This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.2983701, IEEE Transactions on Visualization and Computer Graphics

Immersion and Coherence: Research Agenda and Early Results

Richard Skarbez, *Member, IEEE,* Frederick P. Brooks, Jr., *Life Fellow, IEEE,* and Mary C. Whitton, *Life Senior Member, IEEE*

Abstract—Presence has been studied in the context of virtual environments for nearly thirty years, but the field has yet to reach consensus on even basic issues of definition and measurement, and there are many open research questions. We gather many of these open research questions and systematically group them according to what we believe are five key constructs that inform user experience in virtual environments: immersion, coherence, Place Illusion, Plausibility Illusion, and presence. We also report on the design and results of a study that investigated the effects of immersion and coherence on user experience in a stressful virtual visual cliff environment. In this study, each participant experienced a given VE in one of four conditions chosen from a 2x2 design: high or low levels of immersion and high or low levels of coherence. We collected both questionnaire-based and physiological metrics. Several existing presence questionnaires could not reliably distinguish the effects of immersion from those of coherence. They did, however, indicate that high levels of both *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 responded markedly differently in the two coherence conditions; no such difference was observed across the immersion 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—virtual reality, presence, place illusion (PI), plausibility illusion (Psi), immersion, coherence, user studies, physiological metrics, research agenda

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*. Despite being in common use both among researchers and lay people, there is widespread disagreement even regarding basic issues such as definition and measurement [1]. As a result, the literature regarding presence is quite fragmented. Here we attempt to identify many open research questions regarding presence and related constructs and assemble them into a cogent research agenda, grouped according to what we believe are the five key constructs that inform user experience in virtual environments: immersion, coherence, Place Illusion, Plausibility Illusion, and presence.

We also present the results of a study 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 in presence as measured 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.

Portions of this work appeared as a poster at IEEE Virtual Reality (VR) 2017 [2] and as a conference paper at ACM Virtual Reality Software and Technology (VRST) 2018 [3].

- *R. Skarbez is with the Department of Computer Science and Information Technology at La Trobe University, Melbourne VIC 3086, Australia. E-mail: r.skarbez@latrobe.edu.au*
- *F. P. Brooks, Jr. and M. C. Whitton are with the University of North Carolina at Chapel Hill.*

Manuscript received April 19, 2005; revised August 26, 2015.

2 Background

The presence construct is most commonly defined as the sensation of "being there" [4] [5] [6] [7] [8] [9]. 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 presence using subjective questionnaires [10] [11] [12] [13] [14] [15], physiological metrics [16] [17], and behavioral metrics [18] [19] [20]. All of these measurement techniques have different strengths (and weaknesses), but none is a "silver bullet" for presence measurement. Referring specifically to subjective presence measures, Hendrix and Barfield argued that they should be relevant, sensitive, convenient, nonintrusive, and reliable [21]. In addition, good testing practices suggest that any measure should be reliable, valid, sensitive, and objective [22]. All existing presence measures fall short on more than one of these criteria.

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. (This is known as the "book problem" [23].) 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, for most applications, it has not 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 results have been inconclusive. One study that explicitly looked for a link between presence and task performance can be found in [24], 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 [15]. Welch has argued that there is no inherent logical connection between presence and task performance [25].

One notable exception to the above discussion is the use of VR to distract users from painful or unpleasant experiences. (Examples include burn wound care [26], chemotherapy [27], and dental procedures [28]. Li and colleagues have written a review of this literature [29].) In these cases, it can be argued that effectiveness and sense of presence (in the VE) are synonymous; the VE is effective to the extent that the user feels less presence in the (unpleasant) real world, and more presence in the virtual world.

To address various problems with the presence construct, Slater proposed a new theoretical framework for evaluating VEs [30]. 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). Sanchez-Vives and Slater take this multi-level realistic response to be the definition of presence [31].

Slater theorized that such realistic response arises from two qualia: Place Illusion (PI) and Plausibility Illusion (Psi). (Qualia are subjective mental states.) 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, they experience the VE as a real place (by virtue of experiencing Psi) that they are really in (by virtue of experiencing PI), and so they would be expected to act realistically.

(Returning briefly to the book problem: In his discussion of Place Illusion, Slater argues that "the types of PI that are possible [in different media] are *qualitatively different for each [medium]*." [Emphasis in the original] [30])

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, 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 [32].

One way that color perception is experimentally tested is by showing a user a test patch of color (say, aquamarine), and then asking the user to replicate that color by mixing three different colored lights. There is not one unique combination of three lights that yields aquamarine. In fact, there are infinite such combinations. But by repeating this experiment, one can build up an equivalence class of aquamarines, and use this to develop an empirically-derived function that defines aquamarine.

Slater argues that a given level of presence is like aquamarine; that is, that there are infinitely many combinations of hardware parameters and user characteristics that can combine to yield that level of presence. By enabling a user to combine different "light colors" (system parameters) to yield equivalent levels of presence, we can also develop empirically-derived functions of presence, so that we can say having a virtual body is more important than having realistic physics [33], or that gaze and environmental factors are more important than audio factors [34].

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 8.

3 PI:IMMERSION::PSI:COHERENCE

In his original PI/Psi paper, Slater states, "Immersion provides the boundaries within which PI can occur" [30]. Immersion, here, is defined in terms of the set of sensorimotor and effective 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 [35]. Immersion as so defined is in agreement with, for example, Bowman and McMahan's usage [36], but not with the usage of Witmer and Singer [15]. In the terminology of Lombard et al., we use "immersion" to mean *perceptual immersion*, while Witmer and Singer (and others) use it to mean *psychological immersion* [11].

A parallel argument can be made regarding Psi. Psi arises to the extent that a participant probes the Psi-inducing (or Psi-breaking) characteristics of the environment. While the concept of immersion is well-established in the VE research community, until recently, there did not exist an equivalent concept for reasoning about the degree to which the virtual scenario behaves in a reasonable or predictable way. We use the term *coherence* for this concept [1] [37].

Coherence and Psi, while being closely related, represent distinct constructs. Coherence is a characteristic of the virtual scenario and/or software, while Psi is a subjective mental state that arises in the user in response to their experience of the virtual reality. This connection is perhaps less clear than that of immersion and PI, due to the difficulty of defining what is "reasonable" behavior. One can imagine various ways of measuring reasonableness, but one is that a panel of raters could be employed. Assuming a sufficiently large and representative panel, their combined score might be such that an "average" user would experience plausibility illusion in that scenario. That said, a particular user might

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.2983701, IEEE Transactions on Visualization and Computer Graphics

or might not experience Psi, due to the idiosyncrasies of that individual's life experiences and patterns of interaction.

Note that coherence (and therefore Psi as well) is inextricably dependent on the particular scenario presented in the VE, as well as the particular user's expectations and previous experiences. For example, if one is specifically told that a VE represents a real-world scenario (or, absent any 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.

Coherence as a construct is related to, but distinguishable from, fidelity. Alexander et al. describe fidelity as "the extent to which the virtual environment emulates the real world" [38]. We would argue that coherence is more general; in particular, it makes no assumption that the virtual scenario is attempting to replicate the real world. (Consider the rocket boots discussion from the previous paragraph.)

4 Research Agenda

Despite nearly 30 years of research in (tele)presence in virtual environments [8], the field remains nascent. In the previous sections, we have begun explicating Place Illusion, Plausibility Illusion, immersion, coherence, and presence, as well as the relationships between them. All too often researchers have idiosyncratic views of presence leading to a proliferation of definitions and measures as catalogued in [1]. There are more questions than answers even in the most studied sub-area of presence research, the relationship between immersive characteristics of a system and Place Illusion/presence.

It is our opinion that an experimentally validated framework that links these concepts would have substantial application and value for both research and development of virtual reality environments. We proposed one such framework in Section 4.2 of [1], but others are possible. In this section we attempt to organize some of these questions into a coherent research agenda informed by that framework.

In that framework, the key components are immersion, coherence, Place Illusion, Plausibility Illusion, and presence. (For now, we exclude social presence, which merits an entire publication on its own, and has in fact received an excellent recent survey by Oh, Bailenson, and Welch [39].) In the following subsections, we present a series of important open research questions, grouped according to these components.

4.1 Immersion

As stated in Section 3, we follow Slater in defining immersion as the set of sensorimotor and valid actions supported by a system. Taken another way, it is a function of a system's input devices, output devices, and interface techniques.

The relationship between immersion as so defined and presence has been widely studied. A 2014 meta-analysis by Cummings and Bailenson identified 83 studies in this area, investigating aspects of immersion that included tracking quality, stereoscopy, field of view, update rate, and more. The overall conclusion of the meta-analysis was that technological immersion has a medium-sized (r = .316) effect on presence [40].

Considering immersion as a vector quantity, the elements might be items such as those mentioned in the previous paragraph. Some system components and characteristics are obvious candidates for inclusion and are relatively straightforward to measure. Consider field of view: one can simply measure the field of view (FoV) of one's display, and say that this FoV is wider than that of display X and smaller than that of display Y. However, already there are complications: Does the aspect ratio of the display matter? Should it be a separate element in the vector? Should we consider vertical FoV and horizontal FoV separately? From this example, it can be seen that even if one interprets immersion to be an objective system characteristic—as we do—operationalizing it as a construct is a nontrivial problem.

Considering system immersion as a vector we are faced with several questions:

- What characteristics of the systems' input devices, output devices, and interface techniques, belong in the vector? That is, what system characteristics have been shown to be or should be considered as contributing to immersion? Can we identify a minimal or standard set against which all systems can be evaluated?
- Is it feasible to come up with a single score for each element? As discussed above, this will be relatively straightforward for some elements, but not so for others. (How best to generate a score for travel technique, for example?)
- Is it feasible to come up with a score for each vector? That is, can we generate another vector of weights that can be multiplied with the immersion vector to produce a single "immersion score" for a system? (This assumes that not all immersion characteristics are equally important in the effect they have on a user's experience and that they are not orthogonal—there are interactions among them.)
- Given the ability to compute a meaningful immersion score, can systems be meaningfully categorized based on that score? For example, the score could be calibrated such that a system with a score greater than 100 is considered a fully immersive system, while one with a score greater than 50 but less than 100 might be considered partially immersive, and so on. In this manner, one might operationalize Slater's equivalence classes of immersion [30].

4.2 Coherence

Similarly, above we argued for the existence of a construct we call coherence, which is parallel to immersion, but that describes the extent to which the depicted scenario behaves reasonably. More precisely, we defined coherence to be the set of objectively reasonable circumstances that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.2983701, IEEE Transactions on Visualization and Computer Graphics

can be demonstrated by the scenario without introducing objectively unreasonable circumstances [1]. By objectively reasonable circumstances we mean states of affairs that are self-evident given prior knowledge provided in the context of the virtual experience. We believe that this the extent to which coherence can be meaningfully controlled, since it is not possible to know what states of affairs may or may not be plausible given a user's prior life experiences.

While the relationship between immersion and presence has been widely studied, the same cannot be said for the relationship between coherence and presence. We speculate that there are several reasons for this. One is the novelty of the concept; Plausibility Illusion was first introduced by Slater in 2009 [30], and coherence was introduced more recently than that. Another is a possible historical bias toward engineering solutions in the virtual reality research community; for much of the history of virtual reality, being *au fait* with the hardware and software systems was a requirement to do work in the field. A third reason is perhaps that doing research regarding coherence is simply *hard*. It is relatively easy to design and conduct a study comparing multiple fields of view; it is more challenging to design a study where the coherence of a scenario is manipulated in a controlled fashion.

All that said, the research questions we propose regarding coherence are much the same as those we proposed for immersion. Due to the paucity of existing research, though, there is a much greater need—or alternatively, a much greater opportunity—for any research regarding coherence. We hope that the following research questions can help to guide research in what we perceive to be an area that will only increase in importance.

- What elements belong in a coherence vector? That is, what scenario characteristics should be considered as part of coherence? Elsewhere, we conducted a study that divided coherence into four elements: the virtual body, physics behavior, character behavior, and "gestalt" coherence [30]. Can we identify a minimal or standard set against which all systems can be evaluated?
- Is it feasible to come up with a single score for each element?
- Is it feasible to come up with a score for each vector? That is, can we generate another vector of weights that can be multiplied with the coherence vector to produce a single "coherence score" for a system? (This assumes that not all coherence characteristics are equally important in the effect they have on a user's experience.)
- Can scenarios be meaningfully categorized based on their coherence score? For example, we might categorize a given scenario as coherent or incoherent based on a threshold, as discussed above with respect to immersion.

4.3 Place Illusion

Slater has defined Place Illusion as, "The illusion of being in a place in spite of the sure knowledge that you are not there" [30]. In that same paper, he also claims that, "Immersion provides the boundaries within which PI can occur" [30]. Extending this idea, we argue that Place Illusion is a function of immersion, user traits (including both psychological traits like transportability [41] or immersive tendencies [15], as well as physical traits such as susceptibility to simulator sickness), user state, and user behavior.

Place Illusion is a subjective feeling of the user and there is little agreement about how to measure it. As summarized in [1], measures include self-report, physiological, behavioral, and psychophysical measures of presence. We do not know whether any one of these is most appropriate for measuring Place Illusion.

Our opinion is that existing presence measures can be used for inspiration, but almost universally, they also measure elements that we would more properly consider Plausibility Illusion, immersion, or coherence. (For example, see Table 3 of [1]. Other than perhaps the Slater-Usoh-Steed questionnaire [13], Kim and Biocca's arrival/departure questionnaire [42], and the Sas and O'Hare questionnaire [6], all the listed questionnaires contain subscales that do not refer to a user's sense of place.) We suspect that triangulating using multiple measures can be productively employed.

- How much of Place Illusion variability can be explained by immersion?
- How much of Place Illusion variability can be explained by individual trait differences?
- How much of Place Illusion variability can be explained by individual state differences?
- How much of Place Illusion variability can be explained by individual behavior differences?
- Which, if any, user traits reliably correlate with Place Illusion?
- Which, if any, user behavioral characteristics reliably correlate with Place Illusion?
- Is there a minimum level of immersion required to induce Place Illusion?
- Is there a saturation level of immersion beyond which there is no further benefit to Place Illusion?
- Is it possible for there to be a situation in which improved immersion can result in a decrease in Place Illusion? (That is, is the relationship ever non-monotonic?)
- How do users respond to breaks in immersion? Does Place Illusion recover?
- Is it possible to "context-switch" without a noticeable loss of Place Illusion? That is, can users take actions in the real world—consider the situation where ones phone goes off in the real world and it has to be answered or silenced—without loss of Place Illusion? (Or consider Spagnolli and Gamberini's "hybridity" of presence, where they observed participants interacting with experimenters while simultaneously continuing to act in the virtual environment [43].)

4.4 Plausibility Illusion

Slater has defined Plausibility Illusion as, "The illusion that what is apparently happening is really happening (even though you know for sure that it is not)" [30]. Similar to how immersion enables (but does not guarantee) the feeling Place Illusion, we argue that coherence enables Plausibility

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.2983701, IEEE Transactions on Visualization and Computer Graphics

Illusion. To be somewhat more precise, we argue that Plausibility Illusion is a function of scenario coherence and an individual user's expectations.

- How best to measure Plausibility Illusion? The discussion above regarding Place Illusion measures also applies here, although there are even fewer relevant existing measures. The most relevant questionnaire seems to be Baños et al's Reality Judgment Presence Questionnaire (RJPQ) [44].
- How much of Plausibility Illusion variability can be explained by coherence?
- How much of Plausibility Illusion variability can be explained by users' expectations?
- How can we measure users' expectations?
- How and to what extent can we manipulate users' expectations, for example, by "priming" the user with stimuli that are related to those that will appear in the virtual environment?
- Is there a minimum level of coherence required to induce Plausibility Illusion?
- Is there a saturation level of coherence beyond which there is no further benefit to Plausibility Illusion?
- Is it possible for there to be a situation in which improved coherence can result in a decrease in Plausibility Illusion?
- How do users respond to breaks in coherence? Does Plausibility Illusion recover?
- Can a user experience Plausibility Illusion in multiple virtual or mediated spaces simultaneously?

4.5 Presence

We and others have decried the imprecision of the term *presence* as it applies to user experience in virtual environments, but it remains a useful shorthand for what makes virtual reality experiences unique and powerful. As in [1], we use *presence* to mean, "The perceived realness of a mediated or virtual experience." We do not consider this to be in conflict with a conception of presence as multi-level realistic response, as in [31]. Rather, if one perceives an experience to be real, one is likely to respond realistically, and if one is responding realistically on the subjective, psychological level, there will be a perception that the experience is real.

- How best to measure presence? Here, existing presence measures are very relevant, however there is still the question of which instrument is (or which instruments are) most appropriate.
- How do Place Illusion, Plausibility Illusion, and/or Social Presence Illusion interact to yield presence?
- Is it possible to have immersion with no coherence, and if so, what are users' experiences of Place Illusion, Plausibility Illusion, and presence likely to be in this case?
- Is it possible to have coherence with no immersion, and if so, what are users' experiences of Place Illusion, Plausibility Illusion, and presence likely to be in this case?
- What is the "space" of presence? (For example, is presence a two-pole construct—one can be present only in the real world or in a technology-mediated environment— or a three pole construct—one can be present in the real world, a mediated environment, or in "mental imagery space" [23].)
- If presence is taken to be the realistic response to a virtual experience, what about experiences that are not "realistic?" For example, one can imagine a virtual experience that presents a world governed by physical laws that are constantly changing. Can one feel "present" in such a virtual world? What would that mean?

4.6 Comments on the Research Agenda

This research agenda brings together many of what we perceive to be the important open questions for presence researchers. We are not under the illusion that these investigations will be easy or straightforward. On the contrary, almost every one of the presented research questions represents a challenging research program in its own right. However, we believe that this reflects the level of investment that will be required to place presence on a solid foundation of research.

The remainder of this paper presents a study that was an early attempt at answering some of these questions.

5 User Study

This study used a 2x2 between-subjects design with multiple outcome measures. We chose a between-subjects design because Khanna and colleagues observed [45] that responses were not symmetric across conditions in a visual cliff scenario 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 [46], finding that physiological responses decreased with subsequent exposures, but not to zero.

In this study, the immersion factor was manipulated as follows: In the High-Immersion conditions, the field of view of the HMD was the maximum supported by the device (60° diagonal), passive haptics were used to provide tactile feedback to the participant, and scenario-appropriate spatial sound cues appeared in the environment. (The passive haptics used were styrofoam walls around the Pit room, and wooden planking that corresponded to the walkway around the virtual pit. The spatial sound cues were related to the elevator, doors opening, and balls falling to the floor.) In the Low-Immersion conditions, the effective field of view of the HMD was restricted to 30° by use of a virtual mask, no passive haptics were used, and there was no sound other than the study instructions delivered through the headphones. The choice to employ multiple simultaneous manipulations of both immersion and coherence was made after a pilot study that manipulated only a single factor proved inconclusive. That study is discussed in [2], but is omitted here for space.

(At the time these results were collected, a 60° diagonal field-of-view was considered reasonable. However, we acknowledge that HMD technology has advanced rapidly in recent years, so as to render this a very poor display. We discuss this in Section 8.)

We argue that failures of coherence can 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 being in the space, actions that you are led to believe will cause one event in fact cause a different event.

In this study, then, the coherence factor is manipulated as follows: In the High-Coherence conditions, physical objects (balls) behave as one would expect them to, and the study instructions are in fact valid. In the Low-Coherence conditions, 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 study 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).

5.1 Participants

Thirty-two male participants took part in this study. The average age was 20.1 years. Participants successfully passed screening for uncorrected vision problems, a history of seizures or strong motion sickness, inability to walk without assistance, deafness, and English comprehension. This study was approved by the Behavioral Institutional Review Board of the University of North Carolina at Chapel Hill.

5.2 Materials

The study took place primarily inside 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.

5.3 Metrics

During the exposure to the virtual environment, participants' physiological responses were monitored and recorded. We collected electrocardiogram (EKG), skin conductance (SCR), and skin temperature data. For both SCR and skin temperature, the mean and standard deviation were computed for each stage of the study. 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 study.

Post-test, participants completed the Witmer-Singer Presence Questionnaire [15] and a modified Slater-Usoh-Steed Presence Score [13].

5.4 Study Procedures

Upon arriving, participants first reported to an office, where they were screened by an experimenter, signed informed-consent forms, and completed pre-exposure questionnaires on a PC using the Qualtrics web application[1]. 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 trial.

The virtual environment phase of the study was divided into five stages. (An illustration of the environment is in Figure 1.):

Stage 1 (Simon room). Participants familiarized themselves with the virtual environment, playing a Simon-like memory game. The physiological responses gathered during this stage were used as the baseline for subsequent analyses. In this stage, all behavior was coherent; that is, participants in the Low-Coherence and High-Coherence conditions had the same experience.

Stage 2 (Scoreboard room). Participants took a virtual elevator to a room where they had to pick up balls and drop them in targeted receptacles. In this room, incoherent behavior was introduced in the Low-Coherence conditions, so a difference in behavior or physiological signals observed at this stage would most likely be due to the coherence manipulation.

Stage 3 (Pre-Pit room). Participants took a virtual elevator to an office-like environment, where they were presented with additional balls to drop on targets. This stage was included to acquire a baseline after exposure to the experimental manipulation but before exposure to the Pit.

Stage 4 (Pit room). 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 (Simon room). Participants returned to the elevator, returned to the Simon room, and played the game again. This stage was included to determine whether the experimental manipulations caused different behaviors after exposure to the stressful stimulus. After 3 minutes, the study ended. The total time in the virtual environment was approximately fifteen minutes.

1. https://www.qualtrics.com/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.2983701, IEEE Transactions on Visualization and Computer Graphics

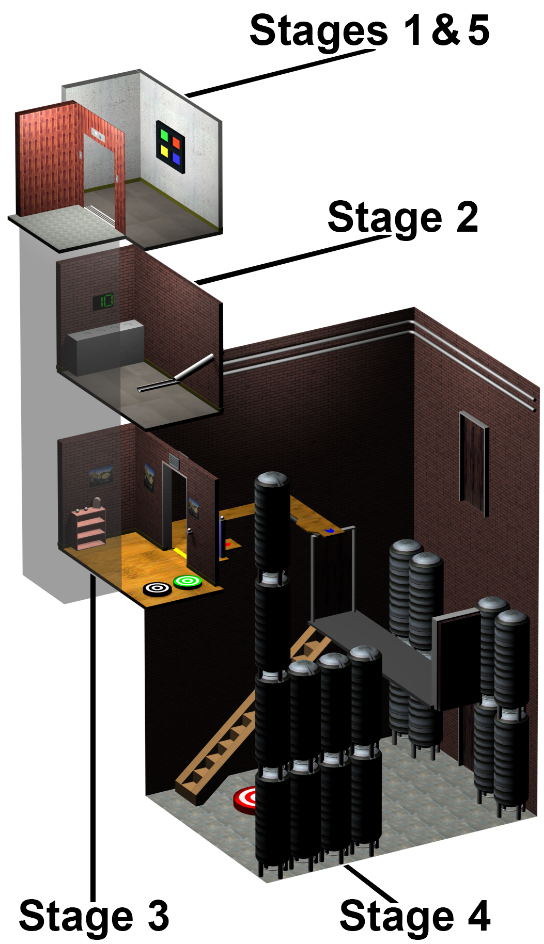


Fig. 1. The virtual environment used in the study.

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

6 Study 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.* [47].

TABLE 1
Mean count of high scores (6 or 7) on the Witmer-Singer PQ for each condition

	Low-Immersion	High-Immersion
Low-Coherence	9.6	10.4
High-Coherence	10.0	12.9

TABLE 2
Mean count of high scores (6 or 7) on the SUS questionnaire for each condition

	Low-Immersion	High-Immersion
Low-Coherence	4.0	2.9
High-Coherence	3.3	4.3

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 the Witmer-Singer PQ explicitly contains items regarding system characteristics (what we refer to as immersion), rather than about subjective experience (what we refer to as Place Illusion). Nonetheless, there is an 80.5% probability that participants in High-Immersion conditions reported higher PQ scores than participants in Low-Immersion conditions.

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

Unlike the Witmer-Singer PQ, the SUS questionnaire contains no questions specifically relating to immersion factors, and instead asks only about subjective experience. The posterior probability that participants in the High-Immersion conditions reported higher scores than participants in the Low-Immersion conditions is 52.1%.

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

The probabilities that participants in the High-Coherence conditions reported higher questionnaire scores than participants in the Low-Coherence conditions are 61.7% and 71.3% on the SUS questionnaire and the PQ, respectively.

6.4 When high levels of immersion and coherence 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 High-Immersion+High-Coherence 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 High-Immersion+High-Coherence 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 High-Immersion+High-Coherence condition scored higher. (See Tables 1 and 2 for mean scores.)

6.5 There is good evidence that SUS questionnaire scores are higher for matched (Low-Immersion+Low-Coherence and High-Immersion+HighCoherence) 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 High-Coherence conditions than the Low-Coherence 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 High-Immersion conditions (High-Immersion+Low-Coherence and High-Immersion+High-Coherence combined) than in Low-Immersion conditions (Low-Immersion+Low-Coherence and Low-Immersion+High-Coherence).

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

In Stage 1 of the study, 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 Low-Coherence conditions were exposed to a series of coherence failures, while those in the High-Coherence 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 2.)

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 High-Immersion vs. Low-Immersion (38.4%, 46.3%), or for High-Coherence vs. Low-Coherence (59.9%, 38.9%).

7 Discussion

As stated above in Section 6.4, we observed that participants have higher PQ and SUS questionnaire scores when both presence and immersion are high;

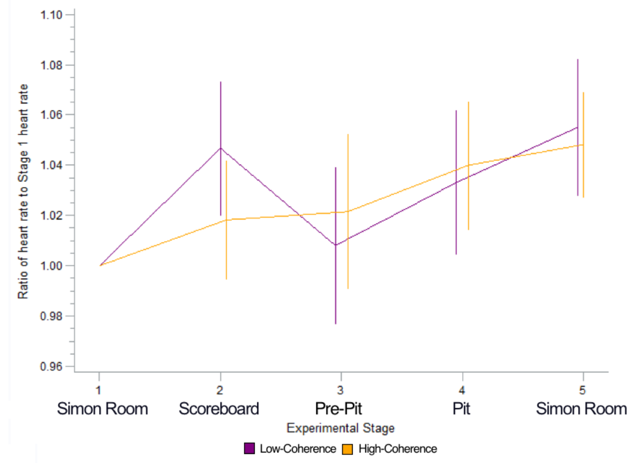


Fig. 2. Comparing coherence conditions by heart rate in each stage of the study. (Bars represent 95% confidence intervals.)

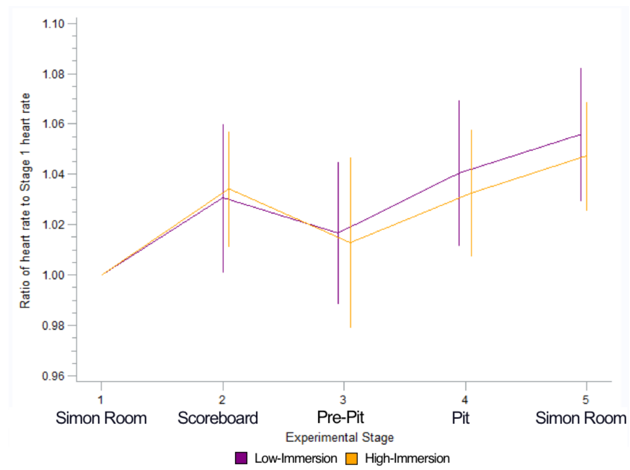


Fig. 3. Comparing immersion conditions by heart rate in each stage of the study. (Bars represent 95% confidence intervals.)

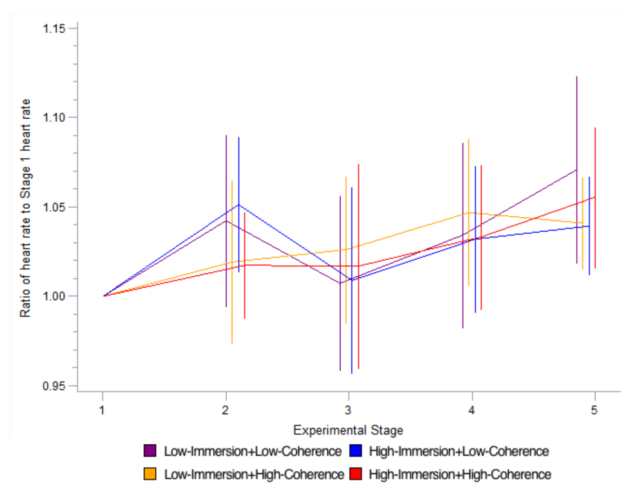


Fig. 4. Comparing all four conditions by heart rate in each stage of the study. (Bars represent 95% confidence intervals.)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2020.2983701, IEEE Transactions on Visualization and Computer Graphics

none of the other conditions is substantially different from another. This demonstrates that when coherence and immersion are present together, participants report significantly higher levels of presence. Further, the scores for Low-Immersion+HighCoherence and High-Immersion+LowCoherence conditions are not substantially different, which may indicate that both immersion and coherence are of roughly equal importance, at least as regards scores on the PQ. Furthermore, neither of these is substantially different from the Low-Immersion+Low-Coherence condition, indicating perhaps that any noticeable failure of either immersion or coherence causes a substantial drop in presence.

We argue that these observations, taken together, indicate that if their experience is "good enough," users will remember and report a high level of presence after the fact, and if it is not, then they will 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 lower a user's feeling of presence and reduce the quality of an experience 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.

We speculate that these results lend credence to an "uncanny valley" effect in virtual environments. In Mori's original presentation of the uncanny valley, it seems to us that the problem is not inherently the humanlike appearance of an entity, but rather the *mismatch* between its appearance and its behavior [48]. Ishiguro has extended the concept of the uncanny valley with a "synergy effect" in this way (essentially, the match between appearance and behavior) [49]. 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.

In comparing our observations (specifically, that the questionnaire scores for the Low-Immersion+High-Coherence and High-Immersion+Low-Coherence conditions are similar) to previous studies measuring the relative influences of different aspects of experience on presence using self-report methods, ours seem to be in line with those of Lessiter *et al.* [10]. That paper suggested that immersion factors and coherence factors contributed roughly equally to presence. (The four factors on the ITC-SOPI are Sense of Physical Space, Engagement, Ecological Validity, and Negative Effects; a factor analysis found that these four factors explained 14.2%, 11.1%, 7.6%, and 5.4% of variance, respectively. We consider Sense of Physical Space to measure Place Illusion and Negative Effects to be a component of immersion; these two together explain 19.6% of variance. The other two seem to measure components of coherence, and explain 18.7% of variance, taken together.) On the other hand, our results agree less with those of Schubert, Friedmann and Regenbrecht [50], which suggested that immersion factors contributed roughly twice as much as coherence factors. (Schubert, Friedmann, and Regenbrecht's factor analysis of a custom seventy-five-item survey yielded eight factors that combined to explain 50.27% of variance: spatial presence, quality of immersion, involvement, drama, interface awareness, exploration of the VE, predictability & interaction, and realness. We group spatial presence, immersion quality, interface awareness, and exploration as immersion factors; explaining 34% of variance. We then group involvement, drama, predictability & interaction, and realness as coherence factors; explaining 16% of variance.) Schubert, Friedmann, and Regenbrecht also performed a replication study [12], and in that study, they identified five factors, explaining 53.0% of the total variance. These factors were spatial presence, exploration of the VE, realness, predictability & interaction, and involvement. Here we group spatial presence and exploration as immersion factors, explaining 39% of variance, and realness, predictability & interaction, and involvement as coherence factors, explaining 14% of variance. Of note is that none of the participants in the ITC-SOPI study, and few of the participants in the Schubert, Friedmann, and Regenbrecht studies, based their responses on experience of an immersive virtual environment. We suspect that these proportions may vary based on the immersiveness and the interactivity of the virtual experience.

Figure 5 depicts the skin conductance response for each condition and each stage. Notable is the fact that the Low-Coherence conditions do not exhibit a spike in Stage 2 as was seen with 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 [46]. 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.

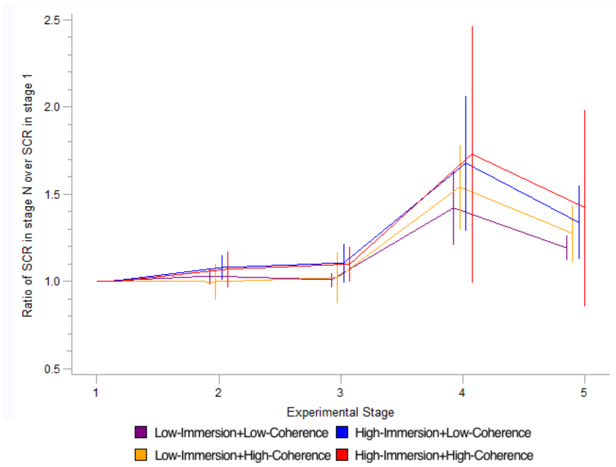


Fig. 5. Comparing all four conditions by skin conductance response in each stage of the study. (Bars represent 95% confidence intervals.)

8 Limitations and Future Work

We commented in Section 5 about the low quality of the HMD used in this study, compared to HMDs that are widely available today. It would be interesting to revisit field-of-view manipulation with the $\approx 100°$ FoV HMDs that are common today, or even with a specialized wide FoV display.

It was an oversight not to include the Reality Judgment Presence Questionnaire of Baños *et al.* in the battery of measures used in this study [44]. This seems to be the existing measure which is most directly related to Plausibility Illusion, and it needs to be specifically evaluated for this purpose in future work.

More generally, there are many presence questionnaires in the literature, and we opted to use only two of them. Other commonly used presence questionnaires include the igroup Presence Questionnaire (IPQ), the ITC-Sense of Presence Inventory (ITC-SOPI), and the Lombard and Ditton questionnaire [51]. Our choice was somewhat arbitrary; it was simply not feasible to incorporate more questionnaires and keep the duration of the experiment reasonable. That said, the use of more or different questionnaires may have yielded different insights.

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 [52]. 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 [53]. 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" [54]. Llobera and colleagues identified narrative coherence as one important dimension of interactive storytelling [55]. Parola and colleagues argue for a new definition of presence, "the sense of feeling real," that seems to correspond in many ways to Psi [56]. Gilbert proposed the concept of authenticity [57], which is quite comparable to coherence; the Bayesian framework he used in that work was adopted by Skarbez in his discussion of coherence in [1]. 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.

The sample in this experiment was chosen to be relatively uniform (young healthy males) with the intention of having more consistent baseline physiological metrics. Moreover, the population at large is certainly WEIRD (Western, Educated, Industrialized, Rich, and Democratic) [58]. As a result, it is unclear how these results will generalize.

9 Conclusion

This paper was written with two purposes in mind. One was to present a research agenda, based on Slater's Place Illusion/Plausibility Illusion model of user experience in virtual environments, for the field commonly known as presence research. This agenda presented many of the open research questions in this field, broken down along the lines of relevant constructs: immersion, coherence, Place Illusion, Plausibility Illusion and presence.

The other was to present the design and results of a study that investigated the effects of changing levels of immersion and coherence on participants' experience of and behavior in a stressful visual cliff virtual environment. One goal of this study was to determine whether existing measures of presence could reliably measure and distinguish between Place Illusion and Plausibility Illusion. This goal was not entirely achieved; that said, there is good evidence that several subscores of the Witmer-Singer Presence Questionnaire respond differently to immersion and coherence, so this may indicate a way forward for study in this area. Other interesting findings were that the Slater-Usoh-Steed questionnaire and the Witmer-Singer questionnaire respond differently to increasing immersion, that high levels of Place Illusion and Plausibility Illusion together result in higher presence scores, that matched levels of immersion and coherence may also result in higher presence scores, and that low coherence can cause increased heart rate. These results advance the theory relating to immersion, coherence, PI, and Psi, as well as offering practical advice to developers of virtual experiences.

Acknowledgments

The authors thank Dr. Benjamin Lok, Dr. Abigail Panter, Dr. Mel Slater, and Dr. Gregory Welch for their guidance and helpful feedback during the writing of the dissertation from which this work is derived. We also thank the anonymous reviewers; the final article was significantly improved by their contributions.

References

[1] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton, "A survey of presence and related concepts," *ACM Computing Surveys*, vol. 50, no. 6, pp. 96:1–96:39, 2017. [Online]. Available: http://doi.acm.org/10.1145/3134301

[2] R. Skarbez, F. P. Brooks, and M. C. Whitton, "Immersion and coherence in a visual cliff environment," in *2017 IEEE Virtual Reality (VR)*, 2017, pp. 397–398. [Online]. Available: doi.ieeecomputersociety.org/10.1109/VR.2017.7892344

[3] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton, "Immersion and coherence in a stressful virtual environment," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '18. New York, NY, USA: ACM, 2018, pp. 24:1–24:11. [Online]. Available: http://doi.acm.org/10.1145/3281505.3281530

[4] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *Journal of Computer-Mediated Communication*, vol. 3, no. 2, 1997. [Online]. Available: http://dx.doi.org/10.1111/j.1083-6101.1997.tb00072.x

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

[6] C. Sas and G. M. O'Hare, "Presence equation: An investigation into cognitive factors underlying presence," *Presence: Teleoperators and Virtual Environments*, vol. 12, no. 5, pp. 523–537, October 2003.

[7] D. W. Schloerb, "A quantitative measure of telepresence," *Presence: Teleoperators and Virtual Environments*, vol. 4, no. 1, pp. 64–80, Winter 1995.

[8] J. Steuer, "Defining virtual reality: Dimensions determining telepresence," *Journal of Communication*, vol. 42, no. 4, pp. 73–93, 1992. [Online]. Available: http://dx.doi.org/10.1111/j.1460-2466.1992.tb00812.x

[9] P. Zahorik and R. L. Jenison, "Presence as being-in-the-world," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 1, pp. 78–89, February 1998.

[10] 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*, vol. 10, pp. 282–297, 2001.

[11] M. Lombard, T. B. Ditton, D. Crane, B. Davis, G. Gil-Egui, K. Korvath, and J. Rossman, "Measuring presence: A literature-based approach to the development of a standardized paper-and-pencil instrument," in *PRESENCE 2000 - 3rd International Workshop on Presence*, Delft, The Netherlands, 2000.

[12] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 266–281, June 2001. [Online]. Available: http://dx.doi.org/10.1162/105474601300343603

[13] M. Usoh, E. Catena, S. Arman, and M. Slater, "Using presence questionnaires in reality," *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 5, pp. 497–503, October 2000.

[14] P. Vorderer, W. Wirth, F. R. Gouveia, F. Biocca, T. Saari, L. Jäncke, S. Böcking, H. Schramm, A. Gysbers, T. Hartmann, C. Klimmt, J. Laarni, N. Ravaja, A. Sacau, T. Baumgartner, and P. Jäncke, "MEC Spatial Presence Questionnaire (MEC-SPQ): Short documentation and instructions for application," Report to the European Community, Project Presence: MEC (IST-2001-37661), Tech. Rep., 2004.

[15] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence: Teleoperators and Virtual Environments*, vol. 7, no. 3, pp. 225–240, June 1998.

[16] C. Dillon, E. Keogh, and J. Freeman, ""It's been emotional": Affect, physiology, and presence," in *PRESENCE 2002 - 5th Annual International Conference on Presence*, Porto, Portugal, 2002.

[17] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, Jr., "Physiological measures of presence in stressful virtual environments," in *SIGGRAPH 2002 - Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*. New York, NY, USA: ACM, 2002, pp. 645–652. [Online]. Available: http://doi.acm.org/10.1145/566570.566630

[18] J. Freeman, S. E. Avons, R. Meddis, D. E. Pearson, and W. IJsselsteijn, "Using behavioral realism to estimate presence: A study of the utility of postural responses," *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 2, pp. 149–164, April 2000.

[19] M. Slater, M. Usoh, and Y. Chrysanthou, "The influence of dynamic shadows on presence in immersive virtual environments," in *Virtual Environments '95*. Springer, 1995, pp. 8–21.

[20] S. Thie and J. van Wijk, "A general theory on presence," in *PRESENCE 1998 - 1st Annual International Workshop on Presence*, Suffolk, England, UK, 1998.

[21] C. Hendrix and W. Barfield, "Presence within virtual environments as a function of visual display parameters," *Presence: Teleoperators and Virtual Environments*, vol. 5, no. 3, pp. 274–289, Summer 1996.

[22] AERA/APA/NCME, *Standards for educational and psychological testing*. Washington, DC: AERA, 2014.

[23] F. Biocca, "Can we resolve the book, the physical reality, and the dream state problems? from the two-pole to a three-pole model of shifts in presence," in *EU Future and Emerging Technologies, Presence Initiative Meeting*, 2003.

[24] 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 *Proceedings of ACM Virtual Reality Software and Technology (VRST)*, 1996, pp. 163–172.

[25] R. B. Welch, "How can we determine if the sense of presence affects task performance?" *Presence: Teleoperators and Virtual Environments*, vol. 8, no. 5, pp. 574–577, October 1999. [Online]. Available: http://dx.doi.org/10.1162/105474699566387

[26] H. G. Hoffman, J. N. Doctor, D. R. Patterson, G. J. Carrougher, and T. A. Furness III, "Virtual reality as an adjunctive pain control during burn wound care in adolescent patients," *Pain*, vol. 85, no. 1, pp. 305–309, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304395999002754

[27] S. M. Schneider and M. L. Workman, "Virtual reality as a distraction intervention for older children receiving chemotherapy," *Pediatric nursing*, vol. 26, no. 6, pp. 593–7, Nov 2000, copyright - Copyright Anthony J. Jannetti, Inc. Nov/Dec 2000; Last updated - 2017-11-08. [Online]. Available: http://ez.library.latrobe.edu.au/login?url=https://search.proquest.com/docview/199391254?accountid=12001

[28] E. Furman, T. R. Jasinevicius, N. F. Bissada, K. Z. Victoroff, R. Skillicorn, and M. Buchner, "Virtual reality distraction for pain control during periodontal scaling and root planing procedures," *The Journal of the American Dental Association*, vol. 140, no. 12, pp. 1508–1516, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002817714615555

[29] A. Li, Z. Montaño, V. J. Chen, and J. I. Gold, "Virtual reality and pain management: current trends and future directions," *Pain management*, vol. 1, no. 2, pp. 147–157, March 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3138477/

[30] 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*, vol. 364, pp. 3549–3557, 2009.

[31] M. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," *Nature Reviews Neuroscience*, vol. 6, no. 4, pp. 332–9, 2005. [Online]. Available: http://discovery.ucl.ac.uk/756/

[32] M. Slater, B. Spanlang, and D. Corominas, "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 92:1–92:9, July 2010. [Online]. Available: http://doi.acm.org/10.1145/1778765.1778829

[33] R. Skarbez, S. Neyret, J. Frederick P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the Plausibility Illusion," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1369–1378, April 2017.

[34] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater, "The plausibility of a string quartet performance in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 4, pp. 1352–1359, April 2017.

[35] 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*, vol. 6, no. 6, pp. 603–616, December 1997. [Online]. Available: http://dx.doi.org/10.1162/pres.1997.6.6.603

[36] D. A. Bowman and R. P. McMahan, "Virtual reality: How much immersion is enough?" *Computer*, vol. 40, no. 7, pp. 36–43, July 2007.

[37] R. Skarbez, "Plausibility illusion in virtual environments," Ph.D. dissertation, The University of North Carolina at Chapel Hill, 2016.

[38] A. L. Alexander, T. Brunyé, J. Sidman, and S. A. Weil, "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," in *The Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, Florida: NTSA, 2005.

[39] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Frontiers in Robotics and AI*, vol. 5, pp. 114:1–114:35, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00114

[40] J. J. Cummings and J. N. Bailenson, "How immersive is enough? a meta-analysis of the effect of immersive technology on user presence," *Media Psychology*, vol. 19, no. 2, pp. 272–309, 2016. [Online]. Available: http://dx.doi.org/10.1080/15213269.2015.1015740

[41] M. C. Green, "Mechanisms of narrative-based belief change," Master's thesis, The Ohio State University, 1996. [Online]. Available: http://rave.ohiolink.edu/etdc/view?acc_num=osu1235579429

[42] 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*, vol. 3, no. 2, 1997. [Online]. Available: http://dx.doi.org/10.1111/j.1083-6101.1997.tb00073.x

[43] A. Spagnolli and L. Gamberini, "Immersion/emersion: Presence in hybrid environments," in *PRESENCE 2002 - 5th Annual International Workshop on Presence*, Porto, Portugal, 2002, pp. 421–434.

[44] R. M. Baños, C. Botella, A. García-Palacios, H. Villa, C. Perpiña, and M. Alcañiz, "Presence and reality judgment in virtual environments: A unitary construct," *CyberPsychology & Behavior*, vol. 3, no. 3, pp. 327–335, July 2000.

[45] 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*, ser. VRST '06. New York, NY, USA: ACM, 2006, pp. 364–367. [Online]. Available: http://doi.acm.org/10.1145/1180495.1180569

[46] M. J. Meehan, "Physiological reaction as an objective measure of presence in virtual environments," Ph.D. dissertation, The University of North Carolina at Chapel Hill, 2001.

[47] I. Bergström, K. Kilteni, and M. Slater, "First-person perspective virtual body posture influences stress: A virtual reality body ownership study," *PLoS ONE*, vol. 11, no. 2, 2016.

[48] M. Mori, K. MacDorman, and N. Kageki, "The uncanny valley," *IEEE Robotics Automation Magazine*, vol. 19, no. 2, pp. 98–100, June 2012.

[49] H. Ishiguro, "Scientific issues concerning androids," *The International Journal of Robotics Research*, vol. 26, no. 1, pp. 105–117, 2007. [Online]. Available: http://ijr.sagepub.com/content/26/1/105.abstract

[50] T. Schubert and H. Regenbrecht, "Decomposing the sense of presence: Factor analytic insights," in *PRESENCE 1999 - 2nd Annual International Workshop on Presence*, Colcester, England, January 1999.

[51] C. Rosakranse and S. Y. Oh, "Measuring presence: The use trends of five canonical presence quesionnaires from 1998-2012," in *ISPR 2014 - 15th International Workshop on Presence*, Vienna, Austria, 2014.

[52] 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: Teleoperators and Virtual Environments*, vol. 10, no. 3, pp. 247–265, June 2001.

[53] S. Beckhaus and R. W. Lindeman, *Experiential Fidelity: Leveraging the Mind to Improve the VR Experience*. Vienna: Springer Vienna, 2011, pp. 39–49. [Online]. Available: http://dx.doi.org/10.1007/978-3-211-99178-7_3

[54] 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*. Eurographics Association, 2014, pp. 21–28.

[55] J. Llobera, K. J. Blom, and M. Slater, "Telling stories within immersive virtual environments," *Leonardo*, vol. 46, no. 5, pp. 471–476, 2016/09/17 2013. [Online]. Available: http://dx.doi.org/10.1162/LEON_a_00643

[56] M. Parola, S. Johnson, and R. West, "Turning presence inside-out: MetaNarratives," *Electronic Imaging*, vol. 2016, no. 4, pp. 1–9, February 2016.

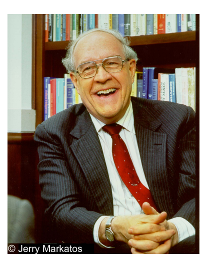
[57] S. B. Gilbert, "Perceived realism of virtual environments depends on authenticity," *Presence: Teleoperators and Virtual Environments*, vol. 25, no. 4, pp. 322–324, Fall 2017.

[58] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?" *Behavioral and Brain Sciences*, vol. 33, no. 2-3, pp. 61–83, 2010.

Richard Skarbez is a lecturer in the Department of Computer Science and Information Technology at La Trobe University in Melbourne, Australia. Prior to arriving at La Trobe, he earned his BS degree in computer engineering from the Pennsylvania State University in 2004 and his MS and PhD degrees in computer science from the University of North Carolina at Chapel Hill in 2010 and 2016, respectively. In 2017-2018, he was a postdoctoral researcher at Virginia Tech. His research interests broadly refer to how humans experience and interact with immersive technologies, and specifically include virtual reality, mixed and augmented reality, human factors, and user experience design. He is a member of the ACM and IEEE.

Mary C. Whitton is research professor (retired) of computer science at the University of North Carolina at Chapel Hill. She co-led the Effective Virtual Environments research group that investigated what makes virtual reality experiences achieve their goals. The group used knowledge of human perception to develop technologies and techniques that make virtual environments more effective for applications such as simulation, training, and rehabilitation. Before joining UNC in 1994, she was a co-founder of two companies (Ikonas—1978; Trancept Systems—1987) that produced high-end user-programmable hardware and libraries of software for graphics, imaging, and visualization. The companies' products were widely adopted in research laboratories for applications including seismic exploration, 3D medical imaging, intelligence, computer animation, and scientific modeling and simulation. The Ikonas RDS-3000 is widely recognized as the precursor of today's general purpose graphical programming units, GPUs.

Professor Whitton has held leadership roles in ACM SIGGRAPH including serving as President 1993-1995. She received the ACM SIGGRAPH Outstanding Service Award in 2013. Whitton is a member of ACM SIGGRAPH, and is a senior member of ACM and life senior member of IEEE. She is chair of the ACM SIGGRAPH History Committee and serves on the ACM History Committee. Whitton earned a BA from Duke University (1970), and an MS in Guidance and Personnel Services (1974) and an MS in Electrical and Computer Engineering (1984) from North Carolina State University.



Frederick P. Brooks, Jr. is Kenan Professor (retired) of computer science at the University of North Carolina at Chapel Hill. He received an AB summa cum laude in physics from Duke and a PhD in computer science from Harvard, under Howard Aiken, the inventor of the early Harvard computers. Brooks has served on the National Science Board and the Defense Science Board. He is a member of the National Academy of Engineering, the National Academy of Science, and the American Academy of Arts and Sciences. He has received the ACM A.M. Turing Award, the IEEE John von Neumann Medal, the IEEE Computer Society's McDowell and Computer Pioneer Awards, the ACM Allen Newell and Distinguished Service Awards, the AFIPS Harry Goode Award, the Eckert-Mauchly Award, and an honorary Doctor of Technical Science from the Swiss Federal Institute of Technology (ETH-Zürich).