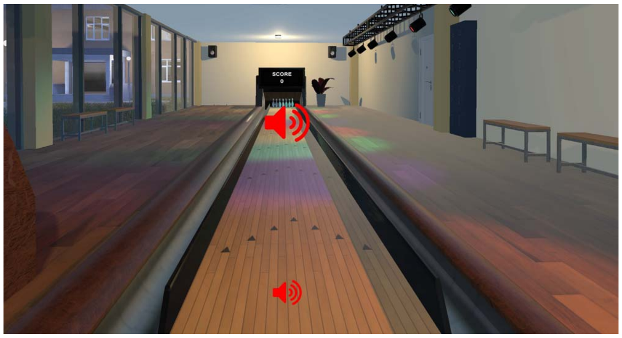

Figure 8: Manipulation *Audio*

4.1.4 Light

Here, the colorful spotlights that illuminate the lane and the ceiling lights were manipulated. Instead of the colorful lights coming from the right they now came from the bottom, seemingly out of nowhere. The ceiling light asset was removed, however, its light spot remained also making it seem like the light comes out of nowhere. The implementation of this manipulation can be seen in figure 9.

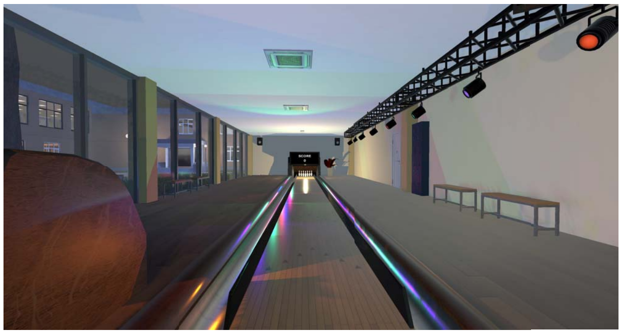

Figure 9: Manipulation *Light*

4.2 Apparatus

The application ran on a high-end computer with an Nvidia Geforce RTX 3080 GPU and an Intel i9-11900K CPU with 64 GB of RAM. The application was developed in the Unity Engine (v2020.3.21f1) using the Steam VR Plugin (v2.7.3). We used the HTC Vive Pro headset in combination with the Valve Index controller to ensure a more natural interaction with the application. Due to their design, the participants are able to open and close their hands completely, which contributed to an easier interaction with the environment.

4.3 Measures

To measure the breaks in plausibility we used the questions from Brübach et al. [1] and modified them to fit our context. This meant that we exchanged the phrase ”object” with ”scenario”. The questions can be seen in table 2. This was done to ensure that participants do not focus on a specific object but rather take the whole scene and what is happening into account. This questionnaire has thirteen items on a 7-point Likert scale ranging from *I do not agree at all*

(1) to *I fully agree (7)*. Additionally, we asked open questions to see whether the participants noticed a manipulation and what they thought was manipulated. Additionally, we asked them three questions as a manipulation-check. We asked whether they noticed a manipulation. If they answered yes we asked what they think was manipulated and what effect it had on them.

To measure spatial presence we used the spatial presence experience scale (SPES) by Hartmann et al. [3]. It measures spatial presence on the two dimensions perceived possible actions and user self-location. Each subscale consists of four items on a 5-point Likert scale from *I do not agree at all (1)* to *I fully agree (5)*.

The Igroup Presence Questionnaire (IPQ) by Schubert et al. [10] was used to measure presence. It consists of three subscales, spatial presence, involvement, and experienced realism, as well as one item that does not belong to a subscale. The questionnaire has fourteen items on a scale from 0 to 6. The wording of the endpoints varies between the questions.

We used the NASA-TLX by Hart et al. [2] to asses the workload of participants during the experiment. It asks for the mental, physical, and temporal demand as well as performance, effort, and frustrations participants feel during a task. The items are measured on a scale from 0 to 100.

Lastly, we used the Virtual Reality Sickness Questionnaire (VRSQ) by Kim et al. [7] to control for virtual reality sickness. The VRSQ measures sickness caused by virtual reality on the two dimensions oculomotor and disorientation. It has four items for oculomotor and five for disorientation on a scale from *not at all* to *strong* to describe the symptoms.

In the end, participants had the option to freely comment on their experience. We asked them if anything, in particular, caught their eye or stood out.

4.4 Procedure

The procedure of the study can be seen in figure 10. The experiment took approximately 1.5 hours. Participants started by filling out the consent forms and their demographical data. To familiarize themselves with the controls participants started with a tutorial scene. They then filled out the pre-VRSQ. Afterward, the participants were given the instructions for the VR part and started with one of the five scenes in a randomized order. They were told to throw the bowling ball 6 times and try to achieve as many points as possible. After throwing the ball 6 times participants left the VR environment to complete the plausibility questionnaire, the SPES, the open questions, the IPQ, the NASA-TLX, and lastly, the VRSQ. This was repeated five times until the participant had completed each manipulation and the control scene. In the end, participants had time for additional comments in an open interview. They were then told about the experiment's intention.

4.5 Participants

Twenty participants took part in the experiment. They received compensation equivalent to 15 $ in the currency of the country where the experiment was conducted in return. The participant pool was divided into 11 females and 9 males. The mean age amounts to $M = 24.5$ ($SD = 9.5$) years. Seventeen participants were students, one was an employee, and two were job-seeking. Six participants had less than one hour of experience with XR technologies, eight had more than one hour, and six had more than 10 hours. Fifteen participants play video games for less than one hour per day, four between one and three hours, and one participant between 3 and 5 hours.

5 RESULTS

Normal distribution and homoscedasticity were violated for almost all of our data. Therefore, we used the non-parametric Friedman test with a significance level of $p < .05$ and Conover's post hoc tests for all our measures.

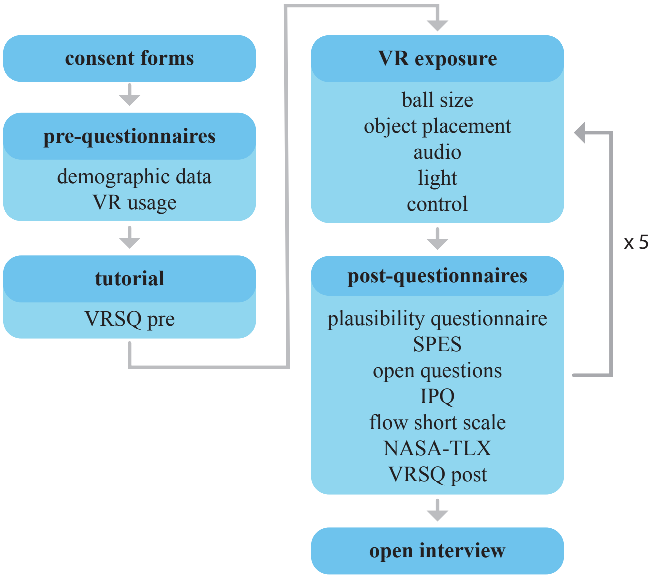

Figure 10: The experiment procedure, consisting of pre- and post-questions as well as five experiment runs and a short open interview.

5.1 Control Variables

The different manipulations had no significant effect on the overall workload measured by the NASA-TLX. Likewise, there was no significant effect on the overall VR sickness measured by the VRSQ. Thus, we could ensure that no simulator sickness or discrepancy in workload did confound our results.

5.2 Plausibility

The different manipulations had a significant effect on the overall plausibility ratings χ^2 (4) = 13.651, $p < .01$. Pairwise comparisons showed a significant effect between the familiar size and the other manipulations, as well as the control scene (all $p < .05$), with means of $M = 3.42$ ($SD = 1.82$) for familiar size, $M = 4.63$ ($SD = 0.37$) for audio, $M = 5.10$ ($SD = 1.03$) for light, $M = 4.70$ ($SD = 1.56$) for object placement, and $M = 5.08$ ($SD = 1.20$) for the control scene.

5.3 Presence

Presence was measured using the IPQ. The overall rating did not show a significant effect between the manipulations with χ^2 (4) = 5.08, p = .28. However, we did find significant effects in all three subscales.

For the *spatial presence* subscale we found a significant effect with χ^2 (4) = 16.05, $p < .01$. Pairwise comparison showed effects between the object placement and the other manipulations as well as the control scene (all $p < .05$), with means of $M = 3.0$ ($SD = 0.53$) for object placement, $M = 3.41$ ($SD = 0.84$) for audio, $M = 3.33$ ($SD = 0.37$) for light, $M = 3.39$ ($SD = 0.95$) for familiar size, and $M = 3.34$ ($SD = 0.14$) for the control scene.

For the *involvement* subscale we found a significant effect with χ^2 (4) = 10.61, p = .03. Pairwise comparison showed effects between the control scene and both familiar size and object placement (all $p < .05$), as well as light and both familiar size and object placement (all $p < .05$), with means of $M = 3.36$ ($SD = 0.74$) for object placement, $M = 3.11$ ($SD = 0.83$) for audio, $M = 3.23$ ($SD = 0.66$) for light, $M = 2.94$ ($SD = 0.98$) for familiar size, and $M = 3.28$ ($SD = 0.65$) for the control scene.

	Control		Familiar Size		Object Placement		Audio		Light		test statistic	p-value	effect size
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Friedman	$p \leq$	
I am used to a scenario behaving this way.	**5.1†**	**1.59**	3.0	1.95	**4.45†**	**2.04**	**4.85†**	**1.87**	**5.0†**	**1.26**	χ^2 (4) = 14.28	0.01	0.188
In everyday life, I expect the scenario to behave this way.	**5.0†**	**1.45**	3.1	2.08	4.3	1.95	**4.6†**	**1.9**	**4.75†**	**1.86**	χ^2 (4) = 9.19	0.06	0.12
I have seen the scenario behave this way in real life.	**5.0†**	**1.75**	3.1	2.2	**4.65†**	**2.11**	**4.55†**	**2.01**	**5.2†**	**1.73**	χ^2 (4) = 17.13	0.01	0.21
The behavior of the scenario is unusual for me. [1]	**4,95†**	**1.61**	3.5	2.12	**4.8†**	**1.7**	**4.85†**	**1.95**	**5.2†**	**1.36**	χ^2 (4) = 14.59	0.01	0.18
I do not know the behavior of the scenario from real life. [1]	**5.15†**	**1.87**	3.5	2.42	**5.0†**	**2.05**	**4.65†**	**2.18**	**5.5†**	**1.36**	χ^2 (4) = 9.76	0.05	0.12
I had a prior expectation of how the scenario would behave.	**4.85†**	**1.27**	3.35	1.98	3.9	1.71	**4.5†**	**1.82**	4.0	0.42	χ^2 (4) = 10.75	0.03	0.13
I expected the behavior of the scenario.	**4.9†**	**1.68**	3.25	1.77	4.3	1.69	**4.3†**	**2.03**	**4.95†**	**1.32**	χ^2 (4) = 15.26	0.01	0.19
I have seen this scenario behavior in movies, games etc. before.	**5.45†**	**1.54**	3.55	2.06	**4.75†**	**2.05**	4.3	2.08	**5.25†**	**2.05**	χ^2 (4) = 13.19	0.01	0.17
I was surprised by the behavior of the scenario. [1]	**4.65†**	**1.5**	3.2	1.96	**4.65†**	**1.69**	**4.55†**	**1.76**	**4.65†**	**1.69**	χ^2 (4) = 10.97	0.03	0.14
I had no idea that the scenario will behave this way. [1]	**4.75†**	**1.89**	3.65	2.08	**4.8†**	**2.02**	**4.6†**	**1.9**	**4.7†**	**1.87**	χ^2 (4) = 10.3	0.04	0.13
The behavior of cause and effect matched the scenario. [1]	5.35	1.42	4.25	2.05	5.15	1.79	**4.7‡**	**1.69**	**5.7†‡**	**1.26**	χ^2 (4) = 9.03	0.06	0.11
The behavior of the scenario made sense.	5.35	1.35	4.25	2.05	5.05	1.76	4.85	1.84	**5.6†**	**1.27**	χ^2 (4) = 5.81	0.21	0.07
I think this behavior of the scenario is impossible. [1]	**5.55†**	**1.47**	3.45	2.24	5.3	1.87	4.85	2.3	**5.75†**	**1.45**	χ^2 (4) = 14.83	0.01	0.19
Overall [1]	**5.08†**	**1.2**	3.42	1.83	**4.7†**	**1.56**	**4.63†**	**1.66**	**5.1†**	**1.03**	χ^2 (4) = 13.65	0.01	0.17

[1] Question is inverted.

†This symbol marks a significant effect between this manipulation and the familiar size manipulation.

‡This symbol marks a significant effect between two manipulations.

Table 2: Main results of the perceived plausibility questions adapted from Brübach et al. [1].

For the *experienced realism* subscale we found a significant effect with χ^2 (4) = 19.26, p < .001. Pairwise comparison showed effects between the object placement and the other manipulations as well as the control scene (all p < .05), with means of $M = 3.09$ ($SD = 0.40$) for object placement, $M = 2.55$ ($SD = 0.65$) for audio, $M = 2.63$ ($SD = 0.72$) for light, $M = 2.55$ ($SD = 0.72$) for familiar size, and $M = 2.84$ ($SD = 0.69$) for the control scene.

5.4 Spatial Presence

The overall spatial presence rating measured with the SPES was not affected by the different manipulations χ^2 (4) = 5.0, p = .29. Neither of the two subscales *self location* and *possible action* showed any significant effect with χ^2 (4) = 3.79, p = .44 and χ^2 (4) = 4.11, p = .39 respectively.

5.5 Open Questions

Twelve participants detected the familiar size manipulation correctly. The audio and light manipulation was identified by seven participants. However, the object placement manipulation was only correctly identified by two out of the twenty participants.

6 DISCUSSION

Hypothesis H1 *Incongruencies in the context of the task have a stronger effect on the perceived plausibility than those not in the context of the task.* can be partially accepted. The familiar size manipulation did cause a significantly greater break in plausibility compared to the other manipulations and the control scene. We argue that because this was the only manipulation where participants actively interacted with the manipulated object it was noticed the most. Twelve out of twenty participants identified this manipulation correctly. Participants watched the ball as it rolled down the lane to see how many points they scored. This is in line with Skarbez et al. [13] that the interaction with an object leads to a higher priority when it comes to its congruence and plausibility. Additionally, this was the only object that had active behavior within the scene. All other objects were static. As the used plausibility questions all contained the word *behavior* the wording might have influenced the participants. In the future, the questions should be re-worded to be more neutral. The familiar size is also the only incongruence that caused a break in plausibility, as we defined them as a significant difference in the perceived plausibility compared to the control scene.

However, we have to reject hypothesis H2 *Incongruencies on the perceptual layer cause a stronger effect on the perceived plausibility than incongruencies on the cognitive layer*. We located the familiar size manipulation on the cognitive layer and it caused a lower rating in perceived plausibility than the other manipulations. The two manipulations on the perceptual layer (audio and light) were not significantly stronger than the cognitive manipulations (familiar size and object placement) or the control scene. This could also be because of the higher interaction with the familiar size manipulation. However, these results should be viewed with caution. The familiar size manipulation was the one that was recognized the most. While the other manipulations were not consistently recognized. This suggests that the manipulations were of different strengths. In the future, it is crucial to find manipulations with more similar strength or use the same manipulations with different intensities to test this hypothesis again. Especially, as it contradicts previous research.

Results show that incongruencies in the context of the participant's task cause a stronger effect on the perceived plausibility. Additionally, incongruencies on the cognitive layer cause a stronger effect on the perceived plausibility than the ones on the perceptual layer. However, only one manipulation, the familiar size, was able to cause a break in plausibility compared to the control scene.

We can mostly accept hypothesis H3 *Incongruencies on different layers have an effect on presence*. We did not find a significant effect overall. However, we found significant effects on all three subscales. Here, the object placement manipulation showed multiple significant effects with other manipulations. However, this manipulation was noticed the least by participants when asked afterward. They often thought other things were manipulated and were actively searching for manipulations. One participant said *I was distracted and paid more attention to my surroundings.* This might be due to the fact that participants knew something was manipulated, however, the object placement manipulation was too subtle. This higher focus on the environment could explain the higher presence scores. As already mentioned, these effects are useful when investigating XR models and their direction of causality. Our results show that some incongruencies have an influence presence. Thus, they may be suitable to study relationships between plausibility and possibly other qualia. Again a larger study with more similar manipulations in terms of their strength could strengthen these results.

We can neither accept nor reject the last hypothesis H4 *Incongruencies on different layers have an effect on spatial presence*. We found a significant effect in the subscale *spatial presence* of the IPQ between the object placement and the control scene. Similar to the presence rating we think that the expectations of the participants to find a manipulation is the reason behind this increase. Participants looked around more and therefore might have felt a higher feeling of spatial presence. However, we could not find significant effects for the SPES. The contradicting results between the two questionnaires leave this question open for future research. AS the SPES subscale *self location* and the IPQ subscale *spatial presence* have a similar wording we would have expected similar results. A larger sample size could provide some clarity in this matter. At this point, it is not possible to make a clear statement regarding this hypothesis.

7 LIMITATIONS AND FUTURE WORK

As remarked by Brübach et al. [1] the questions used to measure the perceived plausibility were not validated before. We changed the wording to make them more fitting for our context. As discussed before it is possible that this specific wording of the questions could have influenced the participant's response and focus. Especially in a within-subject design where participants later know the questions and might pay attention to details in later conditions differently. We tried to reduce this effect with randomization. The within-subject design was chosen to reduce the number of participants. The sample size of N = 20 was still quite small. The small number of participants weakens the significance of the results and the generalizability. Therefore, the results should be confirmed with a larger study in the future. Additionally, a between-subject design could help with possible sequence effects. Another limiting factor is the composition of the sample. The subjects are mostly young adults with a majority of VR experience. A diverse sample would be desirable for future work.

As we have seen from the open question in the end the different manipulations were detected with varying frequency and only one manipulation was able to cause a break in plausibility. This is an indication that the manipulations had different strengths. It is possible that the other manipulations weren't obvious enough compared to the control scene. Future work could either use different manipulations that are equally strong or change the implementation of the current manipulations to match their strength better. Resulting from this, it is hard to tell whether the weaker manipulations had a subconscious effect, that was just too small compared to the control scene or if they had no effect on the perceived plausibility at all. Future work should focus on ensuring that all manipulations are recognized by participants.

8 CONCLUSION

Plausibility has gained more attention in the context of XR experiences and their evaluation. However, systematic studies on this topic have not been carried out enough so far. As we have seen from our results we are just at the beginning of understanding plausibility and its influence. It is, therefore, necessary to find ways to systematically influence plausibility in order to be able to specifically investigate its influence on other qualia. Breaks in plausibility as an analogy to breaks in presence caused by incongruencies on different layers seem to be a good tool for this. Our experiment is a first step in this direction. We looked at four different incongruencies. The results show that incongruencies in the context of the task have a greater effect on the perceived plausibility. Additionally, incongruencies on the cognitive layer seem to have a greater effect than incongruencies on the perceptual layer. However, we were only able to cause a break in plausibility in one condition. The presented study and its results can be the basis for a systematic approach to studying incongruencies and their ability to cause breaks in plausibility. After understanding how we can manipulate plausibility we can begin to research its implications on other XR qualia.

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). We thank our students David Egelhofer, Marius Röhm, Lina Seyfried, Fredrik Stamm, and Niclas Wagner for their contribution to this project, without whom this project would not have been possible. And a special thank you to Andrea Bartl and Sebastian Oberdörfer for their support.

REFERENCES

[1] 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

[2] 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.

[3] T. Hartmann, W. Wirth, H. Schramm, C. Klimmt, P. Vorderer, A. Gysbers, S. Böcking, N. Ravaja, J. Laarni, T. Saari, et al. The spatial presence experience scale (spes). *Journal of Media Psychology*, 2015.

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

[5] 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.

[6] H. Kim and I.-K. Lee. Studying the effects of congruence of auditory and visual stimuli on virtual reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2080–2090, 2022. doi: 10.1109/TVCG.2022.3150514

[7] 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.

[8] 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

[9] D. Martin, X. Sun, D. Gutierrez, and B. Masia. A study of change blindness in immersive environments. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2446–2455, 2023. doi: 10.1109/TVCG.2023.3247102

[10] 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

[11] 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

[12] 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

[13] 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

[14] R. T. Skarbez. *Plausibility illusion in virtual environments*. PhD thesis, The University of North Carolina at Chapel Hill, 2016.

[15] 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

[16] 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

[17] 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

[18] M. Slater and A. Steed. A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9(5):413–434, Oct. 2000. doi: 10.1162/105474600566925

[19] 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