Immersion and coherence in a visual cliff environment

Richard Skarbez* Frederick P. Brooks, Jr.† Mary C. Whitton‡

University of North Carolina at Chapel Hill

ABSTRACT

We report on the design and results of two experiments investigating Slater's Place Illusion (PI) and Plausibility Illusion (Psi) in a virtual visual cliff environment. PI (the illusion of being in a place) and Psi (the illusion that the depicted events are actually happening) were proposed by Slater as orthogonal components of virtual experience which contribute to realistic response in a VE. To that end, we identified characteristics of a virtual reality experience that we expected to influence one or the other of PI and Psi. We designed two experiments in which each participant experienced a given VE in one of four conditions chosen from a 2x2 design: high or low levels of PI-eliciting characteristics (that is, immersion) and high or low levels of Psi-eliciting characteristics. We use the term "coherence" for those characteristics which contribute to Psi, parallel to the use of "immersion" for characteristics that contribute to PI. We collected both questionnaire-based and physiological metrics. Several existing presence questionnaires could not reliably distinguish the effects of PI from those of Psi. They did, however, indicate that high levels of PI-eliciting characteristics and Psi-eliciting characteristics *together* result in higher presence, compared any of the other three conditions. This suggests that "breaks in PI" and "breaks in Psi" belong to a broader category of "breaks in experience," any of which result in a degraded user experience. Participants' heart rates, however, responded markedly differently in the two Psi conditions; no such difference was observed across the PI conditions. This indicates that a VE that exhibits unusual or confusing behavior can cause stress in a user that affects physiological responses, and that one must take care to eliminate such confusing behaviors if one is using physiological measurement as a proxy for subjective experience in a VE.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented and virtual realities; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Evaluation/methodology; I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

1 INTRODUCTION

One construct that many researchers have used to reason about and evaluate user experiences in virtual environments independent of the particular task is *presence*. To address some perceived problems with the presence construct, Slater proposed instead two new constructs, Place Illusion and Plausibility Illusion [25].

We present the results of two studies investigating the relationship between Place Illusion and Plausibility Illusion. These results suggest that having both Place Illusion and Plausibility Illusion together results in substantial improvement using existing presence metrics. We also further develop the theory relating to Place Illusion and Plausibility Illusion, particularly in explicating the term coherence, which is to Plausibility Illusion as immersion is to Place Illusion.

*e-mail: skarbez@gmail.com

†e-mail:brooks@cs.unc.edu

‡e-mail:whitton@cs.unc.edu

2 BACKGROUND

The presence construct is most commonly defined as the sensation of "being there" [10] [13] [18] [19] [20] [33]. Presence, then, is a quale (singular of qualia), part of the subjective character of an experience. As such, it is difficult to measure directly. Researchers have previously studied it with subjective questionnaires [30] [32], physiological metrics [14], and behavioral metrics [28]. In Meehan's dissertation, he describes an ideal measurement of presence that would be reliable, i.e. producing repeatable results, both within and between subjects, valid, i.e. demonstrated to correlate with the subjective feeling of presence, multi-level sensitive, and objective [15]. One could go further and suggest that such an ideal metric should also be measurable in real time and without modification to the scenario. To date, no measure of presence exists that meets all of these criteria and is also generalizable across VEs.

Despite its popularity, there are concerns about using the presence construct to evaluate virtual environments. One problem is that the sense of presence can be elicited from non-immersive, non-virtual scenarios, such as a non-immersive video game, a movie, or a book. Since presence can be elicited by such disparate types and characteristics of experiences, is it by itself a useful tool for evaluating virtual environments?

A second problem is that it has never been established that more presence significantly correlates with improved effectiveness (e.g. improved task performance, faster training, or training to a higher level of competence) of a VE. Several researchers have sought evidence of such a relationship as it pertains to task performance, but the results have been inconclusive at best. One study that explicitly looked for a link between presence and task performance can be found in [26], in which participants had to observe, remember, and replicate a series of moves in a tri-dimensional chess game. The authors did not find that task performance was associated with presence. Witmer and Singer reported mixed results [32]. Welch argued that there is no inherent logical connection between presence and task performance [31].

To address such problems regarding the presence construct, Slater proposed a new theoretical framework for evaluating VEs in [25]. Specifically, he proposed that the goal of a VE should be to stimulate participants to respond to situations in the VE in the same way they would respond to a similar situation in the real world (dubbed "respond-as-if-real", or RAIR).

He also theorized that such realistic behavior arises from two qualia—subjective mental states—in the mind of the participant: Place Illusion (PI) and Plausibility Illusion (Psi). PI is a new term for presence as it is most commonly understood, as "being there." Slater defines PI as, "the strong illusion of being in a place in spite of the sure knowledge that you are not there." Psi is a new construct, defined as "the illusion that what is apparently happening is really happening (even though you know for sure that it is not)." In short, PI is the illusion that, "I am in the place that I perceive," and Psi is the illusion that, "The scenario I am experiencing is real." When a user has both of these, she experiences the VE as a real place (by virtue of experiencing Psi) that she is really in (by virtue of experiencing PI), and so she would be expected to act realistically.

Since PI and Psi are qualia, attempts to measure either construct objectively are subject to the same difficulties that face attempts to measure presence. In his 2010 SIGGRAPH paper [27], Slater uses an analogy to color metamers — colors that appear the same to the visual system but have different spectral power distributions — to propose a method of measuring PI and Psi using VEs simulated within VEs. Conceptually, this is a very appealing approach. However, owing to the complexity of experimental setup, the large amount of data needed, and the inherent dissimilarity to real-world experiences, the metamer technique may not be appropriate for evaluating practical VR applications. Instead, we developed experiments to explore whether any existing techniques for evaluating presence could distinguish between and measure PI and Psi. Some relevant previous efforts are included in the discussion in Section 7.

3 PI:IMMERSION::PSI:COHERENCE

In the original PI/Psi paper, Slater states, "Immersion provides the boundaries within which PI can occur" [25]. Immersion, here, is defined in terms of the set of sensorimotor valid actions supported by the system. Valid actions are those actions that a user can perform that result in changes to his perception or to the state of the VE, such as moving his viewpoint. By this definition immersion is strictly a function of system characteristics and possible user actions [29]. Strictly speaking, then, our experimental factor is not PI, but rather immersion as so defined.

A parallel argument can be made regarding Psi. Psi arises to the extent that a participant probes the Psi-inducing (or the Psi-breaking) characteristics of the environment. While the concept of immersion is well-established in the VE research community, there does not exist an equivalent concept for reasoning about the degree to which the virtual scenario behaves in a reasonable or predictable way. Following Skarbez [23] [24], we use the term coherence for this concept.

In our experiments, we sought to identify measures that could distinguish between the effects of PI and those of Psi on participants. We designed between-subjects experiments where the factors were different levels of immersion and coherence that were expected to elicit differing levels of PI and Psi, respectively. Henceforth, these factors will be referred to as LowPI and HighPI, and LowPsi and HighPsi, respectively.

We sought to identify system characteristics that would affect only (or, at least, mainly) one or the other of immersion or coherence. Broadly, immersion factors deal with the physical interface between the user and the VE, e.g. tracking, display, and input devices or techniques. Coherence factors deal with the appropriateness of the scenario and of interactions between system users, virtual characters, and virtual objects; e.g. world physics, behavior of virtual humans, and "glitches."

Note that coherence (and therefore Psi as well) is inextricably dependent on the particular scenario being represented in the VE. For example, if one is specifically told that the VE represents a real-world scenario (or, absent priming, the user expects "normal" behavior), and attempting to jump sends your avatar soaring hundreds of feet in the air, this would be unexpected and shocking behavior. However, if the user was told that this VE represents a future city on a world with very low gravity, or that he is wearing special rocket boots, this would be normal behavior, or at least plausible behavior. In the former case, this startling behavior would be perceived as a failure of coherence and would decrease the user's feeling of Psi. But in the latter case, the very same behavior would be perceived as a confirmation of the reality of the scenario presented, and would likely *increase* the user's feeling of Psi.

4 EXPERIMENT 1

This experiment used a 2x2 between-subjects design with multiple outcome measures. We chose a between-subjects design because Khanna and colleagues observed [8] that responses were not symmetric across conditions in a visual cliff experiment such as the one used here. That is, the difference in effect between the first exposure and subsequent exposures to the visual cliff stressor cannot be entirely compensated for by counterbalancing order. Also, Meehan exposed participants to a visual cliff environment twelve times over four days [15], finding that physiological responses decreased with subsequent exposures, but not to zero.

In this first experiment, a single system characteristic was varied to create different levels of immersion and coherence. The Immersion factor was manipulated by changing the effective field of view of the head-mounted display (HMD). In the HighPI conditions, the field of view was 60° diagonal, the maximum supported by the HMD. In the LowPI conditions, a virtual mask reduced the effective field of view to 30° diagonal. By construction, changing the field of view must change the level of immersion, because it changes the sensorimotor actions supported by the system. For example, a user in the restricted field of view condition might have to turn his head to see a virtual object that a user in the normal field of view condition could see without any head movement at all. Hendrix and Barfield showed that field of view has a significant effect on presence (defined in that paper as "being there", which corresponds to PI), and Arthur also observed that presence scores trended lower with restricted field of view [6] [1]. Slater and colleagues showed that participants who were explicitly trying to increase PI improved field of view significantly more often than participants who were trying to increase Psi [27]. Note that the behavior of the environment remains unchanged regardless of the field of view condition. Based on the definitions of PI and Psi, this manipulation should therefore have no effect on Psi.

The Coherence factor was manipulated by changing the physical behavior of the environment. In [25], Slater theorized that a key component of Plausibility Illusion was the "correlational principle". That is, events occurring in the environment should react or appear to react in response to the user's actions in the environment. This experiment attempted to directly manipulate the participant's sense of this correlation. Participants were instructed at several points that they must perform a task in order to advance to the next stage of the experiment. In the HighPsi conditions, these instructions were true: advancement through the experiment depended upon participant behavior. In the LowPsi conditions, the instructions were false: the advancement events were controlled by a software timer, and the participant's actions had no effect. There were three such events: participants were told that when they finished dropping ten balls into a receptacle, the elevator would arrive to take them to the next room, that the elevator would descend when they pressed the correct button, and that the door to the visual cliff room would open when they picked up a ball from a particular pedestal.

There is no experimental precedent for modifying the behavior of the environment in response to user's actions in such a way in order to manipulate Psi. As above, however, from the definitions of PI and Psi, we argue that changing the environment's behavior should have no effect whatsoever on PI; if it has any effect at all, it must be on Psi.

A metric that could distinguish between the effects of PI and the effects of Psi would show a substantial difference between measured values in the LowPI-HighPsi condition and the HighPI-LowPsi condition, that is, the cross diagonal in the 2x2 design. In these conditions, the participant is expected to feel a high level of either PI and Psi, and a low level of the other. Differences along the main diagonal (Low-Low, High-High) would indicate only whether a strictly "better" VE differs from a strictly "worse" one. We designed our conditions on the cross diagonal to represent experiences that are overall of approximately the same level of quality. A measure that exhibits a difference between these conditions, then, would be evidence that that measure responds differently to PI and

Psi, and therefore, that PI and Psi are separable constructs.

4.1 Participants

Thirty-two participants (8 female, 24 male) took part in this experiment. All were recruited from an introductory undergraduate psychology class and received course credit. Their average age was 19.5 years. Participants successfully passed screening for uncorrected vision problems, a history of seizures or strong motion sickness, inability to walk without assistance, deafness, self-reported pregnancy, and English comprehension. This experiment was approved by the UNC Behavioral Institutional Review Board.

4.2 Materials

The experiment took place in an immersive virtual environment. Participants wore an nVisor SX HMD with 1280x1024 resolution per eye and native 60° diagonal field-of-view, with attached stereo headphones. The head and right hand of each participant were tracked using the 3rdTech Hiball 3000 optical tracking system. Participant physiological reactions were measured using the ProComp Infiniti wireless telemetry system from Thought Technologies, Ltd. A Pentium D dual-core 2.8GHz computer with an NVIDIA GeForce GTX 280 GPU and 4GB RAM rendered the virtual environment and recorded logs. The application was implemented using the UNC-developed EVEIL2 library that communicates with the Gamebryo software game engine from Gamebase USA. The Virtual Reality Peripheral Network (VRPN) interface handled tracker communication and logging of physiological signals and tracker data.

4.3 Metrics

Participants' experiences were evaluated using both in-test and post-test metrics. During the test, we collected electrocardiogram (EKG), skin conductance (SCR), and skin temperature. For both SCR and skin temperature, the mean and standard deviation were computed for each stage of the experiment. From the EKG data, several measures of heart rate variability (HRV) were computed. Candidate R spikes were identified algorithmically, and the signals were then processed by hand to ensure that the time stamps of those spikes were recorded correctly. These data were then used to compute metrics in both the time domain (mean R-R time interval, mean heart rate, and percentage of R-R intervals that are less than 10/30/50ms) and the frequency domain (power in the low-frequency band (LF), power in the high-frequency band (HF), and the LF/HF ratio) domains. These metrics were also computed for each stage of the experiment.

The frequency domain analysis merits further discussion. The distribution of power as a function of frequency is computed by power spectral density (PSD) analysis of the time series of R spikes. The HRV literature defines the low-frequency band as 0.04-0.15 Hz, and the high-frequency band as 0.15-0.4 Hz. The physiological significance of these bands is that both sympathetic nerve activity (reflecting stress) and parasympathetic nerve activity (reflecting rest/normal conditions) increase LF spectral power, but only parasympathetic nerve activity increases HF spectral power. An increase in the LF/HF ratio, then, indicates that the participant is experiencing increased stress.

Post-test, participants completed the Witmer-Singer Presence Questionnaire [32], the Slater-Usoh-Steed Presence Score [30], the Virtual Experience Tool of Chertoff, Goldiez, and LaViola [5], the Arrival/Departure questionnaire of Kim and Biocca [9], as well as a short experimental questionnaire intended to measure participants' levels of Psi, whose questions are listed in Table 1. Note that the Psi questionnaire is ad hoc and was constructed for this experiment. At this point, we can only argue for its face validity for responding to Psi. All participants were also debriefed by an experimenter.

4.4 Experimental Procedures

Upon arriving, participants first reported to an office, where they were screened by an experimenter, signed informed-consent forms, and completed pre-experiment questionnaires on a PC using the Qualtrics web application (Qualtrics, Provo UT). After completing the questionnaire, participants were equipped with the ProComp Infiniti and escorted to the lab, where they donned the NVIS HMD and started the experiment.

The experiment itself consisted of three stages (Figure 1 shows the common environment.):

Stage 1. Participants familiarized themselves with the virtual environment by picking up, carrying, and dropping balls into receptacles.

Stage 2. Participants took a virtual elevator to an office-like environment, where they were presented with additional balls to drop on targets.

Stage 3. The door to the Pit room opened and participants were exposed to the virtual visual cliff environment, where there were several more balls to drop on targets on the floor below.

The experiment ended when participants re-entered the office-like room from the Pit room. The total time in the virtual environment was approximately ten minutes.

Participants then doffed the HMD and the ProComp hardware, and returned to the office, where they filled out post-test questionnaires on the PC and were debriefed orally.

4.5 Results

Experiment 1 failed to identify any metrics that significantly distinguished between PI effects and Psi effects. Analysis identified several possible reasons why that experiment failed to generate a significant result, even if PI and Psi are in fact distinguishable in outcome measures. The most likely reason is that the factor levels in that experiment were not sufficiently different, so that individual differences obscured any effect. We addressed those concerns in a subsequent experiment, described in the next section.

5 EXPERIMENT 2

In Experiment 1, only a single system characteristic was manipulated to create the two levels of each factor. Immersion and coherence, though, are both multi-dimensional constructs. For example, immersion depends on the combination of multiple sensory modalities (visual, auditory, tactile, etc.), and each of these sensory modalities is dependent on multiple system characteristics (for the visual channel, field of view, display resolution, display size, display latency, etc.). It is possible that the two levels of immersion in Experiment 1 were sufficiently close together in this multi-dimensional immersion space that there was no practical distinction to participants. A similar rationale applies for coherence. For Experiment 2, multiple system characteristics were varied simultaneously to create more sharply distinct high and low levels of immersion and coherence.

We argue that there are several meaningful categories of immersion failures. One, *reduced fidelity*, occurs when the sensory data stream is somehow impoverished—for example, a limited field of view or monophonic sound. Another, *sensory conflict*, occurs when sensory data from different modalities conflict—for example, a participant is experiencing a virtual spacewalk, but is hearing the sounds from the office or lab she is actually in. A third, *missing or invalid cues*, occurs when the sensory data stream is interrupted, or contains invalid data—for example, when tracking is lost. In this experiment, then, the immersion factor is manipulated as follows: In the HighPI conditions, the field of view of the HMD will be the maximum supported by the device (60° diagonal), passive haptics will be used to provide tactile feedback to the participant, and scenario-appropriate spatial sound cues will appear in the environment. In the LowPI conditions, the effective field of view of

Table 1: Ad hoc Plausibility Illusion questionnnaire

The environment's behavior was the same as I would expect in the real world.
I could anticipate how the environment would respond to my actions.
The environment's behavior was surprising or unexpected.
The environment's behavior was inconsistent.
I forgot that the environment was virtual.
The behaviors of the environment were appropriate.
Interacting with the Pit environment was the same as interacting with a real environment.
Interacting with the VE was more like interacting with… (1: a video game, 7: a real room)

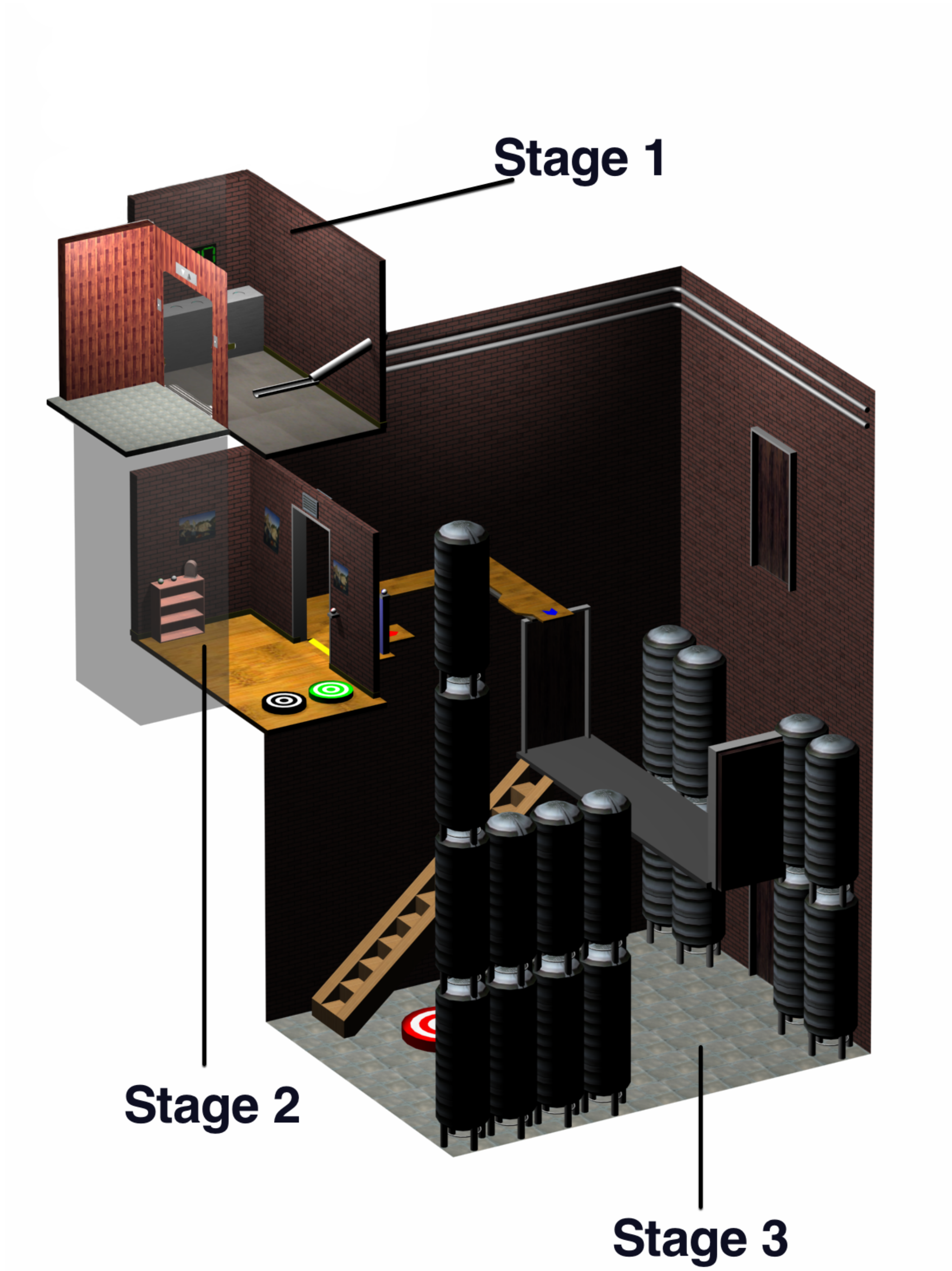

Figure 1: The virtual environment for Experiment 1.

the HMD will be restricted to 30° by use of a virtual mask, no passive haptics will be used, and there will be no sound other than the experimental instructions delivered through the headphones.

Failures of coherence can also be meaningfully categorized. *Physical coherence* can fail—that is, the laws of physics as we know them do not seem to apply, e.g., an object falls through the virtual floor, a rolling ball is never slowed by friction. Also, *narrative coherence* can fail—virtual characters or the scenario itself do not abide by the expected rules of behavior from everyday life, e.g., a character performs repetitive actions or otherwise does not respond meaningfully to your presence, actions that you are led to believe will cause one event in fact cause a different event.

In this experiment, then, the coherence factor is manipulated as follows: In the HighPsi case, physical objects (balls) behave as one would expect them to, and the experimental instructions are in fact valid. In the LowPsi case, physical objects behave in an apparently random fashion (dropped balls can fall with normal acceleration due to gravity, accelerate much faster or much slower than normal, remain stationary, or float slowly upward), and the experimental instructions are false (the scoreboard which claims to show the number of balls you have dropped in fact never changes, the elevator teleports instantly rather than seeming to work as a normal elevator, and the door which claims to open when an object is moved in fact operates on a timer, forcing the participant to wait).

Regarding individual differences among participants, several of the most promising metrics that may be able to distinguish the effects of PI and Psi are heart rate variability (HRV) metrics. These include mean heart rate, power in different frequency bands of the electrocardiogram (EKG) signal, and variability of the beat-to-beat interval. Analysis of these metrics following the Experiment 1 showed no significant main effects. Discussion with an HRV expert, however, indicated that these metrics vary wildly from person to person based on a variety of personal characteristics, including age, sex, weight, physical fitness, among others. As a result, comparing aggregate HRV metrics for a heterogeneous group is unlikely to yield any meaningful result. Therefore, in an effort to reduce individual differences within the participant pool, participation in Experiment 2 was restricted to non-smoking, non-drug-using undergraduate males, ages 18-22, who exercise 3-5 times per week.

5.1 Participants

Thirty-two male participants took part in this experiment. The average age was 20.1 years. Participants were additionally screened as in Experiment 1.

5.2 Materials

The materials are the same as in Experiment 1, except that the virtual environment has an additional room, as described in 5.4.

5.3 Metrics

During the experiment, participants physiological responses were monitored as in Experiment 1. Post-test, participants completed the Witmer-Singer Presence Questionnaire [32] and a modified Slater-Usoh-Steed Presence Score [30].

5.4 Experimental Procedures

Participants underwent pre-experiment screening, filled out consent forms and questionnaires, and donned the VR equipment, all as in Experiment 1.

This experiment consisted of five stages. (An illustration of the environment is in Figure 2.):

Stage 1. Participants familiarized themselves with the virtual environment, playing a Simon-like memory game. Stages 2, 3, and 4 are similar to the three stages in Experiment 1.

Stage 2. Participants took a virtual elevator to a room where they had to pick up balls and drop them in targeted receptacles.

Stage 3. Participants took a virtual elevator to an office-like environment, where they were presented with additional balls to drop on targets.

Stage 4. The door to the Pit room opened and participants were exposed to the virtual visual cliff environment, where there were several more balls to drop on targets on the floor below.

Stage 5. Participants returned to the elevator, returned to the Simon room, and played the game again. After 3 minutes, the experiment ended. The total time in the virtual environment was approximately fifteen minutes.

Participants then doffed the HMD and the ProComp hardware, and returned to the office, where they filled out post-test questionnaires on the PC and were debriefed orally, all as in Experiment 1.

6 Experiment 2 Results

We performed Bayesian data analysis on the study data. In the Bayesian method of analysis, all variables are considered as part of a single overall model, where all the stochastic equations are evaluated simultaneously, rather than one at a time. Unlike traditional analysis, that is, null-hypothesis testing, there is no single value such as a p-value that determines whether the result is “significant”. Instead, we report the posterior probabilities and readers are free to interpret those probabilities for themselves. Posterior probabilities near 50% indicate that both outcomes are approximately equally likely, so we refer to posterior probabilities between 50% and 70% as offering *negligible evidence* for the stated hypothesis. Similarly, for convenience, we refer to probabilities above 70% as offering *little evidence* in favor of a hypothesis, 75% as offering *some evidence*, probabilities above 80% as *good evidence*, and probabilities above 90% as *strong evidence*. (These probabilities can also be less than 50%, providing evidence in the corresponding way for the inverse hypothesis.) This manner of describing the results follows Bergström et al. [3].

In the remainder of this section, we present the results organized as claims about the data grouped with the supporting evidence for each claim.

6.1 There is good evidence that the Witmer-Singer Presence Questionnaire responds to higher levels of immersion as a main effect.

This may be somewhat unsurprising, as Place Illusion is quite closely related to traditional notions of presence, which is what the PQ was designed to measure. Nonetheless, there is an 80.5% probability that participants in HighPI conditions reported higher PQ scores than participants in LowPI conditions.

Table 2: Mean count of high scores (6 or 7) on the Witmer-Singer PQ for each condition

	LowPI	HighPI
LowPsi	9.6	10.4
HighPsi	10.0	12.9

6.2 There is negligible evidence that the Slater-Usoh-Steed presence questionnaire (SUS) responds to increased immersion as a main effect.

The posterior probability that participants in the HighPI conditions reported higher scores than participants in the LowPsi conditions is 52.1%.

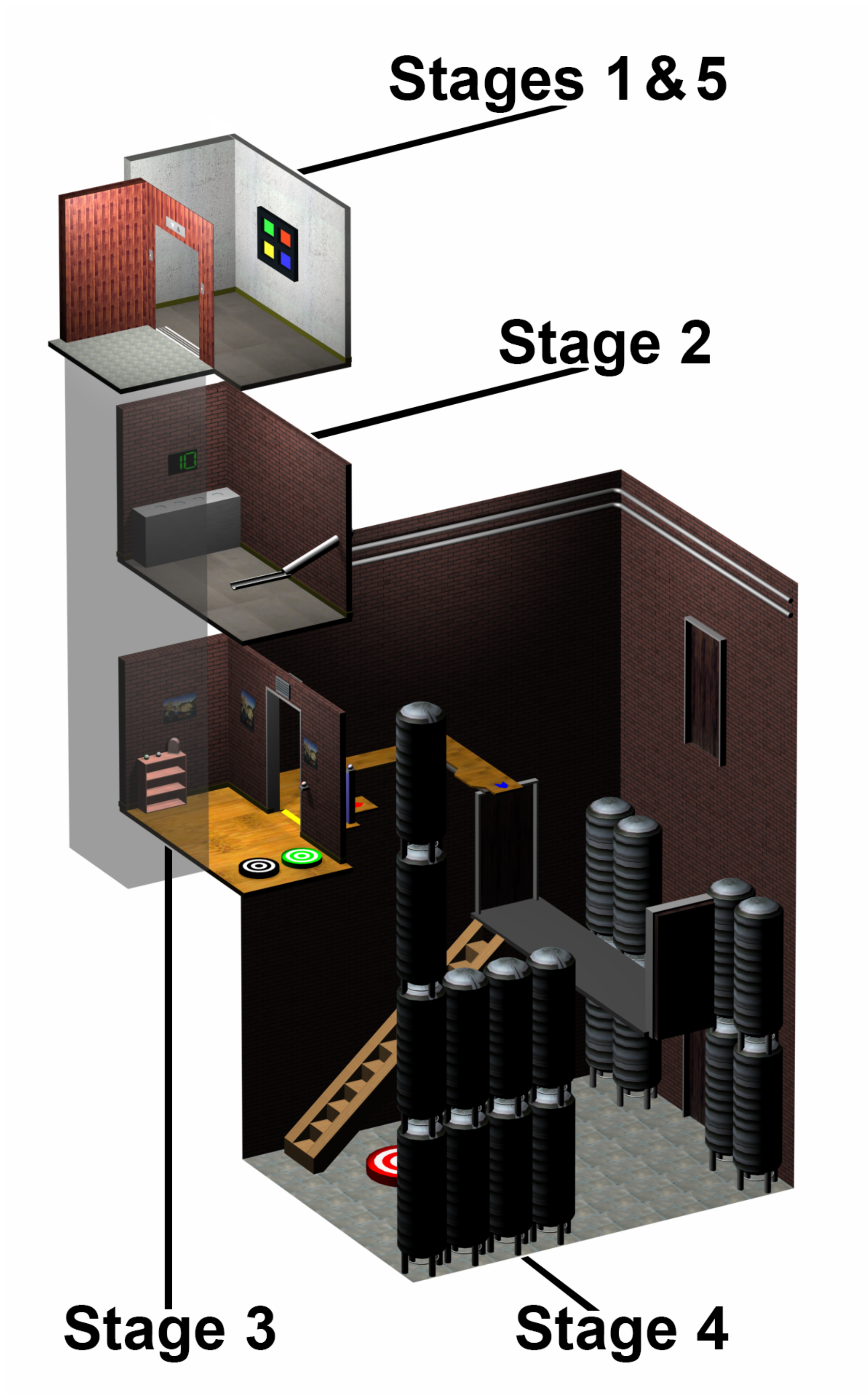

Figure 2: The virtual environment for Experiment 2.

Table 3: Mean count of high scores (6 or 7) on the SUS questionnaire for each condition

	LowPI	HighPI
LowPsi	4.0	2.9
HighPsi	3.3	4.3

6.3 There is little evidence that either questionnaire responds to increased coherence as a main effect.

The probabilities that participants in the HighPsi conditions reported higher questionnaire scores than participants in the LowPsi conditions are 61.7% and 71.3% on the SUS questionnaire and the PQ, respectively.

6.4 When high levels of PI and Psi are present *together*, questionnaire scores increase.

For each of the SUS and the PQ questionnaires, there is at least some evidence that participants in the HighPI-HighPsi condition reported higher scores than in any of the other three conditions. On the SUS questionnaire, the posterior probability is 79.8% that participants in the HighPI-HighPsi condition scored higher than participants in the other three conditions combined; for the PQ questionnaire, there is strong evidence, with a 96.7% probability that participants in the HighPI-HighPsi condition scored higher. (See Tables 2 and 3 for mean scores.)

6.5 There is good evidence that SUS questionnaire scores are higher for matched (LowPI-LowPsi and HighPI-HighPsi) than mismatched conditions.

There is 86.6% probability that SUS scores are higher for participants in the matched conditions than in the mismatched conditions. There is little evidence for this effect on the PQ, with a 66.1% posterior probability.

6.6 There is good evidence that several PQ subscores respond differently to immersion and coherence.

There is good evidence (86.9% posterior probability) that the PQ Naturalness subscore is higher for participants in the HighPsi conditions than the LowPsi conditions. There is negligible evidence (54.2%) that it responds to immersion.

On the other hand, there is good evidence that both the audio (85.1%) and haptic (83.3%) subscores are higher for participants in HighPI conditions (HighPI-LowPsi and HighPI-HighPsi combined) than in LowPI conditions (LowPI-LowPsi and LowPI-HighPsi combined).

6.7 There is strong evidence that exposure to bad coherence (i.e., glitches) causes heart rate to increase.

In Stage 1 of the experiment, coherence was the same for all participants. This stage was used to measure the baseline heart rate for all participants. In Stage 2, though, participants in the LowPsi conditions were exposed to a series of coherence failures, while those in the HighPsi conditions were not. There is strong evidence that LowPsi participants experienced an increase in heart rate in Stage 2, with a posterior probability of 87.1%. (See Figure 3.)

6.8 There is negligible evidence that the increase in heart rate caused by exposure to the Pit is dependent on either PI or Psi separately.

The effect of the Pit on heart rate can be considered either by comparing to the baseline (Stage 4 - Stage 1) or to the previous stage (Stage 4 - Stage 3). In neither case is it probable that the size of the increase is greater for HighPI vs. LowPI (38.4%, 46.3%), or for HighPsi vs. LowPsi (59.9%, 38.9%).

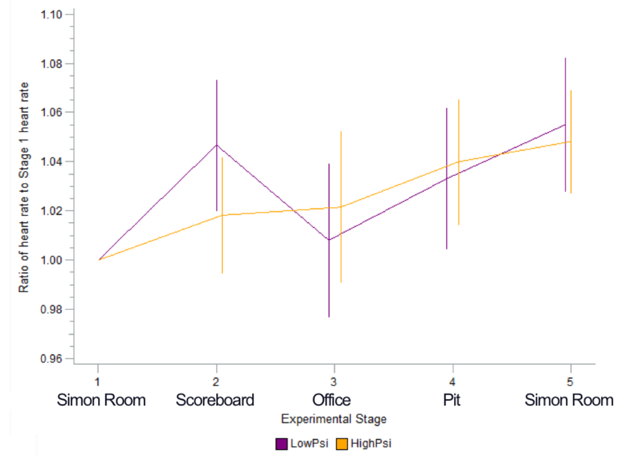

Figure 3: Comparing coherence conditions by heart rate in each experimental stage.

6.9 There is little evidence that the ad-hoc Psi questionnaire administered here responds to increased coherence, but negligible evidence that it responds to increased immersion.

There is only little evidence that the experimental Psi questionnaire (from Table 1) responds to higher coherence (71.0%). That said, it is more probable that it responds to coherence than immersion, for which there is negligible evidence (51.1%). Further, there is some evidence that it does respond to both higher immersion and higher coherence together (79.4%), and some evidence that it responds negatively to lower immersion and lower coherence together (22.1%).

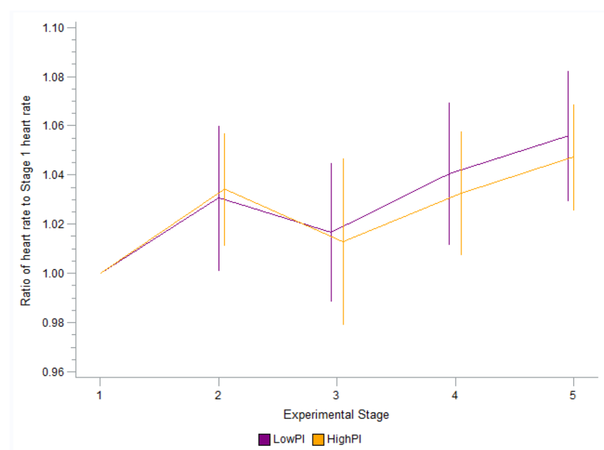

Figure 4: Comparing immersion conditions by heart rate in each experimental stage.

7 Discussion

As stated above, we observed that participants have higher PQ and SUS questionnaire scores when both presence and immersion are high; none of the other conditions is substantially different from another. This demonstrates that when coherence and immersion

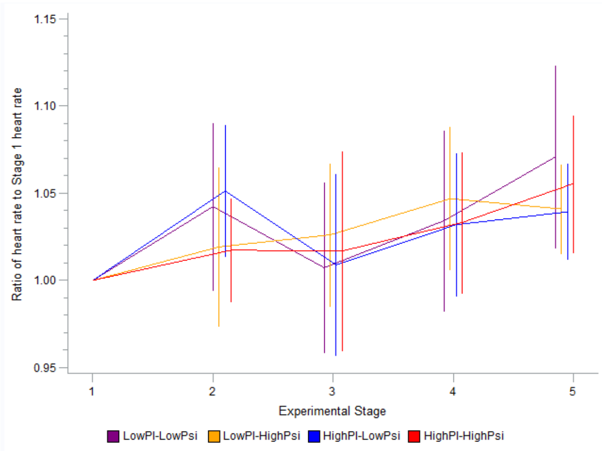

Figure 5: Comparing all four conditions by heart rate in each experimental stage.

are present together, participants report significantly higher levels of presence. Further, the scores for LowPI-HighPsi and HighPI-LowPsi conditions are not substantially different, which may indicate that both PI and Psi are of roughly equal importance, at least as regards scores on the PQ. Furthermore, neither of these is substantially different from the LowPI-LowPsi condition, indicating that any noticeable failure of either immersion or coherence causes a substantial drop in presence.

In comparing these observations to previous studies measuring the relative influences of different aspects of experience on presence using self-report methods, these are in line with those of Lessiter et al. [11]. That paper suggested that immersion factors and coherence factors contributed roughly equally to presence. On the other hand, our results agree less with those of Schubert and Regenbrecht [21], which suggested that immersion factors contributed roughly twice as much as coherence factors.

(Note that neither of the papers mentioned above described their factors in terms of PI and Psi. Each reported a list of factors and the associated amount of variance they explained, and I have characterized those factors as either immersion or coherence as follows. From Lessiter et al., we consider Sense of Physical Space and Negative Effects—explaining a combined 19.6% of variance—to compose immersion, and Engagement and Ecological Validity—18.6%—to compose coherence. From Schubert et al., we consider Spatial Presence, Immersion Quality, Interface Awareness, and Exploration Factors—explaining a combined 34% of variance—to compose immersion, and Involvement, Drama, Predictability and Interaction, and Realness—16%—to compose coherence.)

The observation that there is no substantial difference among any of the non-HighPI-HighPsi conditions echoes previous arguments that presence is a binary construct, that either one has it or one doesn't. Those arguments focused on presence as a moment-to-moment sensation, though: At any given moment, the thinking goes, you are present in exactly one place, whether it is the virtual environment or the real world lab or an imaginary space. What was observed in this study, though, is something different. Here, if your experience is "good enough," participants remember and report a high level of presence after the fact, and if it is not, then they report a lower level. This suggests that self-report and/or post-facto measures of presence, at least, favor experiences that are of a consistent level of acceptable quality, and penalize experiences that have failures, glitches, or breaks that draw a user's attention and can linger in the memory. This provides a piece of practical advice for designers and builders of virtual reality systems: Only build those features into a VE or a virtual environment system which you are capable of delivering at a consistently high level. Adding virtual humans to an environment, for example, might actually reduce the quality of an experience and lower a user's feeling of presence if in the process distracting or unnatural behavior is also introduced.

The evidence that matched conditions result in higher scores on the SUS questionnaire than mismatched conditions may further suggest an effect where users prefer an environment of consistent quality (whether high or low) to one that is inconsistent. In the matched conditions, the sensory representation of the environment and the behavior of objects in it are of the same level of quality (whether good or bad), while in the mismatched conditions, the environment looks realistic but behaves badly or vice versa. This difference is evidence that consistency and predictability are more important to users—at least as far as the feeling of presence is concerned—than having the best possible environment, if that level of quality cannot be maintained throughout.

These results lend credence to an "uncanny valley" effect in virtual environments. In Mori's original presentation of the uncanny valley, it seems to me that the problem is not inherently the human-like appearance of an entity, but rather the *mismatch* between its appearance and its behavior [16]. Ishiguro has extended the concept of the uncanny valley with a "synergy effect" in this way (essentially, the match between appearance and behavior) [7]. If an entity looks exactly like a human and behaves exactly like a human, there should be no loss of affinity. For all intents and purposes, it would be a human.

Following this logic, the uncanny valley theory no longer has to be restricted to humanoid characters. One would likely feel more affinity for a dog character if it behaved like a real dog, for example. And, going further, we can consider the environment itself as a character. If the environment is treated as a character, these results suggest that one would feel greater affinity for it if its behavior matched its appearance. Furthermore, these characteristics of an environment—its behavior and its appearance ("appearance" in all sensory modalities)—map neatly onto coherence and immersion.

Figure 6 depicts the skin conductance response for each condition and each stage. Notable is the fact that the LowPsi conditions do not exhibit a spike in Stage 2 as was seen in heart rate. Skin conductance has generally been considered to be less suitable as a measure of stress in virtual environment due to its slow onset and slow decay [15]. However, these results suggest that it might be useful to gather this information, as heart rate is affected by both stressful and "confusing" situations, whereas skin conductance seems only to respond to stress.

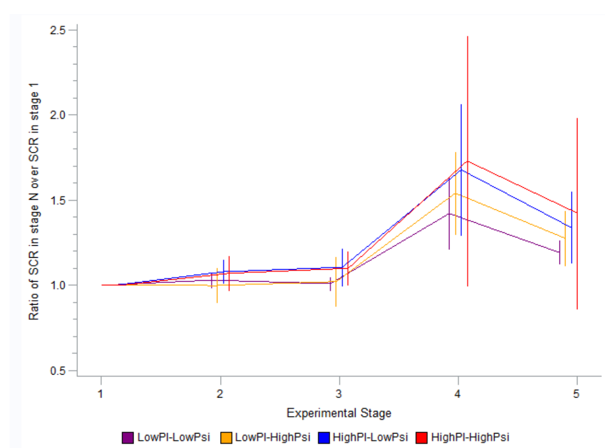

Figure 6: Comparing all four conditions by skin conductance response in each experimental stage.

Despite the fact that coherence, PI, and Psi are relatively new ideas in the VE literature, there have been other researchers that have identified related concepts. Biocca et al. performed an experiment in which one of the factors was the "scenario-appropriateness" of models in a VE task. Participants in that study removed objects from a virtual cadaver; the objects were either modeled as organs or simply as geometric primitives [4]. Experiential Fidelity is a term coined by Beckhaus and Lindeman for the extent to which the stimuli presented by a VE correspond to a user's beliefs and expectations [2]. In their discussion of a study investigating the neural correlates of breaks in presence, Sjölie et al. commented that an important part of maintaining presence is "avoid[ing] anything that 'disproves' it by violating expectations" [22]. Llobera and colleagues identified narrative coherence as one important dimension of interactive storytelling [12]. Parola and colleagues argue for a new definition of presence, "the sense of feeling real", that seems to correspond in many ways to Psi [17]. It is our hope that the coherence-Psi terminology and model can unify some of this existing work. It may also be worthwhile to examine the existing literature to determine if there are other experiments (such as the Biocca et al. study described above) that manipulated coherence variables, even though they may have predated the terminology.

8 Conclusion

These studies began as an attempt to identify metrics that would enable us to reliably measure and distinguish between the effects of Place Illusion and Plausibility Illusion. No such metrics were identified; however, other effects were observed, including that PI and Psi together result in higher presence scores, that matched coherence and immersion levels may lead to higher levels of presence, and that low coherence can cause increased heart rate.

References

[1] K. W. Arthur. *Effects of Field of View on Performance with Head-Mounted Displays*. PhD thesis, The University of North Carolina at Chapel Hill, 2000.

[2] S. Beckhaus and R. W. Lindeman. *Experiential Fidelity: Leveraging the Mind to Improve the VR Experience*, pages 39–49. Springer Vienna, Vienna, 2011.

[3] I. Bergström, K. Kilteni, and M. Slater. First-person perspective virtual body posture influences stress: A virtual reality body ownership study. *PLoS ONE*, 11(2), 2016.

[4] F. Biocca, J. Kim, and Y. Choi. Visual touch in virtual environments: An exploratory study of presence, multimodal interfaces, and cross-modal sensory illusions. *Presence*, 10(3):247–265, June 2001.

[5] D. B. Chertoff, B. Goldiez, and J. J. L. Jr. Virtual experience test: A virtual environment evaluation questionnaire. In *Virtual Reality*, pages 103–110, 2010.

[6] C. Hendrix and W. Barfield. Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators and Virtual Environments*, 5(3):274–289, 1996.

[7] H. Ishiguro. Scientific issues concerning androids. *The International Journal of Robotics Research*, 26(1):105–117, 2007.

[8] P. Khanna, I. Yu, J. Mortensen, and M. Slater. Presence in response to dynamic visual realism: a preliminary report of an experiment study. In *Proceedings of the ACM symposium on Virtual reality software and technology*, VRST '06, pages 364–367, New York, NY, USA, 2006. ACM.

[9] T. Kim and F. Biocca. Telepresence via television: Two dimensions of telepresence may have different connections to memory and persuasion. *Journal of Computer-Mediated Communication*, 3(2), 1997.

[10] K. M. Lee. Presence, explicated. *Communication Theory*, 14(1):27–50, 2004.

[11] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A cross-media presence questionnaire: The ITC-Sense Of Presence Inventory. *Presence: Teleoperators and Virtual Environments*, 10:282–297, 2001.

[12] J. Llobera, K. J. Blom, and M. Slater. Telling stories within immersive virtual environments. *Leonardo*, 46(5):471–476, 2016/09/17 2013.

[13] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997.

[14] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr. Physiological measures of presence in stressful virtual environments. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '02, pages 645–652, New York, NY, USA, 2002. ACM.

[15] M. J. Meehan. *Physiological reaction as an objective measure of presence in virtual environments*. PhD thesis, The University of North Carolina at Chapel Hill, 2001.

[16] M. Mori, K. MacDorman, and N. Kageki. The uncanny valley. *IEEE Robotics Automation Magazine*, 19(2):98–100, June 2012.

[17] M. Parola, S. Johnson, and R. West. Turning presence inside-out: Metanarratives. *Electronic Imaging*, 2016(4):1–9, February 2016.

[18] G. Riva, M. T. Anguera, B. K. Wiederhold, and F. Mantovani, editors. *From communication to presence*, volume 9 of *Emerging Communication: Studies in New Technologies and Practices in Communication*. IOS Press, Amsterdam, 2006.

[19] C. Sas and G. M. O'Hare. Presence equation: an investigation into cognitive factors underlying presence. *Presence: Teleoperators and Virtual Environments*, 12(5):523–537, 2003.

[20] D. W. Schloerb. A quantitative measure of telepresence. *Presence: Teleoperators and Virtual Environments*, 4:64–80, 1995.

[21] T. Schubert and H. Regenbrecht. Decomposing the sense of presence: Factor analytic insights. In *2nd International Workshop on Presence*, January 1999.

[22] D. Sjölie, G. Kalpouzos, and J. Eriksson. Neural correlates of disrupted presence: strange disruptions in a naturalistic virtual environment. In *Proceedings of the 24th International Conference on Artificial Reality and Telexistence and the 19th Eurographics Symposium on Virtual Environments*, pages 21–28. Eurographics Association, 2014.

[23] R. Skarbez. A preliminary investigation of place illusion and plausibility illusion. In *IEEE Virtual Reality (VR) Doctoral Consortium*, 2015.

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

[25] M. Slater. Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364:3549–3557, 2009.

[26] M. Slater, V. Linakis, M. Usoh, R. Kooper, and G. Street. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In *ACM Virtual Reality Software and Technology (VRST)*, pages 163–172, 1996.

[27] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Trans. Graph.*, 29:92:1–92:9, July 2010.

[28] M. Slater and A. Steed. A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9:413–434, October 2000.

[29] M. Slater and S. Wilbur. A Framework for Immersive Virtual Environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6):603–616, 2016/02/18 1997.

[30] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9:497–503, 2000.

[31] R. B. Welch. How can we determine if the sense of presence affects task performance? *Presence: Teleoperators and Virtual Environments*, 8(5):574–577, Oct. 1999.

[32] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 7:225–240, 1998.

[33] P. Zahorik and R. L. Jenison. Presence as being-in-the-world. *Presence: Teleoperators and Virtual Environments*, 7:78–89, 1998.