UNDERSTANDING PRESENCE IN AUGMENTED REALITY: A PSYCHOPHYSICAL STUDY

By

Nishan Gennetian Shehadeh

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Computer Science

August 11, 2023

Nashville, Tennessee

Approved:

Bobby Bodenheimer, Ph.D.

Matthew Berger, Ph.D.

ACKNOWLEDGMENTS

Thank you Dr. Bobby Bodenheimer and the LiVE Lab for your support in my classes and research. Thank you Dr. Richard Skarbez for advising and collaborating on my thesis project and Dr. Matthew Berger for your advice throughout my undergraduate and graduate studies. Finally, thank you to my parents for your unwavering support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

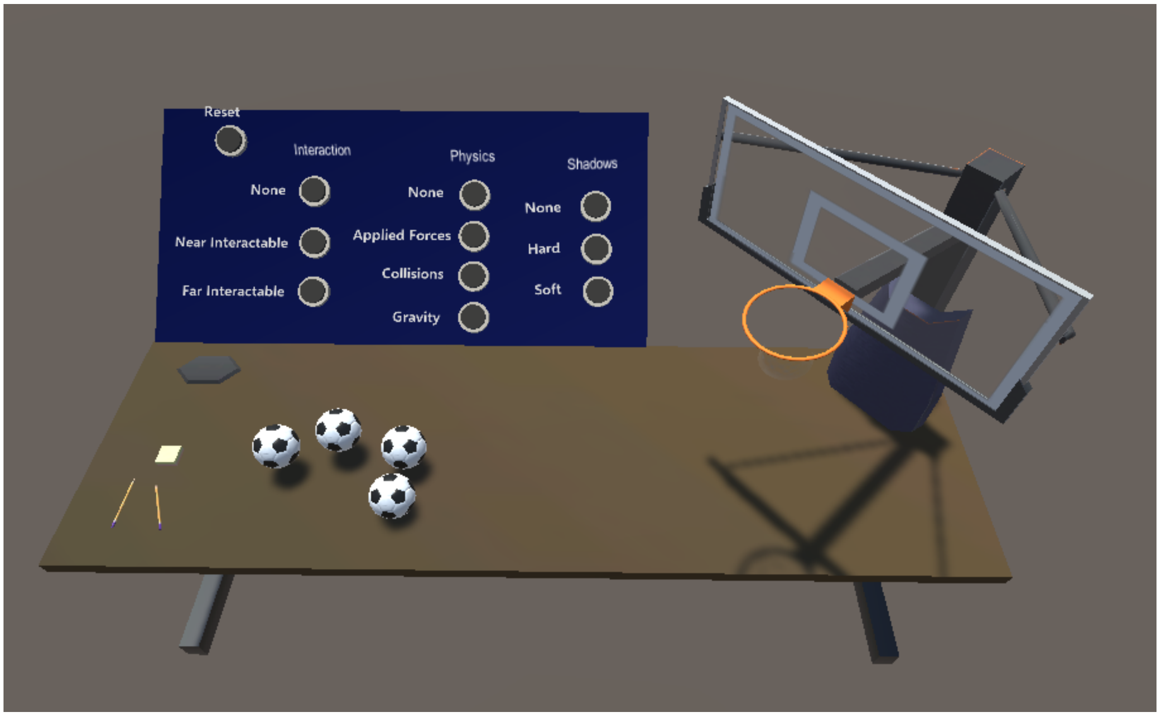

Figure 1.1: User study environment with a virtual desk with virtual pencils, post-it notes, soccer balls, and basketball hoop. A control panel (blue slanted panel behind the desk) adjusts different factors within the scene.

Augmented reality (AR) is displaying virtual content overlayed on the real world (Milgram et al., 1995). AR is a quickly growing field with applications in training, education, surgery, and more (Doolani et al., 2020; Chen et al., 2016; Ma et al., 2017). Strategically improving certain factors that are shown to increase plausibility and realism of the virtual content improves AR experiences and increases user engagement in various fields of application.

An inherent challenge in the field of AR lies in effectively evaluating user experiences when dealing with a blend of real and virtual content. Evaluating user experience within any type of mixed reality (MR) environment involves modeling and measuring how people perceive virtual content. If the virtual objects or environment have certain qualities, it can lead to a higher level of immersion and a better experience (Slater, 2009). These qualities can deal with anything from hardware limitations like head-tracking latency and field of view to software design choices like how interactive the environment is. Design choices often require compromise between hardware limitations and software choices. There are frameworks for understanding how good an AR environment is, but these frameworks need evidence to support whether they properly capture how users perceive AR content. Presence is one dominant framework for understanding the quality of virtual content.

Presence has been defined differently by many researchers. Slater originally used Place Illusion (PI), the illusion of being somewhere you are not, and Plausibility Illusion (Psi), the sense that something happening is really happening, as constructs to build a model for presence (Slater, 2009). Latoschik and Wienrich proposed a different model in which congruence and plausibility are the central components (Latoschik and Wienrich, 2022). In this model, different sensory input to the manipulation space is processed and compared to the user's expectations, resulting in a level of congruence. Plausibility is a function of the weighted congruencies. Their model has some empirical evidence through experiments that study plausibility and presence levels while a scene is adjusted to have certain incongruencies (Brübach et al., 2022) and a study that compares manipulating congruence in AR versus virtual reality (VR) (Westermeier et al., 2023). In this thesis, we seek to further understand presence in AR through a psychophysical user study designed to compare the relative importance of different factors in an AR environment by manipulating interaction levels, physics, and shadows.

These categories each represent important aspects of AR experiences and provide different sensorimotor contingencies for the perception of virtual objects (Slater et al., 2022). Interaction is essential to many AR experiences. Being able to reach out and touch objects is an important sensorimotor contingency for users to perceive them as real (Slater et al., 2022). Physics is an important, omnipresent prior that can have strong effects on the plausibility of virtual objects (Brübach et al., 2022). Lastly, shadows are representative of the rendering level of the experience and affect how users place objects in 3D space.

In this thesis, participants complete a series of activities in a simple AR environment that has a control panel for adjusting certain settings within the environment (Figure 1.1). Their choices while transitioning between and budgeting different environmental configurations, defined by the factors they toggle on or off, showed that realistic interaction is the crux of immersive AR experiences, followed by having an environment that strongly matches a real-world frame of reference through important priors like gravity. These findings also show the preference for bottom-up processing in AR as participants built plausibility through baseline interaction and matching their reference frame.

This thesis is organized as follows: Chapter 2 revisits related work for definitions of AR and MR, presence in MR, and psychophysical methods; Chapter 3 describes the AR system; Chapter 4 reports the results from the user study; Chapter 5 discusses the results in a broader context; Chapter 6 provides limitations and future work; Chapter 7 concludes the thesis.

CHAPTER 2

Related Work

2.1 Augmented and Mixed Reality

In his 1997 paper, Azuma provided a comprehensive overview of state-of-the-art AR research, techniques, and definitions for the time (Azuma, 1997). The survey defined and categorized AR, identified key technical challenges, and outlined future research directions. Importantly, he presented a definition for AR as a "variation of Virtual Environments" where virtual objects supplement a user's experience in the real world. He defined AR systems as having the following three characteristics: "1) combines real and virtual, 2) interactive in real time, and 3) registered in 3-D" (Azuma, 1997). In a more recent survey on the definition of MR, Speicher et al. (2019) name a set of relevant characteristics for AR experiences based on interviews with ten AR/VR experts. The characteristics include "merging of 3D graphics with the real world," "spatial registration in the physical environment," "the user has to be in control," and "the virtual content has to interact with the real world" (Speicher et al., 2019). Clearly, these definitions are aligned with Azuma's original three criteria, and all definitions state the necessity of having virtual, interactive, 3D objects registered in the real world.

Milgram and Kishino's famous taxonomy defined different terms related to the "merging of real and virtual worlds" along the "virtuality continuum" (Milgram and Kishino, 1994). The continuum consists of all environments between two anchor points, an entirely real environment consisting of only real objects and a completely virtual environment consisting solely of virtual objects. This continuum is a foundation for defining MR as any environment along the spectrum that fuses real and virtual elements. They define AR as MR environments where virtual content is added to the real world to some capacity. In order to classify the merging of real and virtual worlds, they define a taxonomy for MR applications based on three axes: Extent of World Knowledge, Reproduction Fidelity, and Extent to Presence Metaphor. Respectively, these axes answer the questions: "How much do we know about the world being displayed?", "How realistically are we able to display it?", and "What is the extent of the illusion that the observer is present within that world?" (Milgram and Kishino, 1994). Skarbez et al. (2021b) revisited and built on this system to account for "the importance of how the real or virtual content is observed". Their new definition expanded the original definition of MR based on visual displays to account for new technology across the AR/VR space. MR was redefined as an environment in which real and virtual objects are presented "within a single percept," which expands the environment to existing across different senses. They also altered the taxonomy of MR environments by replacing two of the axes. Firstly, they combined the Reproduction Fidelity and Extent of Presence Metaphor dimensions into a single dimension, Immersion, based on Slater et al.'s original definition of the set of valid actions supported by a system (Slater, 2009). By basing Immersion on sensorimotor valid actions, actions that change the user's perception of the environment, and effectual valid actions, actions that change the environment itself, Skarbez et al. (2021b). transform the system definition to account for interaction beyond the choice of viewpoint. As a third axis, they define Coherence as the "consistency and predictability of the virtual objects' interactions within the system."

These definitions of AR and MR, provide a framework to study AR applications and help create a language for describing the interactions between users and virtual content.

2.2 Presence in MR

Slater proposed a model for user experience in virtual environments that further defined characteristics of virtual environments and built on the previously discussed definitions of AR (Slater, 2009). He introduced two orthogonal components that contribute to participants' responding realistically to the content of an "immersive virtual reality system": Place Illusion (PI) and Plausibility Illusion (Psi). PI is the "qualia of having a sensation of being in a real place" or the sense of "being there" while Psi is the illusion that "what is apparently happening is really happening" and provides correlations between a user's sensations and the events in the environment (Slater, 2009). Further tying together coherence, immersion, Psi, and PI, Skarbez later affirmed Slater's original idea that immersion represents the boundaries within PI can occur while introducing the idea that coherence represents the boundaries within Psi can occur (Skarbez et al., 2021a). He notes the distinction being that coherence is a characteristic of the environment itself while Psi is based on the user's response to their experience within the environment. Alexander et al. describe fidelity as "the extent to which the virtual environment emulates the real world" (Alexander et al., 2005) and Skarbez argues that coherence is a more general, yet related, term because it is a broader concept that does not necessarily assume that the virtual scenario aims to replicate the real world (Skarbez et al., 2021a). In a recent update on PI and Psi, Slater expanded PI and Psi to AR by determining that, in AR environments, PI becomes inverted (Slater et al., 2022). In this case, PI is aiming to incorporate virtual objects into the real world as opposed to its original goal in VR to place the participant into the virtual world. He also mentions the importance of increasing the number of sensorimotor contingencies for perception. In AR, this means participants being able to perceive virtual objects by interacting with them and the objects themselves exhibiting realistic behavior such as reflecting light from the real world. With the inversion of PI, the line becomes blurred between PI and Psi because both are dealing with the responsibility of virtual content to have personal interactions, be consistent, etc. in order to seem realistic. Whereas before, PI and Psi were mainly split by one dealing with location and the other dealing with events, according to Slater's new definitions, both can be defined by the virtual contents' ability to offer sensorimotor interactions (Slater et al., 2022).

Tangentially, other models of presence expanded on the original Slater definitions. As previously mentioned, Skarbez added Coherence and redefined the taxonomy from the Milgram-Kishino continuum (Skarbez et al., 2021a). In Hofer et al.'s work studying the effect of plausibility violations on spatial presence in VR, they discuss Skarbez's idea that virtual environments feel plausible based on prior knowledge and define two types of plausibility: internal and external (Hofer et al., 2020). Internal plausibility is the "extent to which the environment is consistent within itself or with respect to the expectations raised by its genre," which aligns with Skarbez's principles of coherence and Psi. External plausibility is "how consistent the virtual environment is to users' real-world knowledge," which aligns itself with inverted PI (Hofer et al., 2020; Slater et al., 2022). Similarly, Wienrich et al. presented the concept of a reference frame when dealing with spatial presence in MR Wienrich et al. (2021). Referential power is defined by the "probability of each entity or class of entities to be selected as a referential cue." Concerning realism in MR, they view plausibility as incongruent entities behaving "coherently to the dominant place-illusion," which, in this case, is the primary egocentric reference frame (PERF) or the real world in an AR environment. Incongruent entities are content that does not match the dominant reference frame, for example virtual objects in the real, physical space. In an experiment causing systematic incongruencies along the MR continuum to examine the effect on plausibility and presence ratings, Westermeier et al. found the first empirical indication that AR environments require a different reference setting than VR environments (Westermeier et al., 2023). The users in the AR environment in their study were found to be more focused on the objects than the environment, leading to expectations about how the virtual objects would interact with the environment. Brübach et al. explored a similar idea by introducing breaks in plausibility through adjusting object behaviors to not conform with the laws of physics. They found that this type of manipulation breaks plausibility because of the incoherence between strong priors like gravity and object behavior even in a purely virtual environment (Brübach et al., 2022). Fidelity (coherence), external plausibility, and reference framing all promote the relationship between presence in AR and users' prior knowledge of the real-world environment. Latoschik and Wienrich also proposed an alternative model for XR experiences altogether after debating the validity of Skarbez and Slater's original ideas on presence Latoschik and Wienrich (2022). In this model, congruence and plausibility are the central conditions to describe XR experiences. Congruence describes the "objective match between processed and expected information at the sensory, perceptual, and cognitive layers of processing," and plausibility is the state that results from the information processed in those same layers.

2.3 Psychophysical Methods with Perception

Psychophysical methods investigate the relationship between physical stimuli and the resulting psychological experiences. In terms of perception in MR environments, psychophysical studies can provide empirical evidence for various factors regarding the users' perception of the environment, whether it be congruence, coherence, or immersion. While questionnaires are the most popular method of gathering measurement due to their simplicity and generalizability, they do not truly capture what happens during the experience from the point of view of the participants because of the conceptual framework inherently imposed by the researchers themselves (Slater et al., 2022). Configuration transitions, a practical implementation of psychophysical studies, have the ability to base conclusions on observed data collected during the experience through transition probability matrices, Markov chain analyses, and conditional probabilities. The ideal experiment, according to Slater, should consist of multiple different sources of data to combine their respective strengths such as transition probabilities backed up by qualitative and quantitative questionnaire results (Slater et al., 2022). Slater first performed transition analysis with a study on Psi and PI in VR based on manipulating field-of-view, display type, virtual body, and illumination (Slater et al., 2010). Skarbez built on that to investigate Psi specifically by manipulating factors regarding coherence in a virtual environment, for instance the behavior of virtual humans in the environment, and found that the virtual body is a powerful contributor to Psi (Skarbez et al., 2017b). Additionally, Bergström et al. studied plausibility factors such gaze, sound spatialization, auralization, and environment in a VR string quartet performance (Bergström et al., 2017). These studies show that transition configurations are a helpful tool to quantify how different factors affect plausibility and create a format for studying different elements of AR environments such as sensorimotor contingencies. This thesis extends these studies into the field of AR by adapting the same techniques to study different aspects of virtual content in an AR experience.

CHAPTER 3

Experiment

In this experiment, participants completed two transition choice tasks, two budgeting tasks, and answered questionnaires on how realistic the virtual content felt. Our goal was to investigate what factors are important to make a compelling AR environment. Based on pilot studies, we formulated these hypotheses concerning the system:

H1: High levels of interaction will be the most salient factor in the user experience.

H2: High levels of physical behavior will be more salient than graphical appearance.

We tested these hypotheses by building a Markov chain from the transition choices, collecting budget configurations, and analyzing questionnaire responses.

3.1 System

3.1.1 Environment

The environment used in this study was built in Unity using Microsoft's Mixed Reality Toolkit (MRTK) (Microsoft, 2021). It was deployed on the Hololens 2. Figure 3.1 shows an overall view of the environment including the virtual desk, objects, and control panel.

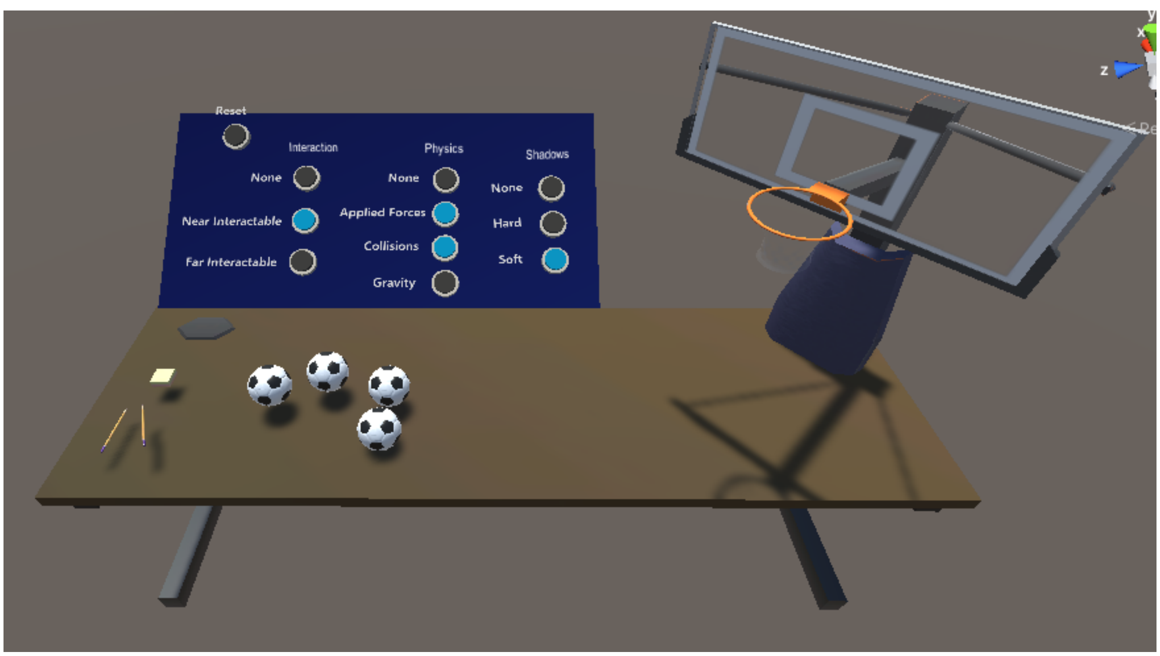

Figure 3.1: Example environment with configuration on the control panel.

3.1.2 Control Panel

The test environment for the study is a virtual desk with an assortment of objects and a control panel. The control panel, the slanted blue board behind the desk in Figure 3.1, consists of virtual toggle buttons for each factor that could be adjusted by a participant. It also has a reset button that both returns all factors to their zero position "None" and resets the position of all virtual objects to the initial configuration.

There are three categories of adjustable factors: interaction (denoted by I), physics (denoted by P), and shadows (denoted by S). Interaction has three options: None, Near Interactable, Far Interactable. Physics has four options: None, Applied Forces, Collisions, and Gravity.

The control panel is controlled via far interaction. Far interaction on the Hololens 2 allows users to interact with virtual objects from a distance through "instinctual" hand tracking and "air tapping," which is essentially just a pinch of the user's fingers or hands (Microsoft, 2021). A pointer extends out of the user's hands that collides with the virtual objects that they are pointing at. To toggle a button the control panel, the user air taps while hovering the pointer over the button. The buttons light up blue when 'on' and are gray when 'off' as seen in Figure 3.1.

3.1.3 Interaction

3.1.3.1 None

When None is selected, the user has no control over any objects in the environment. The user has no way to interact with or manipulate the objects. If None is selected, no other interaction toggles may be selected at the same time.

3.1.3.2 Near Interactable

When Near Interactable is selected, the user may interact with the objects on the desk by grabbing the objects anywhere on their 3D mesh. The same tapping (pinching) motion is used to grab the objects when sufficiently close enough to reach them.

3.1.3.3 Far Interactable

When Far Interactable is selected, the user may use the same far interaction as used to control the control panel to manipulate objects. In this case, the objects can be manipulated in all ways similar to normal, near interaction (picked up, thrown, etc.). The Hololens 2 also supports an intuitive way of adjusting the length of the pointer, or the distance of the object to the user, giving the ability to control an object's position in 3D space completely.

3.1.4 Physics

3.1.4.1 None

When None is selected, the virtual objects on the desk are not influenced by any physics simulations. That is, no forces of any kind, whether they be virtual or from the user, affect the objects. The only exception is if interaction is enabled, in which case objects may be moved while the user is grabbing them (without any other forces).

3.1.4.2 Applied Forces

When Applied Forces is selected, any force applied by the user's hand interaction will impact the objects in a normal manner. For example, if a user picks up a ball and uses a throwing motion, the ball will continue to move in that direction.

3.1.4.3 Collisions

When Collisions is selected, the objects will collide with all other virtual objects in the scene including other objects on the desk, the desk itself, and the hoop. In order to make the control panel easy to use at all times, the collider for the panel is always on. Similarly, MRTK's spatial awareness system is always on to give the walls, ceiling, floor, and other physical objects in the room colliders.

3.1.4.4 Gravity

When Gravity is selected, the virtual objects all adhere to a normal gravitational force. This also displays the bouncy quality of the balls when they are dropped on the floor or thrown against other objects. When Gravity is turned on, the user can immediately see the effects because multiple objects are placed slightly above the desk, so either they drop and bounce off of the desk (if collisions are on) or all objects drop through the desk to the floor (if collisions are off).

3.1.5 Shadows

All shadows are meant to mimic shadows that would exist for the objects based on the physical overhead lights in the room where the study was conducted. There are no explicitly added virtual lights besides a directional light in the Unity scene that simulates the physical, overhead lights.

3.1.5.1 None

When None is selected, the virtual objects do not display any shadows from the physical light in the room.

3.1.5.2 Hard Shadows

When Hard Shadows is selected, the virtual objects display shadows that show the general shape of the object, but are very pixelated and blocky as seen on the left side of Figure 3.2. In Unity, hard shadows are created by the shadow maps by taking the nearest shadow map pixel and adding a dark shade to it. In this environment, the resolution of the hard shadows is manually decreased even further to make the shadows more blocky and contrast more with soft shadows. Hard shadows are less computationally expensive to render in real time (Unity, 2016).

3.1.5.3 Soft Shadows

When Soft Shadows is selected, the virtual objects display shadows that are more realistic and smooth as seen on the right side of 3.2. In Unity, soft shadows are created by the shadows maps by averaging several shadow map pixels, resulting in the smoother look. Soft shadows are more computationally expensive and considered more realistic than hard shadows (Unity, 2016).

Figure 3.2: Hard shadows (left) vs soft shadows (right) on a virtual desk.

3.1.6 Configurations

Based on the toggle buttons, the environment has a configuration that represents the level of each category at that moment. For the first stage of the experiment, the transition analysis, each stage was represented by a corresponding level for each category. A configuration C can be represented as $C = \{I, P, S\}$ with levels for each category given by Table 3.1 where the options that are toggled on for each level are listed. For example, a configuration C_1 where Interaction is on level 2, Physics is on level 1, and Shadows is on level 0 is given by $C_1 = \{2, 1, 0\}$.

For the second phase of the experiment, each toggle button is independent, and a configuration can be given by a 10 integer number with 0 representing an off position and 1 representing an on position for

Level	I	P	S
0	None	None	None
1	Near Interactable	Applied Forces	Hard Shadows
2	Near Interactable, Far Interactable	Applied Forces, Collisions	Soft Shadows
3		Applied Forces, Collisions, Gravity	

Table 3.1: Toggle Options by Level for Categories I, P, and S.

each button. The first three digits represent the interaction buttons, the next 4 represent the physics buttons, and the last 3 represent the shadow buttons. For example, a configuration C_2 with just Near Interactable, Applied Forces, and Soft Shadows would be represented by $C_2 = \{0100100001\}$. The necessity of having two different configurations is further explained in Section 3.3.

3.1.7 Virtual Objects

The common desk items in the environment, the pencil and post-it notes, allow participants to experience interacting with different sized and shaped items. The collection of balls and a hoop gives the participants an intuitive task to complete that helps accentuate the configuration of the environment. The action of throwing the ball in the hoop is designed to give a user a notion of what the interaction level is while controlling the ball, how the physics impact the ball as it is thrown, and how shadows impact the ability to place the ball in 3D space.

3.2 Participants

30 participants took part in this study. This sample size was selected based on previous studies of configuration transitions in VR (Skarbez et al., 2017b). All were current students (undergraduate and graduate) at Vanderbilt University. The demographic data is provided in Table 3.2

Demographics	# Participants (out of 30)
Gender	17 Male; 13 Female
Age	M = 21.57; SD = 1.33
Occupation	30 Student
Duration of AR experience	20 $<$1hr; 3 1-3hr; 5 3-5hr; 1 5-10hr; 1 $>$20hr
Frequency of AR experience	20 never used; 6 used 1-3 times, 2 used 3-5 times, 1 used $>$20 times
Computer usage	M = 6.83; SD = 1.95

Table 3.2: Demographic data for all participants.

3.3 Experimental Procedures

3.3.1 Pre-experiment

The experiment took place in the LiVE space for all participants. When they arrived, participants were informed that the study was completely voluntary. They signed an informed consent document outlining information about the study with IRB approval. They also filled out a payment form from Vanderbilt University to approve a cash payment. The overall structure of the study was then explained to the participants: two rounds of experiments with the headset on and accompanying questionnaires. Given the participants' lack of AR experience, the Hololens 2 head-mounted display (HMD), was adjusted to fit their heads before they were instructed to follow the Hololens iris calibration procedure that automatically initiates when a new user puts on the headset. If the participant was not confident in completing the calibration through air taps, they were instructed to open a pre-saved Microsoft 3D object to practice manipulation and the airtapping control in AR. Once it was clear they understood the basic controls of the Hololens, the study scene was opened, with the virtual desk (Figure 1.1) in the middle of the empty lab room.

Upon entering the scene, the participants were instructed to toggle the reset button on the control panel. If needed, they were given a series of steps regarding using far interaction. The steps were as follows: observe your hand models while they are in the field-of-view; observe the pointer coming out of the hand model; place your hands in a natural state by your side so that they are not necessarily in the field of view and observe the pointer still coming out of your hands and being controlled by your movements; observe how when you tap your fingers or hands together while hovering over any virtual object the pointer activates indicated by the reticle closing in; do this motion over a button to see it light up when you toggle it. Upon successfully clicking the reset button, the participants were walked through the control panel options one by one with verbal descriptions similar to in Figure 3.1.1. Any questions about the specific options were answered, and, after walking through each option, the participants were given an opportunity to interact with the optimal environment $C_o = \{2,3,2\}$. During the walk-through, common questions arose about what would happen if gravity is activated, but one or both of the other factors - applied forces and collisions - are not. They were instructed to note how present and realistic the virtual objects felt in this optimal configuration.

3.3.2 Transition Experiment

The participants were instructed to press the reset button, setting all categories' levels to 0. Following Skarbez's procedures (Skarbez et al., 2017b), participants were then told that they were playing a game in which the goal was to reach a scene with optimal sense of reality as they experienced in the optimal environment with respect to the virtual content. Participants were informed that the rules of the game were that they could only upgrade one category at a time, and one level at a time within each category. The levels given in Table

3.1 were automatically set up to adjust as the participants moved down the column of options. For example, if a participant upgraded from hard shadows (S = 1) to soft shadows (S = 2), the hard shadows toggle would turn off when the soft shadows toggle was selected. They were also told to focus on the choice of improvements that were of the highest priority, in their opinion, to reach the optimal level of reality. Each transition had to be a single-step improvement, and they were not allowed to backtrack on their choices. The participants were allowed to interact with the scene in between choices, and there was no time limit on their transitions. All transition steps were saved out during the trial. The first trial (starting at $C_0 = \{0,0,0\}$) was complete when the participant deemed the environment to have an equal level of realism as the optimal or all possible upgrades had been made. They were then told to choose a number from one to three, which randomly assigned one of the three first upgrades to be the automatic first upgrade for the second trial. They then followed the same procedure of choosing upgrades until an end configuration was met and recorded. When the two transition trials were complete, the participants took off the HMD and completed a plausibility questionnaire from Brubach et al. designed to measure the plausibility of the objects in their final chose configurations 3.3 (Brübach et al., 2022). This questionnaire was modelled after the definitions for internal and external plausibility given by Hofer et al. and discussed earlier (Hofer et al., 2020). The questions in group ”EP” were designed to measure the external plausibility and ”IP” designed to measure the internal plausibility. All questions were taken on a seven-point Likert Scale with endpoints ”not agree at all” (1) and ”totally agree” (7). For all individual users, the final configurations from their two trials were the same.

No	Group	Question
1	EP	I am used to objects having this way
2	EP	In everyday life, I I expect objects to behave this way
3	EP	I have seen objects behave this way in real life
4	EP	The behavior of objects is unusual for me
5	EP	I do not know the behavior of the objects from real life
6	IP	I had a prior expectation of how the objects would behave
7	IP	I expected the behavior of the objects
8	IP	I could not anticipate what would happen next with the objects
9	IP	I was surprised by the objects' behavior
10	IP	I had no idea the objects will behave this way
11	IP	The cause and effect behavior matched the scenario
12		The behavior of cause and effect made sense
13		I think this behavior of cause and effect is impossible

Table 3.3: Questions regarding the plausibility of the objects and environment.

3.3.3 Budget Experiment

After the plausibility questionnaire was complete, participants put the HMD back on and were told to press the reset button, putting all levels back to zero. They were then instructed on the rules of the second experiment, the budget challenge. They were told they were given three changes to make the scene as close to having the optimal level of realism as possible. It was made clear that this time they can choose any combination of toggles, and they do not have to be in order. The one exception was that only one type of shadow can be toggled at a time, which is built into the environment. Their budgeted configurations for three changes were saved. They were then given two more changes to make and their new budgeted configurations for five changes were saved.

3.3.4 Post-Experiment

After the budget experiments, the participants took off the HMD for the final time and were given a final questionnaire consisting of ranking and short answer responses. For each category, they were asked to rank which factor within the category was most important (for example, collisions within physics) with options for no factor being more/less important and any factor besides *None*. They were then asked to explain their rankings. The last two questions were general question on the study inquiring if any specific part of the environment had a large or small affect on the plausibility of the scene and if they thought doing the same experiment in VR would lead to different outcomes. They were then compensated $12 for their participation in a cash payment. The study on average took around 45 minutes end-to-end.

CHAPTER 4

Results

In this section, we present the findings based on the accepted final configurations of the environments considered optimal during the two transition trials, the order of the participants' transition choices, the participants' configurations when given a budget, and two questionnaire responses. The first questionnaire consisted of Likert scale plausibility questions, and the second questionnaire consisted of ranking and free response questions.

As previously done by Skarbez et al., we assume the two trials used for transition analysis were statistically independent (Skarbez et al., 2017b). Although they cannot be truly independent because each participant used their prior knowledge and preferences to make transition decisions, they had to reconsider their priorities for the random starting configuration option due to the investment in one category at the beginning of the scenario.

4.1 Accepted Configurations

There were 4 end configurations accepted by more than one user throughout all 60 trials (2 x 30 participants) with probabilities shown in Figure 4.1. For this section of results, there were effectively 30 trials because participants were always consistent in their end-state decisions for their different transition choices. Unsurprisingly, the most popular accepted configuration was with each category full maxed out with $11/30$ participants using all transitions. The next three end-state choices were all combinations of interaction equal to 1 (I $\geq$ 1), maxed out physics (P $=$ 3), and varying levels of shadows. The accepted end configurations can be seen in Figure 4.2.

4.2 Transition Analysis

Following the technique from Slater et al. (2010), we constructed a probability matrix P from the configurations chosen by each participant. P has a row and column for each unique state reached during the transition analysis, with indexes representing the probability that one state is likely to go to another. For example, for the first row, representing state $\{0,0,0\}$, the corresponding probabilities, as can be seen in the Transition 1 sub-graph of 4.4, are stored in rows 2, 4, and 11 representing the possible next configurations $\{0,0,1\}$, $\{0,1,0\}$, and $\{1,0,0\}$ respectively. This leads to a very sparse matrix as it is size 34 x 34 (34 possible states), with 3 possible next state options for each. The row and column orders are based on integer value for the state (e.g 001 $<$ 010 $<$ 100). Given this matrix P, we can compute the probability distribution over

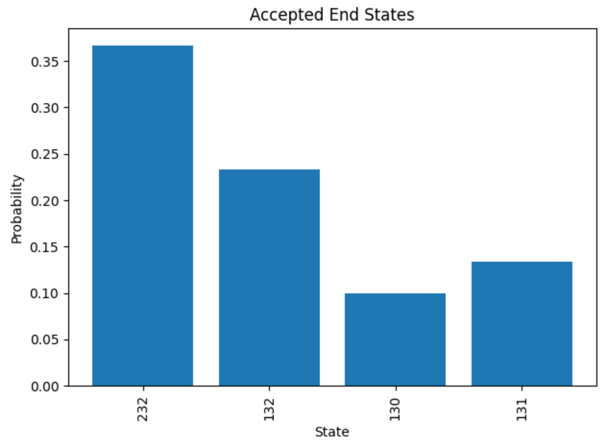

Figure 4.1: End configurations accepted by more than one user. $C = \{I, P, S\}$.

the configurations for any given configuration if we take a vector s of 1 followed by 33 zeros which represents the starting configuration C = $\{0,0,0\}$ and compute sP, yielding the probability distribution after one improvement, sP^2, the probability distribution after two improvements, and so on. The seventh step yields a probability distribution of 100% for the maxed out end configuration $\{2,3,2\}$ (Slater et al., 2010).

Based on the probability distributions for each step, we can compute a most likely path. The most likely path based on all trials was to first upgrade interaction to I = 1, Near Interaction, followed by fully upgrading physics from P= 0 to P= 3, upgrading shadows from S = 0 to S = 2, and, finally, upgrading I = 2 to include Far Interaction. This path can be visualized in Figure 4.3, which also shows all paths chosen by the participants during the trials. The probability of each edge being the next step in the Markov chain is given by 4.4.

As expected, the overall transition probabilities reflect the same likely choices as the optimal path. We observe the two most popular transition routes being upgrading near interaction first (77%) before either adding Hard Shadows followed up by maxing out physics or simply maxing out physics right away. The average number of overall improvements observed by each participant was 5.5 per trial, leading to 330 total observed transitions.

4.3 Budget Analysis

The budget analysis results are largely correlated to the transition analysis results. Because there were not a large number of overall options, participants' preferences for individual upgrades, despite being able to choose them regardless of the transition order, supported the same themes as their transition choices. All states chosen by the participants and their respective percentage of users can be seen in Figure 4.5.

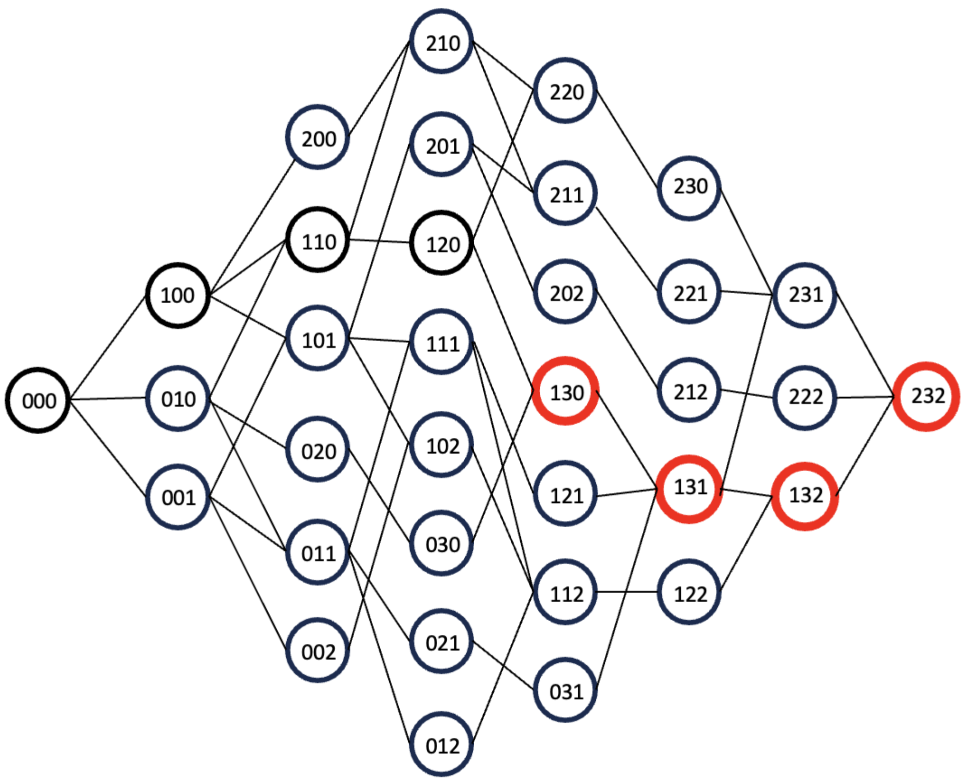

Figure 4.2: The end configurations accepted by more than one user are highlighted red on the graph. The edges in the graph represent transition all participants' choices during the experiment and the nodes are the environment's configuration at a certain point. $C = \{I, P, S\}$

4.3.1 Three Upgrades

When instructed to make three individual improvements as outlined in the study design, participants consistently favored near interaction and gravity as the two most important factors with 20/30 enabling near interaction and 25/30 enabling gravity. Additionally, 16/30 specifically chose the combination of applied forces and gravity. With regards to shadows, 24/30 participants chose to not upgrade shadows at all, which further illustrates the preference found in the transition analysis that enabling physics was seen as a more important prior for real world consistency. 6/30 participants chose to budget a shadow improvement, all of which were soft shadows.

4.3.2 Five Upgrades

When given two additional improvements to make, it becomes evident which improvement was least likely to be chosen in any budgeted configurations. 25/30 participants chose not to select far interaction despite the minimal options. Naturally, hard shadows were not selected by any users as soft shadows are inherently

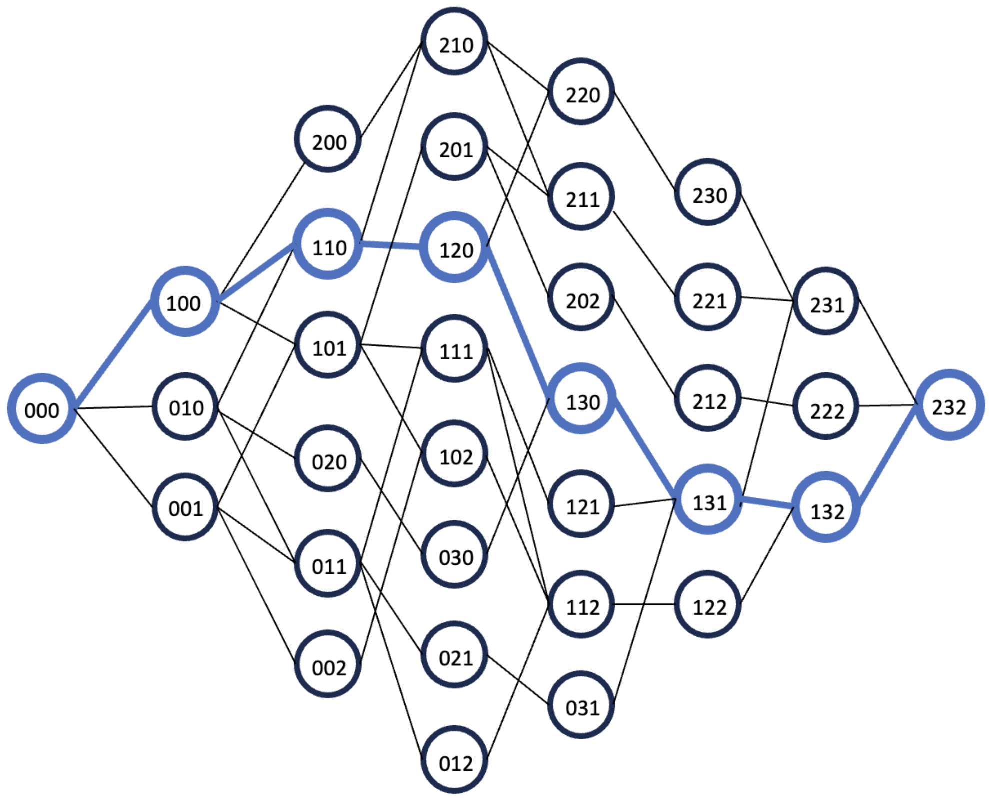

Figure 4.3: The blue path represents the most likely path for participants to chose during the study based on the Markov chain probability distributions. The edges in the graph represent transition all participants' choices during the experiment and the nodes are the environment's configuration at a certain point. $C = \{I, P, S\}$

higher quality.

4.4 Plausibility Questionnaire Results

The responses to Table 3.3 are displayed in Table 4.1. The overall plausibility of the environment reached by the participants was high. We split the participants by gender and whether they ranked interaction or physics higher, but there were no significant results based on a two-tailed t-test.

No	Group	All	Male	Female	Interaction	Physics
1	EP	4.90 ± 1.54	5.06 ± 1.78	4.85 ± 1.28	4.64 ± 1.29	5.28 ± 1.67
2	EP	5.83 ± 1.14	6.29 ± 0.69	5.31 ± 1.38	5.73 ± 1.01	6.00 ± 1.24
3	EP	5.03 ± 1.88	5.59 ± 1.91	4.46 ± 1.71	5.27 ± 1.62	5.17 ± 1.98
4	EP	4.76 ± 1.57	4.65 ± 1.73	5.00 ± 1.35	4.73 ± 1.95	4.78 ± 1.35
5	EP	4.86 ± 1.22	4.94 ± 1.30	4.85 ± 1.14	5.00 ± 1.18	4.89 ± 1.28
	Mean Value EP	5.12 ± 1.44	5.31 ± 1.48	4.89 ± 1.37	5.07 ± 1.41	5.22 ± 1.50
6	IP	5.14 ± 1.83	5.53 ± 1.46	4.69 ± 2.14	5.64 ± 1.86	5.06 ± 1.66
7	IP	5.45 ± 1.48	5.24 ± 1.79	5.77 ± 0.83	6.09 ± 0.83	5.11 ± 1.68
8	IP	4.66 ± 1.08	4.71 ± 1.21	4.62 ± 0.87	5.36 ± 0.50	4.22 ± 1.11
9	IP	4.34 ± 1.54	4.47 ± 1.50	4.23 ± 1.59	5.09 ± 1.22	3.89 ± 1.57
10	IP	4.62 ± 1.50	4.88 ± 1.27	4.31 ± 1.70	4.82 ± 1.47	4.50 ± 1.54
11	IP	6.14 ± 1.09	5.94 ± 1.30	6.38 ± 0.65	6.45 ± 0.69	5.89 ± 1.23
	Mean Value for IP	5.12 ± 1.32	5.13 ± 1.42	5.00 ± 1.30	5.58 ± 1.10	4.78 ± 1.47
12		6.34 ± 0.86	6.41 ± 0.80	6.31 ± 0.95	6.18 ± 0.98	6.44 ± 0.78
13		5.24 ± 1.06	5.53 ± 0.62	4.92 ± 1.38	5.45 ± 0.69	5.17 ± 1.25
	Mean Value for all	5.23 ± 1.31	5.33 ± 1.34	5.05 ± 1.31	5.42 ± 1.18	5.11 ± 1.41

Table 4.1: Likert scale responses regarding the plausibility of the objects and environment.

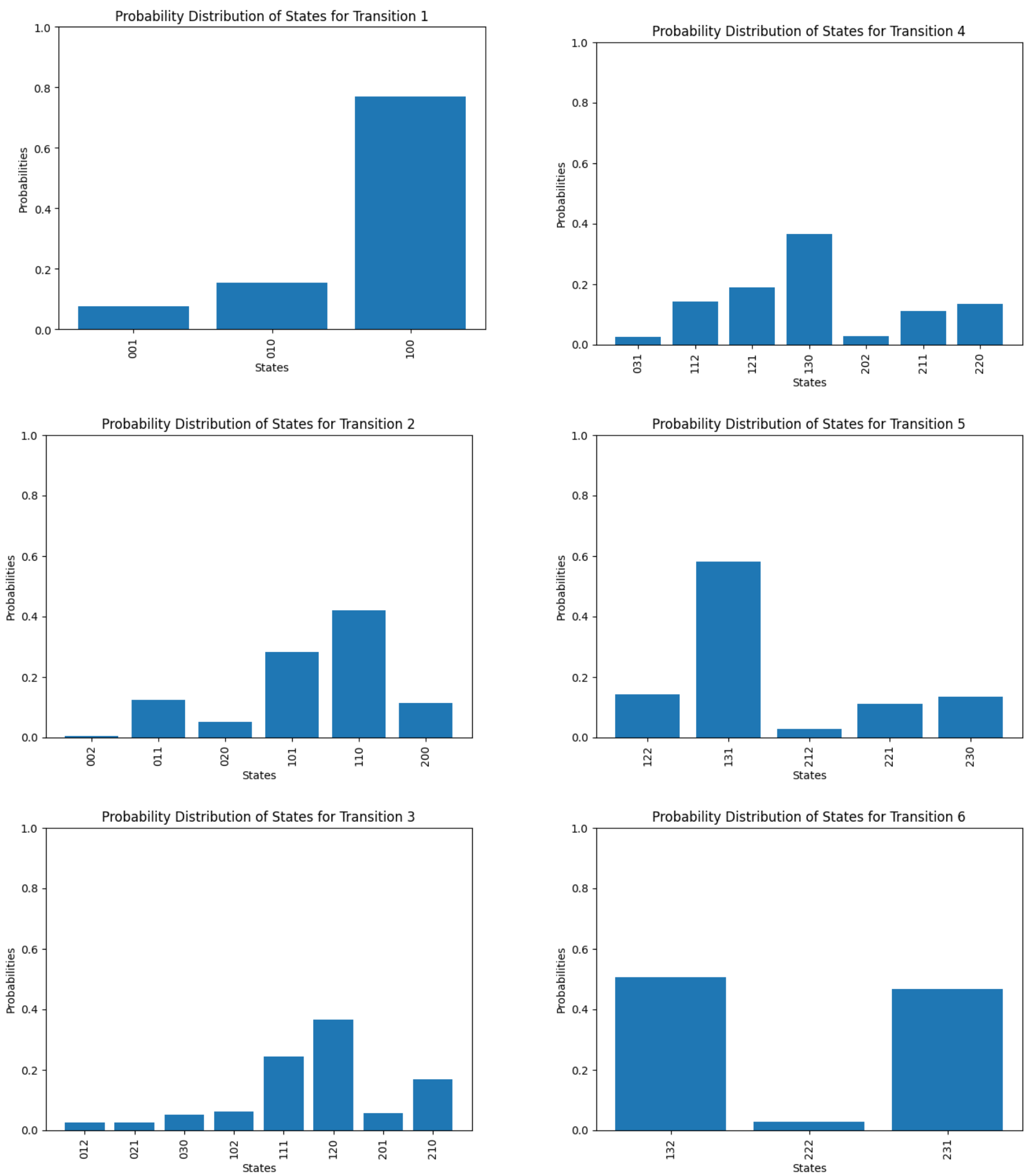

Figure 4.4: Transition probability distributions for each step n, $p = sP^n$. State $C = \{I, P, S\}$.

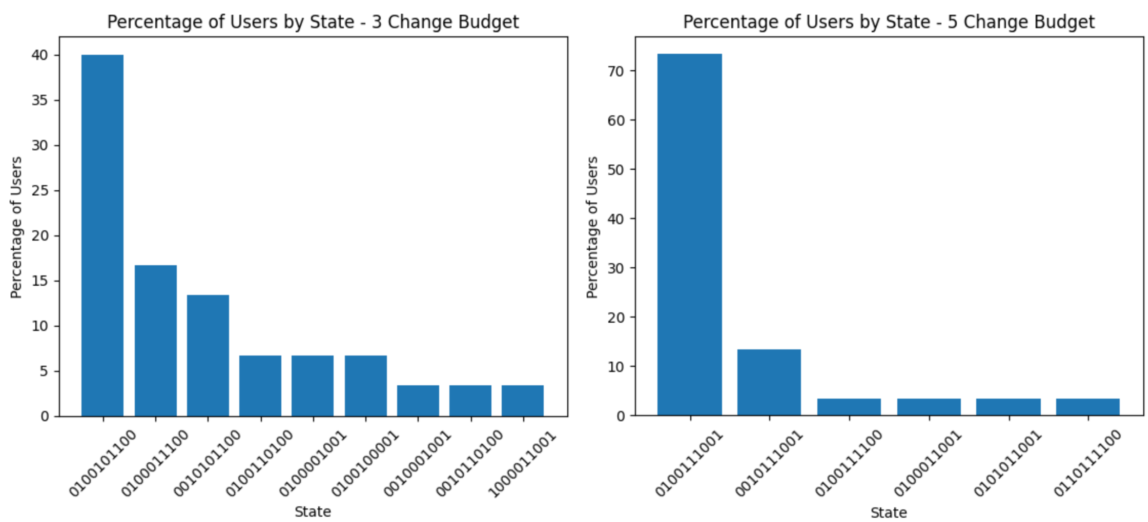

Figure 4.5: All budgeted configurations and the respective percentage of participants that chose them for 3 and 5 changes. A configuration is given by a 10 integer number with 0 representing an off position and 1 representing an on position for each toggle on the control panel.

CHAPTER 5

Discussion

5.1 A baseline interactable component was the most important factor for plausibility

As discussed in the results section, the most prevalent conclusion across all trials and experiments was the importance of having a baseline interactable component to the scene. The most popular first improvement was near interaction with 23/30 participants choosing to upgrade it first. Users seemed to view interaction as a baseline to plausibility while the other categories were more enhancements. One user stated, “Interactions are important for a real-life environment. If there are no interactions, then physics and shadows are not particularly relevant.” Being able to pick up the objects made them tangible, real, and more plausible overall. This lends itself to the idea that behavior is more important than appearance for virtual objects. Without being able to pick up the objects, participants had no way to really interact with the scene besides watching the objects change appearance or move with gravity. One of Azuma’s original claims was AR content needs to be interactive in real time, and this experiment provides some evidence that even being interactive in real time is not enough; virtual content in AR needs to be interactive through a users’ actions, showing the importance of Speicher et al.’s claim that “the user has to be in control” (Azuma, 1997; Speicher et al., 2019). Interestingly, the baseline interactive component that appeared necessary from the transition experiments did not apply to the inclusion of far interaction, which provides more manipulation ability for the user on the virtual content. When questioned on the different types of interactions, a few participants noted the advantages of far interaction and how it, “was easier for [participant] to use,” and, “gave [participant] a better perspective of how [the objects] moved in the environment.” However, far interaction was notably the least likely improvement to be chosen in the budget experiments. 25/30 participants chose not to upgrade it in their budgeted configurations, and almost half (14/30) of the accepted end states did not have interaction on I = 2, which includes Far Interaction. One participant explained, “no interactions would be extremely boring and a false representation of reality,” but, “far interaction is too far on the other side of the spectrum.” Despite multiple participants clearly stating that far interaction was their preferred way of moving the objects, the overwhelming majority determined it unrealistic and unimportant to the plausibility of the scene. Interaction of some type was necessary, supported by every participant choosing interaction of some type in the first budgeting experiment (three changes allowed), but beyond a baseline ability to interact with the objects through grabbing, advanced manipulation, regardless of its described utility, was not important for the plausibility of the experience. H1 is not confirmed by these results because high levels of interaction were not necessary;

only a low level was considered salient by the users. Following Skarbez's definitions, these experiments point to baseline adding an important component of fidelity to the user (Skarbez et al., 2017a). The virtual objects naturally had physical fidelity because they were loaded in 3D space when a participant entered the environment, but being able to touch and pick up the objects provided a functional fidelity. In the inverted definition of PI, the sensorimotor contingencies most important for plausibility seem to be a standard ability to, "perceive the objects by interacting with them," as Slater put it (Slater, 2009).

5.2 The reference setting of the environment had a strong impact on participants' transition choices

Although near interaction was the most important factor, the physics category was significantly voted the most important category across trials and users. After having a baseline interaction level, it appears improving the physics level to the maximum (as realistic as possible) was considered the most important. In the Markov chain probabilities for transitions 2 onward, improving physics was the highest probability transition choice. 18/30 participants ranked physics as the most important category. Regardless of interaction being most likely to be upgraded first and commonly referred to as the single most important improvement, physics was the most important category towards creating a sense of realism. One participant explained, "Physics was definitely the most important component for my sense of reality because I felt that was at the core of how objects would interact." Another echoed this by explaining that, "I think physics makes the world realistic which is the most important part," and "If the balls move and obey physics I'm then more inclined to look at them." Multiple participants mentioned that physics is what made interacting with the objects more intuitive, engaging, and realistic. Within physics, gravity was ranked the single most important improvement. 12/30 voted gravity as the most important closely followed by 8/30 voting for applied forces. Gravity is a strong prior because of its multi-perceptual nature and omnipresence, so it has a major impact on plausibility (Brübach et al., 2022). One user explained, "Gravity is one of the first forces we learn about and persists always. Collisions and applied forces are both important, but neither are as essential as gravity." Applied Forces was still considered important because it enabled the objects to, "move in a realistic ways," and, "the other options had less of an impact without applied forces enabled." Regardless of applied forces enabling users to interact more with the objects, gravity was considered more important, further displaying the need for only a baseline of interaction. These results also show the connection of both immersion and external plausibility in AR environments. In order to maintain high levels of immersion, users preferred improvements like near interaction and applied forces because they increased users' ability to be part of the virtual scene and observe the virtual objects. The overall preference users had towards physics shows that it provided psychological immersion. Here, psychological immersion is heavily tied into fidelity and external plausibility because objects that replicate real-world settings also replicate the psychological factors experienced in the real world, making the overall experience more plausible (Alexander et al., 2005). Although shadows were not found to be an important category, a few participants still noted their ability to give space and depth perception that further helped the virtual objects align with real-world expectations. The preference towards improvements that had the strongest priors, namely gravity, points to the reference frame being the real world for AR experiences, supporting Wienrich et al.'s findings that MR experiences having different reference frames based on their place on the continuum (Westermeier et al., 2023). Because the environment was already set to be the real world, users' expectations were on making the objects match the expected behaviors of those objects in the environment as opposed to building a new reference setting for the experience itself. These results confirm H2 as the physical behavior of the virtual objects proved to be salient to the users perception of their reality while graphical appearance (shadows) were considered far less important.

5.3 Informing Presence Models for AR

As Skarbez and others have pointed out, internal coherence or plausibility is less relevant in AR than in VR because building a consistent environment makes more sense when the environment is virtual itself, as opposed to the real world. With AR, the two important factors are external plausibility, consistency with the real world, and functional fidelity, how a user is able to interact with the virtual objects or tasks at hand. These factors fit into Wienrich and colleagues' model of plausibility coming from a variety of manipulation space factors and leading to qualia such as presence. Within the AR environment, users built their model of plausibility by first setting their reference frame to the real-world, followed by turning on a baseline perception of the objects by being able to grab them and, lastly, improving higher-order cognitive queues like matching the physics of the virtual objects to those of the real-world. This study shows bottom-up processing is used to experience AR environments because of the importance of real-world expectations being met in the perception of virtual content in AR.

CHAPTER 6

Limitations and Future Work

By integrating more virtual content into the study environment, it is possible to study a variety of other factors that define user experience in AR in a more concrete way. This study yielded preliminary findings about users' preferences in AR and how they can inform design choices and influence definitions of presence. It also displayed avenues for further research, including through more psychophysical user studies. Firstly, each main tested category can be further broken down into more individual factors to improve the depth and breadth of findings using transition analysis. For example, including multiple levels of gravity with different amounts of force to study how the realism of world forces impacts overall plausibility. Also, more categories could lead to stronger conclusions because each choice is between more options. One possibility is to include more rendering categories to continue to build a model of what makes virtual content in AR realistic. A rendering category that has levels adjusting how advanced the geometry of the objects are or how high quality the color and texture of the objects would offer new insights. The shadow category proved not to be considered important for this study, but some participants made interesting notes in their free responses that point to questions for future work. When asked about the shadow category specifically, multiple participants said only the soft shadows were helpful, and the hard shadows just lowered realism even compared to no shadows. Future studies can investigate whether the same sentiment applies to color, texture, and other rendering factors that can be adjusted to have options from low to high quality based on the required computation. Another avenue for future psychophysical studies in this field is designing transition paths to give insight towards a specific model of presence. Taking Latoschik and Wienrich's model, one could take each category of the manipulation space (sensation, perception, and cognition) and ask users to transition through improvements with the goal of gaining more insight into how specific areas of the manipulation space affect the qualia space, similar to how Brübach et al. used incongruencies (Brübach et al., 2022; Latoschik and Wienrich, 2022). To broaden these results to comment more on different types of environments within the MR spectrum, side by side psychophysical experiments could be run in AR and in VR similar to Westermeier et al. but including transition analysis experiments (Westermeier et al., 2023). Even within AR, adjusting how much physical content is present during the study could also inform more findings on how the reference frame adjusts based on environment. In this study's environment, it would be interesting to test if the plausibility level is different for users throwing virtual balls into a virtual hoop versus virtual balls into a real hoop, for example. There are also other small design improvements that would show improvements for future studies. Specifically within the physics category, the scene could have been designed in a way that better accentuates which factors are on and off even without user interaction. If there were a virtual avatar interacting with the objects that users could simply watch, it would help further separate interaction and physics.

CHAPTER 7

Conclusion

This is the first example of using a psychophysical study design with transition and budgeting analyses to deduce new information about presence in AR. This study follows previous work by Skarbez showing how user choices can inform design choices for virtual content and help build models of presence (Skarbez et al., 2017b). Based on Slater's idea for ideal studies for user experience, we used multiple sources of data (transition analysis, budgeting activities, and questionnaire responses) to determine that realistic user interaction is the most important baseline feature for AR experiences to feel real (Slater et al., 2022). We also found that strong priors, such as gravity, play a crucial role in AR environments. This attribute is unique to AR when compared to other MR environments. In AR, users are able to build a real-world reference frame for the virtual objects, giving emphasis to strong priors like gravity. Overall, we found plausibility and immersion to be heavily influenced by specific factors, near interaction and gravity, and a bigger overall theme, a real-world reference frame.

References

Alexander, A. L., Brunyé, T., Sidman, J., and Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, 5:1–14.

Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385.

Bergström, I., Azevedo, S., Papiotis, P., Saldanha, N., and Slater, M. (2017). The plausibility of a string quartet performance in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1352–1359.

Brübach, L., Westermeier, F., Wienrich, C., and Latoschik, M. E. (2022). Breaking plausibility without breaking presence - evidence for the multi-layer nature of plausibility. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2267–2276.

Chen, C.-H., Lee, I.-J., and Lin, L.-Y. (2016). Augmented reality-based video-modeling storybook of non-verbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions. *Computers in Human Behavior*, 55:477–485.

Doolani, S., Wessels, C., Kanal, V., Sevastopoulos, C., Jaiswal, A., Nambiappan, H., and Makedon, F. (2020). A review of extended reality (xr) technologies for manufacturing training. *Technologies*, 8:77.

Hofer, M., Hartmann, T., Eden, A., Ratan, R., and Hahn, L. (2020). The role of plausibility in the experience of spatial presence in virtual environments. *Frontiers in Virtual Reality*, 1.

Latoschik, M. E. and Wienrich, C. (2022). Congruence and plausibility, not presence: Pivotal conditions for xr experiences and effects, a novel approach. *Frontiers in Virtual Reality*, 3.

Ma, L., Zhao, Z., Chen, F., Zhang, B., Fu, L., and Liao, H. (2017). Augmented reality surgical navigation with ultrasound-assisted registration for pedicle screw placement: a pilot study. *International Journal of Computer Assisted Radiology and Surgery*, 12(12):2205–2215. PMID: 28779275.

Microsoft (2021). Hololens 2 gestures for authoring and navigating in dynamics 365 guides. Accessed: 2023-07-13.

Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Information Systems*, vol. E77-D, no. 12:1321–1329.

Milgram, P., Takemura, H., Utsumi, A., and Kishino, F. (1995). Augmented reality: a class of displays on the reality-virtuality continuum. In *Proc. SPIE 2351, Telemanipulator and Telepresence Technologies*.

Skarbez, R., Brooks, F. P., and Whitton, M. C. (2021a). Immersion and coherence: Research agenda and early results. *IEEE Transactions on Visualization and Computer Graphics*, 27(10):3839–3850.

Skarbez, R., Brooks, Jr., F. P., and Whitton, M. C. (2017a). A survey of presence and related concepts. *ACM Comput. Surv.*, 50(6).

Skarbez, R., Neyret, S., Brooks, F. P., Slater, M., and Whitton, M. C. (2017b). A psychophysical experiment regarding components of the plausibility illusion. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1369–1378.

Skarbez, R., Smith, M., and Whitton, M. C. (2021b). Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1535):3549–3557.

Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., and Oliva, R. (2022). A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, 3.

Slater, M., Spanlang, B., and Corominas, D. (2010). Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graph.*, 29.

Speicher, M., Hall, B. D., and Nebeling, M. (2019). What is mixed reality? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–15, New York, NY, USA. Association for Computing Machinery.

Unity (2016). Shadows in unity. Accessed: 2023-07-13.

Westermeier, F., Brübach, L., Latoschik, M. E., and Wienrich, C. (2023). Exploring plausibility and presence in mixed reality experiences. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2680–2689.

Wienrich, C., Komma, P., Vogt, S., and Latoschik, M. E. (2021). Spatial presence in mixed realities–considerations about the concept, measures, design, and experiments. *Frontiers in Virtual Reality*, 2.