PLAUSIBILITY ILLUSION IN VIRTUAL ENVIRONMENTS

Richard T. Skarbez

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2016

Approved by:

Mary C. Whitton

Frederick P. Brooks, Jr.

Benjamin Lok

Abigail Panter

Mel Slater

Gregory F. Welch

© 2016
Richard T. Skarbez
ALL RIGHTS RESERVED

ABSTRACT

RICHARD T. SKARBEZ: Plausibility Illusion in Virtual Environments
(Under the direction of Mary C. Whitton and Frederick P. Brooks, Jr.)

Historically, research into subjective user experience in virtual environments has focused on presence, the feeling of "being there" in the virtual environment. Recently, Professor Mel Slater proposed that in addition to this feeling of being in the virtual space, researchers also need to consider the subjective feeling that the events depicted in the virtual environment appear real. He coined the terms *Place Illusion* (PI) and *Plausibility Illusion* (Psi), respectively, to refer to these subjective feelings.

There exists a substantial amount of previous research applicable to PI, but very little regarding Psi. This dissertation fleshes out the concept of Plausibility Illusion by introducing new terminology, and reports on several experiments investigating the factors and effects of Psi. I demonstrate that Psi can be detected using existing presence measures, including questionnaires and physiological metrics. Of particular interest in these results is that factors contributing to Plausibility Illusion affected heart rate, with inconsistent behavior of the virtual environment leading to increased heart rate. I also demonstrate that study participants' individual differences affected how they interacted with a virtual environment, leading to different levels of Plausibility Illusion and, therefore, presence. I further demonstrate that, among the factors tested, the virtual body is the most important factor contributing to users' feelings of Plausibility Illusion, and that the coherence of the virtual scenario is the second most important factor. This shows it is feasible to determine a rank ordering of factors that affect users' sense of Plausibility Illusion in virtual environments, offering guidance to creators and developers.

To my parents, Tom and Joan Skarbez, who believed in me even when I didn't believe in myself.

ACKNOWLEDGMENTS

I could not have completed this dissertation without help, guidance, friendship, and support from a great many people. In particular, I thank my committee members:

- Mary Whitton, for her friendship, her encouragement, and her tireless work in helping me complete this dissertation. She made herself available during nights and weekends, often on short deadlines, and always provided excellent and useful feedback. This document was immeasurably improved by her efforts.

- Fred Brooks, for his guidance, his patience, and his example. Through his tutelage, I feel that I have improved tremendously as a writer, as a researcher, and as a person.

- Mel Slater, for his inspiration and his hospitality. His ideas of Place Illusion and Plausibility Illusion form the very bedrock of this dissertation, and the time I spent at his lab in Barcelona was among my most productive and most enjoyable periods in graduate school.

- Greg Welch, for helping guide me throughout my life as a graduate student at Carolina. In addition to this dissertation, I worked on two separate projects with Greg, one of which—Wide Area Visuals—was my first research project in graduate school. I learned much about computer graphics, tracking, simulation, training, and research from working with Greg.

- Ben Lok, for his good advice, for his help and the help of his lab in designing the Avatar experiment, and for introducing me to virtual humans. The treatment in this dissertation barely scratches the surface of what can be learned about plausibility illusion and virtual humans.

- Abigail Panter, for teaching me how to design experiments, and for being an unfailing source of cheer and support whenever we interacted.

I would also like to thank:

- The Effective Virtual Environments research group as a whole, and particularly Luv Kohli, Tabitha Peck, and Sharif Razzaque for their support and advice during this dissertation, Chris Oates and Chris VanderKnyff for their invaluable work in building EVEIL, and Jeremy Wendt, Eric Burns, Jason Jerald, Jeff Feasel, Jess Martin, Ky Waegel, and Alexis Chan, for their camaraderie.

- Missy Wood, Kelli Gaskill, Janet Jones, and Tim Quigg, for making sure that the Computer Science department was always a kind and welcoming place for me.

- The tremendous technical support staff of the UNC Computer Science department for all their help over the years, particularly David Harrison, John Thomas, David Musick, and Bil Hays.

- Jodie Gregoritsch, for helping make sure that I could actually graduate.

- The many professors I had a chance to study under and learn from during my time at Carolina, particularly Jim Anderson, Russ Taylor, Gary Bishop, Steve Pizer, Ming Lin, and Anselmo Lastra.

- The Wide Area Visuals research group, for giving me my first opportunity to do research in graduate school, thanks to Henry Fuchs, Herman Towles, Florian Gyarfas, and Tyler Johnson.

- Eric Bennett, for helping me when I taught the computer graphics course, which remains a highlight of my time at Carolina.

- The many talented performers and great friends from DSI Comedy Theater, for helping to keep me sane and to find myself, especially Paula Pazderka, Kevin Browning, Jonathan Yeomans, Sarah Eldred, Brandon Holmes, Nick Faber, Andy Lavender, Shane Smith, Liz Sanders, Joe Dawson, Lauren Herget, and Kara Phelps.

- All the other friends who helped me along the way, particularly Tim Thirion, Brad Moore, David Borland, Caroline Green, David Marshburn, Jeff Gaines, Tom Hart, Chris Werner, Laura Kassler, Kyle Moore, Cory Quammen, Katie Harrold, Allison Rackley, Toni Sheppard, Grace Carnes, and Greg Hohn.
- The EVENT Lab at the Universitat de Barcelona. In particular, thanks to Xavi Navarro, Rodrigo Pizarro, and Alejandro Beacco Porres for their invaluable help in implementing my final experiment, Irene Sanjuan Navais for making sure I was taken care of, and Birgit Nierula, Ausias Pomes, Anna Bellido, and Sofia Seinfeld, for their kindness.
- Sameer Kishore and Kanaka Raghavan, for their help and valued friendship while I was in Barcelona.
- Solène Neyret, without whom I could not have completed this dissertation.

Finally, and most importantly, I would like to thank my family, especially my parents, Joan and Tom Skarbez, my sister Katie, and my grandparents. They have always provided love, support, and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Introduction - Evaluating Virtual Experiences

Virtual environments (VEs) are tremendously sophisticated human-computer interfaces that are used for a wide variety of applications. Examples include psychological treatment, psychological research, military and medical training, entertainment, and sociological research. Each of these applications has different task requirements and objectives, and suggests different hardware or software implementations. Additionally, there is no consistent definition for what constitutes an effective VE. For these reasons, there does not yet exist a single, generalizable metric that can be used to determine whether a given VE is a success or not.

It is possible to identify specific measures that determine the effectiveness of a particular VE. For example, if a VE is developed to train participants to complete an assembly task, one could create tests of that training, e.g., number of units assembled per unit time, number of errors per unit time, percentage of units correctly assembled. One could then develop a controlled experiment where some participants are trained using the VE and some are trained using whatever the traditional technique is. Then, when both sets of users are tested post-training, these measures would give some concrete evidence for whether the VE was effective at training. Such formal training transfer studies are rarely done, however, due to the time, effort, and cost required. Even if they were done, though, they would not enable the comparison of results among different VEs designed for different purposes.

The development of generalizable measures of VE effectiveness, then, remains an open research problem. One such concept, that of presence, has been driving VE research for decades.

1.2 Presence

Presence has been defined and operationalized in many ways by different researchers (see Chapter 2 for a more thorough overview of presence research), but it is most commonly defined as something akin to the feeling of "being there" in a virtual place. One example comes from Witmer and Singer, who defined presence as "the subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer and Singer, 1998).

Presence has the distinct advantage of being a metric applicable to any VE. One can reasonably ask how present a user was in any given VE A, and then ask how present the user was in some VE B, and if the user reports more presence in VE A, then that is some evidence that enables the comparison of VEs A and B, though they may represent entirely different scenarios and be designed for entirely different purposes.

While conceptually appealing, the evaluation procedure described in the previous paragraph has several very important flaws. First is that determining "how present" a user is is in itself a very difficult problem. Presence is what the philosophy literature calls a *quale* (plural *qualia*), which is defined as a subjective and internal feeling elicited by sense perceptions. This subjective and internal nature makes measurement of presence (or any quale) extremely difficult. The predominant method has been to use one or more post-experiment questionnaires to measure presence, but this is itself problematic. (See Chapter 2 for an overview of presence questionnaires and the difficulties associated with using them.) There have been efforts to develop objective correlates of presence, including physiological (Meehan et al., 2002) and behavioral (Freeman et al., 2000) measures, but these are also flawed, requiring the addition or modification of elements of the VE to enable measurement of presence.

For example, the most common physiological surrogate for presence is arousal, which can be detected using measures of heart rate or skin conductance. The change in heart rate associated with the onset of a stressful stimulus was shown by Meehan to correlate with presence (Meehan et al., 2002). However, adding a stressful stimulus to a non-stressful training task may violate the ecological validity of the training. Or, if a task also involves physical exertion, it may not be possible to distinguish the effect of the stressful stimulus from the heart rate changes associated with the exertion. Therefore, while physiological and behavioral metrics are promising, they are not one-size-fits-all solutions.

In Meehan's dissertation he posits 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 (Meehan, 2001). I would go further and suggest that such an ideal metric should also be measurable contemporaneously, continually, and without modification to the scenario. No measure of presence yet exists that meets all of these criteria and is also generalizable across VEs.

A second important flaw, beyond the difficulty inherent in measuring presence, is that it has not been conclusively demonstrated that more presence is a good thing. Welch argues powerfully that there is no inherent reason to think that more presence leads to improved task performance in a VE (Welch, 1999), and experimental results linking presence and task performance are mixed (Witmer and Singer, 1998)(Slater et al., 1996).

So, presence is defined and operationalized in many different and occasionally conflicting ways, is inherently difficult to measure (and there are many different measures), and may not actually correlate with what one actually cares about in a VE in the first place. To address some of these concerns, Slater proposed a theory that presence is composed of two orthogonal components, Place Illusion and Plausibility Illusion, and that together these factors will enable a user to react-as-if-real (RAIR) while in a VE (Slater, 2009). For many applications, eliciting RAIR from users is in fact what it means for a VE to be effective.

1.3 Place Illusion and Plausibility Illusion

In 2009, Slater introduced the concepts of Place Illusion (PI) and Plausibility Illusion (Psi) as the two constructs that contribute to realistic response in virtual environments. He defined PI as, "the . . . illusion of being in a place in spite of the sure knowledge that you are not there," and Psi as, "the illusion that what is apparently happening is really happening (even though you know for sure that it is not)" (Slater, 2009). PI, then, corresponds to the traditional conception of presence as "being there", while Psi represents an entirely different conception of presence, that of believing what you are seeing. For example, assume you are in a VE is intended to represent a library. Here, presence would be your feeling of, "I am in a real library." If you turned your head and saw more bookshelves, that would reinforce your feeling of PI. If all the library patrons were being quiet, that would reinforce your feeling of Psi. Contrarily, if you turned your head and the imagery didn't change, that would break PI, and if patrons were yelling loudly in the library, that would break Psi.

1.4 Thesis Statement

In this dissertation,

1. I usefully elaborate Slater's concept of Plausibility Illusion with the auxiliary concepts of virtual scenarios, coherence, and reasonable circumstances, and
2. I describe experiments that have shown the following:
    - Plausibility Illusion can be detected using existing presence measures,
    - Plausibility Illusion is impacted by individual differences,
    - The virtual body is the most important contributing factor to Plausibility Illusion, and
    - The coherence of the scenario is the second most important factor, among those tested.

In this chapter, as well as in Chapters 2 and 3, I define several concepts related to Plausibility Illusion, most notably virtual scenarios, coherence and reasonable circumstances, which enable the discussion of Plausibility Illusion with its own vocabulary as a parallel concept to Place Illusion.

Chapter 2 reviews and summarizes the existing presence literature, and categorizes existing thinking about components of presence using the Place Illusion/Plausibility Illusion framework.

Chapter 3 fleshes out the concepts of Place Illusion and Plausibility Illusion, in particular, commenting on several existing questions regarding virtual environments that could be productively addressed by considering Plausibility Illusion.

In Chapter 4, I present the results of an experiment that strongly suggests that participant motivation is one of the individual differences that impacts Psi.

In Chapter 5, I demonstrate that existing presence measures—namely, the Witmer and Singer Presence Questionnaire (PQ), the Slater-Usoh-Steed questionnaire (SUS), and physiological measures of arousal—can be affected by various factors believed to influence both Place Illusion, Plausibility Illusion, and the interaction of the two. The proposed interaction is illustrated in Figure 1.1.

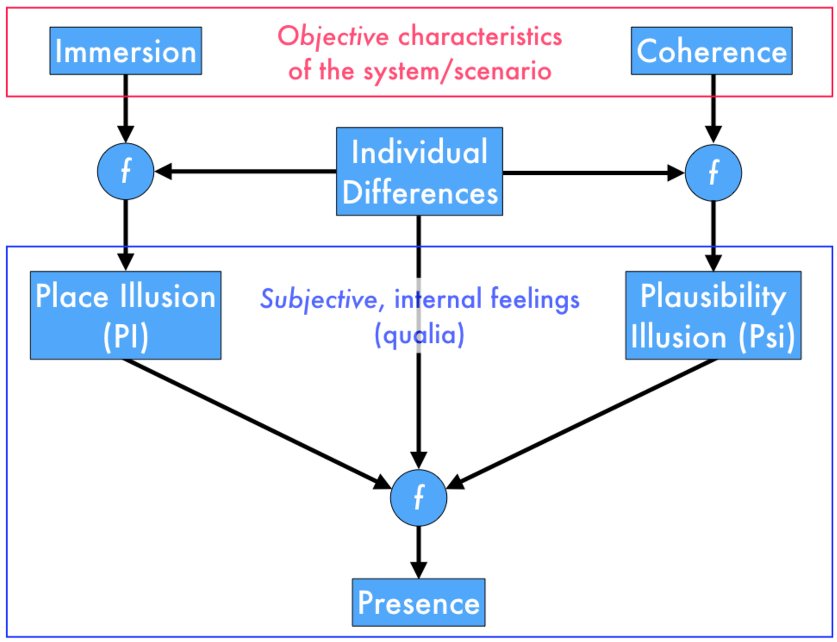

Figure 1.1: Illustration of the proposed interaction of immersion, coherence, Place Illusion, Plausibility Illusion, and presence

In Chapter 6, I demonstrate that participants may interact differently with a virtual environment, specifically in a virtual football-playing task, and that these differences are reflected in participants' order of selected improvements in the virtual environment as well as their reported presence on the SUS questionnaire.

In Chapter 6, I also demonstrate that participants who were directed to improve their level of Plausibility Illusion overwhelmingly chose to improve their virtual bodies over other available improvements to the virtual scenario. This shows that the virtual body is the most important component of a virtual scenario for eliciting Plausibility Illusion, and also demonstrates the feasibility of determining whether some aspects of a virtual scenario are more important than others for the purpose of eliciting Plausibility Illusion.

1.5 Definitions

To aid the reader, this section defines several terms and phrases frequently used throughout this dissertation. Some of these are taken from the literature, but others—virtual scenarios, reasonable and unreasonable circumstances, coherence—are my own.

In this dissertation, a *virtual scenario* (also *scenario*) is specifically the virtual environment and the behavior of that environment. A virtual recreation of a historic site without any characters or interactable objects would be a virtual environment, but not a virtual scenario. If there were also virtual humans populating that environment, or objects in the environment that could be manipulated, it would be a virtual scenario.

Sensorimotor contingencies are the regularities in how sensory stimulation depends on the activity of the perceiver (O'Regan and Noë, 2001). For example, if one leans toward a particular object, it takes up more of their field of vision.

A *valid action* is any action a VE user can take that results in a change to the state of the virtual environment (an *effective valid action*) or to her perception of the virtual environment (a *sensorimotor valid action*) (Slater, 2009). Note that an action can be both effective and sensorimotor: for example, if a participant has a fully-body-tracked avatar and takes a step to change his viewpoint in the environment (a sensorimotor action), he also changes the location of the avatar model in the environment (an effective action). In fact, under normal circumstances in the real world, physical movements are always both sensorimotor and effective. In virtual environments, it is possible for an action to be a valid sensorimotor action but not a valid effective action (moving one's head in a VE where there is head tracking but no user-controlled virtual body), or vice versa (moving one's body when it is controlling a fully-body-tracked avatar but the view is from a third-person perspective on a stationary desktop display).

The *immersion* of a VE system is the set of all valid actions supported by that VE system(Slater, 2009). A VE system that is *less immersive* (or has *lower immersion*) than another, *more immersive* (*higher immersion*) system can be simulated by the second system, but not vice versa. A trivial example would be that a system that provides two buttons for user input is more immersive than a system that provides only one button. Or for a more real-world example, a system that uses a head-tracked HMD for display is more immersive than a stationary desktop ("fishtank") VE system, as the former could be used to simulate the latter, but not vice versa.

In this dissertation, a *reasonable circumstance* is a state of affairs in a virtual scenario that is self-evident given prior knowledge. This knowledge can come from the real world (for example, if objects in the VE seem to be affected by gravity in the same way they would be in the real world), or from within the VE (for example, if two characters start to fight immediately upon meeting each other, after the user has previously been told that the two are feuding). Objects can behave reasonably—a rock that is kicked can move like it would on the real Earth, a teleporter can make a box blink out of existence—and characters can behave reasonably—a man can wave and say hello when you walk past, a wizard can converse with a dragon. *Unreasonable circumstances* are the opposite, that is, they are states of affairs that are inconsistent with prior knowledge.

The *coherence* of a virtual scenario is the set of reasonable circumstances that can be demonstrated by the scenario without introducing unreasonable circumstances. A *more coherent* (or *higher coherence*) virtual scenario can demonstrate the same amount or more reasonable circumstances while demonstrating the same amount or fewer unreasonable circumstances than a *less coherent* (or *lower coherence*) version of the same scenario. That is, all else being equal, it can demonstrate a greater number of reasonable circumstances, fewer unreasonable circumstances, or both. If one assumes that demonstrating reasonable circumstances is more difficult than demonstrating unreasonable circumstances, this enables the same partial ordering relationship for coherence that is defined for immersion. That is, a higher coherence scenario can "simulate" a lower coherence version of the scenario by artificially injecting unreasonable circumstances or "turning off" some reasonable circumstances, but not vice versa.

Place Illusion (PI) is the illusion of being in a place in spite of the sure knowledge that you are not there (Slater, 2009).

Plausibility Illusion (Psi) is the illusion that what is apparently happening is really happening, in spite of the sure knowledge that it is not (Slater, 2009).

A *quale* (plural *qualia*), is a subjective and internal feeling elicited by sense perceptions. Place Illusion, Plausibility Illusion, and presence are all qualia. Immersion and coherence, on the other hand, are not.

1.6 Summary of experimental results

(Note: The first three experiments—described in Chapters 4 and 5—were performed at the University of North Carolina at Chapel Hill. The final experiment—described in Chapter 6—was performed at the Universitat de Barcelona.)

The first experiment (Coherence of virtual human interactions) involved 32 medical student participants who conducted an interview with a virtual human patient. Each participant experienced the virtual human in either a high-coherence or low-coherence condition. In both conditions, the virtual human was controlled by an experimenter in a Wizard of Oz (WoZ) configuration. In the high-coherence condition, the virtual human responded to the participant as quickly and accurately as possible. In the low-coherence condition, the virtual human's responses included a variety of conversational errors introduced according to a script. There was a marked difference between the two groups in their head-tracking logs—participants who interacted with the low-coherence virtual patient looked at the patient less—however, there was no difference in their abilities to perform the task or their ratings of the virtual patient. This suggests that high user motivation may be able to compensate for a deficit in coherence.

The second experiment (Immersion and Coherence in the Pit) involved 32 participants in the Pit, a virtual visual cliff environment. The scenario was presented with one of two levels for each of two factors in a 2x2 design: field of view (as a system characteristic affecting immersion) and behavioral realism, in the form of instructions being true or not (as a characteristic of the scenario affecting coherence). These factors were chosen to elicit differences in PI and Psi, respectively. This experiment failed to identify any metrics that could distinguish between PI effects and Psi effects. Analysis identified several possible reasons why the experiment failed to generate a result, even if PI and Psi are in fact distinguishable in outcome measures. The most likely reason was that the factor levels in the experiment were not sufficiently different, so that individual differences obscured any effect of the experimental conditions. In particular, this experiment varied only one parameter each for immersion (field of view) and coherence (correctness of experimental instructions), while each of these are certainly multidimensional constructs. As justification, consider that immersion must at the very least consider sensorimotor contingencies for each of the sensory modalities supported by the VE, and coherence would be impacted if, for example, virtual characters were introduced to the environment.

A third experiment was designed to address the concern from the previous experiment that the levels of immersion and of coherence were not sufficiently distinct from one another. The overall goal of the experiment was the same, to see if immersion and coherence produce different responses when participants are exposed to a visual cliff. Here, the two factors were again immersion and coherence, but the way they were manipulated was different. The high and low immersion levels were generated by manipulating field of view, passive haptics, and sound cues (either all at a high level, or all at a low level). The high and low coherence levels were generated by manipulating physical behavior of objects, validity of instructions, and the presence of scenario-inappropriate sound (again, either all high, or all low).

Once again, this experiment failed to identify any traditional measures of presence that could distinguish between PI effects and Psi effects. However, presence questionnaires did indicate that high levels of immersion and coherence together result in higher presence, compared to any of the other 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. In addition, participants' heart rates, 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 (as in the case of the low coherence level in this experiment) can cause stress in a user that affects physiological responses.

A final experiment (Factors of Coherence and Psi) involved 20 participants who experienced a virtual bar scenario at a variety of levels of coherence. In this experiment, participants chose the number and order of coherence improvements in order to match the level of Psi they felt when experiencing the bar with all coherence factors at the maximum level, making as few improvements as possible to do so. There were two improvements available for each of four factors: the coherence of virtual human behavior, the coherence of behavior of one's own virtual body, coherence of physical interactions (the physical behavior of a football in the environment), and scenario coherence (the match between the rendered environment and the described scenario).

Participants overwhelmingly chose improvements to their own virtual body first, followed by the scenario, with the other two factors (virtual human coherence and coherence of physical interactions) being improved last, on the whole. Responses to the ball varied substantially among participants, with participants who interacted with the ball more (predominantly males) reporting higher presence and upgrading the ball sooner.

1.7 Overview of the Thesis

Chapter 2 introduces background research relating to presence, focusing on definitions, measurements, and previous experimental results. Chapter 3 presents several theoretical arguments that further develop the theory of PI and Psi. Chapter 4 presents the design and analysis of a study evaluating Psi in the context of training medical students using virtual humans. Chapter 5 presents the design and analysis of an initial study investigating PI and Psi in the Pit, a visual cliff virtual environment, as well as it's follow-up study. Chapter 6 presents the design and results of a study exploring the relative importances of factors of Psi. Finally, Chapter 7 contains a discussion of overall results and thoughts on future work needed to further develop the theory and application of PI and Psi in the field of virtual environments.

CHAPTER 2

Presence and related concepts

In Chapter 1, I introduced the presence construct, along with comments regarding the ways in which it is defined and measured. In this chapter, I explicate many of these definitions and methods of measurement, identify some of the core themes and key differences in how presence is discussed, and introduce related concepts including transportation, agency, and reality judgment. I present the work on presence as three sections: definitions of presence, measurement of presence, and models of presence—that is, works that propose a list of factors that contribute to presence. I present my analysis at the end of each section.

The concepts presented here are the foundation of the work presented in this dissertation. While Place Illusion and Plausibility Illusion were introduced to the literature only in 2009, they build on twenty years of prior work endeavoring to better define and measure the presence construct.

2.1 Defining Presence

There have been many definitions of presence proposed in the literature. I propose that these can be grouped into three categories: being there, non-mediation, and real objects. Those definitions I classify as being there consider presence to be the feeling of being in an environment, while those I classify as non-mediation consider presence to be a lack of attention to the mediating technology. I further propose that the being there definitions can be subdivided into two subcategories, active (in which the ability to act is specifically considered as part of the definition) and passive (in which user actions are not specifically addressed). I also propose that non-mediation definitions consist of two subcategories, internal (in which the focus is on one's thoughts, as in "suspension of disbelief" and external (in which the focus is on the technology, as in the "illusion of non-mediation"). The classification of definitions into these categories and subcategories can be seen in Table 2.1.

Table 2.1: Categories of presence definitions

Being there		Non-mediation		Other
Active	Passive	External	Internal	Real objects
(Steuer, 1992)	(Witmer and Singer, 1998)	(Lombard and Ditton, 1997)	(Slater and Usoh, 1993)	(Lee, 2004)
(Schloerb, 1995)	(Sas and O'Hare, 2003)	(International Society for Presence Research, 2000)		
(Flach and Holden, 1998) (Zahorik and Jenison, 1998)	(Spagnolli and Gamberini, 2004)			
(Mantovani and Riva, 2001)	(Wirth et al., 2004)			
(Biocca, 2001)				
(Slater, 2004b)				
(Carassa et al., 2005)				
(Witmer et al., 2005)				
(Riva et al., 2006)				
(Herrera et al., 2006)				
(Wirth et al., 2007)				

In the remainder of this section, I present these definitions grouped by category, and arranged in chronological order within each category. My comments follow at the end of this section.

2.1.1 Definitions

The notion of presence as it is used in the context of virtual reality can be traced to psychologist James Gibson, who defined *presence* as "the experience of one's physical environment . . . not [one's] surroundings as they exist in the physical world, but [the] perception of those surroundings as mediated by both automatic and controlled mental processes" (Gibson, 1979). In this Gibsonian context presence is explicitly in the context of the real world, but already the idea is in place that presence can't be determined simply by considering the ground truth of the real environment, it is a *subjective* feeling generated by our perception of the real world as *mediated* by our sense organs and the mental processes governing and integrating them.

2.1.1.1 Being there

Steuer introduced Gibson's notion of presence to the field of computer-mediated environments, defining the term *telepresence* as "the experience of presence (in the sense of Gibson) in an environment by means of a communication medium" (Steuer, 1992). Steuer's definition is the start of some significant confusion, as many researchers have been concerned primarily—or only—with the sense of presence in computer-mediated or virtual environments, but common practice has been to refer to the sensation simply as presence, rather than "telepresence" or "virtual presence."

Schloerb introduced a different, "objective" definition of presence (Schloerb, 1995). *Subjective presence* occurs when one perceives oneself as physically present in an environment. However, one is only *objectively present* if one can successfully complete a specified task in the environment. Here, then, we have an explicit, and in fact definitional, link between presence and task performance: If one can successfully complete more tasks more often, one is more present.

Flach and Holden returned to Gibson's research as the basis of presence, but not just in terms of his definition of presence (Flach and Holden, 1998). To Gibson, "the reality of experience is grounded in action"—we see the world in terms of affordances; that is, in how can we interact with the world around us. The important characteristics of the world, then (and in particular, the characteristics of the world that are important for experiencing presence) are behavioral, rather than aesthetic.

Continuing that line of thinking, Zahorik and Jenison described presence as "tantamount to successfully supported action in the environment" (Zahorik and Jenison, 1998). To them, presence is determined by the extent that the perception/action coupling in the virtual world matches our learned perception/action coupling in the real world.

Mantovani and Riva presented a view of the Gibsonian actor in (Mantovani and Riva, 2001). For such an ecologically situated actor there is not a clear separation of the subjective internal model of the world and the objective ground truth of the outer world. Rather, the actor is constantly in a process of adaptation to the estimated (that is, mediated) world in which they exist. In this picture, one's willingness to "react as if real" to the observed stimuli and the world's ability to "react as if real" to my sensorimotor actions are in fact inseparable.

In the paper which presented their landmark Presence Questionnaire, Witmer and Singer defined *presence* as, "the subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer and Singer, 1998).

In 2001, Biocca defined *presence* as "the phenomenal state by which an individual feels located and active in an environment, and, especially in the case of telepresence, the class of experience where the environment is mediated by a technology" (Biocca, 2001). So here, the user must not only be "located" (the traditional sense of "being there"), but must also be "active." This is in keeping with the Gibsonian tradition, as in Zahorik and Jenison (Zahorik and Jenison, 1998) or Flach and Holden (Flach and Holden, 1998). Note also that these authors generally treat telepresence as a special case of presence, and that presence can be (and normally is) felt in the real world.

In 2003, Slater revisited presence terminology, describing presence as a "response" to "an appropriate conjunction of the human perceptual and motor system and immersion" (Slater, 2004b). This is quite similar (albeit using very different terminology) to Zahorik and Jenison's conception of presence as being-in-the-world (Zahorik and Jenison, 1998). Here, if we assume an actor (with a functioning perceptuomotor system) ecologically situated in the world (the precise nature of the world and this situation being defined as *immersion*), then presence arises to the extent that a valid perception/action coupling is supported by the virtual environment (VE) system. Also, note the novel conception of presence as a "response." Presence, in this conception, can occur involuntarily: if the correct set of stimuli are provided (in terms of the immersion of the system and the perceptuomotor characteristics of the individual user), then presence will result. This seems to represent an evolution in thinking from Slater's earlier definition of presence involving suspension of disbelief.

Immersion

An aside about immersion: Immersion is the source of some confusion in its own right. Slater has consistently regarded immersion as an objective characteristic of a VE system (Slater, 1999). This is in contrast to Witmer and Singer, who define *immersion* as, "a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" (Witmer and Singer, 1998). Lombard and colleagues refer to these different conceptions of immersion as *perceptual immersion* (Biocca and Delaney, 1995) and *psychological immersion*, respectively (Lombard et al., 2000). These are clearly related concepts—Slater's immersion is what makes it possible to experience Witmer and Singers immersion—but using them interchangeably has led to a lack of clarity in the VR literature. In this dissertation, unless specifically stated otherwise, immersion is used to mean Slater's conception of immersion as an objective characteristic of a VE system.

Sas and O'Hare offered a slightly different conception of presence: one is *present* in another world (mediated or imaginary) if (1) one's cognitive processes are oriented toward that world to the extent that one experiences "being there", and (2) one's focus of consciousness is on the proximal (body-oriented, perhaps) stimuli in the mediated or imaginary world (Sas and O'Hare, 2003). Here, as in Biocca (Biocca, 2001), we see a definition of presence as "being there plus", in this case, the "plus" being the fact that one is responding to stimuli from the virtual world, rather than the real one. It would seem to me that this definition assumes that one can, at any given moment, only be present in one or the other environment; that your consciousness can have two foci seems unlikely.

Spagnolli and Gamberini, on the other hand, maintained the focus of presence on location (Spagnolli and Gamberini, 2004): "Whenever a person is qualified as 'present' . . . her location is the salient, characterizing feature." So an interaction with a virtual human in the real world might elicit a whole host of feelings, but it would not elicit a sense of presence.

Wirth and colleagues define *spatial presence* as "the subjective experience of being in the mediated environment" (Wirth et al., 2004). Here again we have presence limited to mediated environments. For the authors it has two components: the sensation of being physically situated in the environment (self-location) and the perception of possibilities to act in that environment (possible actions). This is another example of "being there plus", as well as another definition of presence that is largely in keeping with the traditional Gibsonian definition.

Carassa and her colleagues propose a definition of *presence* inspired by situated cognition theory, in which "presence depends on the proper integration of aspects relevant to an agent's movement and perception, to her actions, and to her conception of the overall situation in which she finds herself, as well as on how these aspects mesh with the possibilities for action afforded in the interaction with the virtual environment" (Carassa et al., 2005). In my interpretation, this correctly integrates a user's learned expectations for correct behavior into the Gibsonian model of presence in virtual environments.

Wilmer and Singer clarify their definition of presence in (Witmer et al., 2005). *Presence* is defined as a psychological state of "being there" mediated by an environment that engages our senses, captures our attention, and fosters our active involvement." They also define *involvement* as "a psychological state experienced as a consequence of focusing one's mental energy and attention on a coherent set of stimuli or meaningfully related activities or events."

Riva and colleagues defined *presence* as "the non-mediated (prereflexive) perception of successfully transforming intentions in action (enaction) within an external world" (Riva et al., 2006). This builds on Zahorik and Jenison's conception of presence as successfully supported action in the environment (Zahorik and Jenison, 1998). Here, we have no distinction between real, virtual, or imaginary worlds; you can feel present in any or all of them if you are able to transform your intentions into action. Also, it is a prereflexive, or intuitive, perception: it is again something that "just happens" if the system is sufficiently in tune with your needs (or vice versa), similar to Slater's conception of presence as a response.

Herrera et al. define *presence* as the "conscious awareness of self, as both agent and experiencer, which characterizes the experiencing self of natural environments" (Herrera et al., 2006). Here again, this could apply to real, virtual, and mediated environments (although, the authors state, not imaginary ones—the environments they refer to are "experienc[ed]", not imagined). This conception of self as both "agent and experiencer" again echoes Gibson, in whose view one is always both acting on and being acted upon by the environment.

Wirth and colleagues refine their definition of *spatial presence* as, "a binary experience, during which perceived self-location and, in most cases, perceived action possibilities are connected to a mediated spatial environment, and mental capacities are bound by the mediated environment instead of reality" in (Wirth et al., 2007).

2.1.1.2 Non-mediation

Slater and Usoh introduced the notion that *presence* in a mediated environment is "(suspension of dis-)belief that [one] is in a world other than where [one's] body is located" (Slater and Usoh, 1993). Slater and Usoh's definition introduces several new ideas: one, that we first believe that we are not in the mediated environment, and two, that by some process, we can overcome that belief. This idea of presence involving "suspension of disbelief" is a recurring concept in the presence literature.

Lombard and Ditton proposed that *presence* is "the perceptual illusion of nonmediation" (Lombard and Ditton, 1997). Note that this definition is explicitly for tele- or virtual presence, not presence in a real environment; it assumes the existence of a communication medium that can seem to disappear. One could make the argument, as in Gibson, that our experience of the world is always mediated—that our perception of the world is not the same as the world itself—and so presence in the real world is only a special case of such mediated presence experiences. Lombard and Ditton do not make this argument.

Lombard and Ditton also explicitly defined presence as binary. “It does not occur in degrees but either does or does not occur at any instance during media use.” This is closely associated to the conceptualization of presence as an illusion: either the illusion is in place, or it is broken. Slater’s later conceptions of Place Illusion and Plausibility Illusion are very much in keeping with this school of thought (Slater, 2009).

Presence was officially defined by the International Society for Presence Research as follows:

> “*Presence* (a shortened version of the term *“telepresence”*) is a psychological state or subjective perception in which even though part or all of an individual’s current experience is generated by and/or filtered through human-made technology, part or all of the individual’s perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at *some level* and to *some degree*, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience. *Experience* is defined as a person’s observation of and/or interaction with objects, entities, and/or events in her/his environment; *perception*, the result of perceiving, is defined as a meaningful interpretation of experience”(International Society for Presence Research, 2000).

This definition is clearly indebted to Lombard and Ditton, as the focus is on the illusion of nonmediation rather than the experience of a place. However, it would seem that the ISPR authors reject Lombard and Ditton’s belief that presence is binary, with the language of “part or all” of an individual overlooking the mediating technology to “some level and to some degree.” Note also that the authors are explicit about the fact that they are using *presence* to mean *telepresence*, indicating clearly that this definition is only applicable to technology-mediated interactions.

2.1.1.3 Other - Objects are experienced as real

Kwan Min Lee defined *presence* as "a psychological state in which virtual (para-authentic or artificial) objects are experienced as actual objects in either sensory or nonsensory ways" (Lee, 2004). This is a new definition, that clearly puts the focus on things in the virtual world. If one experiences these things as actual objects, they are present; if one doesn't, they aren't. The "sensory or nonsensory" language is included specifically to account for situations where feelings of presence are elicited by non-immersive media such as text (known as the "book problem"). So "being there" is no longer the primary quality of the experience, making this definition more amenable to usages in applications where one, for example, interacts with a virtual human rather than experiencing a new place.

2.1.2 Analysis

So what, in the end, do we mean when we say *presence*? It seems to me that the shortest and most commonly used definition, "the feeling of 'being there'," actually comes quite close to the heart of the matter. Defining presence as a feeling has some theoretical grounding, as well; Schubert conceptualizes presence as a "cognitive feeling," with all that entails (Schubert, 2009).

I do not agree with some aspects of presence that appear in the definitions in the previous section. Firstly, I disagree with definitions that include an element of the illusion of nonmediation. Spagnolli and Gamberini showed that users were capable of acting simultaneously in the virtual/mediated environment and the real environment (Spagnolli and Gamberini, 2002). It seems clear in this case that the user is aware, at least on some level, that it is a mediated experience, since they are able to speak and act in ways that demonstrate their awareness of the mediation. Similarly, I feel that the very existence of the book problem (Biocca, 2002) is reason to doubt this conception of presence. I am not aware of any study that attempted to demonstrate that readers are present only in the environment presented in the book, but I suspect, on face, that while a user reading a book may report feeling presence, they are always aware of the fact that they are reading a book.

I also feel that the conception of presence as a binary (on/off) construct is not necessarily true. The Spagnolli and Gamberini study cited above provides some evidence to the contrary, and Schubert also argues against this requirement. It may be true that users will report feeling present primarily in one space at any given time, but even so, there is no reason to believe that the strength of this feeling must be constant. It may be that "feeling of presence" can be conceived of as a continuous function that, as it rises and falls, may rise above or fall below a binary threshold.

Definition of presence in this dissertation

In this thesis, then, I define *presence* as the cognitive feeling of being in a place. This feeling can change based on the sensory representation of the place (particularly in the case of a mediated environment, where this is dictated by the immersion of the mediating technology), the affordances available to the user, the scenario in which the user finds himself, and the user's personal history, state and traits. Or, in short, the user who is present is located and active in the space, whether real or mediated.

I claim that presence arises from the immersion of the system (the sensorimotor and effective valid actions it supports), the coherence of the scenario, and the individual characteristics of the user. That is, it arises naturally in a user who experiences Place Illusion and Plausibility Illusion.

2.2 Measuring presence

Welch and colleagues identify self-report, behavioral, and physiological measures as potential means of measuring presence (Welch et al., 1996). I follow that categorization here in discussing the variety of presence measures that appear in the literature.

2.2.1 Self-report

Self-report refers to all techniques in which users report their subjective feelings of presence to the experimenter. An important subset of self-report measures are post-experience questionnaires, which are discussed separately below.

2.2.1.1 Questionnaires

This section briefly describes several existing questionnaires designed to measure presence (typically referred to as presence questionnaires), and concludes with discussion of their history of use.

The first commonly used questionnaire to appear in the literature was the Slater-Usoh-Steed (SUS) questionnaire, which first appeared in some form in (Slater et al., 1993) (later republished as (Slater et al., 1995)). In the 1993 version of the SUS questionnaire there were only three questions. The more common form of the questionnaire has six questions, and can be seen in, for example, (Usoh et al., 2000).

Kim and Biocca introduced a questionnaire based around the constructs of arrival and departure (Kim and Biocca, 1997). *Arrival* is the feeling of being there in a mediated environment, *departure* is the feeling of not being in the real environment.

Witmer and Singer introduced their Presence Questionnaire (PQ), as well as their Immersive Tendencies Questionnaire (ITQ), in (Witmer and Singer, 1998). The PQ is based on the authors' conception of presence as having four major categories of factors: control, sensory, distraction, and realism. Each of the thirty-two questions is designed to address some aspect of one of these four factors.

Lombard et al. discuss their efforts to develop an instrument for presence based on their theoretical model of its components in (Lombard et al., 2000). In the paper, they identify six "dimensions" of presence they found in the literature: presence as social richness, presence as realism (both social and perceptual), presence as transportation, presence as immersion, presence as a social actor within a medium, and presence where the medium is a social actor. The authors claim that the common element among these types of presence is a perceptual illusion of nonmediation. To measure these different conceptions of presence, the authors present a 103-item questionnaire.

Baos and her colleagues argue that presence and reality judgment (the belief that our experiences are real, or, they say, willing suspension of disbelief) are separate constructs and should be treated as such (Baños et al., 2000). They present an initial seventy-seven-item questionnaire, the Reality Judgment and Presence Questionnaire (RJPQ), intended to measure both constructs. They chose questions to address nine factors of experience: reality judgment, presence, emotional involvement, interaction, control, attention/flow, realism, congruence/continuity, and expectations.

Lessiter and colleagues introduce the ITC Sense of Presence Inventory (ITC-SOPI) in (Lessiter et al., 2001).The intent of this 44-item questionnaire is to focus entirely on the users experience with the media, and so there are no questions that address specific properties of either the system (e.g., input devices), or the content (e.g., story elements). It is intended to be usable with a variety of media types, including non-immersive and non-interactive media, such as television programs or movies.

Schubert, Friedmann, and Regenbrecht introduce the igroup Presence Questionnaire (IPQ) in (Schubert et al., 2001). The authors follow Zahorik and Jenison in connecting presence to supported action in the VE (Zahorik and Jenison, 1998). This thirteen-item questionnaire is intended for use in all forms of virtual environments, including immersive VR systems, desktop VR, 3D games, and text-based VEs such as MUDs (multi-user dungeons).

Vorderer et al. present the MEC Spatial Presence Questionnaire (MEC-SPQ) in (Vorderer et al., 2004). This questionnaire assumes that spatial presence is built of nine constructs: four process factors (attention allocation, spatial situation model, spatial presence—self location, spatial presence—possible actions), two psychological state factors (higher cognitive involvement, suspension of disbelief), and three psychological trait factors (domain-specific interest, visual/spatial imagery, and absorption). They offer short, medium, and long versions of the MEC-SPQ, comprised of 4, 6, or 8 questions, respectively, for each construct (thirty-six, fifty-four, or seventy two questions in total).

Sas and O'Hare developed a novel 34-item questionnaire for their experiment in (Sas and O'Hare, 2003). They validated this questionnaire by demonstrating that it was highly significantly correlated with the SUS questionnaire.

Chertoff and colleagues present a survey developed to measure "holistic virtual environment experiences" in (Chertoff et al., 2010). By holistic, the authors seem to mean that the environment incorporates aspects of experiential design; specifically that it includes affective (emotion) and cognitive (engagement) aspects. The survey includes 17 questions addressing five dimensions of experiential design: affective, cognitive, sensory (immersion), active ("personal connection ... to an experience"), and relational (social).

Published use of questionnaires

Rosakranse and Oh identify five canonical presence questionnaires—the Slater-Usoh-Steed (SUS) questionnaire, the Witmer-Singer Presence Questionnaire (PQ), the igroup Presence Questionnaire (IPQ), the ITC-Sense of Presence Inventory (ITC-SOPI), and the Lombard and Ditton questionnaire—and track their histories of use in three academic publishing outlets—*Presence: Teleoperators and Virtual Environments*, the ISPR conference proceedings, and *Cyberpsychology, Behavior, and Social Networking* (Rosakranse and Oh, 2014). It is notable that these three outlets represent different research communities. *Presence* tends to focus on research in immersive virtual environments, while the ISPR conference primarily focuses on media scholarship, and *Cyberpsychology* is an outlet for psychology researchers.

Rosakranse and Oh found that in *Presence*, the PQ and SUS questionnaires have remained dominant, while in ISPR, the ITC-SOPI questionnaire is now most commonly used, and in *Cyberpsychology*, SUS, PQ, IPQ and ITC-SOPI are all used approximately equally often. Note that all of these questionnaires came into use before 2002 and are still in use in 2014 (when the paper was published). In particular, the authors do not consider the usage of the MEC-SPQ.

2.2.1.2 Other self-report measures

Welch et al. reported the results of two studies where participants experienced a simulated driving scene (Welch et al., 1996). In these studies, presence was measured by means of paired comparisons—after every pair of exposures, the participant marked on a scale of 1 to 100 how different their senses of presence were between the most recent exposure and the previous one.

Snow and Williges use the technique of free-modulus magnitude estimation to measure presence in VEs (Snow and Williges, 1998). In free-modulus magnitude estimation, a participant is presented with a series of stimuli and asked to assign a numeric value representing their level of the desired quantity—in this case, presence—to each stimulus. There is no predetermined scale. The participant is instructed to assign any positive number to the first stimulus, and then score all successive stimuli relative to that first number.

Freeman et al. present a novel form of direct subjective presence evaluation and the results of three experimental studies using it (Freeman et al., 1999). They gave users a handheld slider that was continuously sampled during each trial. They instructed users to move the slider depending on how present they felt. However, rather than analyze these slider values as a continuous measure of presence, the mean of the slider value was computed for each trial for each participant, and these means were the values used in their analyses.

Techniques based on measuring breaks in presence, as introduced by Slater and Steed, are important variations on the self-report theme. Here, rather than reporting their level of felt presence, users report the moments when they do not feel present, and this series of events can be analyzed to generate a measure of presence. In the original paper (Slater and Steed, 2000), the breaks in presence were used to generate a Markov chain that continuously modeled the probability that a user felt present at any given time. Subsequent research also evaluated using raw counts of breaks in presence, rather than the more complex Markov chain analysis, and demonstrated that the overall count of breaks in presence is significantly negatively correlated with presence as measured by questionnaire (Brogni et al., 2003).

Breaks in presence have also been investigated in combination with physiological measures (Slater et al., 2003) (Slater et al., 2006), as well as with other types of self-report measures. Garau et al. induced breaks in presence in a virtual environment, then followed up with semi-structured interviews, the transcripts of which were subjected to content analysis (in which researchers define categories of interest before the experiment and then measure them quantitatively by looking for key words or phrases in the transcript) and thematic analysis (which looks for ideas that are not connected to the initial research questions). Participants also were asked to draw graphs corresponding to their sense of "being there", with time on the X-axis and the environment (lab or bar) on the Y-axis (Garau et al., 2008).

Kuschel and colleagues propose a new measure of presence based on perception of conflicting information across multiple sensory modalities (in their specific case, visual and haptic). In this measure, the user is presented with two or more streams of conflicting sensory data in different modalities, and is present in whichever one they perceive as dominant (Kuschel et al., 2007).

Riener and Proffitt propose a means of quantifying spatial presence by comparing the results of visual illusions (specifically the vertical-horizontal illusion (Figure 2.1) and the Ponzo illusion (Figure 2.2) in photographs, the real world, and in virtual environments. Here the measure is the users' estimated sizes of the lines in each illusion. They found that the size misestimations in virtual environments were closer to those in the real world for the vertical-horizontal illusion, while they were closer to photographs with strong perspective cues for the Ponzo illusion (Riener and Proffitt, 2002).

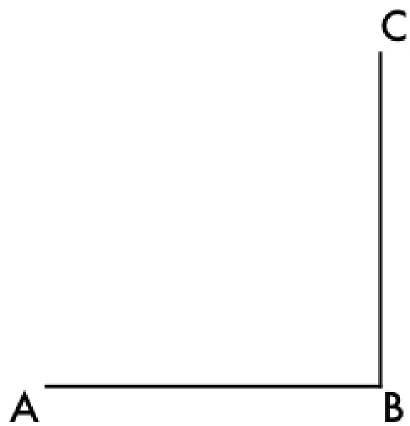

Figure 2.1: **Vertical horizontal illusion**: Even though line segments AB and BC are the same length, participants estimate BC to be longer than AB. (3-6% longer in photographs, 20-40% longer in the real world.)

2.2.2 Physiological metrics

Dillon et al. put forward a plan to compare skin conductance response (SCR) (also referred to as galvanic skin response, GSR, or electrodermal activity, EDA) and electrocardiogram (EKG) data with presence as measured by the ITC-SOPI (Lessiter et al., 2001) in a study where participants view a video stream presented either stereoscopically or monoscopically in (Dillon et al., 2000). The results of that study are summarized in (Dillon et al., 2002),

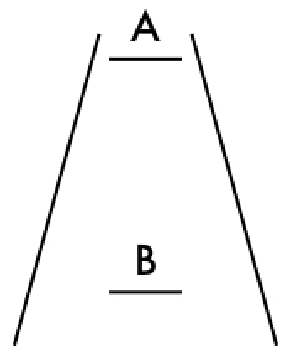

Figure 2.2: **Ponzo illusion**: Even though line segments A and B are the same length, participants estimate A to be longer than B.

and the results of another study investigating the effects of emotional media content and display size on presence and arousal are also presented there. The authors did not find a correlation between physiological metrics and presence. Note that the stimuli in these studies were neither immersive nor inherently arousing.

Meehan explored the same measures as well as skin temperature in an immersive virtual environment that did contain an inherently stressful stimulus, a visual cliff scenario (Meehan et al., 2002). There, Meehan and colleagues did find that a larger increase in heart rate when exposed to the visual cliff significantly correlated with an increase in presence as measured by the SUS questionnaire, indicating that physiological measures such as change in heart rate may be able to serve as an objective proxy measure for presence in such scenarios. (That is, between virtual environments that contain a known arousal-inducing stimulus, such as the visual cliff, and those that do not.)

As mentioned in the previous section, when talking about breaks in presence, Slater and colleagues have used physiological metrics to measure users' responses to breaks in presence in virtual environments, both experimenter-caused (Slater et al., 2006) and incidental (Slater et al., 2003).

2.2.3 Behavioral metrics

Sheridan proposed that in addition to self-report methods, presence could be measured by behavioral methods such as response to threatening stimuli (for example, flinching out of the way of a virtual ball) or to socially-conditioned behaviors (for example, saying, "Gesundheit," in response to a sneeze in the VE) (Sheridan, 1992).

Regenbrecht, Schubert, and Friedmann demonstrated that fear increased with higher presence in a virtual environment designed to elicit fear of heights (Regenbrecht et al., 1998). Presence was measured using Likert-style responses to 14 questions that included questions from (Slater et al., 1994) and (Hendrix and Barfield, 1995); anxiety was measured using the State-Trait Anxiety Index, a 20-item questionnaire (Laux et al., 1981). I classify this with the behavioral metrics rather than the self-report metrics because the self-reported quantity is not presence, as it is in all the other measures classified as self-report.

Freeman et al. proposed the use of behavioral metrics for presence; specifically the magnitude of postural response for seated participants viewing a video (Freeman et al., 2000). The authors conducted a 24 participant study to evaluate this metric. Participants viewed two video clips: one was excerpted from a video recorded from the hood of a rally car, and the other was a still frame from a video taken at the side of the rally track. The soundtrack was the same in both videos, giving the impression of a car off in the distance in the still video case. All participants saw both stimuli in both monoscopic and stereoscopic presentations. Participants rated their presence (as well as involvement, self-motion, and sickness) on a scale from 0 to 100 on a visual analogue scale after each trial, and their postural responses were tracked. Participants' self-rated presence scores showed significantly higher presence for stereoscopic presentation and for the moving video stimulus, with no interaction between the factors. There was no significant relationship between presence and postural response, however.

2.2.4 Analysis

The proliferation of questionnaires adds complexity to presence research. In my previous experience (see Chapter 5), I've used several presence questionnaires, and found their results to be extremely highly correlated. I believe that it is only necessary to use one questionnaire, although which to use may change based on the specifics of the experiment. For example, if one is particularly interested in reality judgment, or one or more of the subscores of the MEC questionnaire, one should certainly use the RJPQ or MEC-SPQ, respectively. For a general presence measure, I would recommend the SUS questionnaire, as it is both the shortest (saving both experimenter and participant time) and the questionnaire that most directly measures the feeling of presence.

I believe that behavioral metrics are a promising area of study that has so far been understudied. Physiological metrics may ultimately be limited in their utility as they have been shown to be useful only in experiments that are known to affect physiological signals in particular ways (e.g., increasing arousal in a stressful environment), but it is likely that appropriate behavioral signals could be found in nearly any virtual scenario.

2.3 Models of presence

In this section, I have grouped together those papers that posit a list of components asserted to contribute to presence. Some of these groupings are purely theoretical, and some were either the basis of questionnaires or factors derived from questionnaires in use. Therefore, some papers from the questionnaire section are repeated here. As well as presenting these theorized components of the presence construct here together, I also intend to demonstrate how these components can be usefully grouped. These models are presented in chronological order, as in Table 2.2. Note that there is no implied relationship among the data columns in Table 2.2; the components for each publication are listed across each row in no particular order. Figure2.3 contains an illustration of these same components, grouped by higher-order concept. Note also that Slater's Place Illusion and Plausibility Illusion are not included in Table 2.2, but they do appear as higher-order concepts in 2.3.

Table 2.2: List of presence models and their components

(Akin et al., 1983)	Ability to act in remote environment	Ability to sense in local environment			
(Heeter, 1992)	Subjective personal presence	Social presence	Environmental presence		
(Sheridan, 1992)	Extent of sensory information	Control of sensors	Ability to modify physical environment	Task dependent characteristics	
(Held and Durlach, 1992)	Sensory factors	Motor factors	Correlation between feedback and actions	Identification with the robot	Familiarity with the system
(Kim and Biocca, 1997)	Arrival	Departure			
(Draper et al., 1998)	Attention to mediated environment	Attention to ignoring distractors			
(Witmer and Singer, 1998)	Control	Sensory	Distraction	Realism	
(Baños et al., 2000)	Reality judgment	Presence	Emotional involvement	Interaction	Control
	Attention/Flow	Realism	Congruence/Continuity	Expectations	
(IJsselsteijn et al., 2000)	Extent and fidelity of sensory information	Match between sensors and display	Content factors	User characteristics	
(Lombard et al., 2000)	Social Richness	Realism (Social)	Realism (Perceptual)	Transportation	Immersion
	Social Actor in a Medium	Medium as Social Actor			
(Sas and O'Hare, 2003)	General cognitive factors	Task specific cognitive factors	Technological factors	Media content	
(Vorderer et al., 2004)	Attention allocation	Spatial situation model	Spatial presence: self-location	Spatial presence: possible actions	Higher cognitive involvement
	Suspension of disbelief	Domain-specific interest	Visual/Spatial imagery	Absorption	
(Chertoff et al., 2010)	Affective	Cognitive	Sensory	Active	Relational

2.3.1 Models

Akin and colleagues define telepresence as the condition that occurs when, "At the worksite, the manipulators have the dexterity to allow the operator to perform normal human functions. At the control station, the operator receives sufficient quantity and quality of sensory feedback to provide a feeling of actual presence at the worksite" (Akin et al., 1983). In other words, I would argue that the authors identify two factors of telepresence: ability to act in the remote environment and sensory fidelity delivered to the user. Note that this definition is specifically referring to telepresence and not presence or virtual presence, hence the references to worksite and control station.

Heeter proposes three dimensions of presence: subjective personal presence (feeling that you are in the virtual environment), social presence (feeling that other beings exist in the world and react to you), and environmental presence (feeling that the environment acknowledges and reacts to you) (Heeter, 1992).

Sheridan proposes three factors of presence: extent of sensory information, control of the relation of sensors to the environment, and ability to modify the physical environment (Sheridan, 1992). He also argues that presence is likely task-dependent, and that "fixed" characteristics of the system (immersion factors and task properties) should affect dependent measures of user experience, such as presence, training efficiency, task performance, and so on.

Held and Durlach speculate on the value of telepresence, as well as its potential causal factors in (Held and Durlach, 1992). They argue that telepresence is most desirable in applications where the tasks are wide-ranging, complex, and uncertain, "because the best general purpose system known to us . . . is us." They go on to speculate on the factors that contribute to telepresence, identifying sensory factors—resolution, field of view, consistency of information across modalities, and displays that are "free from production of artificial stimuli that signal the existence of the display", motor factors—support for movements of sensory organs and of viewed effectors, high correlation between kinesthetic feedback and sensed actions from the remote environment, identification with the robot (visual similarity), familiarity with the system, and "the cognitive representation of the operator's interaction with the world" as factors that are likely to contribute to greater telepresence.

Arrival and departure were identified as the two factors in the presence questionnaire created by Kim and Biocca (Kim and Biocca, 1997). *Arrival* is the feeling of being there in a mediated environment, *departure* is the feeling of *not* being in the real environment, and presence arises from the combination of the two.

Draper and his colleagues review existing conceptions of telepresence, and put forward an attentional resource model for telepresence in (Draper et al., 1998). This model argues that telepresence increases as a function of the sum of attentional resources devoted to processing task-related stimuli from the mediated environment and the attentional resources devoted to overcoming distractors.

Witmer and Singer proposed four major categories of factors that affect presence in the development of their presence questionnaire: Control, Sensory, Distraction, and Realism (Witmer and Singer, 1998). They also claim that factors may influence presence by acting on psychological immersion, involvement, or both. For example, they theorize that control factors impact psychological immersion but not involvement, while realism factors impact involvement but not psychological immersion. Distraction and sensory factors are theorized to affect both. Control factors include predictability, interactivity of the environment, and input controls; sensory factors include richness of the environment, number and fidelity of sensory modalities, and consistency of multimodal stimuli; distraction factors include isolation from the physical environment and interface awareness; and realism factors concern the degree to which the experience is meaningful and coherent with expectations from the real world. Each question on the PQ is intended to address some aspect of one of these factors. The results of a cluster analysis of four studies using the Witmer and Singer PQ identified three subscales in the PQ data—Involvement/Control, Naturalness, and Interface Quality.

Bystrom et al. propose the immersion, presence, performance (IPP) model for interaction in virtual environments. The authors adopt Slater's definition of immersion, and presence is used in the common sense of "being there". The IPP model, in brief, claims that sensory fidelity (resulting from a sufficiently immersive system) causes a user to allocate attentional resources to the VE, and that this allocation of attentional resources enables the user to experience presence in the VE and perform the given task. Furthermore, it claims that there is a feedback loop: more attention causes more presence and more task engagement, and increased task engagement causes the user to allocate more attentional resources (Bystrom et al., 1999).

Bystrom et al. state that this model is based on the two models of presence proposed by Slater and colleagues (Slater et al., 1996) (Slater and Wilbur, 1997) and by Barfield and colleagues (Hendrix and Barfield, 1996b) (Hendrix and Barfield, 1996a). The Slater model as outlined here describes presence as, "determined not only by ... aspects of displays ... but also mediated by the sorts of sensory information required to perform the task at hand ... and individual differences in preferences for information". The Barfield model describes presence as "dependent on the degree to which ... transformations of objects in a virtual environment are similar to ... transformations of objects in the real world."

Schubert et al. present the results of a factor analysis on an experiment in which 246 participants answered a 75-item survey of new questions and questions taken from Carlin et al. (Carlin et al., 1997), Ellis et al. (Ellis et al., 1997), Slater et al., (Slater et al., 1994), Towell and Towell (Towell and Towell, 1997), (Witmer and Singer, 1998), and Regenbrecht et al. (Regenbrecht et al., 1998). The authors extracted eight factors that combined express 50.27% of the total variance. These factors, in decreasing order of importance, were spatial presence, quality of immersion, involvement, drama, interface awareness, exploration of the VE, predictability and interaction, and realness. The authors then performed a second order factor analysis to see how the factors grouped together. In a two-factor solution, the first factor grouped spatial presence, quality of immersion, involvement, drama, and realness, and the second factor grouped interface awareness, exploration, and predictability and interaction. In a three-factor solution, the first factor grouped spatial presence, involvement, and realness, the second factor grouped interface awareness, predictability and interaction, and exploration, and the third factor grouped drama and quality of immersion (Schubert and Regenbrecht, 1999) (Schubert et al., 2001).

IJsselsteijn et al. review the existing presence literature to summarize research into the factors contributing to presence and the methods for measuring it. The authors identify four determinants of presence: (1) the extent and fidelity of sensory information, (2) the match between sensors and display, (3) content factors (a broad category covering most anything else that is part of the virtual scenario), and (4) user characteristics (IJsselsteijn et al., 2000).

Lombard et al. discuss their efforts to develop an instrument for presence based on their theoretical model of its components. In the literature, they identified six "dimensions" of presence: presence as social richness, presence as realism (both social and perceptual), presence as transportation, presence as immersion, presence as a social actor within a medium, and presence where the medium is a social actor. The authors claim that the common element among these types of presence is a perceptual illusion of nonmediation (Lombard et al., 2000).

Sas and O'Hare presented a "presence equation", where 45% of presence variation can be predicted as

$$0.37 \times Willingness\ to\ Suspend\ Disbelief + 0.29 \times Creative\ Imagination/Absorption. \quad (2.1)$$

They later presented a more general form of the presence equation,

$$\begin{aligned} Presence = a \times (General\ cognitive\ factors) + b \times (Task\ specific\ cognitive factors) \\ + c \times (Technological\ factors) + d \times (Media\ content). \quad (2.2) \end{aligned}$$

Note that "General cognitive factors" is the only term that is entirely dependent on the participant, whereas c and d are entirely dependent on the specific VE, and b is at least partially dependent on the specific VE. Furthermore, the authors' own previous discussion regarding immersive vs. non-immersive VEs seems to indicate that they believe $c > a, b$. They argue for the use of a non-immersive VE for their experiment because then presence differences will be due to human factors rather than immersion (Sas and O'Hare, 2003).

Witmer and colleagues revisit their presence questionnaire with a factor analysis in (Witmer et al., 2005). They identified four factors of their presence questionnaire, which combined account for 52.2% of the variance. These factors are Involvement (accounting for 31.9% of variance), Sensory Fidelity, Adaptation/Immersion, and Interface Quality.

Wirth and colleagues present a theoretical model by which spatial presence might be generated in a participant. They propose a two-stage model. In the first stage, one constructs a spatial situation model (SSM), i.e., a mental model of the spatial environment that one constructs based on (1) spatial cues that one processes and (2) relevant personal spatial memories and cognitions. In the second stage, one defines his or her primary egocentric reference frame (PERF), which is either the SSM representing the mediated environment, in which case one is present in the virtual environment, or the SSM representing the real world, in which case one is not present in the VE. Specifically, the authors claim that spatial presence occurs when the medium-as-PERF hypothesis is confirmed repeatedly through processed information and is thus stabilized over time (Wirth et al., 2007). It seems to me that the SSM formalizes Held and Durlach's "cognitive representation of the operator's interaction with the world" (Held and Durlach, 1992).

For Wirth et al., then, an individual's sense of presence in any mediated environment is dependent on both characteristics of the environment—e.g., richness, salience, consistency—and of the individual user—e.g., attention, involvement, suspension of disbelief.

2.3.2 Analysis

Unlike the definitions of presence in Section 2.1.1, the models of presence are strikingly similar. Almost all can be transformed into one another, or into, for example, the more recently developed PI/Psi framework. Akin et al.'s conception of telepresence being composed of the ability to act in the remote environment plus the ability to display sense data in the local environment is very similar to Slater's conception of immersion being composed of effective and sensorimotor valid actions (Akin et al., 1983). The authors do not consider coherence, but they have no need to, since they are explicitly talking about remote real environments as opposed to virtual ones. Heeter's subjective personal presence is precisely Place Illusion, while social presence and environmental presence are components of Plausibility Illusion (Heeter, 1992). Sheridan's factors contributing to telepresence are, again, sensorimotor and effective valid actions, plus the extent of sensor information, which is also an aspect of immersion (Sheridan, 1992). Witmer and Singer's conception of presence as arising from control factors, sensory factors, distraction factors, and realism factors can be restated as immersion (control and sensory) plus coherence (distraction and realism) (Witmer and Singer, 1998).

IJsselsteijn et al., Sas and O'Hare, and Wirth et al. introduce individual differences to the discussion (IJsselsteijn et al., 2000) (Sas and O'Hare, 2003) (Wirth et al., 2007). Specifically, Sas and O'Hare's presence equation

$$Presence = a \times (General\ cognitive\ factors) + b \times (Task\ specific\ cognitive factors) + c \times (Technological\ factors) + d \times (Media\ content) \quad (2.3)$$

consists of individual differences of state and trait, immersion, and coherence. It also consists of the respective coefficients on each of these terms, which might be better restated as

$$Presence = \mathbf{A}[Vector\ of\ cognitive\ factors] + \mathbf{B}[Vector\ of\ task\ specific\ cognitive\ factors] + \mathbf{C}[Vector\ of\ technological\ factors] + \mathbf{D}[Vector\ of\ media\ content\ factors] \quad (2.4)$$

to more accurately represent the difficulty involved in computing a "presence equation."

In Figure 2.3, I have grouped the presence components discussed in this section (and presented in Table 2.2). This grouping demonstrates that most of the components that have previously been proposed as making up the presence construct can in fact be grouped as components of PI, Psi, immersion, or coherence. Several others can be grouped under the heading of attention or distraction, and another subset can be grouped under individual differences. Taken together, these categories account for the overwhelming majority of components that have been proposed as part of the presence construct.

While immersion and coherence (and therefore PI and Psi) are largely under the control of the developer of the virtual environment, attention and individual differences are generally not. Many of these models of presence, then, take into consideration the impact of individual differences on presence, at least implicitly.

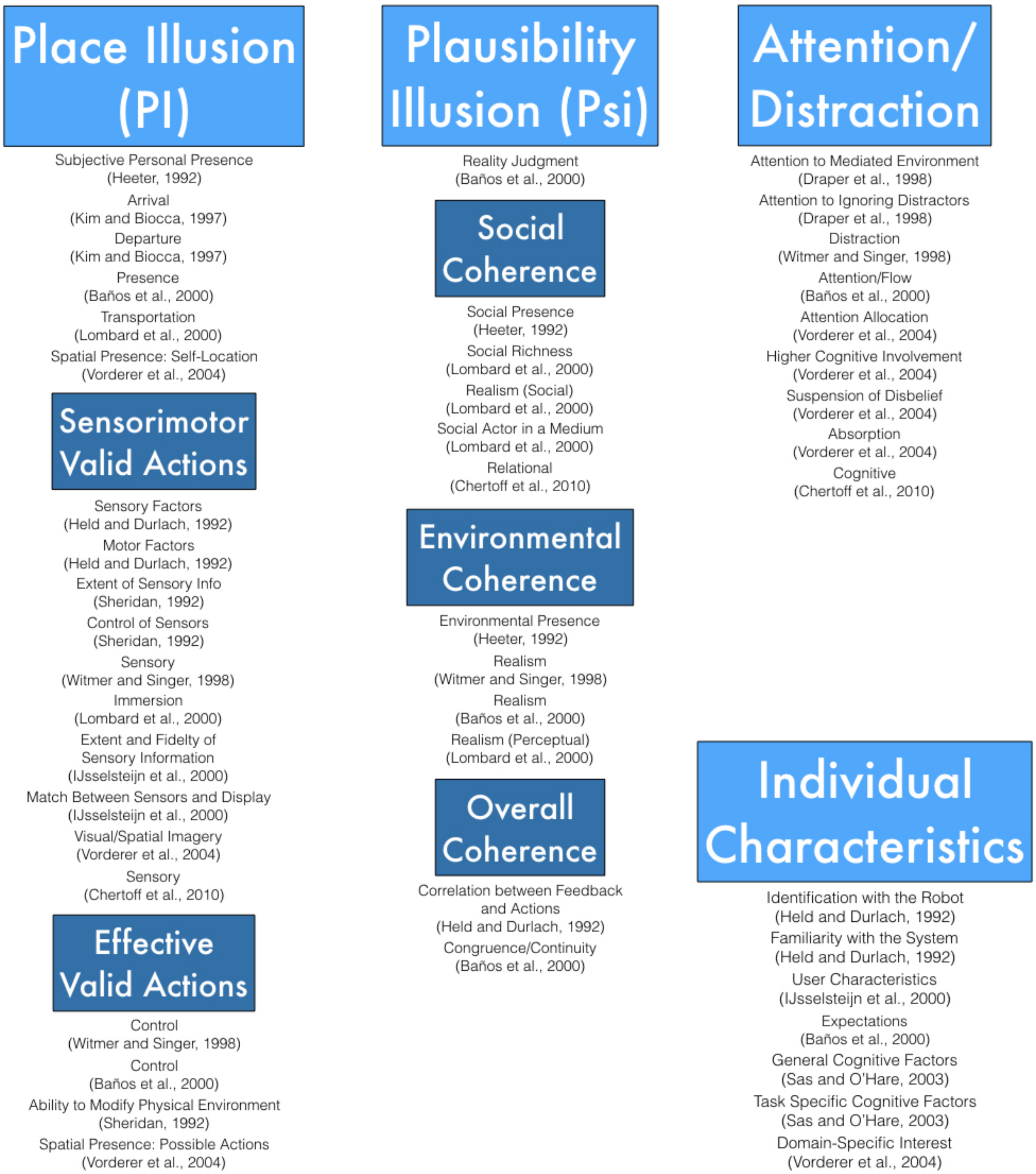

Figure 2.3: Clustering of presence model components

2.4 Factor Analyses

There have been three major factor analyses of presence and presence questionnaires in the literature, performed by Schubert et al., Lessiter et al., and Witmer et al.(Schubert and Regenbrecht, 1999) (Schubert et al., 2001) (Lessiter et al., 2001) (Witmer et al., 2005). Schubert et al. identified eight factors—spatial presence, quality of immersion, involvement, drama, interface awareness, exploration of the VE, predictability and interaction, and realness—that then grouped into three second-order factors—spatial presence, involvement, and realness. Lessiter et al. identified four factors—sense of physical space, engagement, naturalness, and negative effects. Witmer et al. identified four factors as well, which were involvement, sensory fidelity, adaptation/immersion, and interface quality.

An inherent limitation of factor analyses is that they can only group based on the items that were actually used in the measure. So if a questionnaire does not include questions about the coherence of social interactions, for example, there cannot be a factor that addresses that construct. On the other hand, if a questionnaire does ask questions about a construct that others do not—as with the ITC-SOPI and negative effects—that construct is likely to be represented by a factor. The initial selection of items, then, inherently biases the factor analysis that follows.

That said, it is enlightening to look at the similarities and differences among these factor lists. All three include a factor they call "involvement" or "engagement". (I will use involvement going forward.) However, a look at the questions that make up these factors reveals that they may actually represent different constructs. For Lessiter et al., this factor is represented by items such as, "I enjoyed myself", and "My experience was intense". These items seem to represent an overall affinity for the experience, rather than specifically relating to presence. For Witmer et al., the Involvement factor contains items including "How much were you able to control events?", "How much did the visual aspects of the environment involve you?", and "How much did your experiences in the virtual environment seem consistent with your real world experiences?", which don't on face seem to represent any one construct. On the other hand, for Schubert et al., Involvement is represented by items including "I concentrated only on the virtual space", and "I was completely captivated by the virtual world", which seems clearly to represent an attentional component. This discussion demonstrates that these factors are not as similar as one would assume from the names.

From Schubert et al., I classify the Spatial Presence factor as a sub-questionnaire asking directly about the feeling of spatial presence (as the SUS questionnaire asks directly about the feeling of presence), Involvement as an attentional component, Predictability and Interaction, Realness, and Drama as coherence factors, and Quality of Immersion, Interface Awareness, and Exploration as immersion factors.

From Lessiter et al. I classify the Sense of Physical Space factor as a sub-questionnaire asking directly about the feeling of spatial presence, Engagement as an affinity component, Naturalness as a coherence factor, and Negative Effects as a (reverse-coded) immersion factor.

From Witmer et al. I classify their Interface Quality and Sensory Fidelity factors as being immersion factors, and Involvement and Adaptation/Immersion to be primarily coherence factors.

2.5 Related Concepts

2.5.1 Transportation

Broadly speaking, the concept of *transportation* is to narrative worlds as the concept of presence is to technology-mediated worlds (Gerrig, 1993). In a state of transportation, "[T]he reader loses access to some real-world facts in favor of accepting the narrative world that the author has created . . . transported readers may experience strong emotions and motivations, even when they know the events in the story are not real" (Green and Brock, 2000).

Transportability refers to a person's inherent ability to be transported by a narrative. There is not an equivalent term in the field of presence research, although this idea was explored in the form of Witmer and Singer's Immersive Tendencies Questionnaire (Witmer and Singer, 1998). There are transportation and transportability questionnaires that are analogous to the PQ and ITQ, respectively (Green and Brock, 2000) (Green, 1996).

2.5.2 Agency

Agency (Russell, 1996) is "the sense that I am the one who is causing or generating an action," (Gallagher, 2000) or "the satisfying power to take meaningful action and see the results of our decisions and choices" (Murray, 1997). Agency has been identified as a concept that may share some features and factors with presence (Herrera et al., 2006). There is some speculation that the brain mechanisms that give rise to the sense of agency may be related to those that give rise to presence (David et al., 2008) (Seth et al., 2012).

2.5.3 Reality Judgment

Baos and colleagues argue for the existence of *reality judgment* as a construct separate from presence. They point out that one can attribute reality to something without feeling a sense of presence, for example, when watching a news broadcast; or vice versa, as when playing a fantasy video game. So reality judgment (the belief that our experiences are real) should be treated as related to, but distinct from, presence (Baños et al., 2000)(Baños et al., 2004).

2.5.4 Analysis

Too often, virtual reality researchers focus only on the research published in their specific sub-field. Virtual reality researchers are computer scientists, media theorists, and both clinical and experimental psychologists, to name only some of the more prominent disciplines. These researchers then tend to publish in the conferences and journals that they are most familiar with, which may not be frequently read by researchers in other disciplines.

Transportation, for example, is directly analogous to presence, and may become even more relevant as research explores narrative and coherence factors of virtual environments that have previously gone understudied. Agency, likewise, is a close cousin of presence, particularly if one accepts that presence is inherently connected to one's ability to act in virtual environments. Reality judgment represents the most direct effort any researchers have made to study the value of realism in virtual environments, at least prior to the introduction of Plausibility Illusion.

2.6 Conclusions

This chapter reviewed the existing presence literature in order to provide background and context for the development of theory regarding Place Illusion and Plausibility Illusion. I first reviewed many of the definitions of presence in the literature. As a synthesis of these definitions, in this dissertation, I define presence as the cognitive feeling of being in a place. Notably, this feeling does not arise only from immersion, but also from coherence and both state and trait characteristics of the individual user. Secondly, I reviewed existing methods for measuring presence, categorizing them as self-report (predominantly, but not only, post-experiment questionnaires), behavioral, and physiological. Thirdly, I reviewed models of presence in the literature, arguing that most of these models are remarkably similar to the Place Illusion and Plausibility Illusion model, albeit using different terms. Finally, I reviewed some related concepts, specifically identifying reality judgment as the closest cousin to Plausibility Illusion existing in the literature. In the next chapter, I will define terms and develop theory specifically relating to the concepts of Place Illusion and Plausibility Illusion.

CHAPTER 3

Fleshing out the Place Illusion and Plausibility Illusion Concepts

As the concepts of Place Illusion (PI) and Plausibility Illusion (Psi) have not yet been widely adopted by the community of VR researchers, there is a lack not only of experimental research, but also a lack of theoretical development. This chapter extends and clarifies theory of PI and Psi to guide and support experimental research described in Chapters 4, 5, and 6. I also identify some previous concepts and experimental results in VR research that are explained or explicated by the PI/Psi framework.

3.1 Definitions

In Chapter 2, immersion was discussed as an objective characteristic of a VE system (Slater, 1999). Here, to be more precise, I use Slater's definition from (Slater, 2009); namely, that the *immersion* of a VE system is given by the set of all valid actions supported by that VE system. A *valid action* is any action a VE user can take that results in a change to the state of the virtual environment (an *effective valid action*) or to her perception of the virtual environment (a *sensorimotor valid action*) (Slater, 2009). For instance, if the visible view rotates as a user rotates their head, that is a sensorimotor valid action; if a user can reach out with their real hand and push an object in the virtual environment, that is an effective valid action.

I define a *virtual scenario* (also *scenario*) to be specifically a virtual environment as well as the behavior of that environment and any characters or objects in it.

As a parallel to Slater's definition of *immersion*, I define the *coherence* of a virtual scenario as the set of reasonable circumstances that can be demonstrated by the scenario without introducing unreasonable circumstances. I define a *reasonable circumstance* as a state of affairs in a virtual scenario that is self-evident given prior knowledge. *Unreasonable circumstances* are the opposite, states of affairs that are inconsistent with prior knowledge. Note that an unreasonable circumstance may become reasonable if the user is properly prepared. (See Section 3.5 for more discussion of priming.) For example, Interrante's Seven League Boots would seem unreasonable if the user was unprepared, but would become reasonable if they had been informed about the technique in advance (Interrante et al., 2007).

3.2 Immersion and Place Illusion

In (Slater, 2009), Slater states, "Immersion provides the boundaries within which PI can occur." By this Slater means not that there is a direct causal relationship in which more immersion leads to more PI. Rather, the immersion of the VE system itself is only one part of the illusion; other components include the specific virtual environment, the task being performed in the environment, the specific actions taken by a user, and both long- and short-term (that is, both trait and state) characteristics of the user.

The potential influence of each of these factors—the VE itself, the task, and user characteristics—on PI can be illustrated with some examples. An environment with much high-frequency visual detail (representing a library with open books, for example) would benefit much more from an imaging and display system that had high optical resolution than an environment with little high-frequency visual detail (such as the view out of a cockpit window). In the former case, the high optical resolution might well enable users to experience more Place Illusion, or alternatively, might prevent their PI from being broken. In the latter case, though, the higher optical resolution display would not be likely to have an effect on PI. These statements may be true even though the immersion of both systems is the same.

Similarly, a free-form information gathering task would also likely benefit more from high resolution visual display than would a timed navigation task, where the user may be moving quickly and unable to tarry. On the other hand, a user who is not wearing his corrective lenses, or a user who is colorblind, would have different sets of available valid actions than a user with normal color vision. Thus, various characteristics of the environment and of the user can effectively reduce the set of valid actions—and therefore, the immersion—of the system, and this can affect the level of PI a user feels.

Immersion, then, is an objective characteristic of a VE system, defined in terms of the valid sensorimotor actions that the system can support. PI is a feeling, which arises to the extent that a user's attempted actions are supported by the system or, alternatively, is broken to the extent that a user's attempted actions are not supported by the system. And a user's attempted actions may be constrained by the environment, the task, or the individual characteristics of the user.

3.3 Coherence and Plausibility Illusion

Immersion as a concept, and in particular Slater's definition thereof, is well-established in the VE research community. So to say that PI arises to the extent that a user successfully probes the immersive characteristics of a system is sensible, and in fact studying the relationships between many immersive characteristics of VE systems and the presence they induce has been a frequent subject for VE research.

There does not, however, exist an equivalent concept for reasoning about the extent to which a VE is *well-behaved*, which is what is needed for reasoning about Psi. Where PI emerges when a user successfully probes the immersive characteristics of a system, Psi emerges when a user successfully probes the behavioral characteristics of a scenario. There is a need for a term that describes the degree to which the virtual scenario behaves in a reasonable and predictable way. I introduce the term *coherence* for this concept.

The coherence characteristics of a system include, but are not limited to: the extent to which the appearance and behavior of a user's virtual body matches his real body, the extent to which virtual humans are present in the scenario and act appropriately for the scenario, the extent to which objects in the scenario can be interacted with and the degree to which those interactions are correct or predictable, and the extent to which a scenario accurately depicts what it purports to depict. Granted, these characteristics are not as objective on face as the immersive characteristics of a system. Field of view can be measured and objectively verified and given a number, whereas the same is not true of the realism of virtual human behavior, for example. That said, it is necessary to attempt to operationalize coherence to talk about Psi, in the same way that immersion is needed in order to reason about PI. It is not possible as yet to measure the actual feelings inside the heads of users, so instead it is necessary to reason about the characteristics in the real (or virtual) world that give rise to those feelings.

3.4 Orthogonality of PI and Psi

Slater has stated that PI and Psi are orthogonal components of virtual experience. It is important to note that by *orthogonal*, he means "logically distinct", not "statistically independent" (Slater, 2004c). That said, can one experience one without the other?

3.4.1 Can a participant experience PI with no Psi?

From the definitions, this would mean that the user believes that he is in a place, but that what is apparently happening in that place is not really happening. From further discussions of PI and immersion, PI is made possible by immersion, which in turn comes from the valid actions the VE supports.

So, a VE that would enable users to experience PI with no Psi would have to have a high degree of immersion (for example, with real-walking locomotion and a head-tracked HMD), but events in that scenario would seem to be "not really happening". One could imagine this in a "surrealist" world, or one that seems to be as in a dream. In short, characters and objects in the environment would have to behave in unrealistic (and inconsistent, so as not to allow the user to "train" themselves on the logic of the VE) fashion. This could be achieved, I believe, by adding random components to behaviors in the VE. Consider an environment in which the laws of physics work randomly, and characters interact (or do not) in randomly selected languages, where some of the characters will approach you and start a conversation while others do not respond to any stimulus whatsoever. Here one might feel PI, because of the natural sensorimotor contingencies, but one would likely not feel Plausibility Illusion.

3.4.2 Can a participant experience Psi with no PI?

From the definitions, this would mean that the user believes that what is apparently happening really is happening, but not that he is in a place. Based solely on these statements, it seems quite obvious that it is possible to experience Psi with no PI. A text adventure, for example, could satisfy these requirements.

However, Slater goes on to argue that, "[A] key component of Psi is that events in the virtual environment over which you have no direct control refer directly to you." Is it possible for events to refer directly to you in a scenario where you are not receiving "first-person view" stimuli? I argue yes; in a text adventure, events still refer directly to the reader, which is the representation of the user in the environment, and in a simple video game (Super Mario Bros., for example) events refer to Mario, which is the user's avatar.

Slater also argues that theater, as an example, is a medium that is able to induce Psi. This would have at most limited PI (theater generally supports NO effective actions on the part of the viewer, although it has the ability to support some sensorimotor actions). Potential PI here would certainly be different here than in the above examples; whether more, less, or just different depends on how effective and sensorimotor actions contribute to the overall sensation of place.

3.5 PI, Psi, and priming in virtual environments

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 (in the traditional sense of the word) behavior. In the former case, this startling behavior would be perceived as a failure of coherence and would decrease the users feeling of Psi. But in the latter case, the very same behavior would be perceived as a confirmation of the reality of the scenario as presented, and would likely *increase* the user's feeling of Psi.

Because a user's feeling of Psi is so dependent on their expectations, I posit that appropriate priming can be used to increase (or decrease) the feeling of Psi in users. This notion (albeit regarding presence rather than Psi) is not new to the VE community. Nunez and Blake (Nunez and Blake, 2003) performed an experiment where participants were primed by reading a passage that was either relevant to the VE they were about to experience (e.g., a booklet about monastic history before experiencing a virtual monastery), or irrelevant to the VE (e.g., the same booklet before experiencing a virtual hospital room). Their results suggest that priming is not a primary driver of presence, but that it can act as a mediator variable—that is, it amplifies the effect of different levels of immersion on presence. Steinicke and colleagues (Steinicke et al., 2009) performed an experiment where participants were presented with a "transitional environment" (a virtual copy of the real lab) between the real VR lab and the virtual environment. This can also be considered as priming the user for the VE. They found that priming with the transitional environment produced more subjective presence in a virtual airplane scenario, but did not observe a priming related difference in physiological metrics.

I believe that appropriate priming can increase Psi, and therefore presence, by reducing or eliminating "wrong" (or unintentionally surprising) events in the VE which can cause breaks in Psi.

3.6 PI, Psi, and the "Book Problem"

One seemingly paradoxical effect observed by some presence researchers is that people can experience presence in very low immersion environments: in the limit, people have been observed to report presence from reading text. This phenomenon has been dubbed the "book problem" (Biocca, 2002). It is my belief that Psi can at least partially explain the book problem.

In my experiments (Chapter 5), I observed that subjective presence was affected by both immersion and coherence separately, and most strongly when they were both high simultaneously. So coherence, and therefore most probably Psi, has an effect on users self-reported presence. Anecdotally, coherence is generally much better in traditional media (books, films, etc.). These media have many built-in advantages in this respect, in that they normally contain a somewhat linear plot, and represent a singular point of view. Books can go even farther and offer written (but unspoken) justification for the actions of their characters. (A corollary, then, would be that non-traditional films and literature would elicit less presence than more traditional forms, but that is beyond the scope of this dissertation.)

This effect may be more clearly demonstrated in other very low immersion "virtual" environments. Anecdotally, a game such as *Super Mario Bros.* has very low immersion—the set of valid sensorimotor actions it supports consists only of the pressing of a half-dozen or so buttons. However, it and its descendants remain popular and engrossing. I speculate that this is due to the very clear and consistent rules of behavior in the game world. The controls are very precise and responsive, if not entirely "realistic." As one plays more, this matching of stimulus to response, of expectation to realization, leads one to become more present in the world it represents.

Another anecdote comes from my personal experience in improvised theater. Here, actors generally have few, if any, props on stage. So if they desire a table for a scene, they create it by force of imagination, by simply acting as if there is a table. With every action that indicates that the table is real (walking around it, setting their imaginary drinks down on it, scratching their characters' initials into it), it becomes more real to the audience. However, if one of the actors forgets the table is there and steps right through it, the illusion is broken.

There is also some evidence to support the idea that coherence can elicit the feeling of presence in the psychology literature. Seth, Suzuki, and Critchley (Seth et al., 2012) propose a neurocognitive model of presence that is based on the brain's ability to successfully "explain" discrepancies between sensorimotor predictions and observations. If the feeling of presence arises more generally from the matching of predictions and observations, it is quite clear how improved coherence should lead to a greater feeling of presence.

3.7 PI, Psi, and the "Uncanny Valley"

The concept of the "uncanny valley" has spread far beyond its original context of robotics. Mori speculated about the appearance of entities, stating that in general, as an entity becomes more humanlike in appearance, one experiences greater affinity for it In (Mori, 1970) (officially translated into English in (Mori et al., 2012)). However, when an entity is almost—but not quite—humanlike, one experiences a negative affinity to it. This is the "uncanny valley" of the title. He goes on to say that if the entity is moving, the effect is more pronounced. (We may find a corpse disconcerting, but would find a zombie even more so.)

However, Mori does not make any mention of the *quality* or *correctness* of the motion, nor does he mention the *context* in which it occurs. For example, he suggests that one would feel quite high affinity for a bunraku puppet. I suggest that this is true only to the extent that the behavior of the puppet conforms to expectations. Within the context of the theater, and controlled by a skillful puppeteer, this is very likely the case. However, if one were to see a bunraku puppet that appeared to be walking down the street unaided, I suspect it would be a quite different story. Similarly, if the puppeteer were very poor, one would also experience low affinity. I also suggest, perhaps counter-intuitively, that a puppet whose behavior was *too good* (that is, too humanlike), would also have lower affinity than one that behaved just right.

3.8 Extending the Uncanny Valley

The foregoing analysis also suggests a possible extension to the theory of the uncanny valley that would apply generally to virtual environments.

In Mori's discussion, it seems to me that the problem is not inherently the humanlike appearance of an entity, but rather the *mismatch* between its appearance and its behavior. Ishiguro has extended the concept of the uncanny valley with a "synergy effect" in this way (essentially, the match between appearance and behavior) (Ishiguro, 2007). 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. (The question posed by a philosophical zombie is beyond the scope of this dissertation.)

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, I suspect 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.

There exists some supportive experimental evidence. In the Chapter 5 experiment, I observed that matched conditions (low immersion with low coherence, and high immersion with high coherence) resulted in higher scores on the SUS questionnaire than mismatched conditions. 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), whereas in the mismatched conditions, the scenario is highly immersive but behaves badly or vice versa.

While these results use presence rather than affinity, this may not be an actual difference. In (Slater, 2004a), Slater presents the results of an experiment in which participants rated the "colorfulness" of their day, which was associated with being "good", "pleasant", but "not frustrating" in short, feelings of affinity. (His argument is specifically against reliance on presence questionnaires, which, because they are the only tool available, may be confounding affinity with presence. My research supports this notion.)

There is also evidence that users perform better when there is consistency within a virtual scenario. Paul Zimmons's dissertation, presents the results of a study investigating the effects of lighting models on accuracy, speed, and recall in an object recognition task. Participants were first shown a 3D knot object for ten seconds, and then were presented with a table on which were fifteen 3D knot objects. The participant was asked to either identify the knot object that had appeared in the training phase, or to indicate that the training object was not on the table. (Zimmons called these objects the *search object* and the *table objects*; I will call them the *training* and *test objects* here.) Each object (if training) or set of objects (if test) was presented in one of three lighting conditions: ambient, local, or global illumination. These conditions were varied across training and test pairs, so each pair represented a cell in a 3x3 design (Zimmons, 2004). (See Table 3.1.)

Based on the previous discussions about the benefits of consistency in virtual environments, I hypothesize that task performance would be greater in the conditions in which the training object and the test object share the same illumination condition.

Zimmons's study provides a nice test case for this hypothesis, since as a 3x3 study, there

Figure 3.1: A bunraku puppet and puppeteer (image from (Renee, 2002), used under Creative Commons license https://creativecommons.org/licenses/by-nc-nd/2.0/)

Table 3.1: The experimental conditions from Zimmons's dissertation

	Training object lighting model			
Test objects lighting model		Global	Local	Ambient
	Global	Consistent	1-off	2-off
	Local	1-off	Consistent	1-off
	Ambient	2-off	1-off	Consistent

are multiple clearly gradated levels of inconsistency between the training and the test cases. In this case we see, repeatedly, a behavior pattern consistent with this hypothesis.

When the metric is accuracy (Was the test object correctly identified?), participants were correct significantly more often in consistent than inconsistent conditions. (76% to 63%, $p < .001$) Even more tellingly, participants were correct significantly more often in consistent conditions than inconsistent conditions in which the lighting models were one-off [see chart above] (76% to 69%, $p < 0.05$), and participants in one-off conditions were correct significantly more often than participants in two-off conditions (69% to 52%, $p < .001$)

When the metric was search time, a similar ordering was observed; participants in two-off conditions were significantly slower (average search time of 7.7 seconds) than participants in one-off (6.7, $p < .01$) or consistent (6.8, $p < .05$) conditions.

These orderings are consistent with what would my hypothesis predicts: as the training and test conditions became more inconsistent (and therefore the mental model developed in training was a worse predictor of the stimuli received in the test state), performance got both slower and less accurate.

3.9 Summary of Theoretical Results

This chapter has presented new terminology (*coherence* and *reasonable circumstances*, as they relate to Plausibility Illusion, as well as *virtual scenario*), and related PI and Psi to other concepts and problems in virtual reality research. I argued that appropriate priming can increase Plausibility Illusion, and therefore presence, and that the "book problem" is caused by the coherence of events in the text giving rise to Plausibility Illusion. I also presented an updated model of the "uncanny valley" taking into account not only appearance but behavior, as in Ishiguro (Ishiguro, 2007), and then extended it to virtual environments generally by arguing that immersion and coherence should be "matched" in quality, citing some previous experimental results supporting this hypothesis.

CHAPTER 4

Coherence of virtual human interactions

In this chapter, I discuss the design and results of an experiment investigating Plausibility Illusion in virtual human interactions, in particular, the coherence of conversation with a virtual human. This experiment was performed in combination with another experiment evaluating two different display technologies. As that aspect of the study is not relevant to this thesis, it will be mentioned only in the Materials section.

This was my first attempt to manipulate coherence in an experimental context, although I hadn't then defined coherence. This experiment is relevant to the dissertation not only for that reason, but also for the lessons learned by attempting to manipulate coherence. Most notably, the role of individual differences in users' perception of coherence/feeling of Plausibility Illusion is quite clear in this experiment. First, the design of the experiment had to be completely changed to account for the fact that the voice recognition system produced wildly varying levels of coherence based on the different voices and accents of users. Second, the fact that users were highly motivated to use the system may have prevented me from detecting a difference between the high-coherence and low-coherence groups on task metrics. That said, there is some evidence that participants responded behaviorally to the difference in coherence, even though it was not apparent in task completion or the majority of questionnaire responses.

The virtual scenario used for this experiment as well as significant technical support were provided by the Virtual Experiences Research Group (VERG) at the University of Florida, led by Ben Lok. The implementation of the display portion of the experiment was done by the Avatar research group at the University of North Carolina, led by Greg Welch and Henry Fuchs.

4.1 Experiment

The experiment was a between-subjects design. Each participant performed one interview with a virtual human patient who had come to a medical facility complaining of stomach pain. The interviews lasted approximately ten minutes. For all participants, virtual human responses were generated by an experimenter in a Wizard-of-Oz (WoZ) setup. Totally freeform responses were not possible; the experimenter selected responses from a searchable list of responses that had previously been recorded by a voice actor. The participant was not aware that responses were being chosen by a real person; it appeared that the responses were generated by voice recognition.

There were two experimental conditions. Specifically, the WoZ followed two different behavior patterns. In the high-coherence condition, the experimenter responded to the participant as quickly and as accurately as was possible. In the low-coherence condition, the experimenter responded according to a script with a variety of conversational errors. The different types of conversational errors were derived from (Skarbez et al., 2011). Since the exact conversation could not be predicted in advance, the error script was of the form, "On the fourth exchange, ignore the participant. On the ninth exchange, repeat the answer twice in a row," and so on. This ensured that all participants experienced a variety of types of errors, and at a predictable frequency.

I had initially intended for the low-coherence condition to have responses selected by the voice recognition software, and for the high-coherence condition to have responses selected by the experimenter in a WoZ setup. Piloting, however, revealed problems with this experimental setup, owing to the nature of the voice recognition software. Using it, some participants' voices were almost perfectly recognized by the system, leading to very few errors, while some other participants were almost unable to get the virtual patient to respond at all. In short, there was no way to have a standardized amount of unreasonable circumstances occur in the low-coherence condition using the voice recognition software.

Participants

Thirty-two medical school students (18 female, 14 male), with an average age of 25.8 2.3

years, were recruited from the university medical school. They were compensated for their participation.

Materials

The virtual human models, scripts, and voice recordings were provided by the Virtual Experiences Research Group (VERG) at the University of Florida. Participants' eye and head positions were tracked during the experiment using an Optitrack optical tracking system. Depending on the display condition, the virtual patient was rendered on a large 3DTV or was embodied in a physical-virtual avatar (PVA), as shown in Figure 4.1. The PVA was initially developed for use in another experiment (Rivera-Gutierrez et al., 2012). Both displays were present in the room for all participants; whichever was not in use was covered with a black cloth. The displays were placed so that the virtual patient in both display conditions would subtend approximately the same visual angle from the participant's seated position, which was the same for all participants. This arrangement can be seen in Figure 4.2.

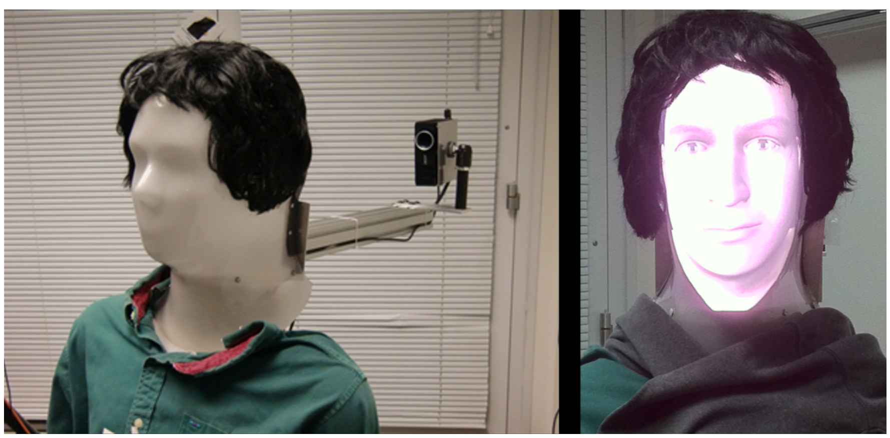

Figure 4.1: At left, the physical-virtual avatar (PVA) apparatus. At right, the participant's view of the PVA when in use. The face is animated, computer-generated imagery projected onto the inside of the plastic face. The projector can be seen in the center of Figure 4.1, Left.

Metrics

Participants filled out both pre- and post-experiment questionnaires, including demographic and medical training information, the Maastricht assessment of the simulated patient (MaSP), a series of questions asking the participant to compare the VH interaction to other types of interactions they may have had in everyday life, and a series of questions intended

Figure 4.2: User performing an interview with the PVA. At left of image is the flat-panel stereo display used in the other display condition

to measure the plausibility of the VH.

The MaSP is an instrument initially designed to assess the quality and authenticity of real simulated patients, i.e., human actors who portray a patient with a given condition (Wind et al., 2004). The MaSP has been modified and used to evaluate the quality of virtual human simulated patients as well, as in (Johnsen et al., 2007) and (Raij et al., 2007).

We also recorded the positions of the eyes and torso for each participant during the interaction, so that we could measure large-scale gaze behavior and postural response to the virtual patient. The eye position was tracked using glasses with retroreflective markers attached, which patients wore regardless of display condition. Note that we did not record gaze behavior or employ eye tracking; the tracked point was roughly the center of the two eyes, a point on the bridge of the nose.

4.2 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. Non-informational priors (normal functions with high variance) have been chosen so as not to bias the results. For more details on this method, please consult Appendix A.

Unlike traditional 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. (Bergström et al., 2016).

For the majority of measures, there is negligible evidence to support an effect of the of the VH response coherence, of the display technology used, or of any covariants (age, semesters in medical school, number of patient interviews, number of standardized patient interviews, or self-rankings of anxiety, comfort, preparedness, or skill). Notably, the MaSP scores do not reveal a difference between the low-coherence and high-coherence conditions. However, there are a few results for which there is some evidence.

The posterior probability that participants in the low-coherence condition moved their heads more than participants in the high coherence condition, as measured by the standard deviation of head position, is 78.6%.

The posterior probability that participants in the low-coherence condition reported lower scores to the question, "How strongly did you sense that the patient was watching you?" than participants in the high-coherence condition is 77.2%.

4.3 Discussion

The results described above indicate that participants detected the experimental manipulation of coherence, at least at some level. Anecdotally, the experimenters noticed that participants in the low-coherence condition were more fidgety, and looked around the room (and not at the virtual patient) more than participants in the high-coherence condition. This observation is supported by the eye-position tracking data, in which the standard deviation of eye position is very likely to be smaller for high-coherence participants than low-coherence participants. Furthermore, the question, "How strongly did you sense that the patient was watching you?" is the only question that directly asked about the behavioral response of the virtual patient. In other words, it was the closest question available to, "Was the virtual patient paying attention?" I speculate that participants used this question as a means to say what they really noticed, which is that the virtual patient seemed less responsive to their statements in the low-coherence condition. Neither of these observations is definitive on its own, but in combination, I believe that this indicates that the lower coherence affected participants, it just did not have an effect on task completion or on the other post-test measures.

This result may have occurred because this group was highly motivated. Of the 32 participants, 30 stated in post-experiment interviews that, regardless of display or coherence condition, they would use the technology if it were available. The overwhelming feeling of participants was that they thought interviewing with virtual simulated patients was useful and they were excited about the potential of the technology. This feeling likely overcame any difficulties or concerns about the implementation. This is good news for the prospect of virtual human simulated patients in general, however, it may make the patient interview an unsuitable use case for differential evaluation of technology.

More generally, this result provides additional evidence that user motivation can supplement the technology, such that a more motivated, invested, or attentive user may feel presence or demonstrate realistic response in a situation where a less motivated user might not (Wirth et al., 2007). This is a boon for developers, because a user who is convinced that new technology can help them do real work is likely to devote more attentional resources, and this in turn will generate more presence, more realistic response, and more motivation in a virtuous circle. It is a challenge for researchers, though, as they face the problem of high user motivation obscuring experimental effects that might be more apparent with naïve users.

The difficulty of measuring a difference in Plausibility Illusion affected the design of later studies, particularly the psychophysical study of coherence factors presented in Chapter 6. The Markov Chain analyses used in that experiment ensure that usable data can be obtained even from users who may "saturate" post-experiment self-report measures.

CHAPTER 5

Immersion and coherence in the Pit

In this chapter, I 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, the Pit. 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, I identified characteristics of a virtual reality experience that we expected to influence one or the other of PI and Psi. I 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 and high or low levels of Psi-eliciting characteristics. I 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.

Because this chapter was originally written as a stand-alone article, it contains some text that repeats ideas presented previously in this dissertation.

5.1 PI:Immersion::Psi:Coherence

In (Slater, 2009), Slater states, "Immersion provides the boundaries within which PI can occur." 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 (Slater and Wilbur, 1997). 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. 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.

5.2 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 (Khanna et al., 2006) 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 (Meehan, 2001), 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 (Hendrix and Barfield, 1996a) (Arthur, 2000). 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 (Slater et al., 2010). 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 (Slater, 2009), 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.

5.2.1 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.

5.2.2 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.

5.2.3 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 (Witmer and Singer, 1998), the Slater-Usoh-Steed Presence Score (Usoh et al., 2000), the Virtual Experience Tool of Chertoff, Goldiez, and LaViola (Chertoff et al., 2010), the Arrival/Departure questionnaire of Kim and Biocca (Kim and Biocca, 1997), as well as a short experimental questionnaire intended to measure participants' levels of Psi, whose questions are listed in Table 5.1. Note that the Psi questionnaire is ad hoc and was constructed for this experiment. At this point, I can only argue for its face validity for responding to Psi. All participants were also debriefed by an experimenter.

Table 5.1: 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)

5.2.4 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 5.1 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.

5.2.5 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.

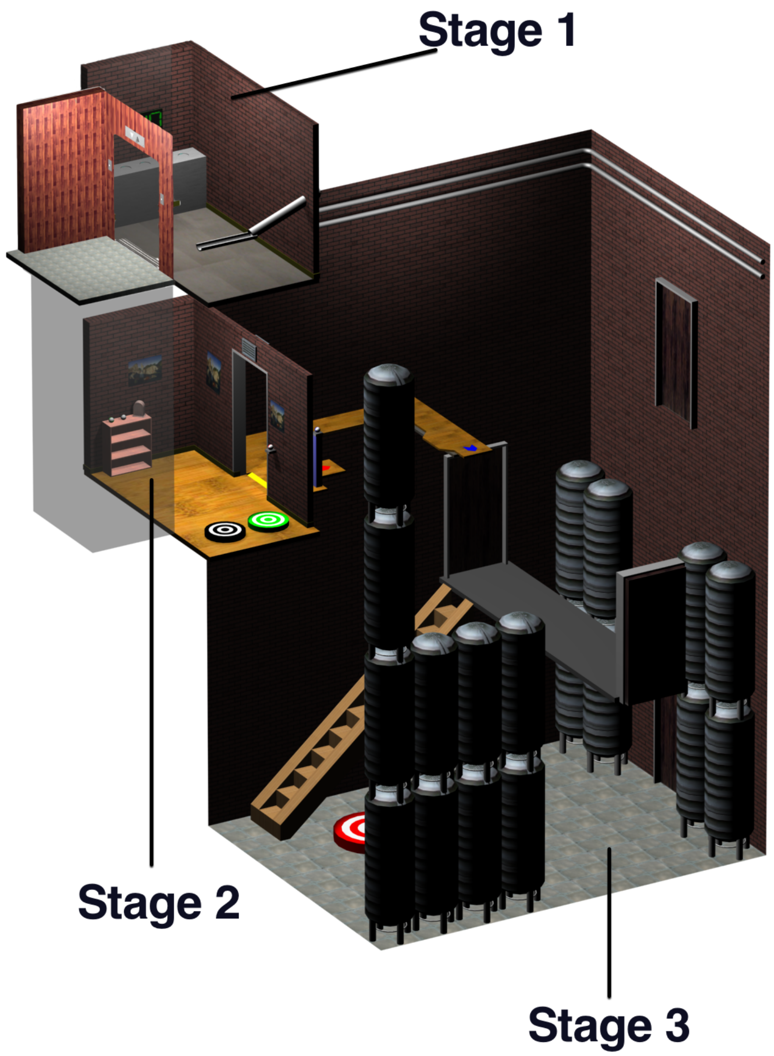

Figure 5.1: The virtual environment for Experiment 1.

5.3 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 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.

5.3.1 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.

5.3.2 Materials

The materials are the same as in Experiment 1, except that the virtual environment has an additional room, as described in 5.3.4.

5.3.3 Metrics

During the experiment, participants physiological responses were monitored as in Experiment 1. Post-test, participants completed the Witmer-Singer Presence Questionnaire (Witmer and Singer, 1998) and a modified Slater-Usoh-Steed Presence Score (Usoh et al., 2000).

5.3.4 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 5.2.):

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

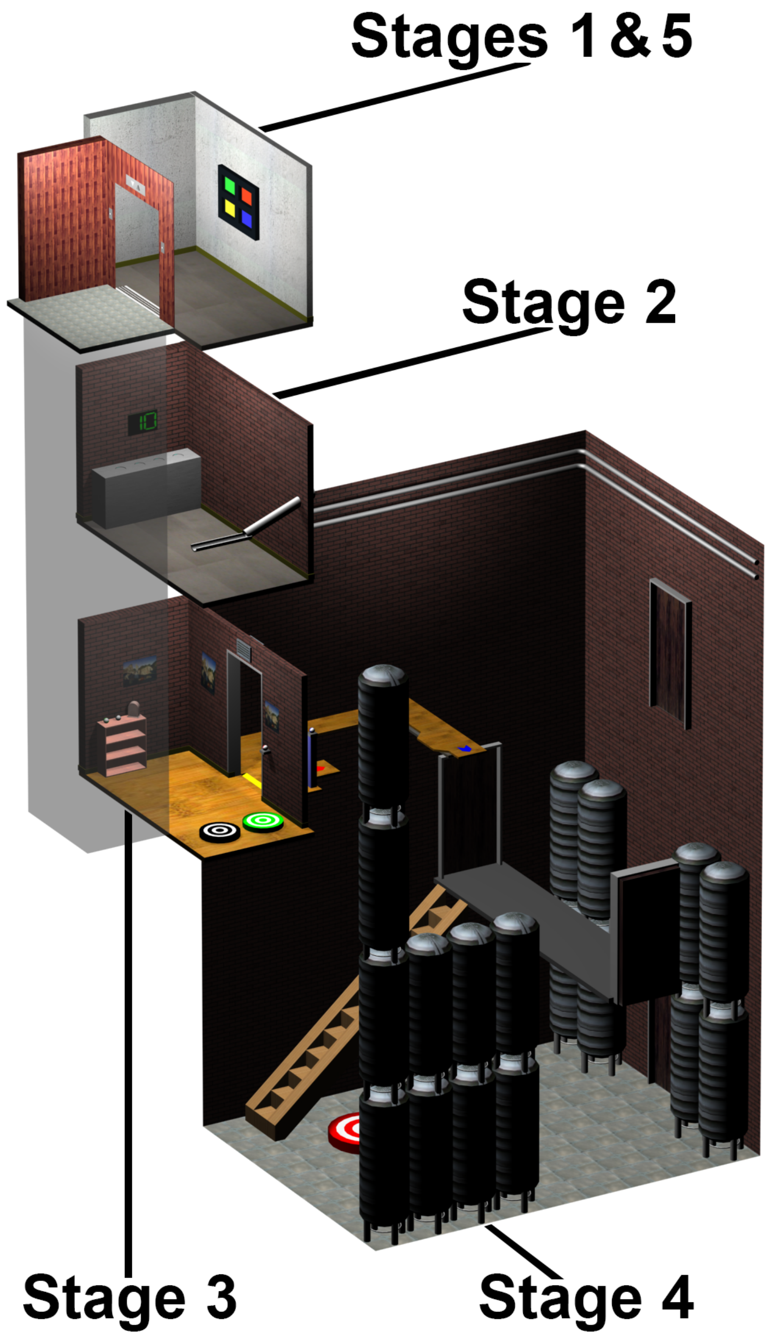

Figure 5.2: The virtual environment for Experiment 2.

Experiment 1.

5.4 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. For justification and more details of the method, please consult Appendix A. 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, I simply report the posterior probabilities.

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

5.4.1 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 5.2: 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

Table 5.3: 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

5.4.2 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%.

5.4.3 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.

5.4.4 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 5.2 and 5.3 for mean scores.)

5.4.5 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.

5.4.6 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).

5.4.7 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 5.3.)

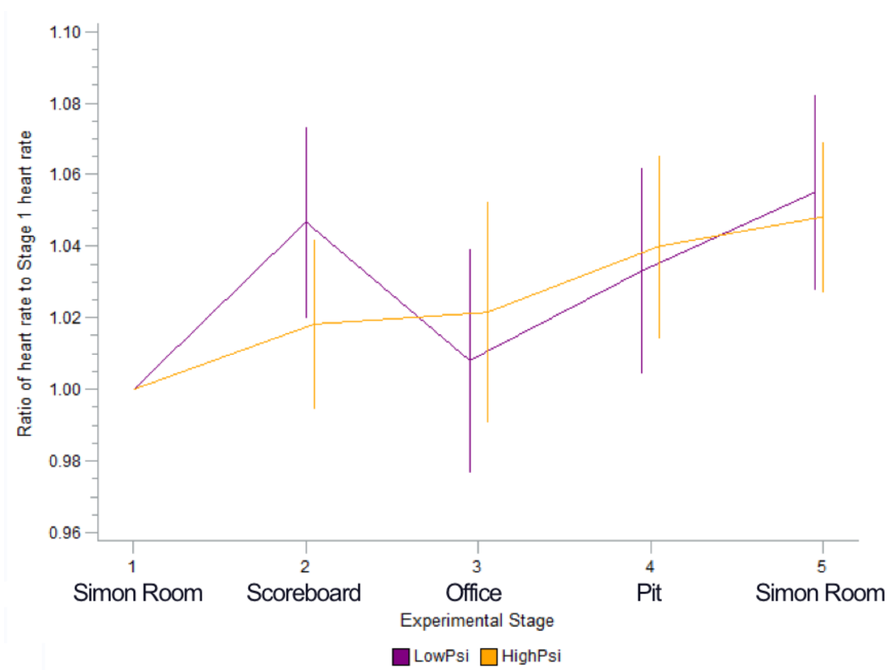

Figure 5.3: Comparing coherence conditions by heart rate in each experimental stage.

5.4.8 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%).

5.4.9 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 higher immersion and lower coherence together (22.1%).

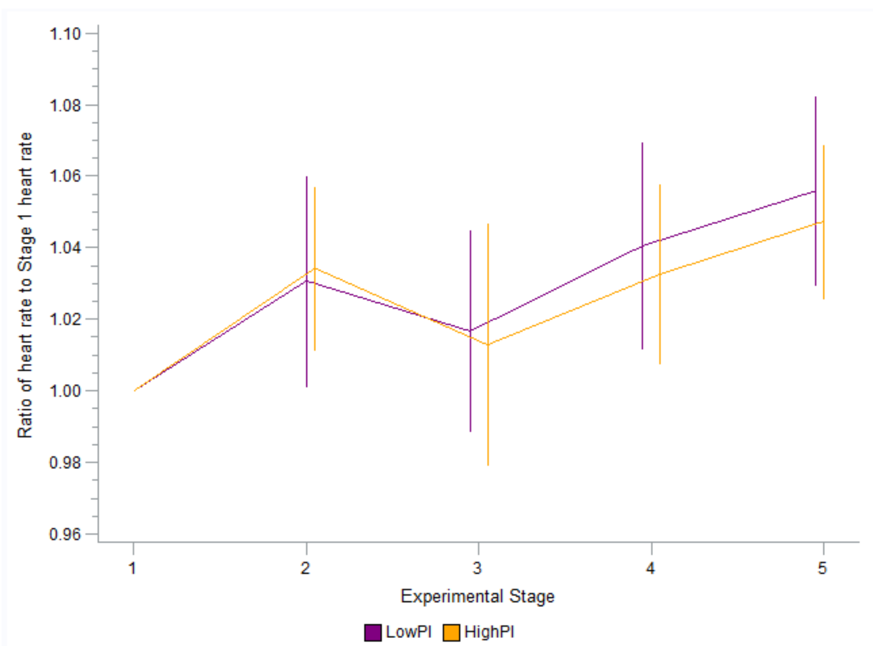

Figure 5.4: Comparing immersion conditions by heart rate in each experimental stage.

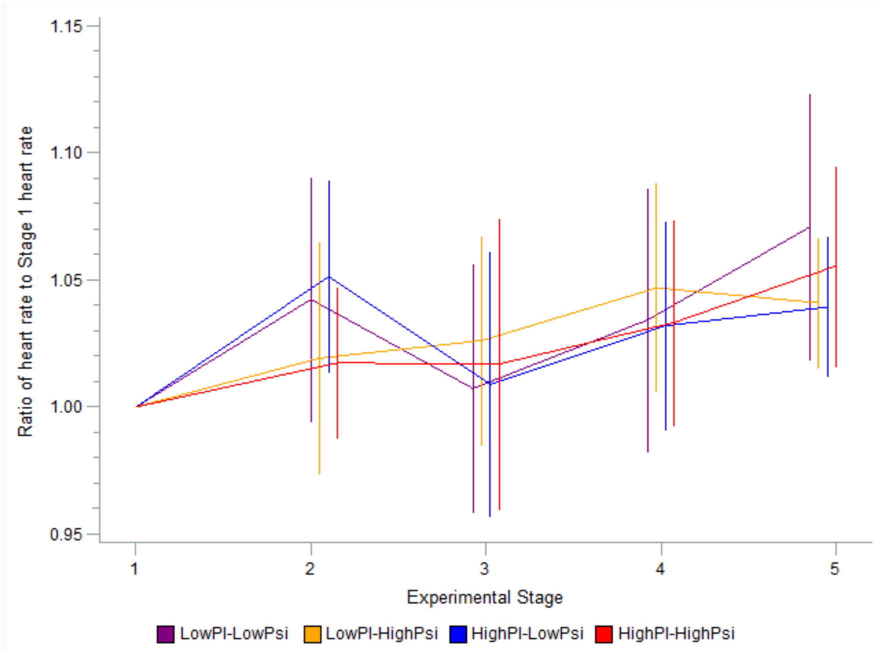

Figure 5.5: Comparing all four conditions by heart rate in each experimental stage.

5.5 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 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. (Lessiter et al., 2001). 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 (Schubert and Regenbrecht, 1999), 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 users 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 doing well. 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 effects lend credence to the "uncanny valley" effect proposed in Section 3.8

Figure 5.6 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 (Meehan, 2001). 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.

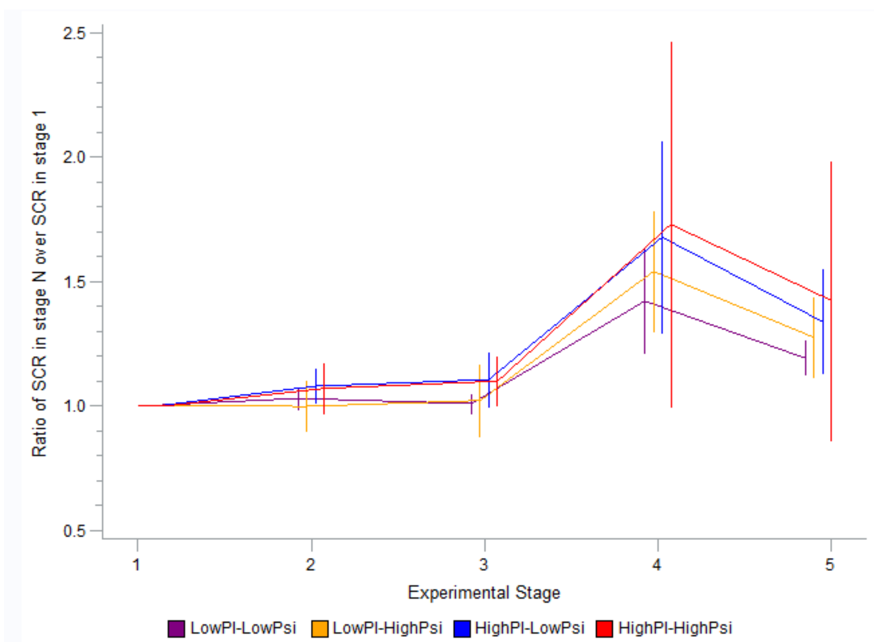

Figure 5.6: Comparing all four conditions by skin conductance response in each experimental stage.

Coherence was also considered by Biocca and colleagues in (Biocca et al., 2001), although not under that name. In the experiment described in that paper, participants performed a similar task (removing objects from a cadaver) in an ecologically-valid environment where the objects were organs, and in a control environment where the objects were geometric primitives. This scenario-appropriateness is an aspect of coherence that was not explored in this paper, but merits consideration in future work.

5.6 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.

As with the study presented in Chapter 4, the difficulty of measuring a difference in Plausibility Illusion informed the design of the psychophysical study of coherence factors presented in Chapter 6.

CHAPTER 6

Factors of coherence and Psi

In this chapter, I report on the design and results of an experiment investigating factors influencing Slater's Plausibility Illusion (Psi) in virtual environments. Slater proposed Psi and Place Illusion (PI) as orthogonal components of virtual experience which contribute to realistic response in a VE. PI corresponds to the traditional conception of presence as "being there," so there exists a substantial body of previous research relating to PI, but very little relating to Psi. I developed this experiment to investigate the components of plausibility illusion using subjective matching techniques similar to those used in color science, and previously used in (Slater et al., 2010). Twenty-one participants each experienced a scenario with the highest level of coherence (the extent to which a scenario matches user expectations and is internally consistent), then in eight different trials chose transitions from lower-coherence to higher-coherence scenarios with the goal of matching the level of Psi they felt in the highest-coherence scenario. At each transition, participants could change one of the following coherence characteristics: the behavior of the other virtual humans in the environment, the behavior of their own body, the physical behavior of objects, or the appearance of the environment. Participants tended to choose improvements to the virtual body before any other improvements. This indicates that having an accurate and well-behaved representation of oneself in the virtual environment is the most important contributing factor to Psi.

This study is the first to focus specifically on coherence factors in virtual environments.

This experiment was conducted at, and with significant support from, the Experimental Virtual Environments (EVENT) Lab at the Universitat de Barcelona, led by Mel Slater.

6.1 Experiment

This experiment is similar in methodology to the experiment described by Slater et al. in (Slater et al., 2010). In that experiment, participants were placed in the system configuration with the highest level of immersion, instructed to remember either their feeling of PI or their feeling of Psi, and then match whichever feeling they were instructed to remember by choosing transitions from lower- to higher-order systems. In this experiment, I followed the same method, but was only concerned with Psi, so no participants were instructed to remember their feelings of PI.

In this experiment, the property vector was C = {VH, VB, P, S}, where VH refers to the coherence of virtual human behavior, VB the behavior of one's own virtual body, P the coherence of physical interactions in the VE, and S the scenario coherence. These are elucidated further below. We refer to each instance of the property vector as a *configuration*.

We chose the factors in the property vector in order to have a reasonable covering of the different types of coherence (and coherence failures) that can be present in a virtual environment. *Virtual humans* were chosen as one of the factors to represent the coherence of interaction with other characters in the virtual environment. The user's *virtual body* was chosen as another factor. In the real world, there is a justifiably strong feeling of agency—the sense that I am the one who is causing or generating an action (Russell, 1996)—especially when it comes to the behavior of one's own body. Therefore, we would expect the presence or absence of the virtual body, and the coherence of its behavior if present, to have a strong effect on one's level of plausibility illusion. *Physical interactions* (kicking a football) were chosen to represent the coherence of allowed interactions with the virtual environment. Finally, *scenario coherence* (the matching of the virtual environment to the represented situation) was chosen to represent any other factors, outside of specific interactions with characters or objects in the virtual environment, that may lead one to disbelieve the virtual environment as a whole. As an example, absent priming, participants are likely to expect the virtual environment to behave according to the rules of the real world. Those expectations can be violated in subtle ways by behavior that is technically valid, but feels "wrong." For example, consider a scenario set in the desert at mid-day, where all behavior is technically perfect, except the other virtual characters are wearing winter coats.

VH (Virtual human behavior coherence)

In all trials, there are three virtual humans in the environment with the participant: a bartender, and two young men having a conversation. After approximately 30 seconds of conversation, one of the men excuses himself to go to the bathroom, requiring him to cross in front of the participant. The specific behavior of these virtual humans depends on the value of VH.

(VH=0) Worst behavior. All virtual humans have only idle animations while talking, and remain in the same place. (That is, neither virtual patron crosses to the bathroom.)

(VH=1) Medium behavior. Virtual humans have realistic conversation and walking animations. When crossing in front of the participant, the crossing VH does not stop or acknowledge the participant.

(VH=2) Best behavior. Virtual humans have realistic conversation and walking animations. When crossing in front of the participant, the crossing VH stops, looks at the participant, and addresses the participant directly about their football playing.

VB (Virtual body behavior coherence)

The appearance and behavior of the participant's avatar could be changed. The different possible levels of the participant's avatar are described below.

(VB=0) Feet-only avatar. In this condition, the participant is represented in the environment only by their feet. (See Figure 6.2.) This condition, rather than having no visible representation in the environment, was chosen to enable participants to meaningfully perform the task of interacting with the football. The feet are fully tracked as, as they are in the other VB levels.

(VB=1) Static avatar. In this condition, the participant is represented in the environment by a gender-appropriate avatar in a seated T-pose. (See Figure 6.3.) The avatar's legs move with the participant's; the torso and arms, however, do not move.

Figure 6.1: Virtual humans in level 2. Note eye contact.

(VB=2) Fully-tracked avatar. In this condition, the participant is represented in the environment by a fully body-tracked gender-appropriate avatar. The avatar pose is driven by real-time input from the optical tracking system, as described below.

P (Physical coherence)

In all trials, participants were directed to play with and control a football between their feet. The behavior of the ball when kicked was determined by the value of P.

(P=0) Null behavior. When the ball was kicked, the force vector applied to the ball was cancelled out by an opposite force vector. In practice, this meant that the ball could be moved while it was in contact with the foot, but it would never roll or maintain momentum once out of contact with the foot.

(P=1) Semi-normal behavior. When the ball was kicked, it would randomly either behave as if it were in level P=0 or level P=2, with equal likelihood. In practice, this meant that the ball would behave normally 50% of the time, and not move 50% of the time.

(P=2) Normal behavior. When the ball was kicked, the physics engine was used to determine

Figure 6.2: Virtual body in level 0 (Only feet)

Figure 6.3: Virtual body in level 1. Torso fixed in T-pose visible in the mirror.

the path of the ball.

S (Scenario coherence) For all trials, the participant was in a virtual bar environment of the same physical configuration (tables and chairs in the same positions, mirror hanging on the wall facing the participant, etc.), but the representation of those objects changed depending on the value of S.

(S=0) Abstract appearance. All models in the environment are replaced with simple geometric primitives. (See Figure 6.4.)

(S=1) Mismatched appearance. The environment model is of an upscale restaurant. (See Figure 6.5.)

(S=2) Matched appearance. The environment model is of a bar. (See Figure 6.6.)

Figure 6.4: Abstract environment (Scenario level 0)

Altogether, there are 81 possible configurations: 3 physical coherence x 3 scenario coherence x 3 virtual human coherence x 3 virtual body coherence.

Figure 6.5: Mismatched environment, appearing to be an upscale restaurant (Scenario level 1)

Figure 6.6: Matched environment, appearing to be a bar (Scenario level 2)

6.1.1 Participants

Twenty-one participants (10 males, 11 females) were recruited from the local university campus. Their average age was 24 ± 5 (S.D.) years and they were compensated for their time.

6.1.2 Materials

The virtual environment was displayed using an Oculus Rift Development Kit 2 (DK2) head-mounted display (HMD) made by Oculus (Figure 7). The DK2 has a nominal 100° field of view, and a resolution of 960 x 1080 pixels per eye. It weighs 440 grams.

Figure 6.7: The Oculus Rift DK2 HMD

For head tracking, the internal tracking of the DK2 was used, with an update rate of 1000Hz. For body tracking, participants wore an Optitrack body suit (Figure 6.8), designed to support real-time whole body tracking of a person. It consists of a black suit with 37 retroreflective markers, which are tracked by 12 infrared cameras. Tracking was handled by the Optitrack Motive software platform.

The experiment was implemented in version 5.2 of the Unity Game Engine. The male and female avatars were created using Mixamo.

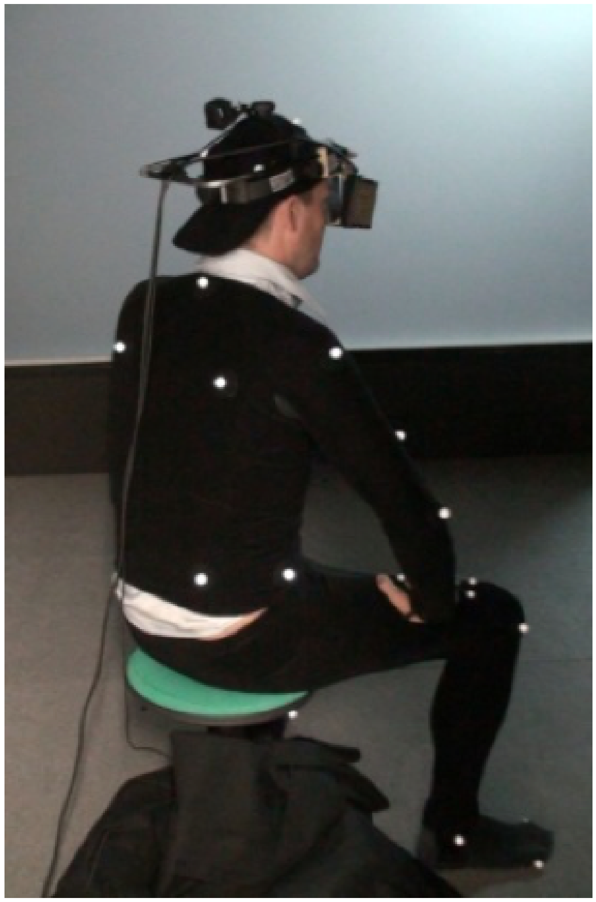

Figure 6.8: The Optitrack body suit with retroreflective markers

6.1.3 Metrics

There were three types of dependent variables. The first was the configuration {VH, VB, P, S} at which a participant declared a matching sensation of reality. The second consisted of the transition set—that is, the specific order of improvements that a user chose to move from one configuration i to another configuration j. The third was the post-experiment questionnaire that was completed by all participants.

6.1.4 Experimental procedures

6.1.4.1 Pre-experiment

Upon arriving at the lab, participants read an information document, signed an informed-consent form, and completed a demographic questionnaire. Participants were informed both verbally and on paper that they were free to withdraw from the experiment at any time without giving any reasons. After completing this process, participants put on the Optitrack suit and underwent a short calibration procedure, after which they donned the Oculus Rift HMD and began the experiment.

6.1.4.2 Experiment

Participants were seated wearing the HMD, through which they were able to see the virtual bar when looking around them and the virtual body from a first-person perspective when looking down. They were also able to see themselves reflected in the mirror in front of them. The virtual body and the scenario were both at the maximum level during this first exposure. Participants were first instructed to look around the room and describe what they saw. They were instructed to move their arms and legs, and to observe these motions both directly and in the mirror in front of them; this was done to establish a sense of embodiment in the virtual body. After that, they were instructed to play with the ball they had between their feet, which was also at the highest level of coherence. Participants were then shown the highest level of coherence of the conversation between the other virtual humans.

During this exposure to the highest levels of coherence for all four factors (virtual humans, virtual body, physical behavior of the ball, scenario) participants were told to focus on the sensation of reality they were feeling at the moment. They were told that this sensation would be used a reference for the rest of the experiment and would be referred to as the "optimal sensation of reality". Then participants were shown all the decreased levels of coherence for each of the elements, in the same order as described above: first, the behavior of the virtual humans, then the behavior of their own virtual body, then the behavior of the ball, and finally the different levels of scenario coherence. After making sure that the participant understood all the improvements they could make to affect their sensation of reality, the experimenter would give the instruction to start the experimental procedure. Participants were told that they would be playing a game in which the goal was to reach the optimal sensation of reality they experienced at the beginning of the experience and that they would earn 5 points each time they would reach this level of reality. They were also instructed to focus on the elements that were their priority for getting closer to this sensation of reality. Participants started each trial in a random configuration presenting different levels of the elements and were able to change one element at a time until they reached the optimal level of reality. The improvements were made by telling the experimenter which factor they wanted to improve. Similarly, they identified when the optimal sensation of reality was reached by saying so to the experimenter. Once they had reached the optimal sensation of reality, the next trial would begin. There were eight trials in total and an average of six changes per trial. Each participant started from configurations {0,0,0,0}, {1,0,0,0}, {0,1,0,0}, {0,0,1,0}, and {0,0,0,1}, and from three configurations randomly chosen from the configurations in which two improvements had already been made. These eight trials were presented in random order. Figure 6.9 illustrates the configuration space and highlights the possible starting configurations.

6.1.4.3 Post-experiment

After completing the virtual reality portion of the experiment, participants completed a short post-experiment questionnaire. The whole procedure including information, consent form signing and questionnaires lasted one hour, and the participants were compensated for their participation.

6.2 Results

6.2.1 Overview

As in (Slater et al., 2010), we make the simplifying assumption that the results of the eight trials are statistically independent. They cannot be truly independent, as the same participant carried out each of them, and may have learned from one trial to the next. However, the design of the experiment was such that each trial started from a different initial configuration, and so participants were presented with a different set of possible upgrades to choose from in each trial, and so had to reconsider their prioritieseach time.

In this section, we report on all three measures: which states were identified as matching the optimal sensation of reality, the order of transitions that each participant chose in each trial, and participants' responses to a post-experiment questionnaire.

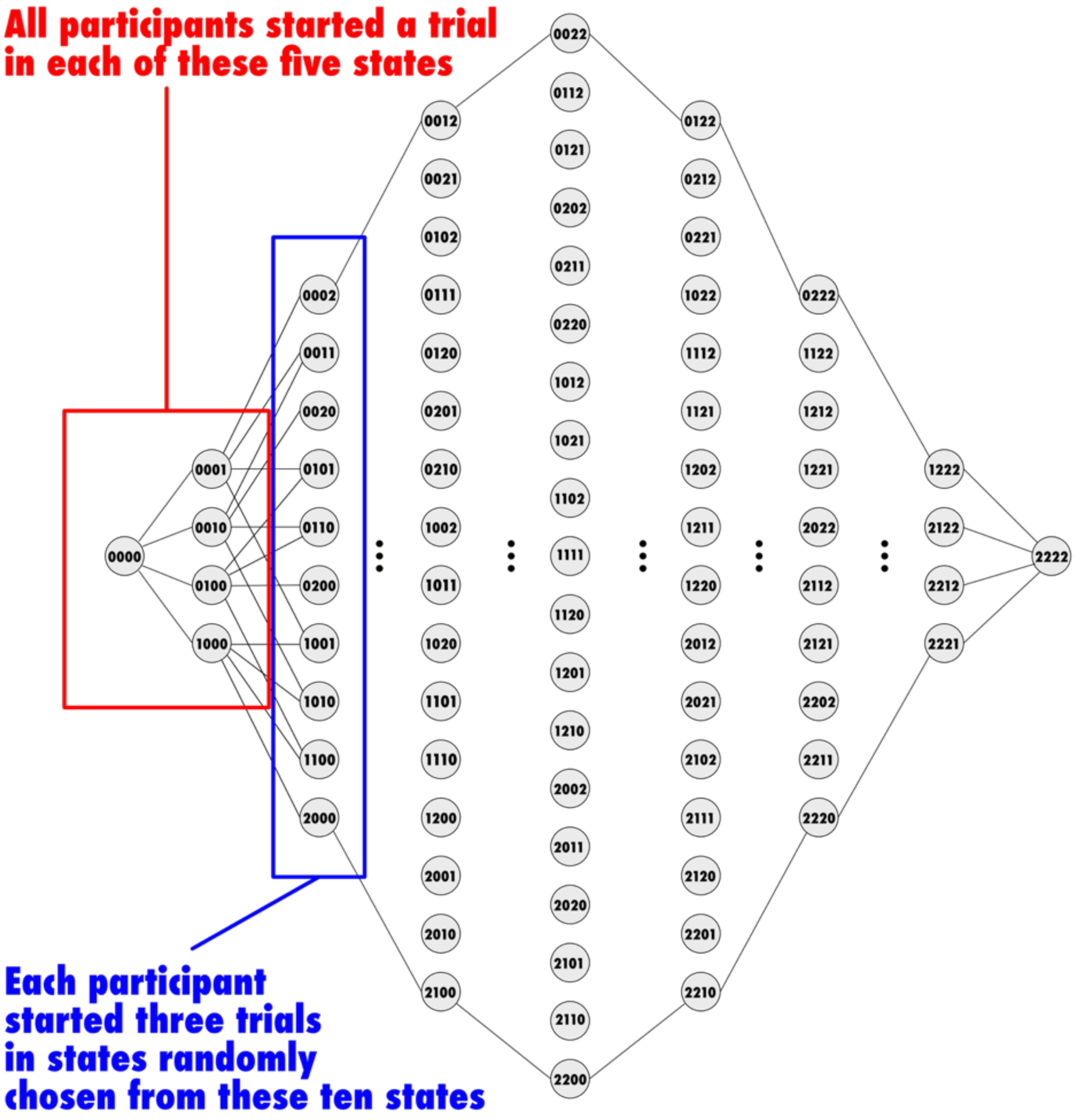

Figure 6.9: Markov chain with starting conditions highlighted

6.2.2 Accepted states

The participants' task was to improve the various factors until they felt that they had reached the same "level of reality" they had felt in the best possible configuration, C = {2,2,2,2}. These accepted states are shown in Figure 6.10. (Only states that were accepted five or more times are included in the figure, for ease of reading.) Included in that figure are both the percentage of total accepted configurations that a given configuration makes up (yellow lines), and also the probability that that configuration was marked as accepted if it was reached (blue lines). (For example, there were 165 total accepted configurations recorded. Configuration {2,2,1,2} was accepted 27 times, so it makes up 16.4% of the total accepted configurations. However, Configuration {2,2,1,2} was only reached 54 times across all participants. So it was accepted 50% (27/54) of the times it was reached.)

Note that the minimum number of improvements (including improvements which were part of the starting configuration for a trial) for any state in Figure 6.10 is 5, and in fact the average number of improvements for all the accepted states included in this figure is 6.90. This can be seen in the Markov Chain in Figure 6.11.

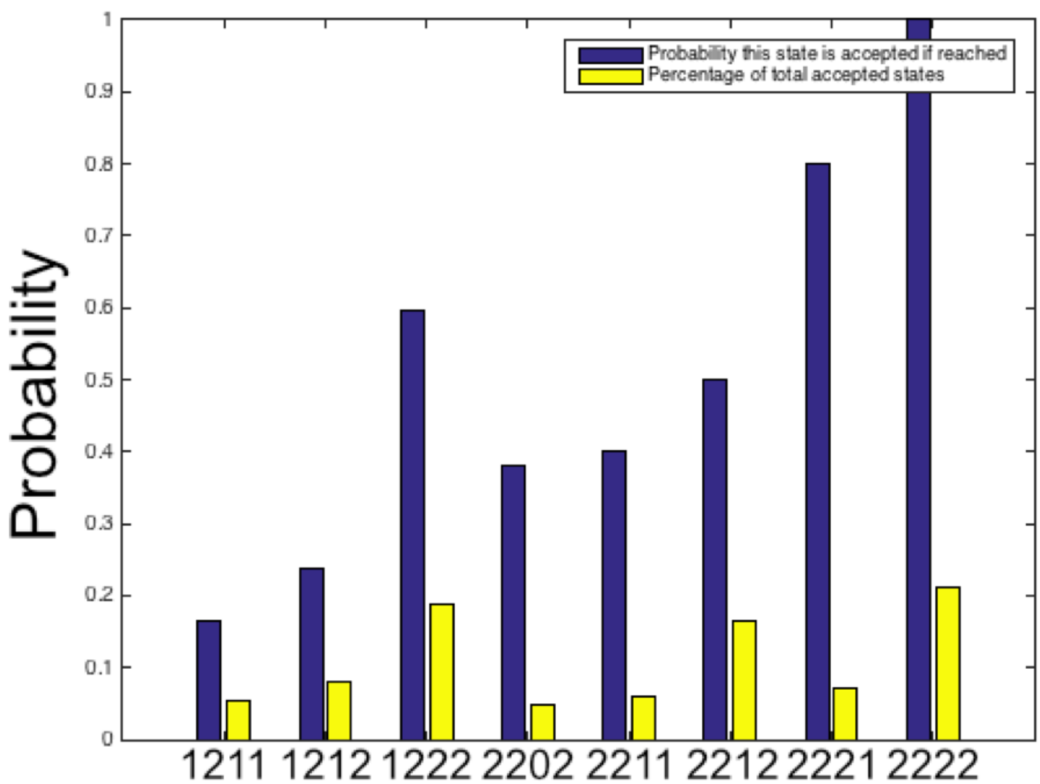

Figure 6.10: Accepted configurations and their related probabilities. The blue bars indicate the probability a configuration was accepted if reached, and the yellow bars indicate the percentage of total accepted configurations the given configuration made up.

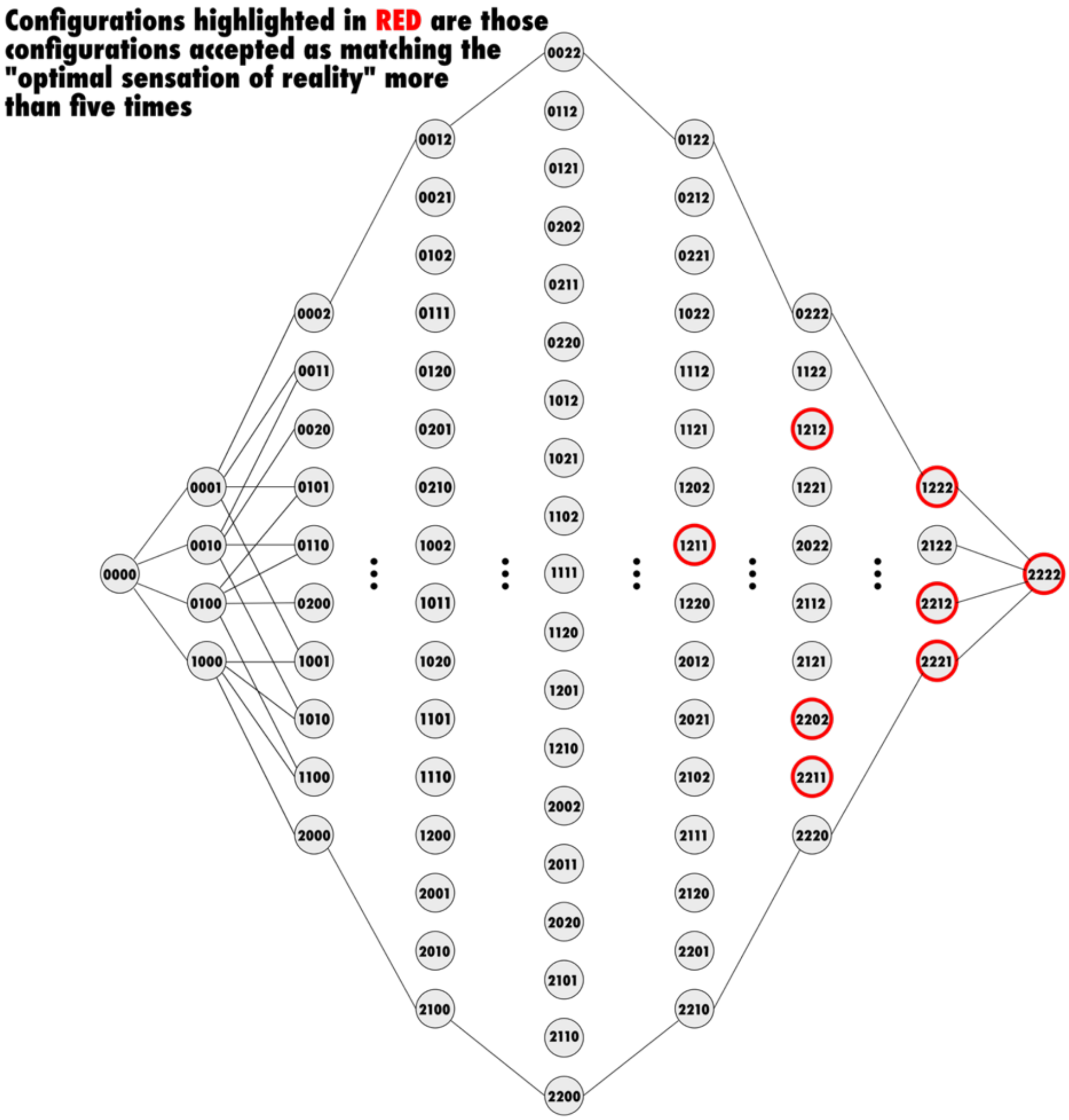

Figure 6.11: The most commonly accepted configurations shown on the Markov Chain

6.2.3 Transitions

We constructed a transition probability matrix P from the orders of states chosen by each participant. Each participant made approximately six improvements in each of eight trials (that is, six improvements from a starting configuration to a configuration accepted as matching the optimal sensation of reality), for a total of 936 observed transitions. Note that by construction (due to the nature of what transitions are allowed at each step) P is a very sparse matrix, with only 125 non-zero entries in an 81 x 81 matrix. (Recall there are 81 possible configurations, 3 virtual human levels x 3 virtual body levels x 3 physical behavior levels x 3 environment appearance levels; this configuration space is illustrated in Figure 6.9.) Given P, we can compute the probability distribution over the configurations for any given state. If we take as the starting configuration C = {0,0,0,0} (the lowest possible levels for each factor), then sP yields the probability distribution after one improvement has been made, sP^2 the probability distribution after two improvements have been made, and sP^n after n improvements. By construction, configuration {2,2,2,2} is absorbing, so the eighth step adds no information, but we can consider the probability distributions over configurations for the first seven steps.

Figure 6.12 shows the estimated probability distributions over the functions at each of the transitions (only probabilities greater than 0.01 are shown, for ease of reading). Figure 6.13 shows the most likely path through the Markov chain. (Note that at transition 4 and at transition 7, there are two approximately equal maximum probabilities, this is reflected by highlighting two nodes in the graph at the 4th and 7th levels.)

A clear majority of users chose to immediately upgrade the virtual body twice, in order to have a fully-tracked virtual body (configuration {0,2,0,0}). Following that, a majority of users upgraded the environment to level 1, moving out of the abstract environment into the mismatched environment (configuration {0,2,0,1}). After that, users tended to upgrade either the behavior of the virtual humans or the behavior of the ball to level 1, followed immediately by whichever one of those wasn't chosen first, restoring symmetry at state {1,2,1,1}. Users then tended to upgrade the environment for a second time {1,2,1,2}, then again were divided over whether to upgrade the virtual humans or the physics behavior to the second level, before finally choosing the other option and reaching state {2,2,2,2}.

6.2.4 Questionnaires

All participants completed a short post-experiment questionnaire. This included a modified SUS presence questionnaire (Usoh et al., 2000), as well as a series of questions asking them to rate the factors in order of which had the most impact on their sense of reality, and to explain why.

6.3 Discussion

6.3.1 The virtual body is the most important factor of Psi.

The importance of the virtual body showed in all measures that were collected: matching configurations, transition probabilities, and questionnaires. 99.4% of the matched configurations (when users declared a configuration to be "equally real" as the initial {2,2,2,2} configuration) had the virtual body at level 2, and 100% had the body at at least level 1. When it was possible to improve the virtual body (that is, VB was 0 or 1), users chose to do so 81.2% (281/346 observations) of the time. When the user had no virtual body (VB=0), that increased to 83.9% (120/143) of the time. Improving the virtual body from VB=0 to VB=1, and then again from VB=1 to VB=2, were the most common first and second improvements to be made. And in our post-experiment questionnaire, 90.9% of participants said that the most important factor to improve was the virtual body. It would seem that having a virtual body that moves with one's own body is extremely powerful for convincing a user that, "This is real."

6.3.2 Regarding the other factors, it is very important to have them in level 1, but not necessarily in level 2.

85.5% (141/165) of accepted configurations have every factor at level 1 or higher. Only three times (out of 165 total accepted configurations) did a participant accept a configuration where the virtual human behavior was at the lowest level (VH=0), only five times did they

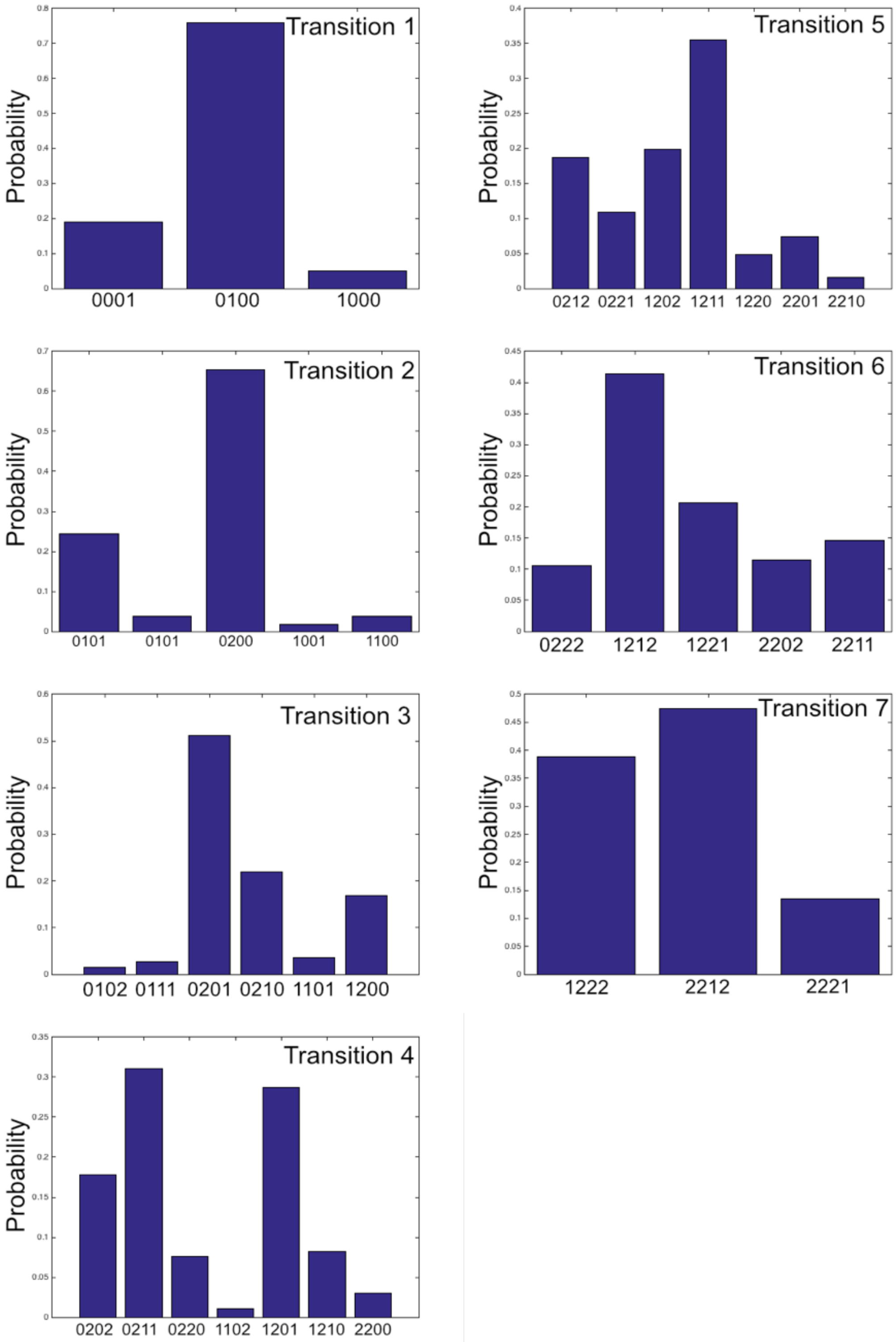

Figure 6.12: Transition probability distributions for each step n, $p = sP^n$

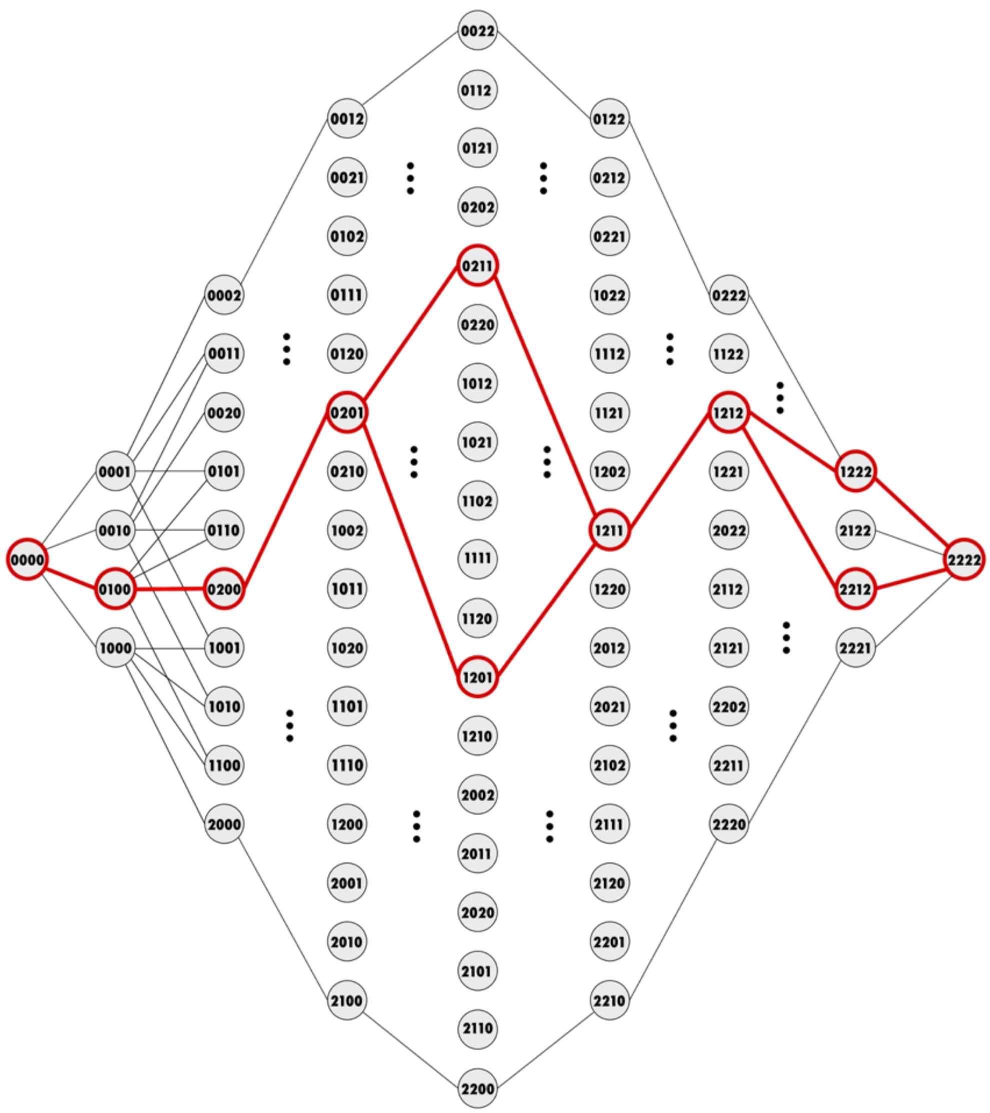

Figure 6.13: The most likely path taken through the Markov chain, based on the probability distributions shown in Figure 6.12

accept a configuration where the environment was at level 0, and only 19 times did they accept with the ball behavior at level 0. So participants very much wanted the virtual humans to move, but they did not necessarily have to interact with the participant. Similarly, participants overwhelmingly rejected the abstract environment, but did not necessarily require the matched environment, and the ball needed to move realistically, but not necessarily perfectly.

6.3.3 The second most important factor seems to be the scenario coherence.

This preference is less strong than the virtual body being most important, but the transition probability distributions show that after improving the virtual body twice, participants then choose to upgrade the environment a slight majority (51.2%) of the time. Also, after reaching configuration {1,2,1,1}, participants tended to upgrade the environment for a second time a plurality of the time (41.4%), more than twice as often as any other option for the sixth improvement (starting from {0,0,0,0}; if the participant started a trial in any other configuration, one or two improvements had already been chosen for them). Note that this is not just about the visual quality or complexity of the environment, but also the appropriateness of the environment to the scenario presented. Levels 1 and 2 of the environment were designed to be of approximately equal visual quality and complexity. Participant comments such as, "Depending on the conversation that I heard, the environment was important to establish it as real," "...[D]epending on the conversation, it was more clear what kind of bar would fit better," and "Level 1 was simply too elegant for some football fans with their team shirts," seem to indicate that this design was at least partially successful, and that participants considered the totality of the scenario and not merely the visual quality.

6.3.4 Response to the ball was not the same for all participants, but was very important for those participants who interacted with it extensively.

Despite the instructions being the same for all participants, there was a wide range of participant behaviors relating to the ball. Some participants barely looked at it, or touched it once or twice, just to see how realistic its motion was, before ignoring it for the rest of the trial. (This was reflected in participant comments such as, "I considered [the ball's] presence irrelevant," or "I had to forget about the other things to focus on the ball because it was at my feet.") However, some participants interacted with the ball extensively and considered the quality of the ball's movement to be very important. One participant actually rated it as the single most important factor, saying "... even if my body could be deformed or incomplete, the movement of the ball gave me the idea that that place was governed by physical laws similar to those of reality."

Anecdotally, male participants seemed to interact with the ball more than female participants did. This is reflected in the accepted configuration data. Males and females accepted the virtual humans, virtual body, and environment at strikingly similar rates. However, males accepted configurations that had the ball's behavior at level 2 63.8% of the time (51/80), while females accepted such configurations only 38.8% of the time (33/85). (Note that this difference was almost entirely in level 1, which females accepted 49.4% of the time, and males accepted only 25% of the time. Both genders accepted the ball at level 0 roughly 11% of the time.) Using Fisher's exact test, this difference is significant, with two-tailed p=0.0018. Also, the most commonly accepted configuration among males was {1,2,2,2}, while the most commonly accepted configuration among females was {2,2,1,2}. For reference, the accepted configurations for both male and female users are split out in Figure 6.14.

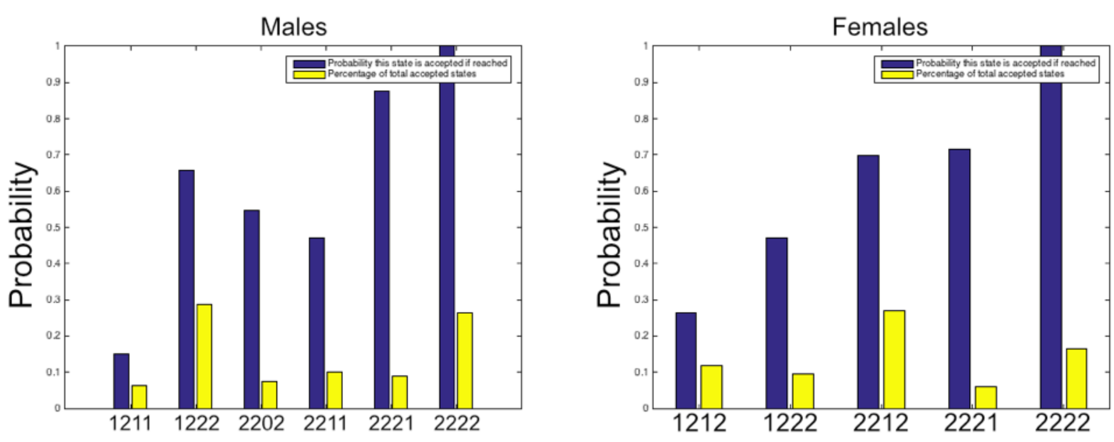

Figure 6.14: Accepted configurations and their related probabilities, split by gender. The blue bars indicate the probability a configuration was accepted if reached, and the yellow bars indicate the percentage of total accepted configurations the given configuration made up.

This difference is also somewhat supported by the questionnaire data, as only two (of eleven) females said that the ball was the first or second most important factor, while five (of ten) males said that the ball was the first or second most important factor.

6.3.5 Participants who reported lower presence were more likely to accept the ball at level 0 or level 1 than participants who reported higher presence.

As part of the post-experiment questionnaire, participants completed a modified Slater-Usoh-Steed questionnaire with five questions. From this information, we divided the population into those who had low presence—0, 1, or 2 responses of 6 or 7 on the Likert scale (11 participants)—and those who had high presence—3, 4, or 5 responses of 6 or 7 (10 participants). As with the male/female split above, these groups accepted configurations with the ball at level 2 at markedly different rates. The low-presence group accepted configurations in which the ball was at level 0 or level 1 57.5% of the time (50/87), while the high-presence group accepted such configurations only 39.7% of the time (31/78). This difference between the groups is again significant, with two-tailed p=0.029.

The design of the experiment does not enable us to say for certain whether these differences are correlation or causation, and if the latter, in what direction. However, we speculate that users who played with the ball more (primarily, but not entirely, male users) interacted more extensively with the scenario and so felt a higher degree of presence. Then, since playing with the ball was important to them, they chose not to accept states in which the behavior of the ball was noticeably unrealistic.

6.4 Conclusion

This study is directly influenced by the study in (Slater et al., 2010), where Slater and colleagues demonstrated the feasibility of such an experiment, investigating qualia (such as PI and Psi) using matching experiments similar to those used to determine metamers in color science. In that paper, they showed that participants given different instructions (to focus on PI or Psi) chose different matching states and made transitions in different orders.

This experiment builds on that one, focusing exclusively on factors thought to influence Psi, namely the behavior of other virtual humans in the scenario, the appearance and behavior of the participant's virtual body, the behavior of other objects in the environment, and the appearance and "scenario-correctness" of the environment itself. Psi has not been previously investigated in this way, nor have the factors contributing to Psi been delineated before.

The results show that the virtual body is the most powerful contributor to Psi of the four factors studied in this experiment, and that result holds across matching states, transition probabilities, and post-experiment questionnaires. This suggests that full-body tracking is the technology that can contribute most to Psi, and that the body may indeed be the "focal point where PI and Psi are fused," as claimed in (Slater, 2009). One thing that is not clear from this experiment, though, is whether the presence of any self-avatar at all is better than none, as participants almost universally increased the virtual body to the maximum level at the earliest opportunity. It is not clear whether participants here actually thought level 1 of the virtual body was substantially better than level 0, or whether it was merely a necessary step to get to the ultimately desired state of having a fully-tracked avatar. More research is needed to answer this question.

This study represents only a first step, intended to further develop our understanding of the factors of coherence and Psi, which have been comparatively understudied compared to immersion and PI. A possible next step is to run a similar experiment focusing only on immersion characteristics of the system, as this experiment focused only on coherence. More interestingly, one could do an experiment in which participants are prompted to maximize their feelings of PI and Psi, in which the factors include both coherence and immersion factors. This could enable us to make practical suggestions as to whether immersion factors or coherence factors might be more or less important for a given type of task.

CHAPTER 7

Conclusions and future work

7.1 Revisiting the thesis statement

In this dissertation,

1. Slater's concept of Plausibility Illusion has been usefully elaborated by the auxiliary concepts of virtual scenarios, coherence, and reasonable circumstances, and

2. I have shown the following:

   - Plausibility Illusion can be detected using existing presence measures,
   - Plausibility Illusion is impacted by individual differences,
   - The virtual body is the most important contributing factor to Plausibility Illusion, and
   - The coherence of the scenario is the second most important factor, among those tested.

In Chapter 1, I presented my novel definitions of several terms related to Plausibility Illusion, namely *virtual scenarios*, *coherence* and *reasonable circumstances*. These terms enable the discussion of Plausibility Illusion as a concept parallel to Place Illusion by giving it its own vocabulary.

In Chapter 2, I summarized previous work and showed how it could be unified using the PI/Psi framework.

In Chapter 3, I fleshed out the concept of Plausibility Illusion and identified some areas where Plausibility Illusion offers a potential explanation for existing phenomena in the study of virtual environments, such as the book problem.

In Chapter 4, I presented an experiment that strongly suggested that participant motivation is one of the individual differences that impacts Psi.

In Chapter 5, I demonstrated that data generated from existing presence measures—namely, the Witmer and Singer Presence Questionnaire (PQ), the Slater-Usoh-Steed questionnaire (SUS), and physiological measures of arousal—can be affected by each of Place Illusion, Plausibility Illusion, and the interaction of the two.

In Chapter 6, I demonstrate that users of VEs, namely my study participants engaged in a virtual football playing task, may interact differently with a virtual environment, and that these differences are reflected in participants order of improvements in the virtual environment as well as in their reported presence as measured by the SUS questionnaire.

In Chapter 6, I also demonstrated that the participants, all of whom were directed to improve their level of Plausibility Illusion, overwhelmingly chose to improve the representation and behavior of their virtual bodies over other improvements to the virtual scenario. This suggests that the virtual body is the most important component of a virtual scenario for eliciting Plausibility Illusion, and the coherence of the scenario is second most important. Taken as a whole, the results of the study also demonstrate the feasibility of ranking aspects of a virtual scenario on their importance for eliciting Plausibility Illusion.

7.2 Future research directions building on the reported work

It must be said that we are in the very early days of research about Plausibility Illusion. Hence the available future work is broad and extensive. It ranges from VR-focused research to further work in cognitive neuroscience to new interactions with content creators. All of these directions, like this dissertation, have the goal of improving the design, development, and evaluation of virtual experiences.

7.2.1 Evaluating immersion and coherence together

The Chapter 6 is, to my knowledge, the first attempt at establishing a priority order for factors of coherence in immersive virtual environments. It, however, says nothing about factors of immersion. A similar order-establishing study should be done, with a theoretically grounded set of immersion factors, to begin developing such an ordering for immersion factors.

Beyond that, though, the more interesting problem is developing a combined ranking of immersion and coherence factors in virtual scenarios. For example, is wider field of view more or less important than a fully tracked virtual body? Are passive haptics more or less important than coherence of physical interactions? The method pioneered by Slater (Slater et al., 2010) and continued in Chapter 6 provides a methodology for answering these questions, and many other similar ones, about virtual environment systems. As more and more elements are ranked against each other, we would develop a comprehensive guide of which system characteristics provide the most "bang for the buck" when designing a VE installation or scenario. Given results such as Zimmons's black-and-white Pit environment (Zimmons, 2004) (Zimmons and Panter, 2003) (in which the Pit produced similar physical stress responses and responses on presence questionnaires regardless of whether it was rendered in low lighting quality, or even when rendered as only a black and white grid with no shading), I suspect that immersion factors are largely "good enough" at the moment, and that coherence factors may actually be more important for improving today's virtual scenarios.

7.2.2 Developing metrics that can measure Place Illusion and Plausibility Illusion separately

In this dissertation (Chapter 5) I have shown that both Place Illusion and Plausibility Illusion affect measures developed to measure presence in virtual environments. In addition, I showed that several subscores of the Witmer and Singer Presence Questionnaire respond differently to high and low levels of immersion and coherence. However, I did not identify a metric that would respond to only Place Illusion or only Plausibility Illusion. Such instruments are necessary to push research forward in this area, and to test some of the hypotheses put forward in this paper. For example, if instruments existed to measure these constructs separately, it would be possible to test my hypothesis about the origin of the book problem (Chapter 3). In addition, the lack of such measures keeps us from measuring the strength of the interaction between PI and Psi.

7.2.3 Priming and Plausibility Illusion

As I argued in Chapter 3, I believe that appropriate priming can increase Psi, and therefore presence, by reducing or eliminating "wrong" (or, I would argue, unanticipated) events in the VE which can cause breaks in Psi. There has been some prior work investigating priming in virtual environments (Nunez and Blake, 2003) (Steinicke et al., 2009), but it has not yet been investigated as a way to improve the perceived coherence in an "unrealistic" virtual environment, such as a science fiction or fantasy scenario. Lessons could be drawn from, for example, the design of theme park rides. As the keynote speaker at SIGGRAPH 2006, Joe Rohde commented on how Disney Imagineers design the waiting areas for attractions such that patrons are being primed to be in the right state of mind to enjoy the ride before they even get on it (Rohde, 2006).

7.2.4 The neurobiological basis for Plausibility Illusion

In (Seth et al., 2012), Seth and colleagues present a neurocognitive model of presence based on minimizing interoceptive prediction error—that is, they propose that presence is elicited when predictions about the internal state of the body, such as balance or proprioception, are accurate. What they label as presence in Figure 7.1 would seem to be Place Illusion, given that it is composed of interoceptive state and predictions about consequences of sensory input generated by motor control signals (roughly, sensorimotor contingencies). Part of their evidence for this model is the brain activity of patients suffering from depersonalization (loss of subjective sense of reality of the self) and derealization (loss of subjective sense of reality of the world). While the authors discuss these both as failures of conscious presence, it seems likely to me that they represent failures of Place Illusion and Plausibility Illusion, respectively. Can this model can be amended to differentiate between Place Illusion and Plausibility Illusion? How?

7.2.5 Collaborations with media creators and game developers

Whereas Plausibility Illusion may be understudied by virtual environments researchers to this point, that does not mean that no one has been thinking about developing coherent scenarios. For game developers, authors, artists, screenwriters, trainers, directors, costume and set designers crafting a coherent world is the life's work. Surely they have intuitions as to which aspects of a scenario are absolutely critical to get right, and which can be (or perhaps even should be) omitted or modified in the effort to create a coherent world. Surely we researchers can use their expertise to guide our research. And similarly, our research can guide the nascent virtual reality game developers and "film" makers.

7.2.6 Storytelling in virtual environments

This dissertation addressed scenario coherence as just one of several factors influencing Plausibility Illusion in virtual scenarios. However, in scenarios where narrative is important, it may in fact be the most important factor. Novels have centuries of narrative convention that can be drawn upon to help structure a story; films have a century of cinematographic and editing techniques that can be used to convey meaning. Such techniques do not yet exist for creating compelling immersive scenarios. Further research into scenario coherence may inform effective means for telling stories in immersive virtual environments.

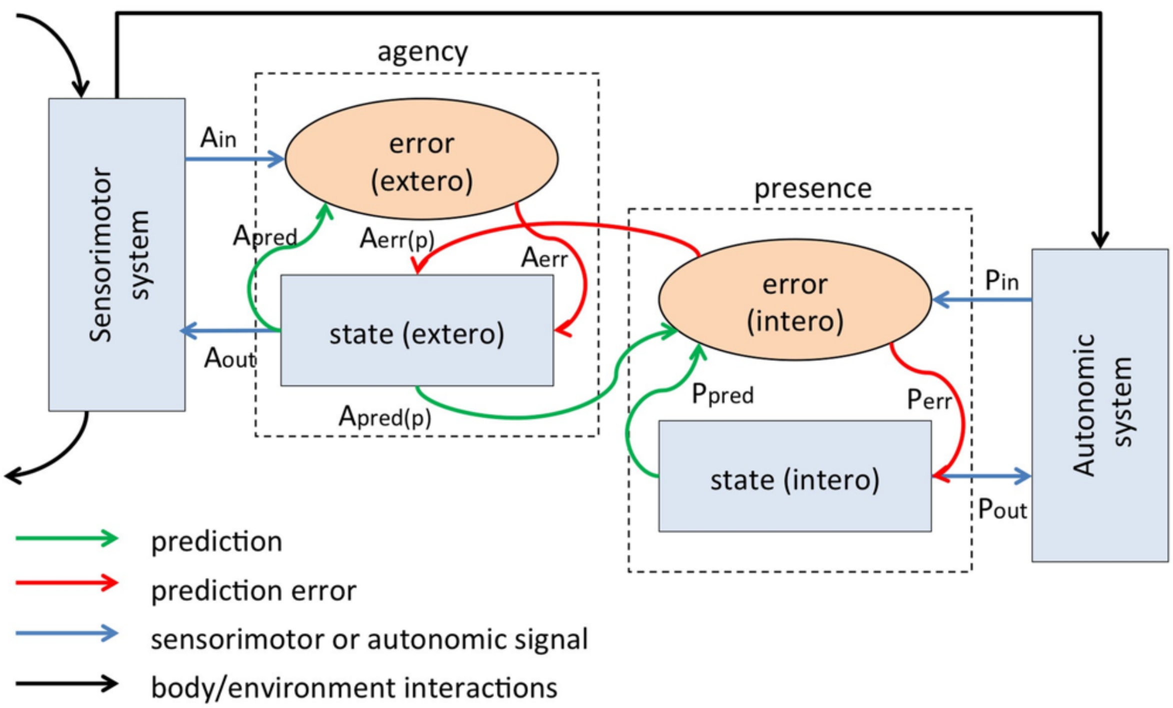

Figure 7.1: An interoceptive predictive coding model of conscious presence. Both agency and presence components comprise state and error units; state units generate control signals (A_{out}, P_{out}) and make predictions [$A_{pred}, P_{pred}, A_{pred}(p)$] about the consequent incoming signals (A_{in}, P_{in}); error units compare predictions with afferents, generating error signals [$A_{err}, P_{err}, A_{err}(p)$]. In the current version of the model the agency component is hierarchically located above the presence component, so that it generates predictions about the interoceptive consequences of sensory input generated by motor control signals. [adapted from (Seth et al., 2012)]

APPENDIX A

Bayesian statistics and JAGS

Throughout this dissertation, I have analyzed (and in several cases, re-analyzed) data using Bayesian statistical methods. This is in contrast to traditional, or frequentist, statistics. For this reason, I do not report p values, preferring instead to report the posterior probability that a hypothesis is true (or false), and allow the reader to make up their own mind about my conclusions.

A.1 A (very brief) introduction to Bayesian statistics

Bayesian statistics is founded upon Bayes' Theorem:

$$P(\Theta|y) = \frac{p(\Theta) \times p(y|\Theta)}{p(y)} \tag{A.1}$$

This can be restated as, the posterior probability of parameter Θ given observed data y equals Prior probability of $\Theta\times$ Likelihood of y given Θ / Marginal likelihood of y. Or, more briefly:

$$Posterior = \frac{Prior \times Likelihood}{Marginal\ Likelihood} \tag{A.2}$$

Or even more briefly (since the marginal likelihood p(y) is just a normalizing constant):

$$Posterior \propto Prior \times Likelihood \tag{A.3}$$

So, in Bayesian statistics, we are computing the posterior probability distribution of a parameter given the prior probability distribution (which may be known or may be estimated) and the observed data from our experiment.

A.2 Advantages of using Bayesian statistics

This section is informed by and adapted in part from (Dablander, 2015).

A.2.1 The result is a probability distribution

This has the advantage of matching the way that we normally reason about probabilities. We may have data that says, for example, that NBA players are 79 inches tall, on average (Basketball Reference, 2016). Intuitively, we know that this doesn't mean that all NBA players are likely to be exactly 79 inches tall. Some will be, like Wesley Johnson (Reddit, 2015). Some, like Stephen Curry, are only 75 inches tall. Others, like, Dirk Nowitzki, are 84 inches tall. In fact, the probability distribution for NBA player height is shown in Figure A.1.

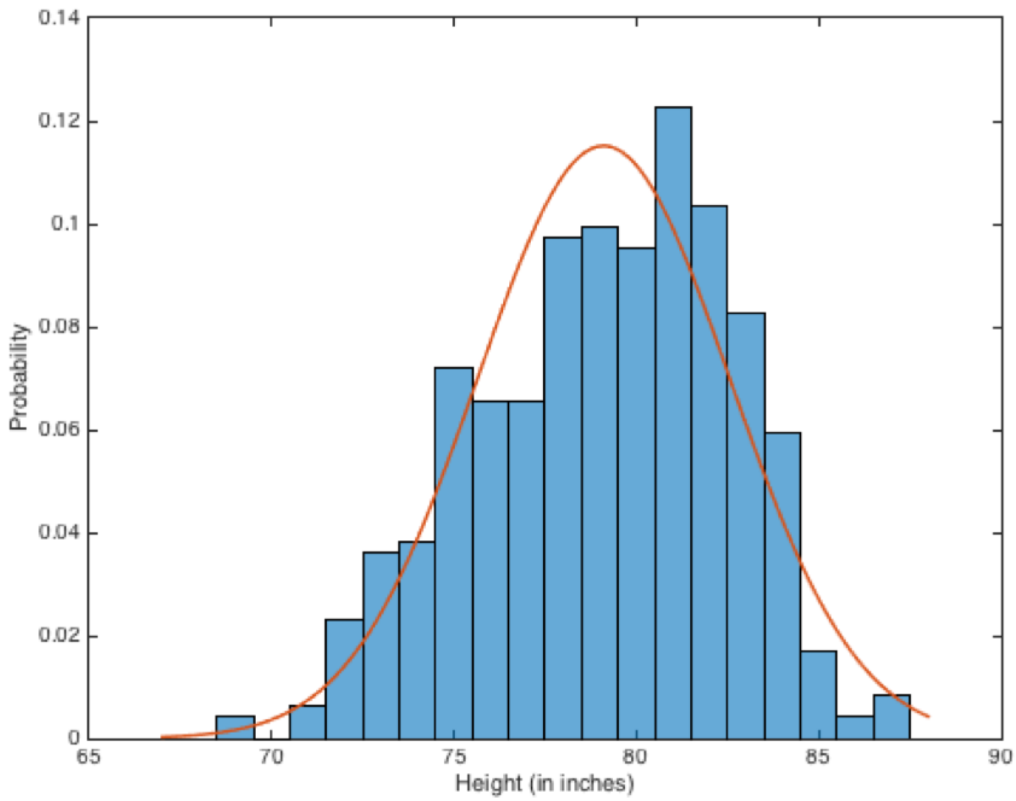

Figure A.1: Distribution of heights for all NBA players in 2015, overlaid with a normal distribution, $\mu = 79.1$, $\sigma = 3.46$ (data from basketball-reference.com)

Since we have this probability distribution, it is very natural to ask questions about it. For example, if you wanted to compute the probability that any given NBA player is more than 78 inches tall, that is simply

$$\int_{78}^{\infty} \frac{1}{3.46\sqrt{2\pi}} e^{-\frac{(x-79.1)^2}{2\times 3.46^2}} dx \tag{A.4}$$

which yields 62.5%. Or if you wanted to know how likely a given NBA player is either less than six feet or more than seven feet tall, you could compute that as

$$\int_{-\infty}^{72} \frac{1}{3.46\sqrt{2\pi}} e^{-\frac{(x-79.1)^2}{2\times 3.46^2}} dx + \int_{84}^{\infty} \frac{1}{3.46\sqrt{2\pi}} e^{-\frac{(x-79.1)^2}{2\times 3.46^2}} dx \tag{A.5}$$

which yields 9.89%.

So, since Bayesian analysis yields a probability distribution, it more closely matches how we tend to think about probability, and it more naturally lends itself to analyses. Given a p value, you can't easily ask how likely it is that your parameter is greater than 12 but less than 37. Given a probability distribution, you can.

A.2.2 Inclusion of prior evidence

The inclusion of a prior makes Bayesian statistics much more well-behaved in the presence of small sample sizes and "extreme" samples, such as those where an outcome is observed in 100% (or 0%) of the samples. If one flipped a coin twice times and it came up heads both times, frequentist statistics would suggest that it is 100% likely to come up heads on the next trial, having no concept of a prior probability distribution. In Bayesian statistics, one might have a prior that heads should come up 50% of the time. In that case, the observed data would result in a posterior probability of 75%: greater than 50% (it might be a weighted coin, after all), but certainly not 100%.

A.2.3 "The illusion of objectivity"

While it may not initially seem so, p values are inherently subjective in a variety of subtle ways (Berger and Berry, 1988). Eric-Jan Wagenmakers (Wagenmakers, 2007) has helpfully categorized p-value problems into three categories.

Firstly, p values depend on data that were never observed. Computing the p value depends on the sampling distribution of the test statistic, given that the null hypothesis H_0 holds.

To obtain this sampling distribution, one assumes that the H_0 holds exactly over many hypothetical replications of the experiment. p is then computed by taking the sum of the observed value as well as all of these hypothetical replications that are at least as extreme as the observed value. As stated by Sir Harold Jeffreys, "What the use of P implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred. This seems a remarkable procedure" (Jeffreys, 1961).

Secondly, p values depend on possibly unknown subjective intentions of the experimenter. For example, consider the case where the test statistic is the number of true responses to exactly ten true/false questions. Now consider the case where the test statistic is the number of true responses to true/false questions obtained before the third false response. Now, the sequence {T, T, F, T, T, T, F, T, T, F} can be observed under either of these sampling plans, but they will generate different p values. If the sampling plan is unknown, p cannot be computed correctly.

Thirdly, p values do not quantify statistical evidence. Two experiments with the same p value but different sample sizes provide different strengths of evidence against the null hypothesis.

Bayesian statistics, on the other hand, is up front with its assumptions. The posterior distribution is based on the data and the prior. As long as the prior is specified, the method is completely open to inspection.

A.2.4 Optional stopping

In traditional statistics, once one has decided on a sample size, one cannot stop early. Conversely, if you've run your full sample size, you cannot simply run more participants. In either case, the p value you calculate for the "new" sample size will not be accurate. In Bayesian statistics, you can recompute the posterior probabilities after every participant, and simply use those as the prior probability for analysis of the subsequent participants. There is no mathematical problem associated with deciding to run more or fewer participants once the experiment has begun.

A.3 Computing posterior probabilities using JAGS

In this dissertation, all Bayesian analyses were performed using the software package JAGS (Plummer, 2003), which was used with MATLAB via the MATJAGS toolkit, `http://psiexp.ss.uci.edu/research/programs_data/jags/`. In this 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. The result is a joint posterior distribution for all model parameters. The prior distributions on all variables were intentionally chosen to be non-informative (normal distributions with variance 1000), so as not to bias the results unduly.

To summarize the method, one begins by expressing the observed dependent variables as a function of the independent variables. For example, there would be an equation for each dependent variable for each participant. The initial values of the model parameters are stochastically sampled from the specified prior distributions for each parameter (sampled from a normal distribution with large variance, in this case), and the system of equations is then simultaneously solved to minimize overall error. This solution generates a single value of each model parameter. The process is repeated a large number of times (here 50,000), with different initial values each time. By doing so, one stochastically generates a posterior probability distribution, composed of the 50,000 solved-for values, for each model parameter.

The following pseudocode illustrates this model.

```
for run = 1 to 50000
  # Populate a system of equations, one per DV per participant
  for i = 1 to number_of_participants
    for j = 1 to number_of_dependent variables
      # y is a stochastic variable, sampled from a normal distribution with
      # mean mu and variance 1/tau
      y[i][j] ~ dnorm(mu[i][j], tau[j])
      # mu is a deterministic node, equal to its RHS
      mu[i][j] <- a[j][0] + a[j][1]*x1 + + a[j][k]*x_k
```

```
    end
  end
  # Initialize all parameters from their prior distributions
  for j = 1 to number_of_dependent_variables
    for k = 0 to number of independent_variables
      # All coefficients are sampled randomly
      a[j][k] ~ dnorm(0, 0.001)
      # The exact value of tau is also sampled randomly
      tau[j] ~ dgamma(0.001, 0.001)
    end
  end

  Sample stochastic variables
  Solve this system of equations by Markov chain Monte Carlo
end
```

Unlike traditional 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. (Bergström et al., 2016).

BIBLIOGRAPHY

Akin, D., Minsky, M., Thiel, E., and Kurtzman, C. (1983). Space applications of automation, robotics, and machine intelligence systems, (aramis) phase ii, vol. 3: Executive summary. Technical report, Massachusetts Institute of Technology, Cambridge, MA, USA.

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

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., and Rey, B. (2004). Immersion and emotion: Their impact on the sense of presence. *CyberPsychology & Behavior*, 7(6):734–741.

Baños, R. M., Botella, C., García-Palacios, A., Villa, H., Perpiña, C., and Alcañiz, M. (2000). Presence and reality judgment in virtual environments: A unitary construct. *CyberPsychology & Behavior*, 3(3):327–335.

Basketball Reference (2016). NBA League Averages.

Berger, J. and Berry, D. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, 76(2):159–165.

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

Biocca, F. (2001). Inserting the presence of mind into a philosophy of presence: A response to Sheridan and Mantovani and Riva. *Presence: Teleoperators and Virtual Environments*, 10:554–556.

Biocca, F. (2002). Presence working group research targets. Presentation at the "Presence Info Day" of the European Commission.

Biocca, F. and Delaney, B. (1995). Immersive virtual reality technology. In Biocca, F. and Levy, M. R., editors, *Communication in the age of virtual reality*, pages 57–124. Lawrence Erlbaum Associates Inc., Hillsdale, NJ.

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

Brogni, A., Slater, M., and Steed, A. (2003). More breaks less presence. In *PRESENCE 2003 - 6th Annual International Workshop on Presence*.

Bystrom, K.-E., Barfield, W., and Hendrix, C. (1999). A conceptual model of the sense of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 8:241–244.

Carassa, A., Morganti, F., and Tirassa, M. (2005). A situated cognition perspective on presence. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society: 2005; Stresa, Italy*, pages 384–389. Sheridan Printing.

Carlin, A. S., Hoffman, H. G., and Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behaviour Research and Therapy*, 35(2):153–158.

Chertoff, D. B., Goldiez, B., and Joseph J. LaViola, J. (2010). Virtual experience test: A virtual environment evaluation questionnaire. In *Virtual Reality*, pages 103–110.

Dablander, F. (2015). Bayesian statistics: Why and how.

David, N., Newen, A., and Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition*, 17(2):523 – 534.

Dillon, C., Keogh, E., and Freeman, J. (2002). "It's been emotional": Affect, physiology, and presence. In *Proceedings of the 5th Annual International Conference on Presence*.

Dillon, C., Keogh, E., Freeman, J., and Davidoff, J. (2000). Aroused and immersed: The psychophysiology of presence. In *Proceedings of the 3rd International Workshop on Presence*, London, UK. Goldsmiths College / Independent Television Commission.

Draper, J. V., Kaber, D. B., and Usher, J. M. (1998). Telepresence. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):354–375.

Ellis, S. R., Dorighi, N. S., Menges, B. M., Adelstein, B. D., and Jacoby, R. H. (1997). In search of equivalence classes in subjective scales of reality. In *Human-Computer Interaction*, volume 2, pages 873–876.

Flach, J. M. and Holden, J. G. (1998). The reality of experience: Gibson's way. *Presence: Teleoperators and Virtual Environments*, 7:90–95.

Freeman, J., Avons, S., Meddis, R., Pearson, D. E., and IJsselsteijn, W. (2000). Using behavioral realism to estimate presence: a study of the utility of postural responses. *Presence: Teleoperators and Virtual Environments*, 9(2):149–164.

Freeman, J., Avons, S. E., Pearson, D. E., and IJsselsteijn, W. A. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments*, 8:1–13.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, 4(1):14 – 21.

Garau, M., Friedman, D., Widenfeld, H. R., Antley, A., Brogni, A., and Slater, M. (2008). Temporal and spatial variations in presence: Qualitative analysis of interviews from an experiment on breaks in presence. *Presence: Teleoperators and Virtual Environments*, 17:293–309.

Gerrig, R. (1993). *Experiencing narrative worlds.* Yale University Press, New Haven, CT.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Houghton-Mifflin, Boston, MA.

Green, M. C. (1996). Mechanisms of narrative-based belief change. Master's thesis, The Ohio State University.

Green, M. C. and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5):701–721.

Heeter, C. (1992). Being there: the subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, 1:262–271.

Held, R. M. and Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments*, 1(1):109–112.

Hendrix, C. and Barfield, W. (1995). Presence in virtual environments as a function of visual and auditory cues. In *Proceedings of the Virtual Reality Annual International Symposium (VRAIS'95)*, VRAIS '95, pages 74–82, Washington, DC, USA. IEEE Computer Society.

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

Hendrix, C. and Barfield, W. (1996b). The sense of presence within auditory virtual environments. *Presence: Teleoperators and Virtual Environments*, 5(3):290–301.

Herrera, G., Jordan, R., and Vera, L. (2006). Agency and presence: a common dependence on subjectivity? *Presence: Teleoperators and Virtual Environments*, 15(5):539–552.

IJsselsteijn, W. A., de Ridder, H., Freeman, J., and Avons, S. (2000). Presence: Concept, determinants and measurement. In *Storage and Retrieval for Image and Video Databases*.

International Society for Presence Research (2000). The concept of presence: explication statement.

Interrante, V., Ries, B., and Anderson, L. (2007). Seven league boots: A new metaphor for augmented locomotion through moderately large scale immersive virtual environments. In *IEEE Symposium on 3D User Interfaces, 2007*, pages 167–170.

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

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press, Inc., Oxford.

Johnsen, K., Raij, A., Stevens, A. O., Lind, D. S., and Lok, B. (2007). The validity of a virtual human experience for interpersonal skills education. In *CHI '07: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1049–1058. ACM.

Khanna, P., Yu, I., Mortensen, J., and Slater, M. (2006). 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. ACM.

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

Kuschel, M., Freyberger, F., Buss, M., and Färber, B. (2007). A presence measure for virtual reality and telepresence based on multimodal conflicts. In *Proceedings of the 10th International Workshop on Presence*.

Laux, L., Glanzmann, P., Schaffner, P., and Spielberger, C. D. (1981). *State-Trait-Angstinventar [State-Trait Anxiety Inventory]*. Beltz, Weinheim, Germany.

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

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

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

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

Mantovani, G. and Riva, G. (2001). Building a bridge between different scientific communities: On Sheridan's eclectic ontology of presence. *Presence: Teleoperators and Virtual Environments*, 10(5):537–543.

Meehan, M., Insko, B., Whitton, M., and Brooks, Jr., F. P. (2002). 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. ACM.

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

Mori, M. (1970). The uncanny valley. *Energy*, 7(4):33–35.

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

Murray, D. (1997). Autism and information technology: Therapy with computers. In Powell, S. and Jordan, R., editors, *Autism and learning: A guide to good practice*. David Fulton, London, UK.

Nunez, D. and Blake, E. (2003). Conceptual priming as a determinant of presence in virtual environments. In *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '03, pages 101–108, New York, NY, USA. ACM.

O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–973.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria.

Raij, A. B., Johnsen, K., Dickerson, R. F., Lok, B., Cohen, M. S., Duerson, M., Pauly, R. R., Stevens, A. O., Wagner, P., and Lind, D. S. (2007). Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Transactions on Visualization and Computer Graphics*, 13:443–457.

Reddit (2015). Your 2014-2015 average NBA player is - Wesley Johnson.

Regenbrecht, H. T., Schubert, T. W., and Friedmann, F. (1998). Measuring the sense of presence and its relations to fear of heights in virtual environments. *International Journal of Human-Computer Interaction*, 10(3):233–249.

Renee, S. (2002). Bunraku (puppets).

Riener, C. and Proffitt, D. (2002). Quantifying spatial presence. In *Proceedings of the 5th Annual International Conference on Presence*.

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

Rivera-Gutierrez, D., Welch, G., Lincoln, P., Whitton, M., Cendan, J., Chesnutt, D. A., Fuchs, H., and Lok, B. (2012). Shader lamps virtual patients: The physical manifestation of virtual patients. *Studies in Health Technology and Informatics*, 173:372–378.

Rohde, J. (2006). *From Myth to Mountain: Insights Into Virtual Placemaking.* Keynote speech at SIGGRAPH 2006, Boston MA.

Rosakranse, C. and Oh, S. Y. (2014). Measuring presence: The use trends of five canonical presence quesionnaires from 1998-2012. In *Proceedings of the International Society for Presence Research 2014*. International Society for Presence Research.

Russell, J. (1996). *Agency: its role in mental development*. Essays in Environmental Psychology. Psychology Press, East Sussex, UK.

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

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

Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281.

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

Schubert, T. W. (2009). A new conception of spatial presence: Once again, with feeling. *Communication Theory*, 19(2):161–187.

Seth, A. K., Suzuki, K., and Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2(395).

Sheridan, T. B. (1992). Musings on telepresence and virtual presence. In *Papers from SRI's 1991 Conference on Virtual Reality on Virtual worlds : real challenges: real challenges*, SRI '91, pages 55–65, Westport, CT, USA. Meckler Corporation.

Skarbez, R., Kotranza, A., Brooks, F.P., J., Lok, B., and Whitton, M. (2011). An initial exploration of conversational errors as a novel method for evaluating virtual human experiences. In *IEEE Virtual Reality (VR)*, pages 243–244.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 8:560–565.

Slater, M. (2004a). How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4):484–493.

Slater, M. (2004b). A note on presence terminology. In *Presence Connect*, volume 3.

Slater, M. (2004c). Presence and emotions. *CyberPsychology & Behavior*, 7(1):123.

Slater, M. (2009). 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.

Slater, M., Brogni, A., and Steed, A. (2003). Physiological responses to breaks in presence: A pilot study. In *PRESENCE 2003 - 6th Annual International Workshop on Presence*.

Slater, M., Guger, C., Edlinger, G., Leeb, R., Pfurtscheller, G., Antley, A., Garau, M., Brogni, A., and Friedman, D. (2006). Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 15:553–569.

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

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

Slater, M. and Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, 9:413–434.

Slater, M., Steed, A., and Usoh, M. (1993). The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In Goebel, M., editor, *First Eurographics Workshop on Virtual Environments*, pages 71–83.

Slater, M., Steed, A., and Usoh, M. (1995). The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In *Selected Papers of the Eurographics Workshops on Virtual Environments '95*, VE '95, pages 135–148, London, UK, UK. Springer-Verlag.

Slater, M. and Usoh, M. (1993). Representations systems, perceptual position, and presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 2(3):221–233.

Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 3:130–144.

Slater, M. and Wilbur, S. (1997). 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.

Snow, M. P. and Williges, R. C. (1998). Empirical models based on free-modulus magnitude estimation of perceived presence in virtual environments. *Human Factors*, pages 386–402.

Spagnolli, A. and Gamberini, L. (2002). Immersion/emersion: Presence in hybrid environments. In *Proceedings of the 5th Annual International Conference on Presence*, pages 421–434, Porto, Portugal. Universitade Fernando Pessoa Press.

Spagnolli, A. and Gamberini, L. (2004). The sense of being 'there': a model for the space of presence. In *Proceedings of the 7th Annual International Worksho on Presence*, pages 48–53, Valencia, Spain. Editorial Universidad Politécnica de Valencia.

Steinicke, F., Bruder, G., Hinrichs, K., Steed, A., and Gerlach, A. L. (2009). Does a gradual transition to the virtual world increase presence? In *IEEE Virtual Reality (VR)*, volume 0, pages 203–210, Los Alamitos, CA, USA. IEEE Computer Society.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4):73–93.

Towell, J. and Towell, E. (1997). Presence in text-based networked virtual environments or "MUDS". *Presence: Teleoperators and Virtual Environments*, 6(5):590–595.

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

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

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14(5):779–804.

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

Welch, R. B., Blackmon, T. T., Liu, A., Mellers, B. A., and Stark, L. W. (1996). The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence: Teleoperators and Virtual Environments*, 5(3):263–273.

Wind, L. A., Van Dalen, J., Muijtjens, A. M. M., and Rethans, J.-J. (2004). Assessing simulated patients in an educational setting: the MaSP (Maastricht Assessment of Simulated Patients). *Medical Education*, 38(1):39–44.

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., and Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media Psychology*, 9(3):493–525.

Wirth, W., Wolf, S., Mögerle, U., and Böcking, S. (2004). Measuring the subjective experience of presence with think-aloud method: Theory, instruments, implications. In *PRESENCE 2004 - 7th Annual International Workshop on Presence*, pages 351–358.

Witmer, B. G., Jerome, C. J., and Singer, M. J. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators and Virtual Environments*, 14(3):298–312.

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

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

Zimmons, P. and Panter, A. (2003). The influence of rendering quality on presence and task performance in a virtual environment. *Virtual Reality*, 0:293.

Zimmons, P. M. (2004). *The influence of lighting quality on presence and task performance in virtual environments*. PhD thesis, The University of North Carolina at Chapel Hill. AAI3140421.