**Investigating the Effects of Plausibility Illusion and Congruency on Declarative Knowledge in VR and AR**

Gabriela Christofoletti

SRN 2076475

Master's Thesis

Communication and Information Sciences New Media Design

Department Communication and Cognition School of Humanities and Digital Sciences

Tilburg University, Tilburg

Supervisor: Dr. M. Shin

Second Reader: Dr. David Peeters

August 2023

## Abstract

The literature on AR and VR lacks conclusive evidence regarding their effectiveness for educational purposes. Some scholars argue that the abundant multisensory information and the high cognitive demands of processing virtual elements in extended reality (XR) may overwhelm learners' mental capacities, hindering effective learning. To explore the potential benefits and drawbacks of immersion, this study compared the effects of VR and AR on a sense of realism. The investigation builds upon the advancements in XR literature and used "plausibility illusion" as a foundational concept for comparing the sense of realism in VR and AR, exploring its impact on cognitive load and declarative knowledge. Additionally, the study explored principles of traditional multimedia instruction design in XR mediations by investigating the moderating role of audio-video congruency on the relationship between XR modalities and plausibility illusion, and plausibility illusion on cognitive load. To test the research question, a mixed 2 (modality: AR vs. VR) × 2 (audio-visual congruency: congruent vs. incongruent) experiment was conducted. Surprisingly, the results revealed no significant difference between VR and AR in inducing plausibility illusion, nor a moderation effect of congruency on this relationship. However, a higher sense of plausibility illusion was associated with increased cognitive load, which negatively affected declarative knowledge acquisition. Moreover, congruency was found to significantly moderate the relationship between plausibility illusion and cognitive load, although the pattern observed contradicted the initial expectation. The implications of these findings suggest the necessity of further exploration into cognitive load and individual differences.


*Keywords*: Augmented reality, virtual reality, plausibility illusion, cognitive load, congruency effects, multimodality processing, declarative knowledge, education

**Table of Contents**

Theoretical Framework ........ 10
Experiencing Immersion: AR and VR Structural Affordances ........ 10
Plausibility Illusion ........ 11
Cognitive Load ........ 13
Psi and CL ........ 14
CL and Declarative Knowledge ........ 15
Multimodality Processing and CL: The Moderating Role of Congruency ........ 16
Methods ........ 20
Design ........ 20
Participants ........ 21
Stimuli ........ 22
Virtual Stimulus ........ 22
Auditory Stimulus ........ 22
Congruent Stimulus ........ 23
Incongruent Stimulus ........ 24
Materials and Equipment ........ 24
VR Modality ........ 24
AR Modality ........ 25
Procedure ........ 26
Measurements ........ 27
Plausibility Illusion ........ 27
Cognitive Load ........ 27
Declarative Knowledge ........ 28

Control Variables ........ 28
Results ........ 28
Measurement Validity ........ 28
Hypothesis Testing ........ 29
Control Variables ........ 34
Model Fit ........ 34
Discussions........ 35
Implications ........ 40
Limitations and Future Directions ........ 42
Conclusion ........ 44
References ........ 46
Appendices ........ 56
Appendix A ........ 56
Appendix B ........ 59
Appendix C ........ 60
Appendix D ........ 62
Appendix E ........ 63
Appendix F ........ 66

**Abbreviations**

| | |
|---|---|
| AV | Audio-Visual |
| AR | Augmented Reality |
| CL | Cognitive Load |
| DK | Declarative Knowledge |

| | |
|---|---|
| HMD | Head-Mounted Display |
| MT | Mouse Trajectory |
| PI | Place Illusion |
| Psi | Plausibility Illusion |
| RVC | Reality-Virtuality Continuum |
| RT | Response Time |
| VE | Virtual Environment |
| VHP | Virtual Human Plausibility |
| VR | Virtual Reality |
| XR | Extended Reality |

# Investigating the Effects of Plausibility Illusion and Congruency on Declarative Knowledge in VR and AR

Virtual Reality (VR) and Augmented Reality (AR) technologies are revolutionizing the way people interact with digital content. These technologies have enormous potential in various industries, including gaming, healthcare, and training (Steffen et al., 2019). Particularly, VR and AR technologies are gaining attention as tools to enhance learning outcomes in education (Bower et al., 2014; Merchant et al., 2014). Although VR and AR are often grouped together as immersive technologies in media and literature, it is important to recognize that their structural affordances fundamentally vary – i.e., VR induces a sense of being there whereas AR elicits a sense of an object being in the environment (Steffen et al., 2019). Consequently, the underlying cognitive mechanisms involved in processing information and acquiring new knowledge through these technologies may also differ. However, there is currently a lack of research exploring whether and how the differences between VR and AR will result in distinct effects on knowledge outcomes. Additionally, it remains understudied the theoretical basis underlying these effects and how VR and AR structural affordances can be addressed to enhance the design of instruction materials.

VR and AR affordances are generally expected to provide an immersive experience, which is believed to enhance engagement and learning outcomes. However, prior literature has yielded mixed results and inconclusive evidence regarding the effects of AR and VR in the field of education (Buchner et al., 2021; Han et al., 2021). One possible reason for having mixed results in Extended Reality (XR) literature may be that previous scholars often used the concept of presence to explain the cognitive mechanisms involved in AR and VR experiences. It is worth noting that the concept of (spatial) presence, which refers to the sense of "being here", does not apply to AR, since users are in contact with their actual physical environment (Latoschik & Wienrich, 2022). Only recently scholars have started to adopt the Plausibility

Illusion (Psi), a sub-construct of presence, to better understand how and why AR could be used for learning and education (Latoschik & Wienrich, 2022). Slater (2009) demonstrates that the concept of Psi, referred to as "the illusion that what is happening is really happening", is equally relevant to both VR and AR research. Therefore, the present study will adopt Psi to bridge the gap between inconsistent findings that have emerged across past studies. Moreover, this study will investigate whether VR and AR might distinctively elicit a sense of Psi.

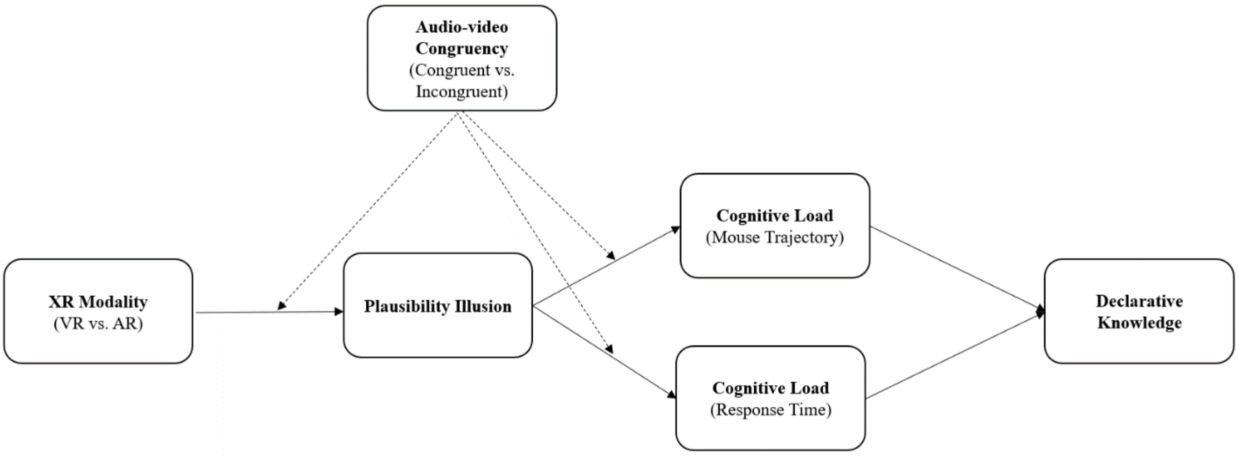
The structural affordances of VR and AR may particularly demand allocations of cognitive resources to process the sense of Psi. To understand the impacts of Psi on learning, this study will draw on Cognitive Load (CL) theory. CL refers to the mental effort required to complete a task and assumes that cognition is a limited resource (Chandler & Sweller, 1992) In educational contexts, CL is of particular importance as the exposure to effortful instructional material can exceed learners' cognitive capacities, leading to diminished recall (Moreno & Mayer, 2000; Paas et al., 2003). Thus, clarifying the cognitive demands involved to process the sense of Psi becomes crucial to comprehend the implications of using AR and VR as educational mediation. For instance, a recent study by Ahn et al. (2022) suggests that the sense of spatial presence in VR negatively affects information recall by depleting the cognitive resources available for processing information. Whereas Psi is considered a sub-dimension of the concept of presence (Latoschik & Wienrich, 2022), it may also impact CL and potentially hinder declarative knowledge outcomes.

While CL remains an active area of research in XR literature, insights from previous studies on traditional learning can aid our understanding. For instance, prior studies in multimedia learning have revealed that congruency effects, derived from the alignment of audio and visual information on a semantic level, can enhance learning outcomes by facilitating information processing and reducing cognitive costs (Patterson et al., 2007; Doehrmann & Naumer, 2008). Considering that XR experiences are multimodal by nature, as they generally encompass visual, auditory, and haptic modalities, exploring congruency in XR instructional design can offer valuable insights into how to improve educational material in VR and AR. As the sense of Psi relies on perceiving visual virtual elements as realistic, it's possible that adding auditory information congruent with the visual stimulus might facilitate the processing of Psi. Thus, congruency may potentially moderate the effects of Psi on CL by smothering the virtual visual elements processing through coherent auditory cues. Additionally, given that VR and AR technologies might impose distinctive demands on cognitive mechanisms to process the sense of Psi, congruency is likely to interact with the effects of technology modalities on Psi.

By addressing Psi as an important aspect of the distinctive effects of VR and AR technologies on cognitive mechanisms, this research aims to provide valuable insights into the suitability of each technology for learning objectives. Additionally, exploring the audio-visual interplay on the perceptual process of Psi and in higher-cognition levels can leverage to optimize the effectiveness of instructional materials in XR educational environments. Taken together, this study aims to address some gaps in the literature, including the minimal exploration of VR versus AR comparisons, the limited evidence on Psi research and its impact on CL (Buchner et al., 2021; Han et al., 2021; Leutner & Leopold, 2021). Moreover, this study will delve into congruency effects to investigate how XR educational mediations can be better designed, contextualizing traditional learning principles to the exclusive psychological mechanisms (such as Psi) that these technologies entail. Specifically, the present research aims to answer the following two research questions: "Will VR and AR engender distinctive effects on learning outcomes through the mediation of Psi?" and "Will congruency moderate the relationship between XR modality and Psi and Psi and cognitive load"? Figure 1 visualizes the conceptual model of this research.

**Figure 1**

*Research Model*




Note. Moderation indicated by dotted lines.

# Theoretical Framework

## Experiencing Immersion: AR and VR Structural Affordances

Virtual Reality (VR) and Augmented Reality (AR) have garnered significant interest within the educational sector over the past few decades (Bower et al., 2014; Merchant et al., 2014). These technologies are revolutionary as they enable users to experience situated learning from their own point of view (Dunleavy & Dede, 2014). Moreover, their technical attributes provide exceptional levels of multisensory and immersive experiences that have the potential to augment students' engagement and involvement with the instructional material (Zhang et al., 2020). However, previous literature has yielded inconclusive evidence regarding the effectiveness of AR and VR for educational purposes (Buchner et al., 2021; Han et al., 2021). This paper aims to bridge inconsistent findings by exploring both instructional design and the underlying psychological mechanisms in XR experiences. Before examining the instruction design, it is paramount to situate some significant implications of XR interactions on information processing, digging into the distinctive attributes of VR and AR that may influence learning.

The most conspicuous difference between AR and VR might be pertinent to their unique structural affordances and the cognitive demands they might impose. VR is characterized by transporting users to a computer-generated environment, providing a fully immersive experience (Huang et al., 2019). In contrast, in AR experiences digital content is superimposed onto the user’s actual physical space, thereby "augmenting" the real environment with virtual objects (Milgram & Kishino, 1994). Such particularities might elicit a distinctive sense of realism and immersion: with VR inducing a sense of being there, whereas AR induces a sense of an object being in the environment (Steffen et al., 2019).

Prior studies have accessed the sense of realism in XR experiences and its implications on cognitive mechanisms by examining the concept of presence. Defined as "the feeling of being there", presence is considered a psychological state of non-mediation, in which the individuals feel fully present in the virtual world (Lombard & Ditton, 2006). However, in AR, the idea of spatial presence is not applicable, as users are in contact with their surroundings and the real environment (Latoschik & Wienrich, 2022). Considering the wide scope of immersive technologies, scholars have recently proposed Plausibility Illusion (Psi), a sub-construct of presence, as a better concept to examine XR experiences (Latoschik & Wienrich, 2022). Following this recommendation, Psi will be adopted in the present research to investigate the potential distinctive effects of VR and AR on information processing and cognitive resources.

**Plausibility Illusion**

Plausibility Illusion (Psi) is defined as the “illusion that what is apparently happening is really happening” (Slater, 2009). Skarbez (2016) states that Psi occurs when sensory stimuli are perceived as coherent, meaning that objects, scenarios, and body representations behave following the user's expectations and previous beliefs. Latoschik and Wienrich (2022) also highlight coherence as key to achieving plausibility in XR experiences, implementing coherence in a broader context compared to the model proposed by Skarbez et al. (2017). According to them, plausibility in XR experiences relies on the coherence found within the sensory, perceptive, and cognitive layers. In other words, when users are exposed to sensory information (e.g., video/audio) in XR technologies, the cues within these stimuli are processed at perceptive and cognitive levels. This processing is influenced by the actual stimuli as well as the users' knowledge and social-cognitive factors. Thus, the plausibility will be determined by the XR experience (sensory stimuli) and the use of this stimulus as cues to evaluate coherence, based on prior knowledge and worldviews. The most important aspect of this perspective is the central role of coherence and plausibility of signals as paramount factors in bridging the sense of realism in XR experiences. The distinctive structural affordances of VR and AR technologies place different challenges concerning coherence and plausibility and may also induce varying levels of Psi (Latoschik & Wienrich, 2022).

In VR experiences, Psi relies on processing elements that are already incorporated into a single and coherent dimension (Steffen et al., 2019). Since the exclusion from the physical environment and detachment from the actual reality are features of VR technologies, VR allows the creation of experiences that do not necessarily follow physical laws, such as linearity and gravity, without necessarily disrupting the sense of realism (Steffen et al., 2019). The most crucial aspect for perceiving VR experiences as realistic is body representation and the sense of body ownership when utilizing virtual hands and controllers (Skarbez, 2016; Skarbez et al., 2017).

On the other hand, the AR’s structural affordances require the integration of virtual objects into the actual physical space. To provide users with a sense that virtual objects and actual surroundings are blended into a single dimension, they must behave coherently (Steffen et al., 2019). Whereas real elements and environment are constrained by physics laws, virtual objects can disregard these laws - leading to non-linear space-time relationships, the absence of physical counterparts, and violations of physical principles like gravity (Steffen et al., 2019). Consequently, AR experiences may be more prone to mismatches between virtual objects and real surroundings, reducing coherence and inhibiting the sense of Psi. On the other hand, in VR technologies users are completely immersed in a virtual environment, blocking out the real world and reducing potential incoherence. Based on this understanding of VR's inherent coherence between the virtual scenario, objects, and the user's body, the first hypothesis is proposed:

H1. Participants will experience a greater level of Psi when a virtual object is visualized via VR compared to when it is visualized via AR.

**Cognitive Load**

Psi might be a key psychological mechanism in XR due to its interrelation with immersion. However, the processing of Psi may impose specific cognitive demands that possibly influence learning. This study will build upon existing cognitive load research and recent investigations into cognitive load within XR technologies to explore the influence of Psi on learning. Cognitive load (CL) refers to the mental effort needed to complete a task (Chandler & Sweller, 1992), and it assumes that cognition is a limited resource (Sweller et al., 2019). CL theory is divided into three types: (1) intrinsic CL: the task's difficulty and the information complexity, (2) extraneous CL: the content and material presentation, and (3) germane CL: the task's relevance and the use of schemas to integrate new information with existing knowledge and worldviews. (Paas & Sweller, 2005).

Among these three types, extraneous CL gained special emphasis in the educational field (Moreno & Mayer, 2000). When instructional material exposes students to a great amount of information, their CL may exceed their working memory capacity, leading to diminished recall performance (Moreno & Mayer, 2000; Paas et al., 2003). Paas et al. (2004) also highlight the importance of instructional strategies that align with the learners' cognitive limitations and capacities to manage CL effectively. According to Paas et al. (2004), instructional design can enhance learning outcomes by reducing extraneous CL (e.g., by eliminating unnecessary information or irrelevant elements) and optimizing germane CL (e.g., by facilitating schema construction and integration of multimodal information).

To retrieve the main findings in VR and AR research on CL, two systematic reviews compiling the main empirical studies were accessed (Buchner et al., 2021; Han et al., 2021). The reviews revealed a lack of consensus among scholars, and the authors recommended future studies focusing on the content of instructional material, suggesting that more weight should be given to the influence of extraneous CL in these experiences (Buchner et al., 2021; Han et al., 2021). This gap in the literature will be further explored in the next session, where I will approach material presentation principles applied to AR and VR's structural affordances. Although these reviews gathered a considerable corpus of XR research on CL, there is a scarcity of studies exploring the distinctive demands of AR and VR on cognitive resources.

An exception is one study conducted by Huang et al, (2019) which evaluated the effects of AR and VR on knowledge retention and used the CL theory to explain their results. They found that while participants in the AR condition presented higher knowledge retention in auditory information, participants in the VR condition retained more visual information (Huang et al., 2019). According to Huang et al. (2019), AR technology elicits weaker responses to process the virtual environment, allowing for more cognitive resources to process auditory information. The findings provide intriguing evidence regarding how AR and VR structural affordances place unique demands to process perceptual illusions from virtual elements. Additionally, the results suggest that the underlying cognitive mechanisms involved in VR and AR will distinctively affect the multimodality (e.g., audio/video) processing. Thus, a comprehensive exploration of the effects of Psi and immersion on cognitive resources is crucial to understand the implications of applying XR as educational mediation.

**Psi and CL**

In traditional educational literature, immersion, and presence have been linked to higher levels of engagement, which, in theory, result in increased attention and positive effects on learning outcomes (Slater & Wilbur, 1997). Studies have indicated that VR and AR can elicit a sense of presence, capturing users' attention and promoting heightened focus and engagement (Kearney & Schuck, 2017). Still, it remains unclear whether this engagement translates into improved learning outcomes, as the sense of presence in XR may lead users to allocate excessive attentional and cognitive resources to process the virtual environment and elements as if they were real. According to Biocca et al. (2001), presence in virtual environments is a product of the mind's attempt to integrate and make sense of sensory cues. This involves a process of intermodal-integration, where users need to integrate information from various sensory channels (Marks, 1978; Stein et al., 1989). During this process, information from one sensory channel may be used to enhance and disambiguate information from another, resulting in perceptual illusions (Biocca et al., 2001). Thus, engagement in XR environments involves not only being exposed to a substantial amount of sensory information but also utilizing and combining them to create a convincing illusion of realism. This illusion processing can result in attentional overload (Biocca et al., 2001). The research of Biocca et al. (2001) and evidence from XR literature (Buchner et al., 2021; Han et al., 2021) suggest that the CL experienced in XR interactions primarily stems from the cognitive processing needed to perceive virtual elements as coherent and realistic. Applying this notion to the concept of Psi, the second hypothesis is proposed:

H2. The increase in Psi will have a positive association with CL.

**CL and Declarative Knowledge**

In XR experiences, it is likely that a significant part of cognitive resources is involved in processing Psi, and the remaining resources might not be sufficient to process the educational content deeply. The relationship between information processing and recall was proposed by Craik and Lockhart (1972) in "The Levels of Processing Theory" and posits that the "depth at which the information is processed during encoding determines how well it will be retained". This is in line with findings from a study conducted by Ahn et al. (2022). According to them, presence in VR might lead to CL and negatively affects information recall. The results were explained based on the Cognitive Theory of Multimedia Learning, which assumes that working memory is a limited cognitive resource that impacts how we process information (Sweller et al., 2019). Accordingly, when cognitive resources are allocated to process presence, the processing of other information might be inhibited, diminishing learning outcomes (Ahn et al., 2022). Considering the implication of CL on recall, the third hypothesis predicts:

H3. The increase in CL will have a negative impact on declarative knowledge.

**Multimodality Processing and CL: The Moderating Role of Congruency**

Hitherto, the theoretical framework presented here might indicate that the structural affordances of VR and AR have distinctive effects on Psi and, consequently, on CL, which might affect declarative knowledge. Huang et al. (2019) findings supported this view and highlighted the different implications of VR and AR for modality (audio/visual) processing. The most intriguing evidence of this study is that AR and VR's distinctive demands on cognitive resources might spill over into audio and visual processing, affecting how different modalities of information will be retained in each XR technology. Moreover, considering the suggestions of Buchner et al. (2021) and Han et al. (2021), the present study will also explore the instructional material to gain a granular understanding of the consequences of XR technologies on education. To address instruction content and the multimodality quality of these technologies, the second part of the theoretical framework will develop further the CL and Psi emphasizing learning material design.

While CL remains an active area of research in XR technologies, insights from traditional learning literature can aid our understanding. For instance, in traditional multimedia learning, combining audio and video make the educational content more attractive and leads to higher engagement (Mayer, 2006). However, this combination imposes an additional burden on cognitive resources. Over the years, investigations into how we process multimodality information have resulted in Mayer's (2005) Cognitive Theory of Multimedia Learning (CTML).

CTML comprises extensive research on multimodality processing, emphasizing the importance of instructional design to help learners construct coherent mental representations from content presentations (Mayer, 2005). The cognitive process of building mental representations from multimedia instruction begins with selecting relevant information (Mayer, 2006). Multimedia instructions can effectively guide learners' attention toward pertinent details, with audio and video reinforcing key aspects of the learning material. Moreover, audio-video congruency allows learners to organize and process information from both modalities into a single mental representation, enabling them to connect new information with prior knowledge and concepts (Mayer, 2006). Therefore, in multimedia instruction, cognitive efforts are greatly influenced by the ease of integrating multimodalities and processing them together, rather than having audio and visual channels competing for working memory resources (Chandler & Sweller, 1992). Chandler and Sweller (1992) argue that when multimedia instruction is coherently designed, information from different modalities acts as cues and facilitates mental integration. Such integration subserves the information process and lessens the CL (Chandler & Sweller, 1992). However, to successfully integrate audio-visual stimuli, their semantics must be congruent (Gottfried & Dolan, 2003). Semantics encompass the entire network of meanings associated with a concept, including properties, usage, and schemas (Doehrmann & Naumer, 2008). When the audio-visual content aligns semantically, this entire network of meanings is activated from both modalities and can be integrated to construct coherent mental representations (Doehrmann & Naumer, 2008). This alignment of different modalities at the semantic level is defined as "congruency effects", and it is known to facilitate information processing by enabling integration across modalities (Doehrmann & Naumer, 2008). The implications of congruency effects are particularly significant in the field of education. Iwamiya (2013) demonstrated that congruence between visual and auditory elements significantly impacts learners' perceptions. When the auditory and visual elements are congruent, they mutually reinforce each other, resulting in a higher level of comprehension and improved information recall (Iwamiya, 2013).

By leveraging the knowledge and principles established in traditional multimedia learning, we can enhance our understanding of XR as a mediated educational environment. The current study explores congruency effects, driven by two key attributes of XR. Firstly, XR experiences are inherently multimodal, combining various sensory modalities (such as visual, auditory, and sometimes haptic feedback) to create immersive and realistic environments. However, the abundance of multimodal information in XR, compared to traditional multimedia (Akçayır & Akçayır, 2017; Antonioli et al., 2014; Cheng & Tsai, 2013; Van Kreveln & Poelman, 2010; Wu et al., 2013), might impose higher cognitive demands. Given that previous studies have demonstrated a stronger congruency effect for complex stimuli compared to simple ones (Van Veen & Carter, 2002), investigating the effects of congruency may be particularly valuable for designing educational content in VR and AR environments. Furthermore, the concepts of coherence in Psi and the congruency effects both propose that the effectiveness of processing multimodal information relies on the extent to which individuals can integrate information from different modalities into a cohesive mental representation.

Secondly, the influence of visual elements in the sense of Psi represents another significant characteristic closely tied to multimedia theory. Mayer (2014) highlights the importance of visual aids to make the relationship between prior knowledge and new information more explicit and emphasizes the role of congruency effects to facilitate learning. In XR environments, where visual stimuli are pivotal to inducing a sense of realism, the inclusion of audio information must be carefully designed to reinforce the visual aids. Although adding audio information might be important to complement educational content, auditory information excessively abstract or semantically divergent from visual stimulus likely prevents their integration, hindering information processing for both modalities. Additionally, the structural affordances of VR and AR may impose different demands on cognitive resources, particularly in how their visual aids induce distinctively the sense of Psi. Hence, the first aspect of congruency this study aims to explore is whether congruency moderates the relationship between XR modalities and Psi.

Different from traditional multimedia, AR technologies might present additional cognitive demands. These demands arise from the need to integrate virtual objects and real surroundings and effectively merge them into a cohesive visual experience. This unique characteristic of AR technology prompts the investigation of whether congruency can also facilitate the integration and binding processing of visual elements in AR’s Psi processing. CTML suggests that multimodality integration subserves information processing and reduces CL (Chandler & Sweller, 1992). It is likely that video-audio congruency in AR, by reducing cognitive burden, can also facilitate the processing of Psi by enhancing the integration of elements within the visual modality. Moreover, previous evidence relates Psi to the stimuli' coherence and to the user’s capability of constructing a coherent mental representation from them. In this sense, the audio information can function as a cue to reinforce the construction of schematic representation from virtual visual stimuli. Hence, I theorize that in AR experiences the congruency effect will benefit the inducement of a sense of Psi, by smothering the integration between virtual and real elements into a single dimension. On the other hand, in VR users are fully immersed in a virtual world where the visual elements are already integrated. In this case, congruency may also benefit the information processing in

VR experiences, but the relationship between this technology and Psi, specifically, may be influenced by congruency to a lesser degree. Thus, the hypothesis is as follows:

H4a. Congruency will moderate the relationship between XR modalities and Psi such that congruency will strengthen the relationship between AR and Psi to a greater extent.

The second aspect of congruency that this study aims to explore is its impact on the relationship between Psi and CL. In XR literature, coherence is widely acknowledged as a critical factor in inducing a sense of Psi and immersion (Latoschik & Wienrich, 2022; Skarbez, 2016; Skarbez et al., 2017). By drawing on findings from CTML, I believe that exploring the interplay between audio-video information can deepen our understanding of XR experiences from a higher-level of cognition. CTML posits that when audio-visual information is semantically congruent, they facilitate the formation of helpful schematics and coherent mental representations, while also reducing cognitive demands (Chandler & Sweller, 1992; Mayer, 2014). Consequently, the semantics alignment of visual stimuli (intensely processed to create the sense of Psi) and audio will reinforce each other, permitting them to be integrated into a single mental representation. This integration may facilitate the processing and might enhance the encoding of information for both modalities. Thus, considering that the congruence between visual elements and audio content permits the multimodalities integration and expanding the notion of Psi as a product of multi-sensory coherence evaluation, the last hypothesis is proposed:

H4b. Congruency will moderate the relationship between Psi and CL such that congruency might weaken the effect of Psi in CL.

# Methods

## Design

A two (modality: AR vs. VR) × two (congruent vs. incongruent) between-subjects design experiment was performed to test the hypotheses. In the experiment, participants were exposed to a congruent or incongruent audio-video stimulus in either VR or AR experience, leading to four conditions: VR-Congruent, AR-Congruent, VR-Incongruent, and AR-Incongruent. The data was collected from April 26th to May 16th, 2023, at Tilburg University Media Design Lab. The Research Ethics and Data Management Committee of the Tilburg School of Humanities and Digital Sciences (TSHD) ethically approved the experiment (REDC2022.17ab_102-60-2023).

**Participants**

Participants were recruited through the Tilburg University pool. Everybody above 18 was allowed to participate in the experiment, apart from people with rabbit phobia. The experiment consisted of 92 participants after removing outliers, divided among four conditions, each containing 22 (VR-Congruent/AR-Congruent) or 24 (VR-Incongruent/AR-Incongruent) participants. The participants' ages ranged between 18 and 41 ($M = 21.60$, $SD = 4.22$), among which 63% identified as female and 37% as male. To test if the congruency groups did not differ significantly on previous experience with VR/AR and knowledge about rabbits, an independent *t*-test was performed. The previous experience with VR/AR and rabbits' knowledge were consequently ($M = 3.87$, $SD = 1.57$) and ($M = 2.48$, $SD = 1.42$) for the congruent condition and ($M = 4.49$, $SD = 1.46$) and ($M = 2.65$, $SD = 1.41$) for the incongruent condition. The difference between conditions were not significant for both experience with VR/AR ($t(90) = 1.95$, $p = .70$), and rabbits' knowledge $t(90) = .57$, $p = .82$), meaning that participants were evenly distributed among the groups. Finally, a self-report questionnaire assessed English proficiency to ensure that participants could understand the auditory information shared during the experiment. 61% of the participants considered themselves intermediate and 39% advanced or higher. Since all participants' scores were adequate, it was not necessary to check for their distribution among conditions.

## Stimuli

### *Virtual Stimulus*

A 3D bunny was selected as the stimuli based on its realism, realistic animated movements, and size, being not too small and not too big. Most importantly, for study purposes, the object should have represented an animal that could be equally found in the wild or at home (as a pet). Furthermore, to control for possible spurious effects on recall, the animal was selected considering familiarity and valence: rabbit is an animal that participants might not have deep scientific knowledge and it is not associated with higher valence, either negative or positive.

**Figure 2**

*The Virtual Stimuli*

The visual stimuli were consistent across all conditions and included the 3D rabbit as well as the environment. Participants interacted with the bunny in a living room, which could be either the actual space (AR) or a virtual recreation (VR) of the lab's living room. For more information on the environment details, please refer to the 'Materials and Equipment' section.

### *Auditory Stimulus – Declarative Knowledge (DK)*

Declarative knowledge (DK) refers to information about facts, concepts, events, and objects that can be consciously recalled and expressed (Anderson, 2013). According to "Levels of Processing theory", encoding DK information is successfully achieved through deeper levels of processing and involves being able to make connections to existing knowledge (Craik & Lockhart, 1972). Thus, considering world knowledge associations and coherent representations as core mechanisms to retain DK, two conditions were created manipulating audio content to be incongruent or congruent with the visual stimulus, detailed below (scripts in Appendix A). The scripts were designed to have the same degree of difficulty regarding the number of scientific concepts, abstract information, and length. Additionally, they were recorded using an auto generator text-to-audio application to avoid prosodics differences between conditions.

Before the experiment, a pretest was performed to ensure that the auditory information and DK test (see Appendix B) were equally challenging in congruent/incongruent conditions. Through an online questionnaire on Qualtrics, participants listened to one of the audios and answered eleven DK questions, followed by the second audio and the corresponding set of questions. The congruent and incongruent audio/questions were randomized within subjects, to compensate for possible learning and fatigue effects. A paired sample *t*-test showed that the congruent ($M$ = 0.69, $SD$ = 0.21) and incongruent ($M$ = 0.66, $SD$ = 0.30) responses' accuracy did not differ significantly, $t(9) = 0.55$, $p = .30$. Therefore, the level of difficulty between conditions could be considered equal.

***Congruent Stimulus***

The participants were exposed to the 3D bunny, while they were listening to audio including information about the bunny as a pet. The auditory content was designed to match and reinforce the visual environmental cues. For instance, conveying information about the bunny's behavior pertinent to a hose/ indoor habitat.

### *Incongruent Stimulus*

Participants were exposed to the same 3D bunny while listening to an audio about the bunny living in its natural habitat. The auditory content was designed to conflict with the visual environmental cues. The information conveyed was pertinent to the behaviors of a bunny living in the wild. Therefore, the audio was created to stimulate a mental representation that was mismatched with what the participant was seeing.

## Materials and Equipment

### *VR Modality*

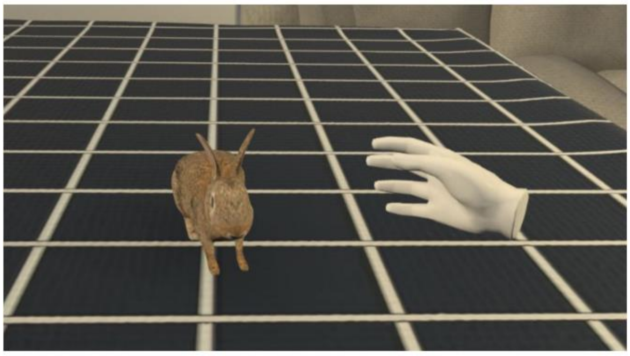
Participants visualized a virtual version of the laboratory's living room, created using Unity 3D. The bunny was displayed on the virtual table in front of the participants. As this study was interested in comparing the differences between AR and VR, the re-creation of the same environment aimed to keep the same conditions for both XR modalities. Thus, as the visual stimulus was the same, it was expected that the experiment's outcomes might have reflected the differences in technology's structural affordances (AR/VR) and audio manipulations (congruency). The bunny moved realistically and had a proportional size compared to the environment. It used an HTC Vive Pro headset with a 5k screen resolution and 120Hz refresh rate. The participants received a controller to interact with the bunny which also allowed them to pick up the animal.

**Figure 3**

*VR Virtual Living Room*

**Figure 4**

*Example of User Virtual Hand Visualized in VR Condition*



***AR Modality***

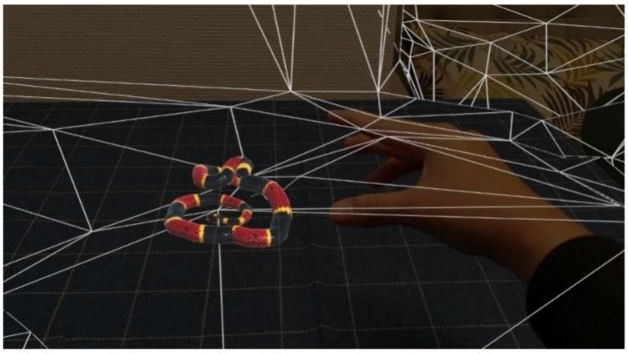
The AR condition used an AR HMD, permitting participants to see the physical space around them, see Figure 5. The bunny was virtually displayed on the physical table in front of the participants. As in the VR condition, the bunny had realistic movements and a compatible size with the environment and objects. It used a HoloLens headset with a 2k screen resolution and a 120Hz refresh rate. Participants in this condition could use and see their own hands to interact with the bunny. The participants were instructed to use only one hand, as the use of both hands could change the bunny's size, and therefore should be avoided.

**Figure 5**

*AR Lab Living Room*

**Figure 6**

*Example of User Hand Visualized in AR Condition*



Note. The darkness present in this figure was not visible during the AR experience.

**Procedure**

The experiment took place at the Tilburg University Media Design Lab. Before partaking in the experiment, participants were conducted in a room with a computer. There, they were given the experiment briefing (Appendix C) and had to complete a consent form (Appendix D). Once agreed, participants were randomly assigned to either the VR or AR room and exposed to either the congruent or incongruent condition. In the VR or AR environment, the HMD was adjusted accordingly to the participant's head. If participants had confirmed the appearance of the virtual rabbit (AR) / virtual scenario and rabbit (VR) and had no questions about the procedure, the audio started. Participants interacted with the rabbit while listening to the audio. Once the audio was done, the HMD was removed, and the participants were conducted back to the computer room. Before the second part of the questionnaire, they first were exposed to the DK test through the mouse-tracking platform. To familiarize the participants with the mouse-tracking usability, they answered four true/false questions, unrelated to the experiment (e.g., The Netherlands is in Europe). Afterward, the researcher left the room, and the participants answered the true or false questionnaire about the auditory information conveyed in the experiment (Appendix B). Once finished, the participant knocked on the door and the researcher presented the second part of the questionnaire, which consisted of questions about Psi and attention allocation. Subsequently, the participant answered the English proficiency and control questions and filled out their demographic information (full questionnaire in Appendix E). Finally, the participants were given a written debrief (Appendix F) and received half a point for taking part in the experiment.

## Measurements

### *Plausibility Illusion*

The Plausibility Illusion was measured using two items (see Table 1) of the Virtual Human Plausibility (VHP) scale (Mal et al., 2022), answered on a 7-point Likert Scale. The internal consistency between items was α = .81.

**Table 1**

*Questions Adapted from VHP by Mal et al. (2022)*

| Variable | Sub-dimension | Questions |
|---|---|---|
| Plausibility illusion | Appearance plausibility | The appearance of the rabbit seemed plausible to me as if the creature is real. |
| Plausibility illusion | Appearance plausibility | The appearance of the rabbit seemed to match well with the appearance of animals in our reality as if the creature is real. |

### *Cognitive Load (CL)*

CL was accessed by computing participant's response time (RT) and participants' mouse trajectories' temporal and spatial dynamics (MT) through a mouse-tracking software, developed by Jon Freeman. The MT and RT were recorded while they were answering the "true or false" DK test, performed immediately after the participants' experiences with AR/VR. Following the assumption of episodic buffer (Baddeley, 2000), CL in the test performance might have reflected the cognitive process in encoding the information, encompassing the CL underwent previously in AR/VR experiences. Since longer RTs and greater MTs are associated with increased CL (Rheem et al., 2018), the CL was measured by computing two variables: (1) participants' means RT and (2) a statistical test comparison of MT over time, examining the onset and timing involved in selecting an answer option.

### *Declarative Knowledge (DK)*

The DK tests were developed based on the audio content (congruent or incongruent script) conveyed during the AR/VR interaction. Alongside the DK script, the congruent and incongruent questionnaires were pretested, and the results indicated that both conditions were equally challenging (refer to the detailed results in the stimuli section). The test included a total of eleven statements that participants were required to classify as true or false, based on the audio narration they listened to while engaging with a virtual rabbit. Examples of statements included, "The rabbits belong to the rodent's order. " (Questionnaires in Appendix B).

## Control Variables

Two control variables were included to ensure that external factors didn't affect the study results. Participants' familiarity with bunnies, as well as their prior experience with VR and AR, were assessed using a brief questionnaire. Questions such as "How much do you familiar with bunnies?" and "To what extent are you familiar with AR technology?" were answered using a 7-point Likert Scale.

# Results

## Measurement Validity

Before testing the hypotheses, the validity of the measurement model was checked using PLS-SEM. According to Kock (2020), the item loadings of reflective indices that have a *p*-value below .001 ensure the suitable reliability of a measurement model in PLS-SEM. All the measured items were reflective indicators in the research model. The construct validity of measures was supported by PLS-SEM results: the item loadings of all constructs were found to be greater than .50 with the significance level below .001 (see Table 2). In addition, as presented in the measures section, the internal consistency reliability of the constructs was good as they were higher than .81.

**Table 2**

*Item Loadings of Reflective Indicators*

| P | IL | C | IL | RT | IL | MT | IL | DK | IL |
|---|---|---|---|---|---|---|---|---|---|
| P1 | .92*** | C | 1.00*** | RT | 1.00*** | MT | 1.00*** | DK | 1.00*** |
| P2 | .92*** | | | | | | | | |

*Note*. P = Plausibility Illusion, C = Congruency, RT = Response Time, MT = Mouse trajectory, DK= Declarative Knowledge.

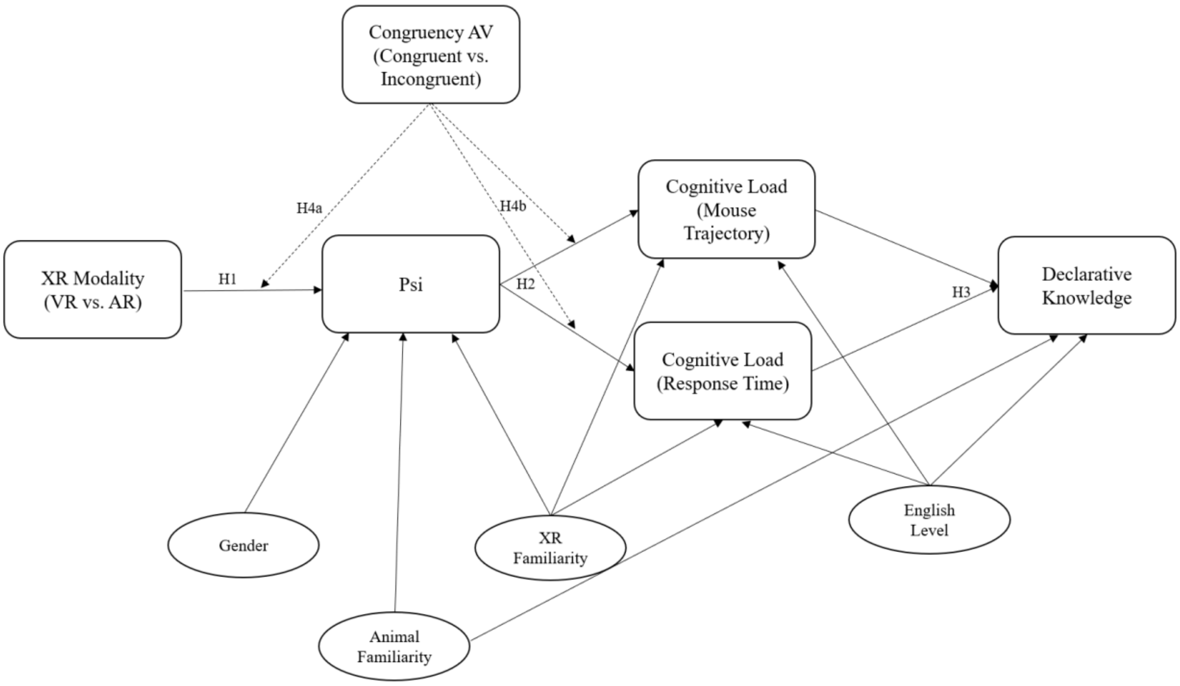
***$p < .001$

## Hypotheses Testing

PLS-SEM was conducted to validate the significance of the causal paths predicted in the structural model (see Figure 7). The summary results can be found in Figure 8.

**Figure 7**

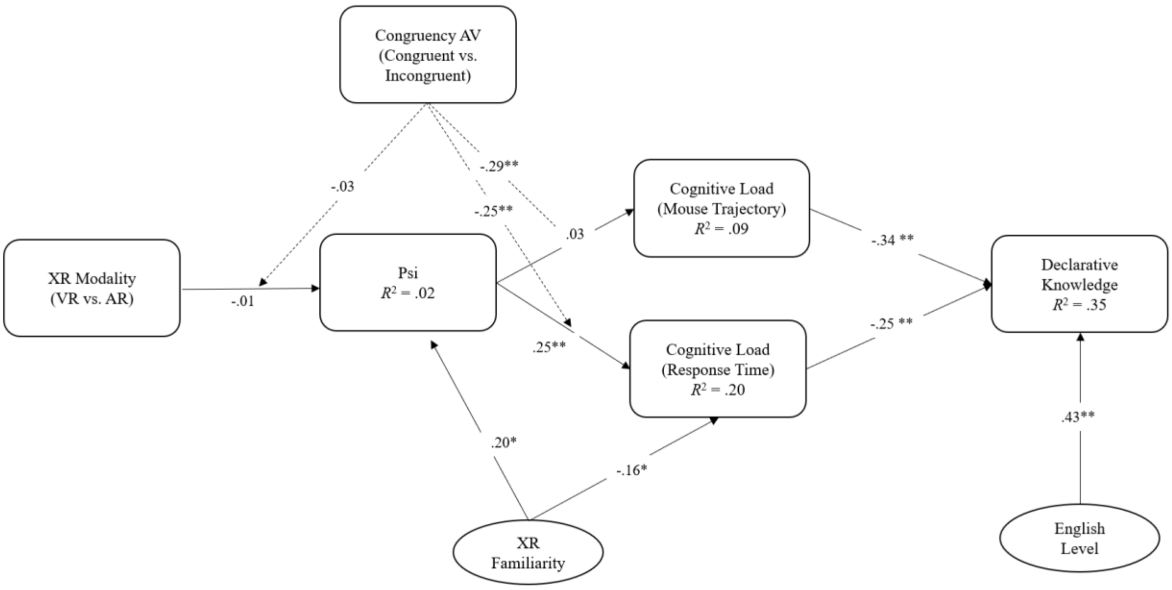
*Research Model*




*Note*. Dashed lines indicate the moderation hypotheses.

**Figure 8**

*PLS-SEM results*



*Note*. * $p < .05$, ** $p < .01$

H1 posited that participants in the VR condition would report a higher sense of *Psi compared to participants in the AR condition.* Results of PLS-SEM showed that VR modality (*M* = 5.28, *SD* = 1.28) did not significantly induce a greater sense of Psi than AR modality (*M* = 5.19, *SD* = 1.17), $\beta = -.01$, $p = .47$, $R^2 = .02$. Therefore, H1 was not supported.

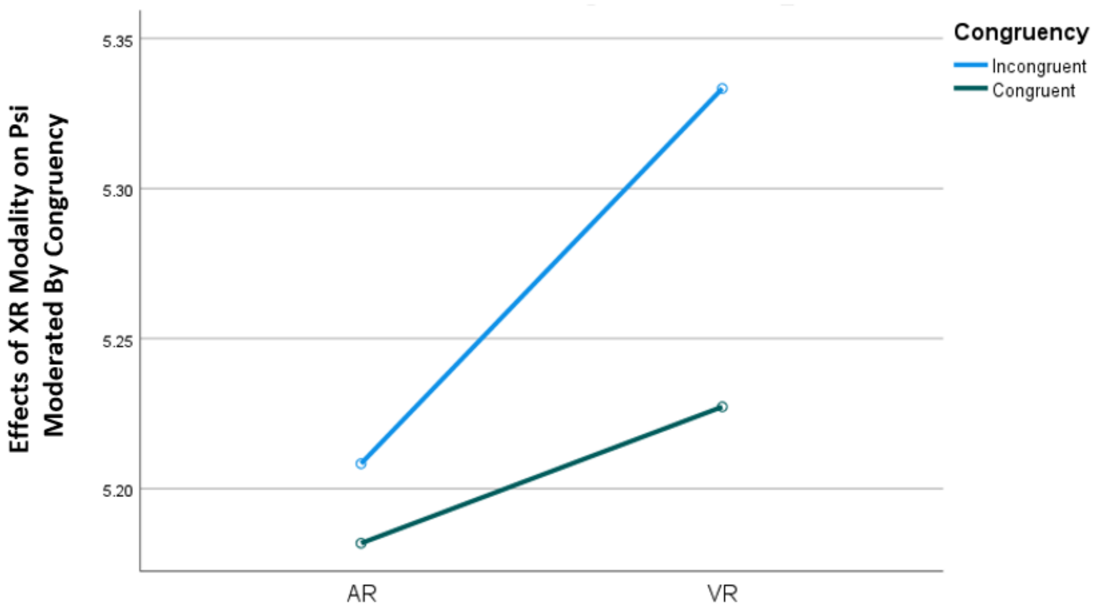
H2(a/b) predicted that experiencing a higher sense of Psi would lead to increased CL. The CL was computed by measuring the participant's a) response time (RT) and b) mouse trajectory length (MT), while they completed the declarative knowledge test. Psi had a significant direct effect on RT, suggesting that a higher sense of Psi increased participants' CL, $\beta = .25$, $p < .01$, $R^2 = .20$. However, Psi did not significantly affect participants' MT, $\beta = .03$, $p = .39$, $R^2 = .09$. Thus, H2 was partially supported by the RT variable.

H3(a/b) predicted that a) RT and b) MT will be negatively associated with DK. As predicted, RT and MT were found to significantly decrease the DK score, $\beta = -.25$, $p < .01$, and, $\beta = -.34$, $p < .01$, respectively. In total, 35% of the variance in the DK score was explained by the antecedent variables ($R^2 = .35$). Hence, H3 was supported.

Finally, H4(a/b/c) predicted that congruency would moderate the influence of XR modality technology on Psi and, Psi on CL, respectively. Specifically, H4a posited that the congruency between audio and video stimuli would enhance the effects of XR modality on Psi, such that the relationship between XR modality and Psi would get stronger in the AR condition. On the other hand, H4(b/c) posited that the congruency between audio and video stimuli would weaken the relationship between Psi and b) MT and c) RT, such that Psi would affect CL in a lesser degree when the audio and video stimuli were congruent. Inconsistent with hypotheses H4a, results showed that congruency did not significantly moderate the effect of XR modality on Psi, $\beta = -.03$, $p = .38$, thereby not supporting H4a (see Figure 9).

**Figure 9**

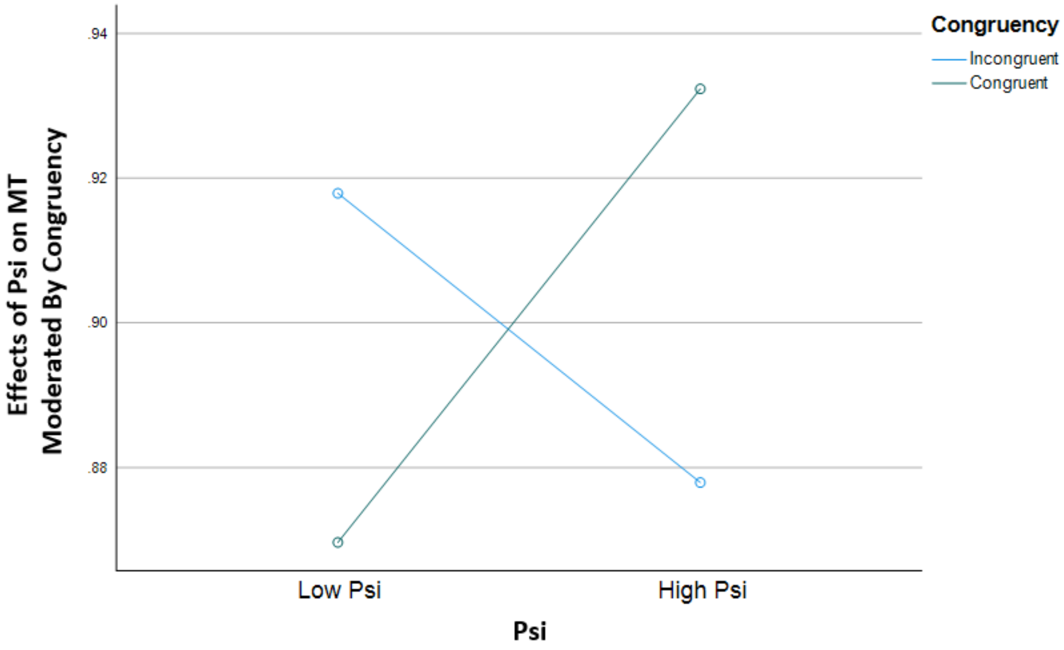
*The Moderating Effects of Congruency on the Relationship Between XR Modalities and Psi*




Regarding hypothesis H4(b/c), the results suggested that congruency significantly moderated the effects of Psi on both b) MT and c) RT, $\beta$ = -.29 $p$ < .01, $R^2$ = .09, and $\beta$ = -.25, $p$ < .01, $R^2$ = .20, respectively. However, contrary to H4(b/c), the results for MT and RT indicated an unexpected pattern of congruency moderation. For participants reporting a high sense of Psi, MT decreased when audio and video stimuli were incongruent, compared to the participants' MT in the congruent condition. Conversely, participants reporting a low sense of Psi displayed decreased MT in the congruent condition, compared to those exposed to incongruent audio and visual stimuli. Thus, the results suggested that congruence strengthened the effects of Psi on MT. Specifically, the congruence might have contributed to a reduced MT only for participants who experienced a low sense of Psi, whereas for participants experiencing a high sense of Psi, the congruence between audio and visual stimulus led to an increased MT (see Figure 10).

**Figure 10**

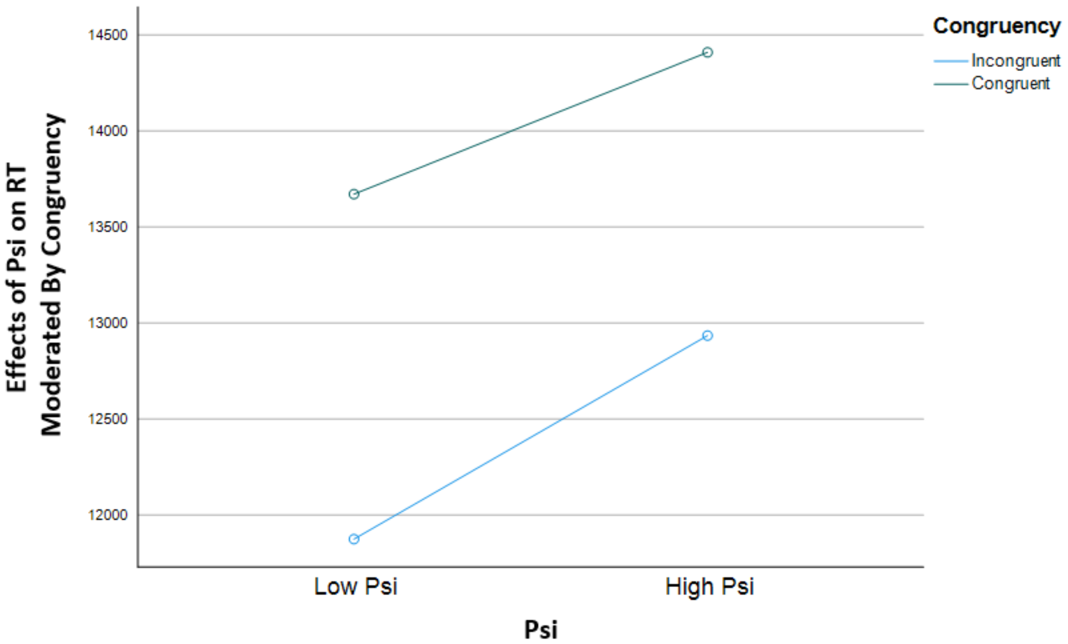
*The Moderating Effects of Congruency on The Relationship Between Psi And CL: MT*




The RT results also indicated that congruency strengthened the effects of Psi on CL, even for participants experiencing a low sense of Psi. When the audio and video stimuli were congruent, the RT was longer for all participants (see Figure 11). Therefore, H4(b/c) was not supported as the effects of congruency moderation were contrary to this thesis prediction.

**Figure 11**

*The Moderating Effects of Congruency on The Relationship Between Psi And CL: RT*




**Control Variables**

Gender, familiarity with rabbits, and familiarity with XR technologies were included as control variables in the statistical model. The results suggested that technology familiarity, $\beta = .20$, $p = .02$, had a significant influence on Psi. However, gender, $\beta = -.12$, $p = .13$, and rabbit familiarity, $\beta = .15$, $p = .06$, did not have a significant effect. Technology familiarity was also used as a control variable to the CL outcome, in addition to the participant's level of English. There was a significant negative effect of Technology-Familiarity on CL measured through RT, $\beta = -.16$, $p = .05$. Nonetheless, CL measured through MT, $\beta = .11$, $p = .14$ and English level, $\beta = .06$, $p = .27$, did not have a significant effect. Finally, the DK was controlled for English level and rabbit familiarity. The participant's level of English, $\beta = .43$, $p < .01$, was the only significant control variable on DK outcome. Rabbit familiarity, $\beta = .12$, $p = .13$, had no significant effect. Figure 3 presents the summary of the PLS-SEM results.

**Model Fit**

The research model's goodness of fit was verified by examining the statistics such as average path coefficient (APC), average R-squared (ARS), average variance inflation factor

(AVIF), and average full collinearity VIF (AFVIF) calculated by WarpPLS 8.0 (Kock, 2020). As established by Kock (2020), the validity of a structural model can be guaranteed when APC and ARS have a significance level below .05 ($p < .05$) and AVIF and AFVIF values are equal to or lower than 3.3. Results from PLS-SEM supported the validity of the research model: APC = .17 $p$ = .02; ARS = .17, $p$ = .02; AVIF = 1.59; and AFVIF = 1.21.

## Discussions

The findings revealed that the XR modalities did not have a significant influence on Psi. However, Psi had a positive direct effect on RT (such as experiencing a high sense of Psi increased participants' RT), whereas no significant effect of Psi was observed on MT. Both MT and RT had a significant negative effect on DK. Although congruency did not significantly moderate the relationship between XR modality and Psi, congruency significantly affected the relationship between Psi and RT/MT. However, contradicting the expectation, congruency strengthened the relationship between Psi and RT, whereas the MT variable indicated that congruency reduced the effects of Psi on MT only among low-Psi participants. The detailed discussion follows below.

The prediction that VR would evoke a higher sense of Psi compared to AR was not supported by the results. The lack of significant effect can be a consequence of the experiment's length. Since the study aimed to explore CL and its impact on DK, participants' interaction with XR modalities was deliberately kept short, lasting less than 3 minutes. This decision assumed that task performance reduces over time, and longer tasks are more susceptible to being influenced by motivation and individual differences, such as attentional vigilance capacity (Brünken et al., 2010; Hockey, 1997). Thus, the experiment was designed to be brief, controlling for these factors. Still, the limited duration may have hindered the sense of Psi in the VR condition, as participants did not have enough time to become accustomed to the technology. Specifically, the short time for participants to familiarize themselves with the controller might have diminished the perception of body ownership in VR conditions. Existing literature suggests that body representation plays a crucial role in eliciting Psi in VR experiences (Skarbez, 2016; Skarbez et al., 2017). Conversely, in AR, the more naturalistic interaction (where participants could interact with the rabbit using their own hands) might have contributed to a sense of Psi despite the short duration. Therefore, it is plausible that the experimental conditions suppressed the potential effect of VR on Psi. This explanation aligns well with the observed effect of the "XR-familiarity" control variable. The results indicated that XR familiarity significantly influenced Psi. Hence, it is possible that the lack of familiarity, particularly in VR conditions (which requires understanding how to control virtual hands), may have masked the distinctive effects of VR within the limited interaction timeframe. Future research can investigate the relationship between familiarity/ time of exposition, and the perception of body ownership in VR experiences to ascertain this assumption. Additionally, it would be interesting to explore how distinct interaction opportunities influence body ownership and the sense of Psi. For example, future studies can compare VR technologies like Leap Motion (which allows users to use their own hands) with technologies that make use of controllers to examine further the relationship between body representation and Psi and the possible moderator effect of technology familiarity.

The influence of Psi on CL was partially supported by the RT data, as high-Psi participants displayed longer RT. However, no significant effect of Psi was found on the MT variable. Individual differences may have influenced this outcome. Mouse-tracking, as a method for measuring CL, is grounded in the concept that cognitive processing continually affects motor movements - including mouse movements. (Spivey & Dale, 2006). MT provides insights into the activation level of a specific response during the decision-making process. Generally, a minimal deviation toward an option indicates a higher degree of activation for that option (Schoemann et al., 2019), and consequently, it suggests a lower demand for cognitive resources (Rheem et al., 2018). Though, a study conducted by Szaszi et al. (2018) proposed that individual differences in the decision-making process may influence mouse-movement behavior. According to them, MT may reflect individuals' distinct thinking dynamics, such as the adoption of heuristic strategies or revising the choices before a final answer is given (Szaszi et al., 2018). Thus, the study by Szaszi et al. (2018) suggests that MT may have been influenced by individual differences, particularly regarding low/high reasoning capacities. In the context of this experiment, it is possible that low-reasoning participants, facing a high cognitive effort, adopted heuristic strategies. This might have resulted in less deviation in their MT, despite experiencing high Psi and CL. In future studies, it would be valuable to address these individual differences by including the measurement of the MT baseline to control for personal decision-making differences.

Finally, consistent with prior research suggesting negative effects of CL on learning (Moreno & Mayer, 2000; Paas et al., 2003), the negative effect of CL on DK was supported by MT and RT variables (35% of the variance in the DK was explained by the antecedent variables). Additionally, participants' English level also had a significant effect on DK outcomes. The participants of the present study were students from Tilburg University and most of them are not English native speakers.

Regarding the moderating role of congruency, the results showed that congruency did not influence the relationship between XR modality and Psi. As discussed previously, it's possible that the timeframe of the experiment didn't permit that the distinctive affordances of VR contributed to a higher sense of Psi. It might be that participants' familiarity with XR experiences, and the experimental conditions also influenced the potential moderating effect of congruency.

The second moderation hypothesis examined the effect of congruency on the relationship between Psi and CL. Surprisingly, the results showed a significant but unexpected effect of congruency. Drawing upon the principles of multimodal design (Mayer & Fiorella, 2021), it was anticipated that when participants listened to audio information congruent with the visual stimulus, the audio-video information would be integrated into a unified mental representation. This integration was expected to improve the processing of Psi and the audio-conveyed, weakening the relationship between Psi and extraneous CL.

However, the results indicated the opposite effect in the RT variable, with congruency strengthening the effects of Psi on RT. Sweller (1994) emphasizes that the combination of audio-visual elements plays a crucial role in determining extraneous and germane CL. While extraneous CL relates to the effort required to process poorly designed instruction, germane CL contributes to the construction of schemas and beneficial mental representations (Martin, 2014). Effective instructional designs aim to minimize extraneous CL and enhance germane CL, as long as the overall CL remains within acceptable limits (Martin, 2014). In the present experiment, it's possible that the congruence between audio-video may have facilitated the processing of Psi, implicating reduced effects of Psi on extraneous CL. Consequently, by reducing the cognitive burden to process Psi, participants may have been able to process the information deeper, creating meaningful connections and schemas. However, this deeper processing of information likely increased the germane CL, leading to a higher overall CL. Hence, the longer RT observed in the congruent condition may have paradoxically stemmed from the cognitive availability which allowed participants to further process and deeply encode the information. Conducting future studies that differentiate between the types of CL in XR experiences can shed light on whether congruence between audio-visual stimuli truly facilitates schematic formation or merely represents extraneous CL. On the other hand, in the incongruent condition, the mismatch between audio and video likely hindered the integration of multimodalities, disrupting the process of Psi. Whereas audio-video incongruency might have strengthened the impact of Psi on extraneous CL, this overwhelming processing of Psi may have inhibited the deeper processing of information and schematic formation. Thus, even with augmented extraneous CL from additional effort to process Psi, the overall cognitive load in the incongruent manipulation might have been reduced by the deficiency of germane CL.

The MT results in the congruent condition yielded the same pattern as the RT variable, where the congruency strengthened the relationship between Psi and CL. Accordingly, it is possible that the germane CL, led by the cognitive availability from reduced extraneous CL, was responsible for the observed augmented overall CL. The most intriguing MT pattern emerged in the incongruent condition. Low-Psi participants exhibited higher MT compared to high-Psi participants, contradicting the expected positive effect of Psi on CL. One possible explanation for this pattern is the individual differences in processing information from incongruent multimodal stimuli. Zilka and Bonneh (2022) found that the tendency to bind conflicting multisensory information is an individual trait. Therefore, the moderation role of congruency might have influenced the relationship between Psi and CL distinctively, depending on individuals' tendencies to bind conflicting information when exposed to incongruent stimuli. Specifically, participants with high binding tendencies might have tried to make sense of the multimodal information, leading to a reduced sense of Psi. At the same time, this attempt in creating a coherent representation from incongruent audio-visual elements might have implicated a great burden on cognitive resources. On the other hand, those with lower-binding tendencies likely disregarded the audio information and relied solely on visual cues to perceive virtual elements as realistic, facilitating the Psi and reducing the demands on cognitive resources. Thus, for individuals with lower tendencies to bind conflicting information, congruency weakened the relationship between Psi and CL, whereas this relationship was strengthened for individuals with higher tendencies to bind multisensory information. In both cases, the incongruency in audio-video may have inhibited the formation of schemas at a higher level of information processing. As a result, the overall MT in the incongruent condition might have mainly reflected the moderation effect of congruency between Psi and extraneous CL, rather than Psi and germane CL. Considering the importance of germane CL to learning outcomes, this explanation becomes plausible when considering that the MT variable had a greater negative impact on DK, compared to RT. Future research, investigating audio-visual binding in XR experiences and different types of CL can potentially provide further evidence on how individuals are influenced by mismatches in XR experiences and the impacts on the relationship between Psi and germane/extraneous CL.

**Implications**

The findings of this study contributed to the existing literature on XR by examining how VR and AR structural affordances induce a sense of Psi, the cognitive demands involved in these experiences, as well as the impacts of CL on DK. Although the study did not find a significant effect of XR modality on Psi, it is noteworthy that familiarity with XR had a significant positive effect on Psi and a negative effect on CL. Since one of the main challenges in XR experiences is providing a sense of reality without overwhelming cognitive capacities, familiarity should be taken into consideration when utilizing VR and AR as educational tools. For instance, it might be beneficial to expose students to entertaining XR content before introducing instructional material through these technologies. While this may seem intuitive, it is important not to underestimate the role of familiarity when employing mediated technologies as learning environments, especially considering XR particularities and their high demands on cognitive resources. Additionally, while participants in the AR condition had the opportunity to interact with the virtual rabbit using their own hands, the use of a controller in VR might have potentially undermined the sense of body representation, affecting Psi. Future studies can delve into how different opportunities of interaction and body ownership impact Psi and affect CL.

This study provided further evidence of the relationship between the perceived sense of immersion - utilizing a more recent construct (Psi) - and CL. Intriguingly, the analysis of the CL through MT and RT mouse records yielded different results. Previous research examining individual differences suggest that mouse behavior might vary across individuals concerning reasoning capacities during the decision-making process. This finding is interesting because, although individual differences are addressed in learning literature, the cognitively demanding nature of VR and AR may amplify the significance of individual differences in XR educational contexts. Therefore, future research on XR should explore the individual differences that are known to impact learning outcomes. It is worth noting that not measuring the MT baseline was a limitation of this experiment, and future studies utilizing mouse-tracking techniques might include this measure to control for individual differences.

Over and above, the present research addressed the content to try to gain a more nuanced understanding of information processing and learning in XR experiences. The results suggested that congruence between audio and video significantly moderated the effects of Psi on CL. Although, contrary to what was expected, the congruency moderation strengthened this relationship. The CL theory seeks to elucidate how multimedia learning design principles influence different types of CL and it's possible that the results paradoxically reflected increased CL from enhanced information processing. Specifically, the congruency likely reduced the extraneous CL associated with processing Psi. This cognitive availability might have allowed deeper information processing, which conversely, led to an increment in the overall CL. However, in the present research, it is not possible to distinguish what type of CL the RT and MT reflected. Implementing real-time brain measurements to monitor germane and extraneous CL may be paramount in gaining valuable insights into the cognitive processes in XR educational interventions.

Lastly, the inconsistencies in RT and MT also suggested that individual differences might also influence the relationship between Psi and CL, considering audio-video congruency and individual distinctive tendencies to bind conflicting stimuli. Future research can explore the implication of individual differences in processing conflicting stimuli in educational XR mediations to further the understanding of the relationship between Psi and CL.

Finally, it is important to highlight that a significant portion of the participants consisted of individuals from Western, Educated, Industrialised, Rich, and Democratic (WEIRD) backgrounds. Therefore, the generalization of these findings to the population at large should be done with caution (Henrich et al., 2010).

## Limitations and Future Directions

The are some methodological limitations in the present study that should be considered in future exploration. The length of the experiment and the less naturalistic interaction in VR (which required the use of a controller) might have reduced the potential distinctive sense of Psi in VR conditions. Future studies could address this limitation by using technologies such as Leap Motion, which allows VR users to engage with virtual objects using their own hands. Similarly, the AR condition also had important technical limitations. For instance, when participants lifted the rabbit from the table, it floated in the air, infringing the physics principles of gravity. Additionally, AR participants visualized the environment with scanner lines, which might have reduced the perceived realism. Finally, although participants used their own hands to interact with the virtual rabbit, the interaction was not fully naturalistic, as they needed to pinch the rabbit to move and grab it. Thus, in AR, neither the hand movement nor the way the rabbit behaved upon release was coherent with participants' real-life expectations, possibly influencing the Psi outcomes as well. Thus, to explore the impacts of XR's different structural affordances, it is important to investigate the impact of technological limitations and distinctive interaction opportunities on users' sense of Psi.

The present study attempted to provide a more objective measurement of CL through mouse-tracking, however not including the MT baseline was a limitation of the experimental design. Future investigations applying mouse-tracking techniques should include the MT baseline to control for individual differences (e.g., decision-making strategies) that can influence mouse-movement behavior. Moreover, even though mouse-tracking is a reliable assessment of CL, it is limited in distinguishing between extraneous and germane CL. This limitation hindered the understanding of the effect of congruency in the relationship between Psi and CL. Future research can clarify how XR users experience different types of CL implementing real-time brain measurements. Distinguishing different types of CL might elucidate the cognitive processes underlying XR experiences and clarify whether congruency effects and other design principles from traditional learning can benefit XR educational mediations. Lastly, the inconsistencies in RT and MT also suggested that individual binding tendencies might have influenced how congruency moderated the effects of Psi on CL. To further understand the implication of congruency in educational XR mediation, it’s important to expand the notions of idiosyncratic tendencies to bind conflicting sensory information and how it influences the relationship between Psi and CL.

Another limitation of this study was the utilization of English for audio information and the DK test among participants who are not native English speakers. As observed, the English level had a significant effect on the DK outcomes. In future studies, it might be important to avoid factors that might influence DK results, restricting only variables that are pertinent to test the research hypothesis. Finally, considering that most of the participants consisted of individuals from Western, Educated, Industrialised, Rich, and Democratic

(WEIRD) backgrounds, the generalization of the results to the wider population might be done with caution (Henrich et al., 2010).

## Conclusion

The present study has added to the XR literature by examining the AR and VR influence on perceived sense of immersion (Psi), the cognitive demands involved in these experiences, and DK outcomes. This study provided unique contributions by comparing the effects of AR and VR technologies and by investigating the influence of instruction design on the underlying psychological mechanisms at play in these experiences. Furthermore, this paper introduced a novel metric to assess the perception of realism (Psi) and utilized mouse-tracking records to provide a less subjective measurement of CL. Although no significant effect of XR modality on Psi was found, the results suggested that technology familiarity might be an important factor to consider when applying XR for educational purposes. Additionally, the inconsistencies found in mouse-tracking variables called attention to the influence of individual differences in XR experiences.

In line with previous research, the results of this study emphasized the importance of balancing CL in the use of XR as an educational tool (Ahn et al., 2022; Buchner et al., 2021; Han et al., 2021). Understanding how audio-visual material can be better designed to provide immersion, while it keeps the cognitive demands at an acceptable level, remains challenging. The unexpected pattern of congruency moderation in the relationship between Psi and CL suggested that the increased CL could be related to a beneficial increased germane CL. Thus, discriminating different types of CL might be paramount to investigating the value of XR experiences for education. Finally, individual tendencies in dealing with conflicting sensory information might also expand the knowledge about the role of congruency in XR educational design, as multisensory binding tendencies might affect Psi and its effects on CL.

In conclusion, considering the complexity of XR experiences and the significant impact of Psi on CL, discriminating extraneous and germane CL might be the key to getting clear insights into the effects of modality, the relationship between Psi and CL, and how immersion and cognition trade-offs can be addressed in XR instructional design. Conducting more investigations into the material design and the impacts of individual differences can provide valuable insights into the challenges of utilizing XR technologies as educational tools.

**References**

Ahn, S. S., Nowak, K. L., & Bailenson, J. N. (2022). Unintended consequences of spatial presence on learning in virtual reality. *Computers & Education*, *186*, 104532. https://doi.org/10.1016/j.compedu.2022.104532

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, *20*, 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

Anderson, J. R. (2013). The Architecture of Cognition. In *Psychology Press eBooks*. https://doi.org/10.4324/9781315799438

Antonioli, M., Blake, C., & Sparks, K. M. (2014). <p>Augmented Reality Applications in Education</p> *The Journal of Technology Studies*, *40*(2), 96–107. https://doi.org/10.21061/jots.v40i2.a.4

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, *21*(6), 34–47. https://doi.org/10.1109/38.963459

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Barfield, W., Hendrix, C. M., Bjorneseth, O., Kaczmarek, K. A., & Lotens, W. (1995). Comparison of Human Sensory Capabilities with Technical Specifications of Virtual Environment Equipment. *Presence: Teleoperators & Virtual Environments*, *4*(4), 329–356. https://doi.org/10.1162/pres.1995.4.4.329

Biocca, F., Kim, J., & Choi, Y. H. (2001). Visual Touch in Virtual Environments: An Exploratory Study of Presence, Multimodal Interfaces, and Cross-Modal Sensory

Illusions. *Presence: Teleoperators & Virtual Environments*, *10*(3), 247–265. https://doi.org/10.1162/105474601300343595

Bower, M., Howe, C., McCredie, N., Robinson, A. T., & Grover, D. (2014). Augmented Reality in education – cases, places and potentials. *Educational Media International*, *51*(1), 1–15. https://doi.org/10.1080/09523987.2014.889400

Bowman, D. A., & Hodges, L. F. (1997). *An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments*. https://doi.org/10.1145/253284.253301

Bowman, D. A., Kruijff, E., LaViola, J. J., & Poupyrev, I. (2001). An Introduction to 3-D User Interface Design. *Presence: Teleoperators & Virtual Environments*, *10*(1), 96–108. https://doi.org/10.1162/105474601750182342

Brünken, R., Plass, J. L., & Moreno, R. (2010). Current Issues and Open Questions in Cognitive Load Research. In *Cambridge University Press eBooks* (pp. 253–272). https://doi.org/10.1017/cbo9780511844744.014

Buchner, J., Buntins, K., & Kerres, M. (2021). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), 285–303. https://doi.org/10.1111/jcal.12617

*CEF levels*. (n.d.). LearnEnglish Teens. https://learnenglishteens.britishcouncil.org/cef-levels

Chandler, P., & Sweller, J. (1992). THE SPLIT-ATTENTION EFFECT AS A FACTOR IN THE DESIGN OF INSTRUCTION. *British Journal of Educational Psychology*, *62*(2), 233–246. https://doi.org/10.1111/j.2044-8279.1992.tb01017.x

Cheng, K., & Tsai, C. C. (2013). Affordances of Augmented Reality in Science Learning: Suggestions for Future Research. *Journal of Science Education and Technology*, *22*(4), 449–462. https://doi.org/10.1007/s10956-012-9405-9

Craik, F. I. M., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/s0022-5371(72)80001-x

Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, *1242*, 136–150. https://doi.org/10.1016/j.brainres.2008.03.071

Drascic, D., & Milgram, P. (1996). <title>Perceptual issues in augmented reality</title> In *Proceedings of SPIE*. SPIE. https://doi.org/10.1117/12.237425

Dunleavy, M. (2014). Design Principles for Augmented Reality Learning. *TechTrends*, *58*(1), 28–34. https://doi.org/10.1007/s11528-013-0717-2

Dunleavy, M., & Dede, C. (2014). Augmented Reality Teaching and Learning. In *Springer eBooks* (pp. 735–745). https://doi.org/10.1007/978-1-4614-3185-5_59

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/bf03203267

Gonçalves, G., Coelho, A., Barbosa, L. S., Vasconcelos-Raposo, J., & Bessa, M. (2021). Evaluation of the impact of different levels of self-representation and body tracking on the sense of presence and embodiment in immersive VR. *Virtual Reality*, *26*(1), 1–14. https://doi.org/10.1007/s10055-021-00530-5

Gottfried, J. A., & Dolan, R. J. (2003). The Nose Smells What the Eye Sees. *Neuron*, *39*(2), 375–386. https://doi.org/10.1016/s0896-6273(03)00392-1

Han, J., Zheng, Q., & Yunan, D. (2021). Lost in Virtual Reality? Cognitive Load in High Immersive VR Environments. *Journal of Advances in Information Technology*, *12*(4). https://doi.org/10.12720/jait.12.4.302-310

Hehman, E., Stolier, R. M., & Freeman, J. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, *18*(3), 384–401. https://doi.org/10.1177/1368430214538325

Henrich, J., Heine, S., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29. https://doi.org/10.1038/466029a

Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, *45*(1–3), 73–93. https://doi.org/10.1016/s0301-0511(96)05223-4

Holloway, R. H. (1997). Registration Error Analysis for Augmented Reality. *Presence: Teleoperators & Virtual Environments*, *6*(4), 413–432. https://doi.org/10.1162/pres.1997.6.4.413

Huang, K., Ball, C. T., Francis, J., Ratan, R. A., Boumis, J. K., & Fordham, J. (2019). Augmented Versus Virtual Reality in Education: An Exploratory Study Examining Science Knowledge Retention When Using Augmented Reality/Virtual Reality Mobile Applications. *Cyberpsychology, Behavior, and Social Networking*, *22*(2), 105–110. https://doi.org/10.1089/cyber.2018.0150

Iwamiya, S. (2013). Perceived congruence between auditory and visual elements in multimedia. In *Oxford University Press eBooks* (pp. 141–164). https://doi.org/10.1093/acprof:oso/9780199608157.003.0007

Kalyuga, S., & Singh, A. (2016). Rethinking the Boundaries of Cognitive Load Theory in Complex Learning. *Educational Psychology Review*, *28*(4), 831–852. https://doi.org/10.1007/s10648-015-9352-0

Lang, A., Bradley, S. D., Park, B., Shin, M., & Chung, Y. (2006). Parsing the Resource Pie: Using STRTs to Measure Attention to Mediated Messages. *Media Psychology*, *8*(4), 369–394. https://doi.org/10.1207/s1532785xmep0804_3

Latoschik, M. E., & Wienrich, C. (2022). Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach. *Frontiers in Virtual Reality*, *3*. https://doi.org/10.3389/frvir.2022.694433

Leppink, J., Paas, F., Van Der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Lombard, M., & Ditton, T. B. (2006). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), 0. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Mal, D., Wolf, E. J., Döllinger, N., Botsch, M., Wienrich, C., & Latoschik, M. E. (2022). Virtual Human Coherence and Plausibility – Towards a Validated Scale. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. https://doi.org/10.1109/vrw55335.2022.00245

Marks, L. B. (1978). The Unity of the Senses. In *Elsevier eBooks* (pp. 181–192). https://doi.org/10.1016/b978-0-12-472960-5.50011-1

Martin, S. G. (2014). Measuring cognitive load and cognition: metrics for technology-enhanced learning. *Educational Research and Evaluation*, *20*(7–8), 592–621. https://doi.org/10.1080/13803611.2014.997140

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In *Cambridge University Press eBooks* (pp. 31–48). https://doi.org/10.1017/cbo9780511816819.004

Mayer, R. E. (2006). The Cambridge handbook of multimedia learning. *Choice Reviews Online*, *43*(10), 43–5643. https://doi.org/10.5860/choice.43-5643

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In *Cambridge University Press eBooks* (pp. 43–71). https://doi.org/10.1017/cbo9781139547369.005

Mayer, R. E., & Fiorella, L. (2021). *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, *70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Milgram, P., & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Information and Systems*, *77*(12), 1321–1329. https://cs.gmu.edu/~zduric/cs499/Readings/r76JBo-Milgram_IEICE_1994.pdf

Minsky, M. (1980). Telepresence. *OMNI Magazine, 44-52.* https://philpapers.org/rec/MINT

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, *92*(1), 117–125. https://doi.org/10.1037/0022-0663.92.1.117

Nilsson, N. C., Nordahl, R., & Serafin, S. (2016). Immersion Revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, *12*(2), 108–134. https://doi.org/10.17011/ht/urn.201611174652

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/s15326985ep3801_1

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, *32*(1/2), 1–8. https://doi.org/10.1023/b:truc.0000021806.17516.d0

Paas, F., & Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In *Cambridge University Press eBooks* (pp. 27–42). https://doi.org/10.1017/cbo9781139547369.004

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987. https://doi.org/10.1038/nrn2277

Rheem, H., Verma, V., & Becker, D. V. (2018). Use of Mouse-tracking Method to Measure Cognitive Load. *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting*, *62*(1), 1982–1986. https://doi.org/10.1177/1541931218621449

Schnotz, W., Böckheler, J. F. N., & Grzondziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education*, *14*(2), 245–265. https://doi.org/10.1007/bf03172968

Schoemann, M., Lüken, M., Grage, T., Kieslich, P. J., & Scherbaum, S. (2019). Validating mouse-tracking: How design factors influence action dynamics in intertemporal decision making. *Behavior Research Methods*, *51*(5), 2356–2377. https://doi.org/10.3758/s13428-018-1179-4

Skarbez, R. (2016). Plausibility illusion in virtual environments. *(Doctoral Dissertation, the University of North Carolina at Chapel Hill).*

Skarbez, R., Brooks, F. P., & Whitton, M. C. (2017). A Survey of Presence and Related Concepts. *ACM Computing Surveys*, *50*(6), 1–39. https://doi.org/10.1145/3134301

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B, 364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Usoh, M., & Steed, A. (1994). Depth of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, *3*(2), 130–144. https://doi.org/10.1162/pres.1994.3.2.130

Slater, M., & Wilbur, S. (1997). A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators & Virtual Environments*, *6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Spivey, M. J., & Dale, R. (2006). Continuous Dynamics in Real-Time Cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

Steffen, J. H., Gaskin, J., Meservy, T. O., Jenkins, J. L., & Wolman, I. (2019). Framework of Affordances for Virtual Reality and Augmented Reality. *Journal of Management Information Systems*, *36*(3), 683–729. https://doi.org/10.1080/07421222.2019.1628877

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral Indices of Multisensory Integration: Orientation to Visual Cues is Affected by Auditory Stimuli. *Journal of Cognitive Neuroscience*, *1*(1), 12–24. https://doi.org/10.1162/jocn.1989.1.1.12

Stroop, J. R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Szaszi, B., Palfi, B., Aczel, B., Kieslich, P. J., & Aczel, B. (2018). Thinking dynamics and individual differences: Mouse-tracking analysis of the denominator neglect task. *Judgment and Decision Making*, *13*(1), 23–32. https://doi.org/10.1017/s1930297500008792

Van Krevelen, D. W. F., & Poelman, R. (2010). A Survey of Augmented Reality Technologies, Applications and Limitations. *International Journal of Virtual Realit*, *9*(2), 1–20. https://doi.org/10.20870/ijvr.2010.9.2.2767

Van Veen, V., & Carter, C. S. (2002). The Timing of Action-Monitoring Processes in the Anterior Cingulate Cortex. *Journal of Cognitive Neuroscience*, *14*(4), 593–602. https://doi.org/10.1162/08989290260045837

Wallen, E. S., Plass, J. L., & Brünken, R. (2005). The function of annotations in the comprehension of scientific texts: Cognitive load effects and the impact of verbal ability. *Educational Technology Research and Development*, *53*(3), 59–71. https://doi.org/10.1007/bf02504798

Witmer, B. G., & Singer, M. B. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators & Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wu, H. K., Lee, S., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41–49. https://doi.org/10.1016/j.compedu.2012.10.024

Zhang, H., Cui, Y., Shan, H., Qu, Z., Zhang, W., Tu, L., & Wang, Y. (2020). *Hotspots and Trends of Virtual Reality, Augmented Reality and Mixed Reality in Education Field*. https://doi.org/10.23919/ilrn47897.2020.9155170

Zilka, R., & Bonneh, Y. (2022). Individual differences in audio-visual binding can predict the varied severity of motion sickness. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2022.05.09.491170

# Appendices

## Appendix A

## Audio Scripts

### Congruent Condition: Rabbits as Pets (423 words)

Rabbits are a popular choice for pets, known for their cute appearance, and low-maintenance care requirements. They are usually kept indoors, and some owners may allow their bunnies to roam free in the house. Rabbits make wonderful indoor companions - in part because they can be litter box trained, just like cats! By nature, rabbits choose one or a few places (usually corners) to deposit their urine and most of their poop. Poop-training requires only giving them a litter box that they know will not be invaded by others. Caring for a bunny requires attention and dedication. They can live for 5 to 12 years, depending on their environment and breed. In contrast to what most of us might think, rabbits are not rodents; they are lagomorph mammals, and they are also herbivores. Therefore, they should never be without vegetables, and pellets, and they require plenty of clean water to drink. Bunnies need regular exercise to stay healthy, and their living space should be spacious enough to allow them to hop, jump, and play. Domestic rabbits come in various sizes and breeds, with some weighting as little as a pound, while others can weight up to 20 pounds s. They have long ears that they can move in different directions, sharp claws for digging, and powerful hind legs that allow them to hop and jump. Rabbits have several unique behaviors, including binkying, where they jump and twist in the air, and flopping, where they suddenly collapse onto their side. These behaviors are believed to be signs of happiness and relaxation. They can form strong bonds with their owners and may even show affection through nudging, licking, or purring. Another curious behavior is that rabbits eat their own poop. This behavior is known as coprophagy, and is essential for a rabbit's health, as this unique digestive system allows them to extract maximum nutrition from plant material. Rabbits have a cecum - a specialized pouch in their digestive tract, where food is fermented and broken down by bacteria, and then passed through the intestines to be expelled as fecal pellets. These pellets are then re-ingested by the rabbit to extract additional nutrients. It's important to clean a rabbit's living space regularly to prevent the accumulation of fecal pellets, as well as to ensure that the rabbit is not re-ingesting contaminated feces. Rabbits make great pets for those willing to provide the necessary care and attention. They are unique creatures with interesting behaviors and characteristics, and their presence can bring joy and companionship to many households.

**Incongruent Condition: Wild Rabbits (425 words)**

Rabbits are fascinating creatures that play an important role in many ecosystems. Wild rabbits are ground dwellers that live in environments, ranging from deserts to tropical forests, and grasslands. The European rabbit occupies open landscapes such as fields, parks, and gardens, where they create their own homes by tunneling into the ground. These tunnel systems are called warrens and include rooms for nesting and sleeping. They also have multiple entrances for a quick escape. Warrens can be as deep as 3 meters underground. In contrast to what most of us might think, rabbits are not rodents; they are lagomorph mammals, and they are also herbivores. Therefore, in the wild, rabbits primarily feed on vegetation, such as grasses and weeds. Wild rabbits vary in size, depending on their species, with some weighting as little as a pound, while others can weight up to 20 pounds. They have long ears that they can move in different directions, sharp claws for digging, and powerful hind legs that allow them to hop and jump. Wild rabbits are also known for their unique behaviors, such as thumping their hind legs to warn other rabbits of danger. Rabbits are very social creatures and live in large groups called colonies. Their insatiable reproductive fame has a good reason. They breed three to four times each year. This is because only 15 percent of baby rabbits make it to their first birthday. So, to ensure that the population grows, rabbits have more babies. Another curiosity about rabbits is that they eat their poop. This behavior is known as coprophagy and is essential for a rabbit's health. Rabbits have a cecum, a specialized pouch in their digestive tract, where food is fermented and broken down by bacteria, and then passed through the intestines to be expelled as fecal pellets. These pellets are then re-ingested by the rabbit to extract additional nutrients. Rabbits are prey animals and have evolved several mechanisms to avoid being caught by predators. They have large ears that can rotate 270 degrees, allowing them to detect sounds from all directions. They also have a specialized muscle in their neck that allows them to quickly turn their head to avoid danger. Rabbits that live in the wild can also swim (when they're in danger). The main predators of rabbits are the stoat and the fox, although young rabbits also fall prey to birds of prey and weasels. Wild rabbits play an important role in many ecosystems, as they help to distribute seeds and provide food for predators such as owls, and foxes.

## Appendix B

## Declarative Knowledge Test (True or False)

### Congruent Statements

| |
|---|
| The rabbits belong to the rodent's order. |
| Coprophagy is the Rabbit's behavior of reingesting their fecal pellets. |
| Seeds are the primary food source for domestic rabbits. |
| Providing plenty of exercise is the primary way to ensure a pet rabbit's health. |
| It is important to clean a rabbit's living space regularly\| to prevent the accumulation of fecal pellets. |
| Binkying is the rabbit's ability to rotate its ears. |
| The primary purpose of a rabbit's litter box\| is to encourage the rabbit to eliminate in a designated area. |
| Rabbits have a relatively short lifespan\| usually living only 2-4 years. |
| Rabbits are animals with antisocial behavior\| and human interaction can cause them great levels of distress. |
| Flopping behavior is believed to be rabbits signs of happiness and relaxation. |
| Domestic rabbits come in various sizes and breeds\| with some weighing as little as a pound while others can weigh up to 20 pounds. |

### Incongruent Statements

| |
|---|
| The rabbits belong to the rodent’s order. |
| Coprophagy is the Rabbits’ behavior of reingesting their fecal pellets. |
| Insects are the primary food source for wild rabbits. |
| Wild rabbits are found in a variety of habitats such as forests and grasslands. |
| The purpose of a wild rabbit's thumping behavior\| is to signal to other rabbits to come closer. |
| The primary function of a rabbit's cecum is to break down plant material. |
| Rabbits are known for their reproductive habits, and they usually breed 5 to 6 times a year. |
| The main function of rabbit’s ears rotation is to notice sounds to detect predators. |
| Wild rabbits are social animals that live in large groups. |
| The main predators of rabbits are the Hawks and the wolves. |
| Wild rabbits vary in size depending on their specie\| with some weighing as little as a pound while others can weigh up to 20 pounds. |

## Appendix C

## Briefing

Dear participant,

You are invited to participate in the study "Psychological Responses to Virtual Stimuli in AR and VR" that is being carried out by a Communication and Cognition Master student at Tilburg University. This information letter will describe the purpose and procedure of this study, along with an explanation of your rights as a participant. Participation in the study is completely voluntary. If at any time during the experiment you would like to terminate the process, you can do so without consequence. If you have any questions after reading this letter, please contact the researcher via the contact information at the bottom of the letter.

**What is the purpose of the study?**
The use of immersive technologies such as Augmented and Virtual Reality is becoming increasingly prevalent today. Immersive experiences have been used in numerous studies focused on the elicitation of emotions. However, comparative studies on the effects of VR and AR on learning are scarce. Therefore, this study will explore how individuals react to interacting with different digital objects in a VR and AR setting. Furthermore, we will explore how realistically these objects are perceived and how users may process the information provided by VR and AR technologies.

**Who can participate?**
The minimum age for participation in this experiment is 18 years. During the experiment you will interact with digital objects in the immersive experiences. The objects are designed to elicit a certain emotional response. It is therefore important that you **do not** sign up for this study if you have a strong fear of rabbit, insects, reptilians, and imaginary animals as you will be interacting with a simulated digital insects, reptilians, cute bunny or imaginary animals.

**What does participation in the study entail?**
After signing up for this study and picking a time slot, you will receive an email with information on where and when you will need to be. The experiment will take place on the campus of Tilburg University and will take approximately 30 minutes.

Before the experiment, you will receive a briefing about the study and will be asked to read and sign a consent form. During the experiment, you will be asked to fill out a questionnaire.

You may withdraw from the experiment if you are not willing to disclose such information without any negative consequences, and without providing any explanation. We ensure that the information you provide will only be used for academic purposes.

**What are the benefits and risks of participating in this study?**
There is minimal risk in participating in this study. However, there could be situations where you may find the experimental conditions to be discomforting. As stated before, it is important that you do not have a strong phobia for insects, reptilians or imaginary animals as you will be interacting with a simulated animals within the simulation. However if you would like to participate and feel fearful at any point during the experiment, you may withdraw any time without providing an explanation.

This is also the case if at any point you experience discomfort from simulator sickness. Simulator sickness might affect your feelings of orientation, and make you feel nauseous. Again, if you do not feel comfortable when participating in the experiment, you may withdraw at any time, without providing an explanation.

Participants in this study are entitled to earn 0.5 credit upon completion of the experiment and will help to further our knowledge on emotion elicitation through immersive technologies.

**What will happen to your data?**
The data you provide will be anonymized before the data analysis takes place. Your personal information (name, email address, phone number) will be deleted after the data collection phase. The researchers at the bottom of the page are the only individuals with access to the data collected in this study.

The data will be anonymized after the data collection process is completed. However, during data collection period (until 20/05/2023) you as a participant have the ability to request access to or rectification, erasure or restriction of your personal data. The purpose of this study is to contribute to scientific research. The results that will be shared in the study are general results, not individual results that can be traced back to participants.

**Do you have questions?**
If you have any questions about the study or the experiment in general, feel free to contact us:

Gabriela Christofoletti,
(g.christofoletti@tilburguniversity.edu),
Master Student, Communication and Cognition

Taegyun Sim
(t.sim@tilburguniversity.edu),
Master Student, Communication and Cognition

Dr. Mincheol Shin,
(m.shin@tilburguniversity.edu),
Assistant Professor at Tilburg University

**Do you have a complaint?**
This study has been approved by the Ethical Review Board of Tilburg School of Humanities and Digital Sciences. If you have any remarks or complaints regarding this research, you may also contact the "Research Ethics and Data Management Committee" of Tilburg School of Humanities and Digital Sciences via tshd.redc@tilburguniversity.edu.

# Appendix D

## Consent Form

**By checking "Yes", I acknowledge the following statements**:

I am at least 18 years old.

I have read the information letter about the study. I have been able to ask questions about the study and I have been able to think long enough about whether I want to participate in the study.

I do not have a strong fear of rabbit, insects, reptilian or imaginary animals.

I know that participation in the study is voluntary. I can withdraw from the study at any time, without it having negative consequences and without having to tell why I want to stop.

I know that my research data is processed anonymously. My personal data will be deleted once the data collection has ended. Only the researcher has access to my personal data (e.g., my name, age, phone number and my email address). This data is stored in a secure place, with a strong password.

I know that the anonymous research data can be used for scientific research now and in the future. The anonymous data is examined for all participants at the same time, and not separately for me.

I know that only the anonymous research data can be shared with other researchers. My personal data will never be shared with other researchers.

I know that the coded (anonymous) research data will be kept for at least ten years.

I know that I have the right to request access, rectification, erasure or restriction of my personal data.

# Appendix E

# Questionnaire

***Section A. Plausibility illusion***

**To what extent do you agree or disagree with the following statements?**

*On a 7-point Likert Scale: 1 strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = agree, 7 = strongly agree).*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| The appearance of the rabbit seemed plausible to me as if the creature is real. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| The appearance of the rabbit seemed to match well with the appearance of animals in our reality as if the creature is real. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

***Section B. Attention Allocation***

**Please choose the most appropriate answer that best describes your experience:**

*On a 7-point Likert Scale: 1 strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = agree, 7 = strongly agree).*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| I devoted my whole attention to the bunny. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I concentrated on the bunny. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| The bunny captured my senses. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I dedicated myself completely to the bunny. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

***Section C. English Level (Listening)***

**Select one answer that best matches your current English listening level:**

- o I can recognize familiar words and very basic phrases concerning myself, my family, and immediate concrete surroundings when people speak slowly and clearly.
- o I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.

- I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programs on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.
- I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programs. I can understand the majority of films in standard dialect.
- I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signaled explicitly. I can understand television programs and films without too much effort.
- I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.

***Section D. English Level (Reading)***

**Select one answer that best matches your current English reading level.**

- I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.
- I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus, and timetables and I can understand short simple personal letters.
- I can understand texts that consist mainly of high-frequency everyday or job-related language. I can understand the description of events, feelings, and wishes in personal letters.
- I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.
- I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialized articles and longer technical instructions, even when they do not relate to my field.
- I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialized articles, and literary works.

***Section E. VR and AR Familiarity***

**Please choose the most appropriate answer that best describes your experience:**

*On a 7-point Likert Scale: 1 strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = agree, 7 = strongly agree).*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| I am familiar with Augmented Reality | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I am familiar with Virtual Reality | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

***Section F. Previous Knowledge on Rabbit and Animals***

**Please answer the three questions below regarding your knowledge on animals.**

*On a 7-point Likert Scale: 1 strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = agree, 7 = strongly agree).*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| I have a lot of knowledge about animals. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| I have a lot of knowledge about rabbits specifically. | 1 | 2 | 3 | 4 | 5 | 6 | 7 |

***Section G. Demographics***

**Please answer the three questions below regarding your knowledge on animals.**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Age: | | | | | | | |
| Gender: | Male | Female | | Non-binary | | Prefer not to say | |
| Ethnicity Please state your ethnicity: | | | | | | | |
| Education Level: | High School | College | Associate Degree | Bachelors degree | Masters Degree | Doctorate | Other |

## Appendix F
## Written Debrief

Thank you very much for participating in this experiment.

There were two studies involved in this experiment. For the first study, you were assigned to one of two conditions, to interact with a simulated bunny either in virtual reality or augmented reality. For the second study, you were assigned to one of four conditions, to interact with either audio-visual congruent or incongruent stimuli in either AR or VR immersive environment.

By doing so, the aim of the studies is to compare groups and explore if different modalities elicit a different level of knowledge retention. Furthermore, we will explore if the level of perceived realism of the immersive experience correlates with the level of knowledge retention.

In the second study, audio-visual congruency was not briefed prior to the experiment in order to prevent preconditioning through the briefing. This was added to the study to measure the impact of congruency effect and plausibility illusion on knowledge retention.

If you have any additional questions regarding the study or the purpose, please feel free to ask them. If there are no questions, I would once again like to thank you for participating. You will receive 0.5 credit on the Sona participant portal for participating in this experiment.

Thank you very much for your time and effort!