Breaking Plausibility Without Breaking Presence - Evidence For The Multi-Layer Nature Of Plausibility

Larissa Brübach *, Franziska Westermeier *, Carolin Wienrich *Member, IEEE*, and Marc Erich Latoschik *Member, IEEE*

(a) normal object behavior in the container ship   (b) manipulated object behavior in the spaceship

Fig. 1: Environment of the second experiment showing the participant's interaction with the circuit breakers in order to fix the ship after the crash.

Abstract— A novel theoretical model recently introduced coherence and plausibility as the essential conditions of XR experiences, challenging contemporary presence-oriented concepts. This article reports on two experiments validating this model, which assumes coherence activation on three layers (cognition, perception, and sensation) as the potential sources leading to a condition of plausibility and from there to other XR qualia such as presence or body ownership. The experiments introduce and utilize breaks in plausibility (in analogy to breaks in presence): We induce incoherence on the perceptual and the cognitive layer simultaneously by a simulation of object behaviors that do not conform to the laws of physics, i.e., gravity. We show that this manipulation breaks plausibility and hence confirm that it results in the desired effects in the theorized condition space but that the breaks in plausibility did not affect presence. In addition, we show that a cognitive manipulation by a storyline framing is too weak to successfully counteract the strong bottom-up inconsistencies. Both results are in line with the predictions of the recently introduced three-layer model of coherence and plausibility, which incorporates well-known top-down and bottom-up rivalries and its theorized increased independence between plausibility and presence.

Index Terms—plausibility, coherence, presence, XR, experience, evaluation

1 Introduction

"Whoever wants to deceive people must first of all make absurdity plausible." This quote from Goethe [5] shows that the concept of plausibility as a key factor for convincing people of something is not new at all. It also emphasizes the connection of plausibility to something unknown, hard to conceive, or even paradoxical. In contrast to Goethe's time, however, it is not just theater plays and novels trying to convince people of something imaginary or even let them dive mentally into new worlds. Today, the technology of augmented, virtual and mixed reality (AR, VR, and MR, short XR for extended reality) is capable of generating much more holistic experiences, utilizing the stimulation of users' sensorimotor and perceptive information processing with egocentric stimuli mimicking stimuli as experienced from natural sources of the real world around them.

Today, XR technology is used to realize many meaningful and serious applications. Prominent XR use-cases now encompass education and training, health and rehabilitation, collaboration and social VR, entertainment and recreation, and many more. For example, XR applications help people overcome bad habituates [26], recover from surgery [8], learn remotely [15], exchange interculturally [10], or even understand Artificial Intelligence [27]. Overall, the XR use-cases comprise some very different goals as well as partly unique requirements. Still, they all profit from a variety of XR-specific qualities to certain but potentially different degrees. These specific XR qualities of presence, body ownership, co-presence, social presence, and plausibility are building the cornerstones of contemporary XR theories [11, 16, 20, 22, 24].

The theories and models about the very nature of XR experiences are necessary to understand the general modes of action and why XR applications are so effective for specific use-cases. However, most models still theorize around the connection and interrelation of important but also high-level constructs and qualia. These often are hard to operationalize, i.e., broken down into easily manipulable variables and conditions, as required for a user-centered approach to application improvements and appropriate design guidelines. For example, in the past, plausibility was often used synonymously to the term of presence [24]. Presence was seen as the overarching factor crucial for an effective XR experience. Thus, researchers have tried to explain its emergence, measurement, impact, and interrelation to other concepts in plenty of models and studies [22, 28, 29], now leading to an overall picture of

- *Larissa Brübach * is with the Human-Computer Interaction (HCI) group and the Human-Technology Systems (HTS) group at the University of Würzburg. E-mail: larissa.bruebach@uni-wuerzburg.de*
- *Franziska Westermeier * is with the HCI group and the HTS group at the University of Würzburg. E-mail: franziska.westermeier@uni-wuerzburg.de.*
- *Carolin Wienrich is with the HTS group at the University of Würzburg. E-mail: carolin.wienrich@uni-wuerzburg.de*
- *Marc Erich Latoschik is with the HCI group at the University of Würzburg. E-mail: marc.latoschik@uni-wuerzburg.de*
  ** the authors have contributed equally to this work and share first authorship*

Manuscript received 6 Sept. 2021; revised 3 Dec. 2021; accepted 7 Jan. 2022. Date of publication 15 Feb. 2022; date of current version 29 Mar. 2022. Digital Object Identifier no. 10.1109/TVCG.2022.3150496

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

many (sometimes mutually) dependent and correlated factors.

Recent approaches tackle this complexity and try to integrate the existing body of knowledge of XR experiences into models of less complexity but still explain the empirical evidence from past research. Specifically, coherence and plausibility were discussed as important factors of such models. While Skarbez et al. [20] integrate coherence as an objective characteristic and plausibility as a subjective quale into an extended but still presence-oriented model, Latoschik and Wienrich [16] proposed a more radical approach. The latter hypothesizes coherence as the one defining condition for controlling, shaping, and predicting an XR experience and its related constructs and qualia. The model incorporates well-known top-down and bottom-up rivalries by assuming coherence activation on three layers (sensation, perception, cognition) as the potential sources leading to a condition of plausibility and from there to other XR qualia such as presence or body ownership. However, empirical evidence on the role of coherence and plausibility for XR experiences are still rare, in terms of coherence incomplete, or the results are inconsistent [11].

We, therefore, conducted the first empirical evaluation of the newly proposed model [16] in order to understand the connection between coherence, plausibility, and presence better and to evaluate the significance of influencing factors. Two experiments utilize breaks in plausibility (in analogy to breaks in presence) by inducing incoherence on the perception and cognition layers. The behavior of objects was manipulated to not adhere to the laws of physics, thereby inducing incoherence on the perceptional and cognitive layer, while a cognitive manipulation through framing engenders incoherence on the cognitive layer only. The studies contribute to systematic investigations on the concepts of plausibility and coherence within the newly proposed model, potentially showing its advantages over existing presence models.

2 Related Work

One of the first presence models that included plausibility as an important factor is from Slater [24]. He postulates that presence is a construct that results from the two orthogonal factors place illusion (PI) and plausibility illusion (Psi). The PI is described as a perceptual phenomenon that is directly influenced by immersion and subjectively experienced by the user testing the limits of sensorimotor contingencies. Immersion is seen to be the frame within which PI can occur. In this model, the PI is used in analogy to the term *spatial presence*, which is seen to be the feeling of “being there” [17]. As the second factor, Slater [24] defines Psi as the “credibility of events in comparison with what would be expected in reality in similar circumstances” [24, p. 3556]. In this view, presence is defined as “being there plus” – the sense of being in the virtual world *plus* the feeling that the events are plausible within this world [22]. The authors stated that both PI and Psi contribute to the sense of presence. Violating the Psi should therefore also affect presence. This division refers to a long debate about influencing factors leading up to the emergence of presence - (1) bottom-up factors such as the sensorimotor contingencies or object behavior, and (2) top-down factors such as the narrative of a virtual experience.

More recently, Skarbez et al. [21] tried to unify different presence models and terminologies using the model by Slater [24] as the basis. The Skarbez-model (see figure 2), postulates that in analogy to PI that arises from immersion (bottom-up or sensorimotor level), Psi arises from coherence (top-down or cognitive level). They describe coherence as the “set of reasonable circumstances that can be demonstrated by the scenario without introducing unreasonable circumstances, where a reasonable circumstance is a state of affairs in a virtual scenario that is self-evident given prior knowledge” [21, p. 6]. As an additional third factor, they introduce the social presence illusion influenced by the copresence illusion and Psi. Similar to Slater, Skarbez et al. [21] assume that breaks in plausibility diminish the sense of presence.

Hofer et al. [11] interpret the idea of a sensorimotor (bottom-up) and cognitive (top-down) distinction differently by understanding the emergence of presence as a psychological dual-system information processing approach. In their view, the sense of presence results from effectively provided sensorimotor contingencies. Thus, *presence* is

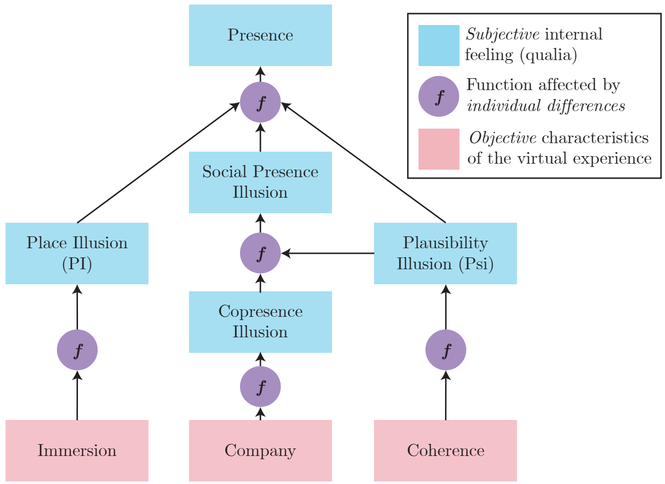


Fig. 2: Model of relationships between presence concepts proposed by Skarbez et al. [21].

defined as the output of a fast and automatic information processing (bottom-up). In terms of the Skarbez-model, this process might refer to the link of immersion and PI. Hofer et al. [11] further conceptualize plausibility as a result of a higher-order cognitive process requiring cognitive resources. This view implies that plausibility impacts the sense of presence on a central, elaborated, or reflective processing route. Looking at the Skarbez-model, this process might refer to the link of coherence and Psi. However, Hofer et al. [11] categorize plausibility further into two sub-components, internal and external plausibility.

In media and communication research, two constructs correspond to this categorization which are referred to as internal and external realism [3, 18].

When reading a narrative, we build a mental model based on the information presented and our real-world knowledge. External realism results from a consistency of these two information sources. When the information from the narrative does not match with our real-world knowledge anymore, the external plausibility is violated.

As long as this violation is either coherent with the mental model constructed through the narrative or explained within the narrative, internal realism or internal plausibility continues. In this context, coherence means that actions, events, and states align within the story. If something does not align, the narrative can give an explanation to maintain internal realism. In this case, the mental model is changed to fit with the explanation [6]. Nonetheless, if the new piece of information is neither coherent nor explained in a way that creates coherence, the internal plausibility is violated. This disruption leads to a loss of engagement and consumes more cognitive resources [3, 18].

In other words, judging the internal plausibility starts a reflective process , which occurs every time a new piece of information needs to be fitted into the established mental model. Hence, mismatches or violations of the internal plausibility should cause an increase in cognitive load.

On the other hand, external plausibility refers to prior real-world knowledge, intuition, and habituate behavior, which usually is processed automatically without any reflective process [13]. So we can assume that the judgment of external plausibility is an automated process with low cognitive demand. This expectation is supported by the results of Hofer et al. [11] which show an effect of the external plausibility manipulation on the subjective plausibility judgment. The interaction with cognitive load manipulation was not significant. The authors conclude that the external plausibility is processed over the automatic route contradicting their assumption that plausibility, in general, is the result of a higher-order cognitive process only. Instead, the results give another hint to involved bottom-up and top-down processes in the emergence of plausibility.

More recently, Latoschik and Wienrich [16] proposed a new model combining the ideas of coherence, bottom-up and top-down processing, and internal and external plausibility. In contrast to the discussed presence models, they assume "plausibility as a state or condition during an XR experience that subjectively results from the — potentially unconscious — processing and evaluation of any information from the sensory, perceptual, and cognitive layers" [16, p. 6]. Hence, plausibility is conceptualized as a quale which is "defined as a subjective and internal feeling elicited by sense perceptions [22, p. 3]. These sense perceptions can be evoked by real world experiences or illusions. However, the nature of the sense perception is irrelevant for the quale itself. Moreover, plausibility and coherence become the central components. Plausibility is conceptualized independently from presence, leading to the assumption that breaks in plausibility must not lead up to breaks in presence.

Latoschik and Wienrich [16] postulate an interplay between two bottom-up layers (i.e., sensation and perception) and one top-down layer (i.e., cognition). The bottom-up layers entail, loosely speaking, our world knowledge which is well learned. Since humans are heavily adapted to it and experience it at the sensorial, perceptional, and cognitive level every day, it is fair to assume that a manipulation in one of the bottom-up layers will impact upwards and, therefore, affect the higher layers. In other words, sensational or perceptual changes in our environment also trigger higher-order processes. Consequently, contradicting this knowledge should have strong effects on the evaluation of plausibility. In contrast, manipulations in the top-down layer should impact this layer only. Violations should have less impact since they contradict within the cognitive level, while sensorial and perceptional cues could still indicate a high degree of plausibility.

Internal and external plausibility as described by Hofer et al. [11] were not introduced in the model of Latoschik and Wienrich [16] since plausibility is seen as a quale and not as an illusion. However, the bottom-up layers might refer to the external plausibility, while the cognitive layer might be linked to the internal plausibility, although both can be processed via the automatic and the reflective route. Again, that would mean that plausibility is manipulable by bottom-up and top-down incoherencies, which contradicts the assumption of Hofer et al. [11] that plausibility results from top-down processes only.

Investigating assumptions about plausibility empirically resembles challenges occurring in studies examining presence, i.e., finding appropriate measurable indicators for the latent construct. Inducing "Breaks in Presence" is one established approach to systematically investigate the occurrence of presence. Similarly, "Breaks in Plausibility" might be an appropriate approach to investigate assumptions about plausibility.

Hofer et al. [11] researched which effect those plausibility violations have on the sense of plausibility and spatial presence with a 2x2x2 between factors design. The factors were 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). Breaks in plausibility originated when the object behavior or the environment were alienated from what participants would expect. A manipulation check was conducted to ensure that the implausible condition had a lower external plausibility using questions like "The rooms I walked through are very similar to rooms in real life", "The environment shows rooms that are like real rooms" or "If I were to go to a real room I would not expect it to be like the rooms I just explored". As a result, Hofer et al. [11] stated that spatial presence and plausibility are two separate, orthogonal concepts. These conclusions are based on their findings that the plausibility factor did not affect the spatial presence rating of the participants even though the check was successful. These results contribute to the idea that plausibility and presence are different qualia, and manipulations of plausibility must not have an effect on presence [16].

Skarbez et al. [23] examined how the participants prioritize different coherence factors. Four coherence factors were defined: virtual human behavior (VH), virtual body (VB), physical interaction (P) and scenario (S). They exposed participants to an ultimate plausible virtual scenario. Subsequently, the participants were placed in a downgraded version of the virtual scenario, e.g., simple geometric models instead of a high-fidelity scenario. The participants' task included the prioritization for level-wise upgrades of the four factors. Participants upgraded their virtual body at first and most often indicating the virtual body coherence as the most crucial factor for Psi. The physical interaction coherence depends highly on the actual interaction with the virtual objects. Participants that did interact with an object prioritized the physical interaction coherence higher. In contrast to Hofer et al. [11], the feeling of presence is also affected by the breaks in plausibility hinting to a more synonymous usage of those constructs as proposed by Slater [24] and Skarbez et al. [22].

Hofer et al. and Skarbez et al. both manipulated the object's physical behavior to create breaks in (external) plausibility. Of course, the design space of breaks in plausibility is vast. Breaking the external plausibility should be induced by a decisive break of our world knowledge as described above. The manipulations of Hofer et al. [11] or Skarbez et al. [23] are good examples since we are experiencing the physical behavior of objects every day. Gravity, for example, is a very strong prior because of its multi-perceptual nature and omnipresence [12]. Contradicting this knowledge should have strong effects on the evaluation of plausibility, as it was shown by Hofer et al. [11] and Skarbez et al. [23]. In contrast, internal plausibility violations should have less impact since they contradict at the cognitive level *only*. Sensorial and perceptional cues could simultaneously indicate plausibility. However, breaking the internal plausibility in virtual reality is less researched yet.

In sum, coherence and plausibility are core concepts for evaluating XR experiences. Different conceptional approaches have conceptional overlaps and discrepancies, particularly regarding the bond strength between plausibility and presence or the level (bottom-up vs. top-down) of influencing factors on plausibility. In comparison to the large corpus of research related to presence, empirical studies investigating the concepts of coherence or plausibility are rare and brought mixed results [11].

3 Approach

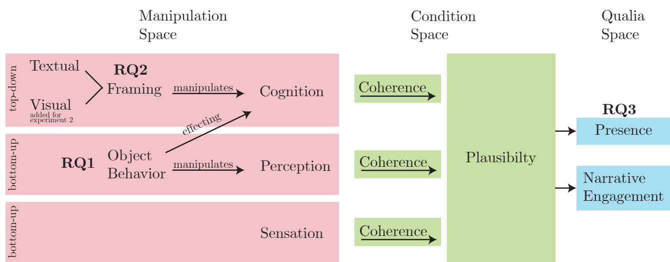


Fig. 3: New theoretical model by Latoschik and Wienrich [16] including the research questions and manipulations of this article.

Latoschik and Wienrich [16] proposed a model focusing on the relation between coherence, plausibility, and XR-related qualia such as presence. They suggested three levels of the manipulation space, allowing for systematic prediction and empirical testing of the resulting coherence and plausibility conditions. The present studies are the first experimental realization of testing the assumption of this new XR experience model.

Similar to presence research [25], we introduce breaks in plausibility on the bottom-up layer (perception) and the top-down layer (cognition). We examine the validity of the model from Latoschik and Wienrich [16] by answering three research questions

- RQ1: Does the manipulation of the bottom-up layer (perception, object behavior) affects the sense of plausibility?
- RQ2: Can a manipulation of the top-down layer (cognition, narrative) counteract the break in plausibility engendered by the perceptional layer?
- RQ3: Do breaks in plausibility have an effect on presence?

RQ1 looks at the impact the manipulation space has on the condition space. We manipulate the perceptual layer and thus also influence the cognitive layer by violating the gravity of objects in the virtual environment. The objects will either adhere to gravity and fall or float around with a random direction and acceleration. In the following, this will be referred to as *object behavior* or the *perceptual manipulation*. We decided to manipulate gravity as it is an important prior for orientation and estimation of space and time [12]. People are very habituated to gravity, and it is relevant for their interactions with the environment. Therefore, it is very likely that people will (automatically) notice the manipulated object behavior. We expect this manipulation to cause a strong break in plausibility.

RQ2 investigates if an additional cognitive manipulation can make up for the effects the manipulation of the perceptual layer has on the cognitive layer. If RQ1 can be seen as successful, e.g., the participants notice the break in plausibility, we further want to investigate if an explanation of the manipulated object behavior through the cognition layer will change the break in plausibility. Textual, and in experiment 2 additional visual, framing is used to achieve this. The participants will be told a story that explains the manipulated object behavior. This will be referred to as *framing* or *cognitive manipulation*. As assumed above, we expect a weaker effect on plausibility compared to the bottom-up manipulation of object behavior. However, we also expect an interaction effect showing higher plausibility values when the framing corresponds to the object behavior.

RQ3 focuses on the influence different breaks in plausibility in the condition space have on the qualia space. Assuming that we were able to create these plausibility breaks that the participants notice , we want to examine their effects on presence. Following the model of Latoschik and Wienrich [16], presence is a quale that is influenced by plausibility to a certain extent, but not necessarily. In contrast, Slater [24] and Skarbez et al. [22] would assume a similar effect on the sense of presence than on the sense of plausibility.

4 First Experiment

4.1 Method

4.1.1 Participants

Forty students (11 male) participated in our study with a mean age of $M = 22.33$ ($sd = 1.79$) ranging from 18 to 26 years. The students had a gaming usage of $M = 1.0125$ ($sd = 1.73$) hours per day. They used a computer $M = 5.56$ ($sd = 2.67$) hours a day on average. They watched TV for $M = 0.74$ ($sd = 0.95$) hours and used streaming services for $M = 1.59$ ($sd = 0.97$) hours per day. Most participants ($N = 13$) rated their fiction movie consumption between two and five hours per month. Half of the participants had VR experience ranging from one to three hours, seven participants had less VR experience, of which three had never experienced VR before. Twenty-one participants have already used VR one to three times before.

4.1.2 Material

The system ran on a high-end, VR-capable PC composed of an Intel Core i7-9700K, an Nvidia Geforce RTX2080 Super, and 16GB RAM. As a high immersive VR display system, we used an HTC Vive Pro HMD [1], providing the user a resolution of 1440×1600 pixels per eye with a total field of view of 110 degrees running on a refresh rate of 90 Hz. The application was developed using Unity 2020.3.5f1 and the Steam VR Plugin for Unity v2.7.3 (SDK 1.14.15).

Variables We measure presence with the Igroup Presence Questionnaire(IPQ, [19]) after each VR exposure. The items of this questionnaire are divided into the three subscales *spatial presence*, *involvement*, and *experienced realism*. The tendency of immersion of the participants in the first place is evaluated by the Immersive Tendencies Questionnaire (ITQ, [29]), including an overall score and the subscales *focus*, *involvement*, and *games*. To avoid confounds caused by simulator sickness, we ask participants to answer the Simulator Sickness Questionnaire (SSQ, [14]) before and after each VR exposure. It contains three subscales: *nausea*, *oculomotor*, and *disorientation*. To get an insight on the mental and physical workload demand, the participants were asked to answer the NASA-TLX [9] after each VR condition. The NASA-TLX provides an overall workload score based on six subscales: *mental demand*, *physical demand*, *temporal demand*, *own performance*, *effort*, and *frustration*. Additionally, we measured the time needed to complete each VR condition.

no	group	question
1	EP	I am used to objects behaving this way.
2	EP	In everyday life, I expect objects to behave this way.
3	EP	I have seen objects behave this way in real life.
4	EP	The behavior of objects is unusual for me. [1]
5	EP	I do not know the behavior of objects from real life. [1]
6	IP	I had a prior expectation of how the objects would behave.
7	IP	I expected the behavior of the objects.
8	IP	I have seen this object behavior in movies, games etc. before.
9	IP	I was surprised by the behavior of the objects. [1]
10	IP	I had no idea that the objects will behave this way. [1]
11	IP	The behavior of the objects matched the scenario. [2]
12	-	The behavior of the objects made sense.
13	-	I think this behavior of the objects is impossible. [1]
14	-	In which scenario did the objects seem more plausible to you? [3]

[1] Question is inverted.
[2] Question 11 is added in second experiment.
[3] Question 14 is only used in the first experiment.

Table 1: Questions regarding the plausibility of object behavior.

Plausibility of Object Behavior Similar to the definition from Hofer et al. [11] of external and internal plausibility, we distinguish between the consistency with the expectations derived from the real-world experience and the consistency with expectations derived from the internal story or what we call framing (on the cognitive level). We came up with questions inspired by the manipulation check from Hofer et al. [11]. Five questions aim to measure the *external plausibility* (EP). As Hofer et al. [11] solely examined external plausibility, two more general questions and five questions targeting the *internal plausibility* (IP) were developed. These aim to check for the expectation that was elicited by the cognitive framing. The 13 questions are rated on a seven-point Likert Scale with the endpoints "not agree at all"(1) and "totally agree"(7). All questions can be seen in table 1. The external and the internal plausibility question groups showed high internal consistency: EP: 0.950; IP: 0.852.

Lastly, the participants were asked to decide which scenario they overall perceived as more plausible.

4.1.3 Design

To answer our research questions, we are using a counterbalanced, randomized 2x2 mixed subject design. The within-factor is the manipulation on the perceptual layer that contains two conditions, a normal object behavior and a manipulated object behavior, where cubes are affected by gravity or floating, respectively. We refer to this factor as the *perceptual manipulation*.

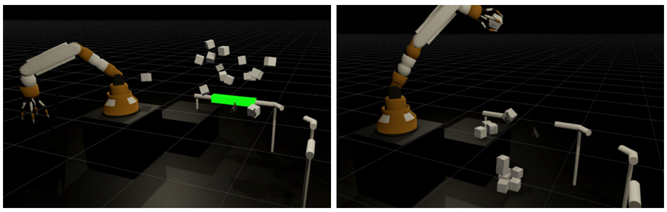
Between the participants, we alternate the textual framing on the cognitive layer. In one condition, participants were given a text that says they are in weightlessness. Therefore, they would possibly expect that no gravity would impact the virtual objects. In the control condition, we omitted this text so that participants would have no specific *new* expectation towards the object behavior but refer to their world knowledge. We refer to this factor as the *cognitive manipulation*.

4.1.4 Procedure

The 40-minutes-experiment (see figure 5) starts with filling out consent forms (including COVID-19 regulations). The participants answer demographic questions and questions about media usage and VR experience. Then the pre-experimental questionnaires assessing presence and simulator sickness are answered. Afterward, the participants read the instructions, and half of them additionally read the cognitive manipulation given by a textual storyline framing.

In the VR part, the participants' task is to grab cubes and attach them to a predefined place (see figure 4). We use small cubes that are given to the participants by a robotic arm. They are told that these cubes are part of a broken circuit and need to be placed in the correct position in order to fix the electricity. Once the robotic arm releases the cube, it either falls down to the floor or floats with a random direction and acceleration. We kept the environment as simple as possible to avoid confounds. Skarbez et al. [23] mentioned that physical coherence highly depends on the degree of the participants' interaction. To ensure that every participant interacts with the objects, we introduced this interaction task.

When all cubes are correctly attached, the first VR part is over, and the participants fill out additional questionnaires, including the IPQ, the set of questions about the plausibility of the object behavior as a manipulation check, the NASA-TLX, and the SSQ. The participants reread the instructions and framing text before starting the second VR part with the other within-condition (perceptual manipulation). The follow-up questionnaires repeat as well. In the last question, we ask participants in which scenario the objects are more plausible.



(a) manipulated object behavior (b) no manipulation in the object behavior

Fig. 4: The experimental environment in which a robotic arm hands the electric cubes to the participant so that they can be put in the right place.

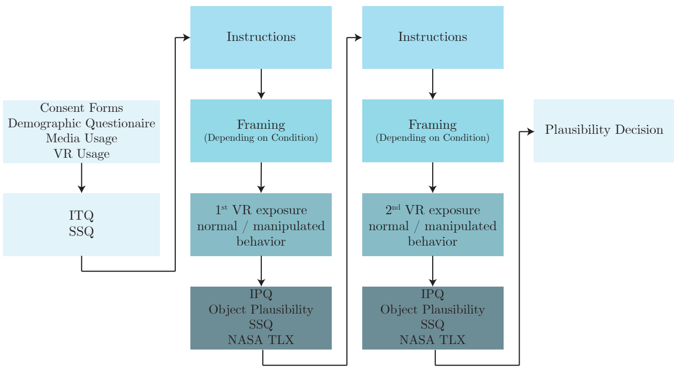


Fig. 5: The experiment procedure consisting of pre- and post-questions as well as two experiment runs.

4.1.5 Hypotheses

Based on the related work, we developed the following hypotheses:

- H1: The manipulation of object behavior on the perceptual layer leads to a break in plausibility.
- H2: The cognitive manipulation neutralizes the break in plausibility induced by the perceptual manipulation.
- H3: Breaks in plausibility lead to break in presence.

no	normal behavior (mean)	manipulated behavior (mean)	F	p $(< .05)$	η_p^2
total	5.45	2.64	203.33	$< .001$	.843
EP	5.4	2.1	138.09	$< .001$	.784
IP	5.2	2.79	101.65	$< .001$	.728
1	5.33	1.92	98.11	$< .001$	.721
2	5.15	1.80	107.55	$< .001$	.739
3	5.38	1.95	126.86	$< .001$	.770
4	5.60	2.08	112.19	$< .001$	.747
5	5.55	2.77	44.11	$< .001$	.537
6	4.30	2.62	22.71	$< .001$	.374
7	5.05	2.17	96.53	$< .001$	.718
8	5.47	4.28	14.96	$< .001$	.282
9	5.75	2.20	153.25	$< .001$	.801
10	5.42	2.70	43.62	$< .001$	.534
12	6.12	3.33	76.64	$< .001$	.669
13	6.30	3.85	44.40	$< .001$	.539

Table 2: Significant results from plausibility questions regarding the perceptual manipulation in the first experiment.

4.2 Results

We evaluated our independent variables using a repeated-measures ANOVA with cognitive manipulation as the between factor and the perceptual manipulation as the repeated measures factor. The alpha-level was set to 0.05 for significant results.

In line with expectations, the questions about the plausibility of object behavior were found significant for all questions and question groups and total within the perceptual manipulation factor (see table 2). Therefore, these questions can be seen as a valid manipulation check on the perceptual layer. The question group aiming for the *external plausibility* with $F(1,38) = 138.09, p < .001, \eta_p^2 = .784$ had the means $M = 5.4$ for the normal object behavior condition and $M = 2.11$ for the manipulated object behavior condition. The question group aiming for the *internal plausibility* had an $F(1,38) = 101.65, p < .001, \eta_p^2 = .728$ and means $M = 5.20$ for normal object behavior and $M = 2.80$ for manipulated object behavior. The total score of plausibility ($F(1,38) = 203.33, p < .001, \eta_p^2 = .843$) had mean values $M = 5.45$ for normal object behavior and $M = 2.64$ for manipulated object behavior. Contrary to expectations, no significance was found for either the cognitive manipulation or the interaction between cognitive and perceptual manipulation. We asked the participants to decide which scenario was more plausible for them, and 39 out of 40 chose the scenario with normal object behavior, regardless of the cognitive manipulation in advance.

The evaluation of presence ratings revealed significances in the IPQ subscales for *experienced realism* and *involvement*. The *experienced realism* became significant with $F(1,38) = 5.28, p = .03, \eta_p^2 = .122$ within the cognitive manipulation factor. Comparing the means for both conditions showed that with $M = 3.95$, the cognitive manipulation reached higher values than the omitted cognitive manipulation ($M = 3.41$). The *involvement* had a significance of $F(1,38) = 5.94, p = .02, \eta_p^2 = .135$ in the perceptual manipulation factor. The *involvement* with manipulated object behavior had a mean value of $M = 4.88$ whereas the mean of the *involvement* with normal object behavior was $M = 4.38$. We could not find any interaction effect between the cognitive and perceptual manipulation. The mean values for the total score were 4.29 (normal behavior × no framing), 4.69 (normal behavior × framing), 4.53 (manipulated behavior × no framing), and 4.70 (manipulated behavior × framing) with a maximum score of 7.

Besides those two main questionnaires, SSQ-data was collected. The evaluation of the SSQ revealed a significant discrepancy in the *nausea* ($F(1,38) = 4.45, p = .04, \eta_p^2 = .105$) and the *oculomotor* ($F(1,38) = 9.44, p < .001, \eta_p^2 = .199$) subscale and total score ($F(1,38) = 6.64, p = .01, \eta_p^2 = .149$) between the normal and the manipulated object behavior. While the means in the normal behavior condition decreased with regard to a pre-measure (*nausea* $M = -4.53$; *oculomotor* $M = -5.31$; Total $M = -2.71$), the means in the manipulated behavior condition increased (*nausea* $M = 2.15$; *oculomotor* $M = 3.22$; Total $M = 2.81$).

We measured the time needed for the task completion. Due to technical problems, we did not collect this data from three participants. In the normal object behavior condition, participants needed approximately $M = 244.50$ seconds. In contrast, participants needed an average of 30 seconds more ($M = 275.39$ seconds) in the manipulated object behavior condition. This time delta is significant (F(1,35) = 62.72, $p < .001, \eta_p^2 = .642$).

With the NASA-TLX, we measured the task-load. All items showed a significance within the object behavior factor. The total score was significant ($F(1,38) = 108.38, p < .001, \eta_p^2 = .740$) with a mean of $M = 3.07$ for normal object behavior and $M = 9.15$ for manipulated object behavior.

4.3 Discussion Experiment 1

Looking at the results, we can accept H1 *"The manipulation of object behavior on the perceptual layer leads to a break in plausibility."*. We successfully caused a break in the external plausibility on the perceptual manipulation layer. With regards to the proposed model from Latoschik and Wienrich [16], we assume that changes of the perceptual layer in the manipulation space do affect the condition space. Previous studies concerning unusual object behavior and the effects on plausibility support this assumption as well [11,23]. Similarly, we would have expected the cognitive manipulation to affect the condition space. However, H2 *"The cognitive manipulation neutralizes the break in plausibility induced by the perceptual manipulation."* must be declined as we did not see a difference in the manipulation check when the cognitive and perceptual manipulation were coherent. Almost all participants (39 out of 40) chose the normal object behavior to be more plausible. Thus, it can be presumed that the participants did not see the two manipulations as coherent. We conclude that the manipulation of the perception, and its influence on the cognitive layer, was much stronger than the cognitive manipulation, which, therefore, could not neutralize the perceptual manipulation. H3 *"Breaks in plausibility lead to break in presence."* can partly be accepted as we found a significance in the *involvement* subscale concerning the perceptual manipulation. The differences in interaction times and task load support the theory that participants in the manipulated object behavior condition had a more intense interaction and therefore, a higher *involvement* . The cognitive manipulation was significant in the *experienced realism* subscale. We assume this might result from more context participants being given in the framing condition, and thus, they are more engaged.

5 Second Experiment

From the results of the first experiment, we drew conclusions for a second experiment. Two approaches are possible here. We can either (1) strengthen the cognitive manipulation, or we can (2) alter the perceptual manipulation by choosing a weaker prior. One concern is that a weaker prior might not lead to a break in plausibility and thus violate H1. Therefore, we decide to rather strengthen the cognitive manipulation in a second experiment.

5.1 Method

5.1.1 Participants

Seventy-six participants (23 male) took part in the second experiment. We excluded five participants due to technical issues or reaching the maximum time, leaving 71 participants (21 male) that could be evaluated further. The remaining participants had a mean age of $M = 28.73$ ($sd = 9.31$) ranging from 19 to 60 years . Fifty-one participants were students. The gaming usage of participants was $M = 0.63$ ($sd = 0.93$) hours per day. They used the computer $M = 5.37$ ($sd = 3.34$), the TV $M = 0.82$ ($sd = 1.18$) and streaming services $M = 1.37$ ($sd = 1.05$) hours per day. Almost half of the participants (N=35) rated their fiction movie consumption as less than two hours per month. Forty-three participants had less than three hours of experience in VR, and 45 participants had used VR less than three times.

5.1.2 Material

The application was set up on a PC with an Intel Core i7-9700K processor with 32GB RAM and an Nvidia Geforce RTX 2070 Super. Again, the HTC Vive Pro HMD [1] was used as the high immersive VR display, with a resolution of 1440×1600 pixels per eye and a total field of view of 110 degrees. The application was built with the Unity version 2020.3.5f1 and the Steam VR Plugin for Unity v2.7.3 (SDK 1.14.15).

Variables We adopted the questionnaires from the first experiment and added a narrative engagement questionnaire [3] to control for the engaging character across the conditions. We used validated translations to the common language [2], which contains twelve items on the subscales *attentional focus*, *narrative understanding*, *emotional engagement*, and *narrative presence*. For our own set of questions concerning the plausibility of object behavior, we added another question that aims for internal plausibility (question 11 in table 1). Additionally, the participants can choose from a list of which scenario components were plausible and implausible.

5.1.3 Design

We use a counterbalanced 2x2 between-subject design. Like in the first experiment, the two factors are the cognitive and the perceptual manipulation. Again, the perceptual manipulation is done through the object behavior, and the cognitive manipulation is the combination of textual and visual framing.

The framing story was lengthened to a logbook text. It is written in first-person perspective and describes feelings and personal observations from an engineer working on a ship. Depending on the condition, the story either mentions a spaceship or a container ship. We included emotional and empathy-evoking language to create a high involvement and identification with the engineer [4, 7].

Second, as it can be seen in the figure 3, we added a visual component to the framing. I.e., the environment was no longer a minimalistic setting but rather a more realistic-looking ship. The outside was set to either show a galaxy or water with waves and a stormy sky. The plain cubes were replaced by real breakers and a circuit box. The new environments can be seen in figure 1.

5.1.4 Procedure

The second experiment takes approximately 30 minutes. Like in the first experiment, we ask for consent (including the COVID-19 regulations) and let participants fill out the demographic questionnaire. We ask about media usage, the ITQ, and SSQ. In contrast to the first experiment, we do not instruct the participants about their specific tasks in advance. The participants read the textual storyline framing and start the VR exposure afterward.

The VR part follows a certain narrative storyline that affiliates with the previously read story. During the VR exposure, the instructions are given to the participants via audio files. Participants first complete a short tutorial that is covered as part of the story. In this tutorial, they get to know the environment and the interaction techniques needed for the task. Afterward, participants are told to watch a monitor. After 50 seconds, a crash happens that leads the story on and explains why the power supply of the ship needs to be repaired. When the participants receive the spaceship story, an additional alert says that the gravity module is down. All participants are told via audio instructions to repair the power supply. Their task is to put the circuit breakers into the circuit boxes to restore the power.

While the power supply is not working, the circuit breakers move depending on the perceptual manipulation condition. They either fall down or float around. After the participants successfully put all breakers back into the right circuit boxes, they are told that the power supply is working again and, if applicable, that the gravity module is up and running again. The VR exposure ends, and another block of questionnaires follows.

The second block of questionnaires includes the IPQ, the plausibility of object behavior manipulation check (see table 1), the narrative engagement, the NASA-TLX, the SSQ, and a list to choose which scenario components were plausible or implausible, as well as a free text field for comments.

5.1.5 Hypotheses

We formulated the following hypotheses for our second experiment as follow-up from our first study:

- H1: The manipulation of object behavior on the perceptual layer leads to a break in plausibility.
- H2.1: A combination of textual and visual framing on the cognitive layer can neutralize the break in plausibility and create an internal plausibility.
- H3.1: Perceived coherence between the cognitive and perceptual manipulation has an effect on presence.

5.2 Results

To evaluate our independent variables, we conducted ANOVA calculations again. We adopted the manipulation check about the plausibility of object behavior from the first experiment. We found significant values for the overall score in the perceptual manipulation space ($F(1,67) = 16.17, p < .001, \eta_p^2 = .194$) with means of $M = 4.69$ for the normal object behavior and $M = 3.79$ for the manipulated object behavior. In the *external plausibility* question group, significance was measured ($F(1,67) = 15.10, p < .001, \eta_p^2 = .184$) in the perceptual manipulation dimension with means $M = 4.12$ for normal behavior and $M = 3.02$ for manipulated object behavior. The *internal plausibility* was significant ($F(1,67) = 13.50, p < .001, \eta_p^2 = .168$) with means of $M = 5.03$ for normal behavior and $M = 4.13$ for manipulated object behavior. A majority of the individual questions was significant (see table 3) in the perceptual manipulation dimension. We furthermore found some slight significances in the interaction of question 11 and 12 (see figure 6). Question 11, *"The behavior of the objects matched the scenario."* ($F(1,67) = 3.58, p = .063, \eta_p^2 = .051$) had mean values in the container ship scenario of $M = 5.82$ for normal object behavior and $M = 5.06$ for manipulated object behavior. Respectively, in the spaceship scenario the normal object behavior had a mean $M = 5.61$ and the manipulated behavior had a mean of $M = 5.94$ (see figure 6 (a)). For question 12, *"The behavior of the objects made sense."* ($F(1,67) = 3.90, p = .053, \eta_p^2 = .055$) we measured mean values in the container ship scenario of $M = 5.12$ for normal object behavior and $M = 4.28$ for manipulated object behavior. Respectively, in the spaceship scenario the normal object behavior had a mean $M = 4.67$ and the manipulated behavior had a mean of $M = 5.22$ (see figure 6 (b)).The medium effect sizes underline the assumption that a larger sample of participants potentially leads to a significance. Again, the *external* and the *internal plausibility* question groups showed high internal consistency: EP: 0.799; IP: 0.743.

We asked participants to mark components of the scenario that they would evaluate as plausible and implausible. From these components, the plausibility of the transition from the textual logbook to the visual components of the scenario was significant within the perceptual manipulation factor ($F(1,67) = 4.57, p = .036, \eta_p^2 = .064$) with means of $M = 0.742$ for the normal object behavior and $M = 0.50$ for the manipulated object behavior. In the cognitive manipulation dimension, participants chose physics as an implausible component significantly more often ($F(1,67) = 4.42, p = .039, \eta_p^2 = .062$) in the container ship framing ($M = 0.513$) than in the spaceship framing ($M = 0.278$).

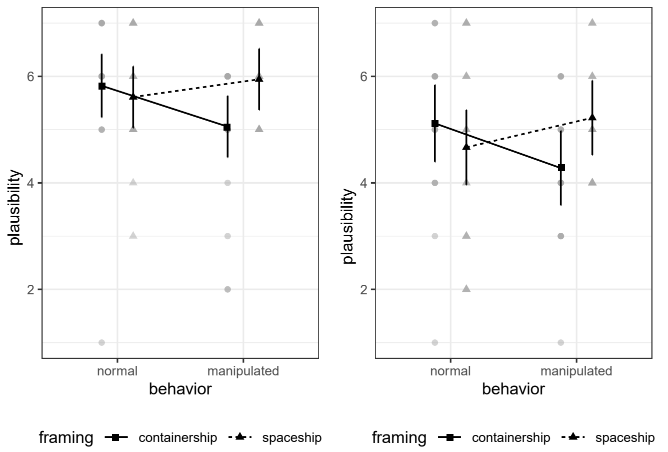
In contrast to experiment 1, the IPQ scores did not show any significance. However, the total scores for the conditions were high (normal behavior × container ship = 4.98, normal behavior × space ship = 5.28, manipulated behavior × container ship = 4.97, manipulated behavior × space ship = 4.88 with a maximum score of 7).

We could not find significant differences in the narrative engagement scores within the conditions. However, the means across all condition combinations were relatively high (normal behavior × container ship = 5.14, normal behavior × space ship = 5.20, manipulated behavior × container ship = 4.96, manipulated behavior × space ship = 4.87 with a maximum score of 7).

Again, the overall NASA-TLX score was significant in the perceptual manipulation dimension ($F(1,67) = 10.70, p = .002, \eta_p^2 = .138$; mean for normal object behavior $M = 5.57$ mean for manipulated object behavior $M = 8.24$) as well as in an interaction between perceptual and cognitive manipulation ($F(1,67) = 5.57, p = .019, \eta_p^2 = .079$). The container ship condition had a mean task load score of $M = 6.70$ for normal object behavior and a mean of $M = 7.40$ for manipulated object behavior while the spaceship condition had a mean $M = 4.44$ for normal and $M = 9.06$ for the manipulated object behavior.

The total SSQ score, which was calculated by the delta between the ratings before and after the VR exposure, resulted in a significant interaction effect ($F(1,67) = 4.33, p = .041, \eta_p^2 = .061$). The container ship condition had a mean SSQ score of $M = -0.88$ for the normal object behavior and a mean of $M = -5.40$ for the manipulated object behavior, while the spaceship condition had a mean $M = -4.36$ for the normal and $M = 5.40$ for the manipulated object behavior.

For the time measurement, we could only evaluate 52 of 71 participants due to technical issues. Similar to experiment 1, participants needed significantly longer ($F(1,48) = 43.84, p < .001, \eta_p^2 = .477$) for the task completion when the object behavior was manipulated



(a) question 11 (b) question 12

Fig. 6: Plausibility questions 11 and 12 indicating a slight interaction effect of the object behavior and framing.

no	normal behavior (mean)	manipulated behavior (mean)	F	p ($< .05$)	η_p^2
total	4.69	3.79	16.17	< .001	.194
EP	4.12	3.02	15.10	< .001	.184
IP	5.03	4.13	13.50	< .001	.168
1	3.94	2.86	7.44	.008	.100
2	3.80	2.78	7.24	.009	.098
3	3.71	2.75	6.22	.015	.085
4	4.92	3.39	14.17	< .001	.175
5	4.24	3.31	4.28	.042	.060
6	4.84	3.56	9.80	.003	.128
7	4.56	3.58	7.30	.009	.098
8	-	-	-	-	-
9	5.04	3.72	13.00	< .001	.163
10	5.19	3.89	14.34	< .001	.176
11	-	-	-	-	-
12	-	-	-	-	-
13	-	-	-	-	-

Table 3: Significant results from plausibility questions regarding the object behavior of experiment 2.

($M = 447$ seconds; normal behavior: $M = 241$ seconds).

The free comment field did not lead to any further interesting findings.

5.3 Discussion Experiment 2

In line with experiment 1, H1 *"The manipulation of object behavior on the perceptual layer leads to a break in plausibility."* can be accepted. The perceptual manipulation can be seen as successful since the overall score of the plausibility questions was significant between the groups. However, the expected effect of the cognitive manipulation remained absent. Thus, H2.1 *"A combination of textual and visual framing on the cognitive layer can neutralize the break in plausibility and create an internal plausibility."* has to be declined as well as H2. The distance between the mean values, however, is no longer as great as in the first experiment. In contrast to the first experiment, the breaks in plausibility did not affect presence. H3.1 *"Perceived coherence between the cognitive and perceptual manipulation has an effect on presence."* can neither be declined nor accepted as the participants did not perceive the coherence between the two manipulations. Thus, the prerequisite for this hypothesis was not satisfied.

Further details and an in-depth comparison between the two experiments will be shown in the overall discussion part.

6 Discussion

Our first hypothesis, H1 *"The manipulation of object behavior on the perceptual layer leads to a break in plausibility."* can be accepted. Participants rated the scenario where objects were adhering to gravity as more plausible regardless of whether they received a cognitive manipulation through framing beforehand or not. As Jörges and López-Moliner [12] pointed out, gravity is a very strong prior that we experience our whole life, and it can hardly be not experienced (except for parabolic flight or space excursions). It is an important factor in our everyday life, e.g., to estimate and anticipate an object's position and behavior, and we automatically notice if something is wrong as it is a bottom-up factor. Therefore, we assume that bottom-up factors are capable of causing breaks in plausibility. Regarding the newly proposed model by Latoschik and Wienrich [16], we can support the assumption that changes of the perceptual layer in the manipulation space affect the condition space, i.e., the coherence and plausibility, which confirms the correct direction of causality in the model. The results are also in line with previous studies investigating unusual object behavior on plausibility [11, 23]. However, they contradict the assumption of Hofer et al. [11] that plausibility is generally the exclusive result of a higher-order cognitive process. Instead, the results give another hint to involved bottom-up and top-down processes in the emergence of plausibility. The results from experiment 2 further support these assumptions and the acceptance of H1.

Equally, we would have expected that the cognitive layer in the manipulation space (top-down factor) has an effect on the condition space resulting in internal plausibility. However, H2 *"The cognitive manipulation neutralizes the break in plausibility induced by the perceptual manipulation."* must be declined as almost all participants (39 out of 40) chose the normal object behavior to be more plausible in experiment 1. When we transfer the results into the new model, we assume that the perceptual manipulation has consequences on the perceptual and cognitive layer. If we additionally manipulate the cognitive layer, there is a conflict inside the cognitive layer. Since the perceptual manipulation of the object behavior is bottom-up, the cognitive framing manipulation, as a top-down factor, has a limited effect. These findings indicate that bottom-up factors that cause breaks in plausibility are too strong to be counteracted by top-down factors. Alternatively, the top-down factor in this experiment might have been designed too weakly compared to the bottom-up factor since the framing consisted solely of a not very detailed, written text. Also, the VR application was kept minimal to avoid possible confounds between the framing and no-framing condition and thus, had little relation to the story.

To address these issues, a second experiment was conducted. This time, the distinction in the cognitive manipulation was between two different framing stories and not between no-framing and framing. Additionally, the stories had a visible connection to the design of the virtual environment and extended into the task. We kept the stories similar to avoid confoundsand controlled this by requesting the narrative engagement. The narrative engagement scores did not show significant differences, and the mean values were relatively high in all conditions indicating an equally engaging character of both framing stories. However, H2.1 *"A combination of textual and visual framing on the cognitive layer can neutralize the break in plausibility and create an internal plausibility."* has to be declined as well. Looking at the model [16], we assume that the three manipulation layers have different influences on the condition space. It strengthens our previous assumption that the bottom-up factors have more impact on the condition space than top-down factors considering that the bottom-up layers affect higher-order processes. This makes it harder for top-down factors to counteract their effects. The plausibility ratings were less significant in the second experiment compared to the first experiment (see figure 7). We found a slight interaction effect in the second experiment in some individual questions, which lets us assume that the cognitive manipulation was stronger in the second experiment than in the first experiment. This could have an overall effect on the plausibility, causing the less extreme ratings. The study design could be another explanation. In the first experiment, participants were exposed to both perceptual manipulations consequently, allowing to compare the two conditions directly. The plausibility ratings of the second exposure could have been influenced by the knowledge of the first exposure, leading to a bigger gap in the ratings. In our second experiment, we used between conditions giving the participants no anchor for comparison.

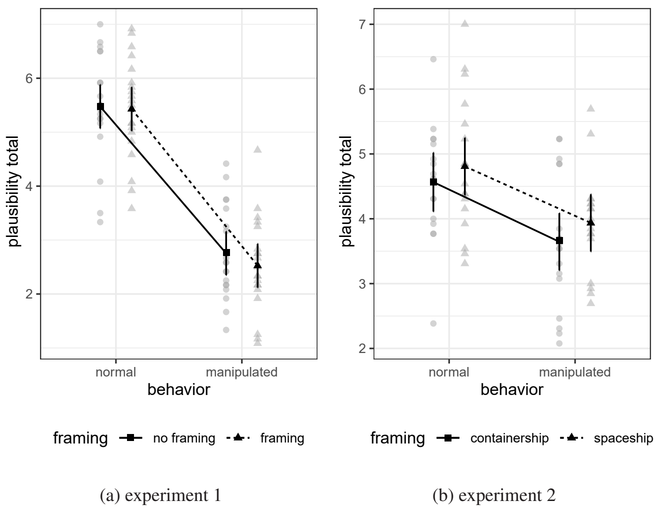


Fig. 7: Plausibility total scores from both experiments which have converged between the behavior conditions in the second experiment.

H3 *"Breaks in plausibility lead to break in presence."* provides information about the transition from the condition space to the qualia space in the proposed model. In experiment 1, it could partly be accepted. The subscale *involvement* of the IPQ was significant. This might result from a higher amount of interaction in the manipulated object behavior condition, which is supported by Skarbez et al. [23] who found that higher interaction leads to a higher presence. The higher interaction can be substantiated by the high values of the NASA-TLX and the longer time needed to fulfill the task. We further found a significance in the IPQ subscale *experienced realism* for the cognitive manipulation factor. We suspect that this significance results from the different cognitive manipulations we used. In one condition, we provided a short textual framing that might have arisen an engaged feeling . In the other condition, we omitted this short text, so participants had no prior expectation and thus, potentially felt a lower engagement. We wanted to rule out the possibility that the effect was caused by a difference in narrative engagement, so we controlled for this factor in the second experiment. However, we did expect that a stronger framing would have a more substantial effect on the cognitive layer, and therefore, cause internal plausibility. This should have led to more significant results and revealed an interaction effect between the cognitive manipulation and the perceptual manipulation's effects on the cognitive layer.

Consequently, for the second experiment, we altered the hypothesis to H3.1 *"Perceived coherence between the cognitive and perceptual manipulation has an effect on presence."*. For testing this hypothesis, coherence between the cognitive and perceptual manipulation has to be given. As we already discussed in H2 and H2.1, an internal plausibility could not be established, so we do not fulfill this requirement. Hence, this hypothesis can neither be accepted nor declined. In the second experiment, the participants rated the manipulated object behavior as significantly more implausible than the normal object behavior. As stated before, we conclude that there is a break in plausibility. In contrast to the first experiment, we did not find any significant values in the presence measures. Hence we conclude that the break in plausibility did not, in fact, cause the effect in the presence scores that we observed in the first experiment. In the first experiment, we assume that the measured *involvement* was mainly due to the interaction with the objects. In experiment 2, on the other hand, the *involvement* was more likely a combination of the interaction, the story, and the environment itself which possibly leads to more similar *involvement* scores between the conditions. As we did not check for the narrative engagement in the first experiment, we can only assume that there might have been a difference in the engagement between receiving framing or not. The narrative engagement did not show significant results in the second experiment. Thus, we argue that the significance in the subscale *experienced realism* of the IPQ in experiment 1 was due to the discrepancy in narrative engagement.

When we try to fit these results into the model of Skarbez et al. [21], we come up against our limits. They argue that presence is influenced by three factors: place illusion, plausibility illusion and social presence illusion. As the participants did not interact with anyone, the social presence illusion could not have influenced our presence score. The environment, technical setup, and interaction possibilities are consistent between the perceptual manipulation conditions, keeping the immersion similar. So we do not assume that the place illusion has a different effect on presence between the conditions leaving the plausibility illusion as a last remaining influence factor on presence. From the manipulation check, we can see that we did create a break in plausibility. Following Skarbez' model, plausibility illusion is an outcome from coherence and has, in return, a direct influence on presence. Hence, we would have expected that a break in plausibility would lead to a lower presence score, which is not the case. However, they do not explicitly state how the place illusion and plausibility illusion are quantifiable and if both factors are mandatory for presence, leaving an explanation gap.

The novel theoretical model by Latoschik and Wienrich [16] postulates a different approach. They propose a three-layer manipulation space with top-down and bottom-up layers. Bottom-up manipulations, e.g., through the object behavior on the perceptual layer, always impact the higher-order layers. Therefore, these cues operate on multiple layers that postulate their influence upwards. The result is a composition of coherencies that build a weighted activation function defining plausibility. Thus, the plausibility of an XR experience can be influenced by different combinations in the manipulation space. We successfully manipulated the perception layer, with impact on the cognitive layer, which led to an activation function that broke plausibility and did not include an influence on presence. Nonetheless, with a successful cognitive manipulation, we would get a different activation function and thus, might have influenced presence. This concern should be subject to further research.

6.1 Limitations and Future Work

As discussed before, the questions for the plausibility of the object behavior have not been validated before. Even though the results for the *external plausibility* were significant, we cannot say with certainty that no *internal plausibility* was established. These questions should be validated in a separate process. However, it was the first time that the internal plausibility component proposed by Hofer et al. [11] was explicitly assessed. The internal consistency of the *external* and *internal plausibility* question group was satisfying, which is promising.

In the first experiment, H3 was declined. However, we did see a significance in the subscales of the IPQ. We argued that these significances might be due to a difference in interaction with the objects (subscale *involvement*) as well as a difference in narrative engagement between the cognitive manipulation groups (subscale *experienced realism*). Nonetheless, in the second experiment, we still had a discrepancy in interaction, overall VR exposure time, and the task load. Moreover, we did not collect data about the narrative engagement in the first experiment. Hence, these are just speculations and need to be investigated systematically. The discrepancy in the task load poses another limitation as the significantly higher task load is another confound that needs to be controlled in future research.

The evaluation of the SSQ revealed significance in both experiments. However, the means calculated by the change over time are not alarming, as their absolute values are relatively small.

Another concern is the uncertainty of whether the new environments in experiment 2 brought any confounds. While they were kept similar between the conditions, they are quite different from the first experiment. Many things happen simultaneously after the crash, and the participants might have been easily distracted from their task and the object behavior. A more engaging environment has the potential to create more confounds.

Moreover, the mainly female and young participant pool could have affected the results. As there is no empirical evaluation on the correlation between plausibility and gender and age to the best of our knowledge, a systematical examination would be insightful. Furthermore, it would be interesting to have a more balanced participant pool to better represent society in the future.

Lastly, this first empirical evaluations only looked at a small part of the newly proposed model. Even though we can make some assumptions from the results, there is still a lot unknown. For one, future research should take other qualia into account. Furthermore, the internal relations within the model need to be tested more in-depth for more reliable assumptions and predictions through the model. One way would be to vary the kind of manipulations used to influence the coherence activation. E.g., we could use a weaker prior than gravity or try and strengthen the cognitive manipulation even further. An additional manipulation of the sensational layer would also be imaginable. However, this is hard to achieve as neuro-signals cannot be influenced easily.

7 Conclusion

While the concepts of plausibility and coherence have recently gained more attention in the XR community, research in this area is still on thin ground. Latoschik and Wienrich [16] proposed a new theoretical model that strengthens the role of these concepts in influencing the overall XR experience. The results of our two studies, which pose the first empirical evaluations of this model, show that this new model is capable of explaining the effect plausibility and coherence have on other qualia that make up the XR experience. We examined whether we could cause breaks in plausibility and which effects they have on the sense of presence. Results show that a break of a bottom-up factor in the manipulation space, i.e., the perceptual layer, could indeed break the plausibility. However, this did not affect the presence, and we were not able to counteract this break with a manipulation of the top-down factor, i.e., the cognitive layer. These results look promising to substantiate a better understanding of XR experiences and provide a good starting point for further research of this theoretical model, its internal relations, and its advantage over existing presence models.

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] HTC corporation, https://www.htc.com/, 2021.

[2] P. Bacherle. Eintauchen in narrative welten - theoretische und empirische zugänge zum rezeptionserleben. 2015.

[3] R. Busselle and H. Bilandzic. Fictionality and Perceived Realism in Experiencing Stories: A Model of Narrative Comprehension and Engagement. *Communication Theory*, 18(2):255–280, 2008. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2885.2008.00322.x. doi: 10.1111/j.1468-2885.2008.00322.x

[4] A. de Graaf, H. Hoeken, J. Sanders, and J. W. J. Beentjes. Identification as a mechanism of narrative persuasion. 39(6):802–823, 2012. Publisher: SAGE Publications Inc. doi: 10.1177/0093650211408594

[5] J. W. v. Goethe. *Maxims and Reflections*. Beck, 2006.

[6] A. C. Graesser, B. Olde, and B. Klettke. How does the mind construct and represent stories? In *Narrative impact: Social and cognitive foundations*, pp. 229–262. Lawrence Erlbaum Associates Publishers, 2002.

[7] M. C. Green. Transportation into narrative worlds. In L. B. Frank and P. Falzone, eds., *Entertainment-Education Behind the Scenes: Case Studies for Theory and Practice*, pp. 87–101. Springer International Publishing, 2021. doi: 10.1007/978-3-030-63614-2_6

[8] N. Hamzeheinejad, D. Roth, S. Monty, J. Breuer, A. Rodenberg, and M. E. Latoschik. The impact of implicit and explicit feedback on performance and experience during vr-supported motor rehabilitation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 382–391. IEEE, 2021.

[9] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, vol. 52, pp. 139–183. Elsevier, 1988. doi: 10.1016/S0166-4115(08)62386-9

[10] R. M. Hein, C. Wienrich, and M. E. Latoschik. A systematic review of foreign language learning with immersive technologies (2001-2020). *AIMS Electronics and Electrical Engineering*, 5(2):117–145, 2021.

[11] M. Hofer, T. Hartmann, R. Ratan, and L. Hahn. The Role of Plausibility in the Experience of Spatial Presence in Virtual Environments. *Virtual Reality*, 1:1–9, Apr. 2020. doi: 10.3389/frvir.2020.00002

[12] B. Jörges and J. López-Moliner. Gravity as a Strong Prior: Implications for Perception and Action. *Frontiers in Human Neuroscience*, 11:203, 2017. doi: 10.3389/fnhum.2017.00203

[13] D. Kahneman. *Thinking, fast and slow*. Penguin Books, London, 2012. OCLC: 798805166.

[14] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, July 1993. doi: 10.1207/s15327108ijap0303_3

[15] F. Kern, P. Kullmann, E. Ganal, K. Korwisi, R. Stingl, F. Niebling, and M. E. Latoschik. Off-the-shelf stylus: Using xr devices for handwriting and sketching on physically aligned virtual surfaces. *Frontiers in Virtual Reality*, 2:69, 2021.

[16] M. E. Latoschik and C. Wienrich. Coherence and Plausibility, not Presence?! Pivotal Conditions for XR Experiences and Effects, a Novel Model. *arXiv:2104.04846 [cs]*, Apr. 2021. arXiv: 2104.04846.

[17] M. Minsky. Telepresence. pp. 45–52, 1980.

[18] L. Popova. Perceived reality of media messages: Concept explication and testing - ProQuest, 2010.

[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

[20] R. Skarbez, F. Brooks, and M. Whitton. Immersion and Coherence: Research Agenda and Early Results. *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, Apr. 2020. doi: 10.1109/TVCG.2020.2983701

[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*, VRST '18, pp. 1–11. Association for Computing Machinery, New York, NY, USA, Nov. 2018. doi: 10.1145/3281505.3281530

[22] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton. A survey of presence and related concepts. 50(6):1–39, 2017. doi: 10.1145/3134301

[23] R. Skarbez, S. Neyret, F. Brooks, M. Whitton, and M. Slater. *A Psychophysical Experiment Regarding Components of the Plausibility Illusion*. Apr. 2017.

[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. Publisher: Royal Society. doi: 10.1098/rstb.2009.0138

[25] M. Slater and A. Steed. A virtual presence counter. 9(5):413–434, 2000. doi: 10.1162/105474600566925

[26] C. Wienrich, N. Döllinger, and R. Hein. Behavioral framework of immersive technologies (behavefit): How and why virtual reality can support behavioral change processes. *Frontiers in Virtual Reality; Special Issue: User Experience in Virtual Reality Application design*, 2:84, 2021. doi: 10.3389/frvir.2021.627194

[27] C. Wienrich and M. E. Latoschik. extended artificial intelligence: New prospects of human-ai interaction research. *Frontiers in Virtual Reality; Special Issue:The Role of Virtual Reality in Human-AI Interaction (HAII)*, 2:94, 2021. doi: 10.3389/frvir.2021.686783

[28] A. Witmer and M. Slater. Measuring presence: A response to the witmer and singer presence questionnaire. 8. doi: 10.1162/105474699566477

[29] B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, June 1998. doi: 10.1162/105474698565686