Joon Hao Chuah*
Andrew Robb
University of Florida
Gainesville, FL 32611

Casey White
University of Virginia
Charlottesville, VA 22903

Adam Wendling
Samsun Lampotang
University of Florida
Gainesville, FL 32611

Regis Kopper
Duke University
Durham, NC 27708

Benjamin Lok
University of Florida
Gainesville, FL 32611

Exploring Agent Physicality and Social Presence for Medical Team Training

Abstract

Mixed reality and 3D user interface technologies have increased the immersion, presence, and physicality of user interactions. These technologies can also increase the physicality of embodied conversational agents (ECAs) by making the ECAs occupy and interact with the physical space. We propose that increasing the physicality of an ECA can increase the ECA's social presence, that is, the feeling that the ECA is a real person. In this paper, we examine existing research and formalize the idea of ECA physicality. We also explored the relationship between physicality and social presence by conducting two user studies ($n = 18$ and $n = 29$). Both user studies took place in a medical team training context and involved virtual human ECAs as fellow team members. The first study's results suggested that increasing physicality increased social presence and elicited more realistic behavior. The second study's results suggested that individual dimensions of physicality affect social presence to different extents.

1 Introduction

Mixed reality (MR) and 3D user interface technologies, such as see-through head-mounted displays (HMDs) and motion-based controllers, have increased the physicality of user interactions (Izadi, 2012). These technologies can also increase immersion and presence, drawing the user into the environment, that is, the space that the user inhabits (Meehan, Insko, Whitton, & Brooks, 2002; Usoh et al., 1999). The user's space can also be inhabited by embodied conversational agents (ECAs). ECAs have a special relationship with human users because ECAs are social entities rather than just inanimate objects. As social entities, ECAs create social presence (the treatment of ECAs as real human beings; Blascovich, 2002), rather than presence (the sensation of being in a real place; Slater, 2009). Like users, ECAs can also leverage mixed reality technology to increase their physicality.

We use the concepts of immersion and presence in virtual environments to create an analogy with ECAs. We propose that the mixed reality ECA analogs of immersion and presence are physicality and social presence. In the context of mixed reality ECAs, we define physicality as the the objective level of fidelity of an ECA's occupancy of and interaction with the physical space. This definition of physicality derives from Slater's definition of immersion as "the

Presence, *Vol. 22, No. 2, Spring 2013,* 141–170
doi:10.1162/PRES_a_00145

---

*Correspondence to jchuah@cise.ufl.edu.

objective level of fidelity of the sensory stimuli produced by a technological system," summarized in McMahan, Gorton, Gresock, McConnell, and Bowman (2006, p. 108). Rather than studying the level of immersion of MR technology used for displaying ECAs, we propose to study the level of ECA physicality enabled by MR technology.

Physicality, as with immersion, can be objectively quantified and does not depend on the user's subjective feelings. Physicality, again as with immersion, also has many dimensions. Immersion's dimensions include field of view, texture resolution, and haptic feedback fidelity. Physicality's dimensions include, but are not limited to, an ECA's size, the form of the display, and the ECA's awareness of changes to the physical space. These physicality dimensions have been explored by several researchers (Dow, Mehta, Harmon, MacIntyre, & Mateas, 2007; Kotranza, Lok, Pugh, & Lind, 2009; Lee, Jung, J. Kim, & S. R. Kim, 2006; Johnsen & Beck, 2010; Rivera-Gutierrez et al., 2012). In Section 2 we formally describe the dimensions and categorize existing research in the context of physicality.

We believe that increasing physicality tends to increase social presence similarly to the tendency that increasing immersion leads to increased presence (Meehan et al., 2002). Social presence, like presence, is a subjective feeling and depends on the user. As presence refers to the user's sense of "being there," social presence relates to the user's sense of an ECA "being there." More formally, Blascovich defined social presence as "the extent to which individuals treat embodied agents as if they were other real human beings" as paraphrased in Bailenson et al. (2005, p. 380). Social presence is often measured using self-report surveys (Bailenson, Blascovich, Beall, & Loomis, 2003) or by observing users (Bailenson et al., 2004) and looking for realistic behavior.

To explore the relationship between physicality and social presence, we ran two user studies comparing ECAs with different levels of physicality. Both user studies took place in a medical team training context and involved virtual human ECAs. As an initial exploration into the relationship between physicality and social presence, the first study ($n = 18$) was a holistic evaluation. The holistic evaluation varied several dimensions of physicality at once and compared a pair of high-physicality ECAs to a pair of low-physicality ECAs. The holistic evaluation's results suggested that increasing physicality increased social presence and elicited more realistic behavior. The holistic evaluation also suggested that physicality may have an interaction effect with plausibility (behaving and reacting as a real human would).

In order to better understand the role of the specific dimensions of physicality in increasing social presence, we conducted an independent-components evaluation ($n = 29$). The independent-components evaluation individually varied two dimensions of physicality: form fidelity and concordance (see Section 2). The results from this independent-components evaluation suggested not only that individual dimensions of physicality affect social presence to different extents, but also, that matching levels of physicality are important for increasing social presence.

This paper is an extended version of our IEEE Virtual Reality Conference publication (Chuah et al., 2012), in which we initially proposed that physicality and social presence are the ECA analogs of immersion and presence; and we presented the results of the holistic evaluation. In this paper, we formalize the dimensions of ECA physicality, provide a more thorough examination of existing research into physicality, and present the results of the independent components evaluation.

The contributions of this paper are the following: it provides a formalization of the dimensions of physicality through a taxonomy; and it provides insight into the correlation of physicality with social presence through comparative evaluations in realistic settings.

1.1 Motivating Scenario: Medical Team Training

ECAs that elicit realistic behavior from users could increase the effectiveness of medical team-training simulations. Team-training is increasingly necessary for safe and effective health care. Many patient care errors can be attributed directly to the inability of clinicians to function as a team (Kohn, Corrigan, & Donaldson, 2000).

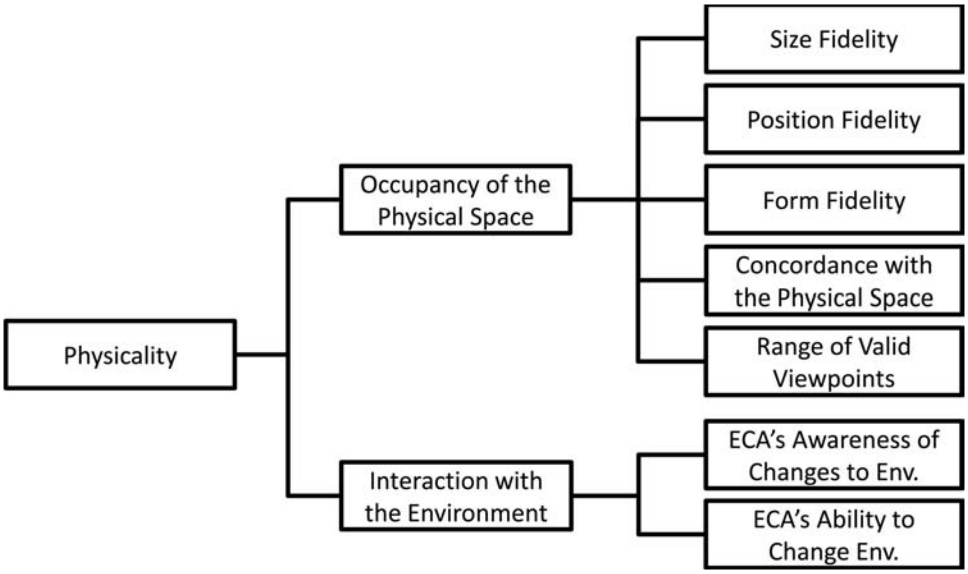


Figure 1. *A taxonomy for the dimensions of ECA physicality.*

Commonly reported errors can be tracked to communication failures, poor coordination, and fragmented care (Zwarenstein & Reeves, 2002). These problems result in medication errors, patient deaths, and many other patient safety issues (Reader, Flin, Mearns, & Cuthbertson, 2009).

Team-training simulations are used by many medical and health professional schools to prepare students to work on interdisciplinary teams after graduation (Wallin & Meurling, 2007). However, it is extremely difficult to bring all of the team members together for training post-graduation in busy health care systems, due to conflicting schedules and clinical demands (Rothschild & Lapidos, 2003). Without all of the team members, training becomes fragmented and even more artificial. These issues could be addressed by using ECAs as readily available and effective substitutes for team members, essentially allowing on-demand interdisciplinary team training. For ECAs to be effective substitutes, the ECAs would need to be considered to be socially present and elicit realistic behavior from the users.

In this paper, we offer the results of two formal evaluations in the context of medical team training. These evaluations compared ECAs of different levels of physicality and provided initial evidence to the requirements for ECAs that are socially present and that elicit realistic behavior.

2 Dimensions of Physicality

MR technology can increase the physicality of an ECA by blending real and virtual elements. Different MR technologies make various trade-offs with regard to things, such as the position of virtual elements, or the use of haptic feedback. To formalize these trade-offs, and their relationship to physicality, we propose a taxonomy of physicality dimensions (see Figure 1). The taxonomy divides physicality into two categories of dimensions: occupancy of the physical space and interaction with the environment. In this section, we detail the dimensions and review existing research that has explored the dimensions in different contexts. We note that the taxonomy we propose is not exhaustive and that other dimensions may exist; here we describe the dimensions we have gathered from personal experience and reviews of existing research. Some of these dimensions also exist in pure VR environments, and findings related to these dimensions may apply in both MR and in pure VR.

Figure 2. *High-physicality condition. The pair of ANDI units display the characters at life-size where they should stand, and use a combination of a see-through display and physical legs to create the sense that the ECAs share the physical space.*

Figure 3. *Low-physicality condition. The projection screen displays the characters at larger-than-life-size against a wall, and uses a virtual background to create the sense that the ECAs occupy a distinct virtual space.*

Note that visual realism and rendering quality are not directly dimensions of physicality. We acknowledge that these may affect social presence, but they generally do not affect the extent to which an ECA occupies the physical space or interacts with the environment. However, some aspects of visual realism and rendering quality can contribute to dimensions of physicality, and we explain this in the corresponding descriptions of dimensions in Section 2.1.4.

2.1 Occupancy of the Physical Space

Occupancy of the physical space describes how accurately an ECA appears to share the physical space with the user. We identify several dimensions within the occupancy of the physical space related to an ECA's size, position, form, and concordance with the physical space. Conforming to our definition of physicality, occupancy of the physical space can always be measured objectively, although this may be done at the perceptual level. For example, on a see-through HMD, an ECA may only be physically an inch tall, but the visual size of the ECA is objectively perceived as life-size, and thus it has high size fidelity (see Section 2.1.1).

2.1.1 Size Fidelity. Size fidelity refers to how closely an ECA conforms to life-size. Examples of high size fidelity ECAs include ECAs displayed at life-size on projection screens and see-through HMDs. Examples of low size fidelity ECAs include displaying ECAs at smaller than life-size on computer monitors.

Size fidelity was explored in a study by Johnsen and Beck (2010), who examined the effects of display size on user respect shown toward a virtual patient. The study compared a life-size virtual patient on a TV to a smaller-than-life-size virtual patient on a monitor. Their results suggested users showed more respect toward the high size fidelity ECA than the low size fidelity ECA.

2.1.2 Position Fidelity. Position fidelity refers to how closely an ECA's position conforms to its expected position in physical space. Position fidelity can be increased by using head-coupled perspective (also known as fish-tank VR; Ware, Arthur, & Booth, 1993) or stereoscopic displays to render ECAs as if they were closer to or farther from the display's physical position. The display's physical position can also be changed to increase position fidelity. Figures 2 and 3 illustrate high and low position fidelity, respectively. In Figure 2, the displays are placed in the physical environment, where real people would be standing. In Figure 3, the ECAs are displayed on the wall, farther back from where real people would be standing.

To the best of our knowledge, no researchers have looked at position fidelity with ECAs. However, interpersonal distance plays an important role in the level of intimacy of an interaction (Argyle, 1967). Bailenson et al. (2003) demonstrated that people generally respect an ECA's personal space and react to an ECA violating their own personal space. As a result, we believe that position fidelity may play an important role in an ECA's social presence.

2.1.3 Form Fidelity. Form fidelity describes how closely the physical form of the display conforms to the intended physical form of the ECA's body. Low form fidelity ECAs include ECAs displayed on monoscopic monitors, televisions, and projection screens; these displays and the images on them are all physically flat. Stereoscopic displays can increase form fidelity by making the image perceptually three-dimensional.

Physically three-dimensional displays that match the ECA's physical form further increase form fidelity by making the ECA solid in addition to being perceptually three-dimensional. For example, robots and mannequins can match the intended physical shape of the ECA. Spatial augmented reality (Bimber & Raskar, 2005) can also be used to display images on 3D surfaces that match the physical shape of the ECA (Lincoln et al., 2009).

High form fidelity ECAs have been created by researchers, and some have explored the effect of form fidelity on social presence. Lincoln et al. (2009) created shader lamp avatars, a display system that projects an avatar onto a styrofoam head. The styrofoam head is a 3D projection surface that allows multiple viewers to see a perspective-correct, 3D image without special glasses. Additionally, the styrofoam head is mounted on a motorized platform that allows it to turn and face people. At the ISMAR 2009 conference, Lincoln et al. added a mannequin body and demonstrated a full life-size avatar; however, they did not run any formal studies looking at social presence or perception of the ECA.

Rivera-Gutierrez et al. (2012) applied shader lamp avatars in a medical context, creating shader lamp virtual patients, which they used in a medical eye exam simulation. Medical students used a Wii controller as an ophthalmoscope and other exam tools. Although Rivera-Gutierrez et al. did not compare the shader lamp virtual patient to a low form fidelity version, they received feedback suggesting that the shader lamp virtual patient's physical presence increased the realism of the patient. This is consistent with our hypothesis that physicality (of which form fidelity is a component) is positively correlated with social presence.

Lee et al. (2006) explored the effect of both low and high form fidelity on social presence. The authors compared a physical Sony Aibo robotic dog to videos of the Aibo displayed on a monitor. While the Aibo is not a conversational agent, the Aibo is a social robot, and as such, it shares a similar relationship with users. Their results showed that people evaluated the high form fidelity agent more positively and that the high form-fidelity agent elicited a higher sense of social presence. These results suggest that form fidelity may be an important factor that affects social presence.

2.1.4 Concordance with the Physical Space. Concordance with the physical space describes how closely an ECA's environment is associated with the user's physical environment. Low-concordance ECAs occupy distinct virtual environments. Examples of low-concordance ECAs include typical video-game characters that occupy a purely virtual environment that is completely dissociated from the user's physical environment. High-concordance ECAs leverage MR technology to blend the ECA with the user's physical environment. For example, a see-through HMD can overlay an ECA on the user's physical environment.

High concordance can also be achieved by using spatial augmented reality (Bimber & Raskar, 2005). Spatial augmented reality techniques were used in Pair et al.'s (2003) FlatWorld project, in which displays were placed within physical doors. When users opened the doors, the displays showed life-size ECAs that appeared to be standing on the other side of the doorway.

Concordance can be increased by using rendering techniques that incorporate the physical environment's lighting. Both physical and virtual objects can cast correct shadows onto each other. Techniques that capture the physical lighting, including diffuse reflections off objects, can blend the physical and virtual elements better.

Dow et al. (2007) explored concordance with the physical space in a study comparing an AR version of the game Facade to the desktop version of the game. ARFacade used a see-through HMD to overlay the game's virtual characters on a physical environment. The results showed that immersive AR increased presence and engagement. Although Dow et al.'s study did not measure social presence, we believe that there are situations in which it is important to focus on the ECAs and to study social presence.

2.1.5 Range of Valid Viewpoints. The range of valid viewpoints is the analog of the so-called field of regard in virtual reality. The range of valid viewpoints captures the range of physical positions from which the ECA can be correctly viewed. An ECA on a flat display with no head tracking can only be viewed correctly from one physical position, the one corresponding to the static virtual camera. Head-coupled perspective increases the range of valid viewpoints by moving the virtual camera to match the user's physical position. A physically three-dimensional display may enable a full 360° range of valid viewpoints. Note that stereoscopic displays without head tracking still afford only one valid viewpoint, because any movement of the user would cause a perspective mismatch.

Increasing the range of valid viewpoints increases the occupancy of the physical space, because it allows the user to perceive the ECA as a 3D entity that can be viewed from multiple angles.

To the best of our knowledge, no researchers have examined the effects of the range of valid viewpoints on the perception of ECAs.

2.2 Interaction with the Environment

We identify two dimensions to the interaction with the environment category: (1) the ECA's awareness of changes to the environment, and (2) the ECA's ability to change the environment. We consider the environment to include the physical and virtual spaces and anything contained in these spaces, including entities such as objects, other ECAs, and users. To classify entities, we leverage Milgram and Kishino's (1994) virtuality continuum and apply it to individual entities. Entities may be purely virtual, purely physical, or a mix of virtual and physical.

Note that we consider only where the entities fall on the virtuality continuum; we do not consider the user's interface for manipulating entities. Technologies such as the Nintendo Wii remote, PlayStation Move, and Microsoft Kinect can increase the physicality of the user's interaction with the objects, but do not affect the physicality of the ECA's interaction with the objects.

Interaction with the environment only includes physical changes to the environment, that is, changes to properties such as position and orientation. Nonphysical interactions such as conversation are not included as dimensions of physicality. While conversation is clearly an important part of an ECA and may also affect social presence, conversation is beyond the scope of this work.

We draw on research into the physicality of user input interfaces to form hypotheses about ECA's interaction with the environment physicality. Usoh et al. (1999) found a positive relationship between user interface physicality and presence. Their research compared real walking, walking in place, and flying as travel methods in a virtual environment. They found both walking methods had statistically higher presence than flying. They also found that real walking had higher presence than walking in place, although the difference was not always statistically significant. Similarly, we expect that increasing the physicality of an ECA's interaction with the environment may increase social presence.

2.2.1 Awareness of Changes. Awareness of changes refers to an ECA's ability to sense and respond to changes in another entity's position, orientation, or any other aspect of physical state. Low-awareness ECAs are only aware of changes to purely virtual entities. High-awareness ECAs are aware of changes to at least some purely physical entities. Purely physical entities include the user and tracked physical objects. Research shows awareness of changes to the environment can increase social presence and enrich communication.

Creating ECAs with high awareness has historically had many barriers, which include setting up and calibrating expensive tracking systems, attaching tracking fiducials to objects and users, and instrumenting objects with sensors. However, recent advances in tracking technology have lowered the entry barriers. For example, the Microsoft Kinect can track not only the user's physical position but also the position of various body parts without the aid of tracking fiducials. The Kinect can be used with gesture recognition libraries such as FAAST (Suma, Lange, Rizzo, Krum, & Bolas, 2011). Gesture recognition could facilitate additional nonverbal communication in social interactions with ECAs. For example, Fujie, Ejiri, Nakajima, Matsusaka, and Kobayashi (2004) created a conversational robot that could use head gestures (e.g., nodding) to supplement verbal communication and correctly interpret ambiguous verbal input.

Other sensors, such as accelerometers, gyroscopes, and cameras, have become cheap and commonplace. This has enabled the spread of motion-sensing devices such as the Nintendo Wii remote, PlayStation Move, and most smartphones. These devices provide new methods for users to manipulate objects. Because of these newly lowered entry barriers, we believe many future ECAs will be aware of changes to the users and objects and it is important to study the effects of awareness on social presence.

Existing research shows that awareness of the user's physical position can increase an ECA's social presence. Bailenson, Blascovich, Beall, and Loomis (2001) compared virtual avatars that maintained eye contact with users (demonstrating awareness of the user's position) to virtual avatars that looked straight ahead. They found that when avatars maintained eye contact, users maintained a larger interpersonal distance (a positive measure of social presence related to not violating personal space).

Object manipulation can enrich ECA interactions and provide alternative communication channels. For example, Kotranza, Johnsen, et al. (2009) used a Nintendo Wii remote to manipulate various virtual tools in a virtual eye exam. They found that nonverbal interaction was effective, efficient, and satisfactory. Additionally, it mitigated dissatisfaction with errors in the voice-recognition system.

2.2.2 Ability to Manipulate. Ability to manipulate refers to an ECA's ability to "physically" change another entity's position, orientation, or any other aspect of state. By "physically," we mean any sort of physics-based contact between the ECA and another entity, including a purely virtual ECA touching a purely virtual object.

Because most ECAs exist as purely virtual entities, even within MR environments, most ECAs can only manipulate virtual objects or at most the virtual components of MR objects. Manipulating physical objects typically requires actuators or motors, which can add significant complexity. Cheap ECAs with manipulation abilities equal to a real human are still a long way off, but limited manipulation of physical objects is achievable.

For example, Kotranza, Lok, et al. (2009) created an ECA that could physically touch the user. The ECA was a mannequin equipped with a basic motorized arm (two servo motors replaced the original mannequin arm's joints). This ECA was used in an MR medical breast exam where an HMD overlaid a virtual patient on the mannequin. To study the effect of the physical touch, Kotranza et al. compared a purely virtual touch (where the motorized arm did not move) to a virtual and physical touch (where both the virtual and motorized arms moved). They found that the higher physicality touch communicated more effectively and simulated the patient better.

2.3 Intertwining of Technology and Dimensions of Physicality

Many dimensions of physicality are tightly coupled because the technoloiges that increase physicality in one dimension may simultaneously affect other dimensions. In some cases, this effect on other dimensions will be an increase while in other cases it will be a decrease. For example, a video see-through HMD increases concordance by blending the real and virtual. The HMD also

Table 1. *Examples of How Technologies May Affect More than One Dimension of Physicality. X indicates that a technology has an effect on a dimension. Awareness of changes to the physical space has been left off this chart because technologies affecting awareness typically do not affect other dimensions*

Technology	Size	Position	Form	Concordance	Viewpoints	Ability to Manipulate
See-through HMD[1]	x	x		x	x	x[2]
Head-coupled perspective		x			x	
Stereoscopic display[3]		x	x			
Spatial AR		x	x	x	x	

[1] With head position and orientation tracking

[2] Manipulation of virtual objects

[3] Without head tracking

supports rendering an ECA at life, size anywhere in the user's field-of-view and from any angle, hence providing high position fidelity, high size fidelity, and a wide range of valid viewpoints.

As another example, consider using a standard department-store mannequin as an ECA. The mannequin has high occupancy of the physical space, but the mannequin cannot move and is unable to manipulate either physical or virtual objects.

Table 1 lists examples of technologies and which dimensions of physicality they affect. This is by no means an exhaustive list of enabling technologies, and many of these technologies can be combined.

The intertwining of technologies and different dimensions of physicality can make it difficult to select the appropriate technology for an application. We believe this selection can be aided by identifying the dimensions of physicality and conducting research examining each dimension's effects on social presence.

3 Related Research

3.1 Social Presence

Social presence is a term shared by both the virtual-reality and computer-mediated-communication fields. There are many varying definitions of social presence, such as the definition of Slater, Sadagic, Usoh, and Schroeder, "the sense of being and acting with others in a virtual place" (Slater et al., 2000, p. 138) and Nowak's definition of "the perceived ability of the medium to connect people" (Nowak & Biocca, 2003, p. 486). Both of these definitions focus on connections to other people, and the people may be images or avatars of real people. Because this paper focuses on ECAs, rather than the avatars of real people, we will use Blascovich's definition of social presence as "the extent to which individuals treat embodied agents as if they were other real human beings" (as paraphrased in Bailenson, 2005, p. 380) Blascovich's definition captures a key foundation for agent-based training applications, treating agents as if they were real.

In a review of social presence research, Bailenson et al. (2005) found that research could be divided into two groups—one focusing on perception of ECAs, and the other on social responses to ECAs. Perception is typically evaluated using self-report surveys, while social responses are evaluated using analysis of user behaviors, physiological data, and task performance.

Bailenson et al. (2003) used both self-report surveys and social responses in a study that examined the effects of gaze behavior and perceived agency on social presence. Gaze behavior determines how and where an ECA looks around its environment. Perceived agency is what the user thinks is controlling the ECA, a human or a computer program. Bailenson et al. found that mutual gaze behavior was correlated with increased self-reported social presence. They also found that with computer-controlled ECAs, mutual ECA gaze behavior was correlated with more users demonstrating realistic behavior. However, with human-controlled ECAs, users demonstrated realistic behavior regardless of whether the ECA demonstrated mutual gaze behavior or not. Informed by these results, our studies used ECAs with mutual gaze behavior that the users believed to be controlled by computer programs. This allowed us to study the effect of physicality on realistic behavior without the influence of perceived agency.

Bailenson et al. (2004) also compared self-report surveys with social response behavioral metrics. Behavioral metrics showed differences even when self-report surveys did not. As a result, they suggest using behavioral metrics to complement self-reports.

3.2 Visual Realism

Visual realism has made great advances in recent years as hardware continues to get faster and cheaper. Polygon counts have gotten higher, and virtual humans can be more photorealistic than in years past. However, much research suggests low levels of visual realism are sufficient to invoke high levels of presence and social presence. In an experiment without virtual humans, Zimmons and Panter (2003) examined different levels of texture and lighting quality and failed to find any differences in presence. In an experiment with virtual humans, Pan, Gillies, Barker, Clark, and Slater (2012) noted that participants responded quite realistically to the virtual woman in their experiment, despite both the virtual woman and the virtual environment not looking realistic.

Increasing levels of visual realism may also hurt rather than help. Extremely high levels of visual realism may cause virtual humans to enter the uncanny valley. The uncanny valley is a phenomenon where increasing realism increases positive responses up to a point; after that, increasing realism suddenly causes a feeling of revulsion (Mori, 1970).

Even without entering the uncanny valley, increasing visual realism may result in lower social presence. For example, Nowak and Biocca (2003) found that participants reported more social presence interacting with a less anthropomorphic image than a more anthropomorphic image.

3.3 Place Illusion and Plausibility Illusion

Slater (2009) proposed that realistic behavior in virtual environments requires both place illusion and plausibility illusion. Place illusion, commonly referred to as presence, is the "illusion of being in a place in spite of the sure knowledge that you are not there" (Slater, p. 3551). Place illusion relies on sensory perception. Plausibility illusion is the "illusion that what is apparently happening is really happening" (Slater, p. 3553). Plausibility illusion depends on credible interactions between participants and their environment. While Slater proposes that both place and plausibility illusion are necessary for participants to respond realistically within virtual environments, little research has explored plausibility illusion.

In the context of ECAs, plausibility may depend heavily on the ECA's ability to respond appropriately to what people say and what happens. If an ECA cannot answer a question or answers the question with a nonsensical answer, the ECA's plausibility is reduced. Such problems are possible and sometimes even frequent with artificial intelligence (AI) systems. Moreover, an ECA may need to respond to events in addition to the conversation. For example, in a medical simulation, a doctor ECA must respond to the deteriorating vital signs of a patient ECA. If an ECA ignores events that a human would not, the ECA might be considered less plausible.

An ECA's nonverbal behavior may also be an aspect of plausibility. Several experiments have found relationships between nonverbal behavior and social presence. Von der Pütten and Gratch (2010) explored whether higher behavioral realism (nodding in acknowledgment) or lower behavioral realism (no nodding) affected social presence. They concluded that the virtual human's behavioral realism was a good predictor of the amount of words spoken by the user, which is one behavioral metric of social presence. Similarly, Bailenson et al. (2001) compared virtual humans that maintained eye contact with a participant to those that did not. They found that female participants maintained more interpersonal distance to those that maintained eye contact. Maintaining interpersonal distance (i.e., not violating another's personal space bubble) is a behavioral measure of social presence.

In our evaluations, we did consider the effects of plausibility and how those effects may have influenced our results. However, a deep exploration of plausibility is beyond the scope of this work.

4 Holistic Evaluation: Exploring Occupancy as a Whole

The holistic evaluation explored whether occupancy of the physical space had an effect on social presence. As an initial exploratory study, the holistic evaluation varied several dimensions in the occupancy of the physical space category simultaneously. The holistic evaluation used a between-subjects design with two conditions, high physicality and low physicality. The holistic evaluation took place in the context of a medical team-training exercise. The team-training exercise involved an anesthesia provider (the participant) working with a nurse ECA to treat a patient. The exercise also required the participant to interact with another ECA, the patient's adult daughter.

4.1 Conditions

The two conditions varied the physicality of the nurse and daughter. The nurse and daughter always had matching physicality (i.e., high and high or low and low). Table 2 summarizes the physicality differences between the conditions. Note that while we use the adjectives high and low to describe the conditions, the conditions do not represent the highest and lowest possible physicalities. Instead, the conditions represent high and low physicality relative to each other.

Table 2. *Physicality Levels by Dimension for the Nurse and Daughter*

	High physicality	Low physicality
Size	Life-size	Near life-size
Position	Correct position	On wall
Concordance	See-through display	Virtual environment
Form	TV and physical legs	Projection screen

In both conditions, the patient was a mannequin patient simulator. A mannequin patient simulator is similar to a department store mannequin but consumes oxygen, produces carbon dioxide, and has a palpable pulse, among other features. These features allow unmodified medical equipment to be used directly with the simulator. Mannequin patient simulators are commonly used to teach physiology (Hyatt & Hurst, 2010) but have also been used in team-training exercises (Hunziker et al., 2011).

4.1.1 High-Physicality Nurse and Daughter.

The high-physicality condition was designed to create a sense that the virtual humans shared the physical space with the user rather than inhabiting a separate virtual space (Figure 2). To achieve high physicality at a reasonable cost, we developed a hybrid virtual-physical virtual human system called ANDI (animatronic digital avatar; Chuah & Lok, 2012). ANDI combines MR technology with physical animatronic props to create a high-physicality virtual human that occupies the physical space. ANDI is not the highest physicality ECA possible, but rather, a balance between cost, high physicality, and flexibility. ANDI uses entirely off-the-shelf components and costs approximately $3,000, including a laptop. Higher-physicality ECAs such as the fully-articulated RoboThespian robot cost approximately $80,000 ("RoboThespian can't," 2011). Apart from having lower cost than robots, ANDI is also more flexible and can display a wide variety of genders, races, and appearances with only minimal physical modifications.

ANDI achieves high size fidelity by displaying the virtual human's upper body at life-size on a 40-in, 1,920 × 1,080 LCD TV in portrait orientation. The TV is mounted on a wheeled cart that allows ANDI to be positioned as appropriate in the physical space, achieving high position fidelity.

To achieve high concordance with the physical space, ANDI creates the illusion of a see-through display. The see-through display combines a Microsoft Kinect tracking system, head-coupled perspective, and a panoramic photo of the real environment. The panoramic photo captures the environment behind ANDI and is taken from the user's expected position in the environment. The photo becomes the texture on a virtual portion of a cylinder placed behind the virtual human. When the cylinder is rendered using the head-coupled perspective, the photo is approximately registered with the real environment around the TV. This creates the illusion of a see-through display and places the virtual human in physical space.

Head-coupled perspective (also known as Fish-Tank VR; Ware et al., 1993) increased physicality by providing an illusion of depth on the 2D displays. Head-coupled perspective placed the virtual camera at the user's physical position and modified the projection matrix. This warped the image and allowed the user to see different parts of the ECA as he or she moved around the room. Stereoscopic 3D was not used, because the active LCD shutter glasses did not work with the TV in portrait orientation.

ANDI increases form fidelity by including physical, animatronic prop legs. The legs consist of a pair of pants filled with pillow stuffing. The pants also physically move when the ECA moves. When the ECA turns to administer drugs to the patient or talk to the other ECA, a stepper motor rotates the top of the pants. This rotation matches the physical leg movements to the virtual upper body's movements. The physical legs and virtual upper body were kept visually aligned by the head-coupled perspective.

4.1.2 Low-Physicality Nurse and Daughter.

The low-physicality condition was designed to create a sense that the virtual humans inhabited a virtual space, rather than the user's physical space (Figure 3). The low-physicality condition also attempted to use only equipment commonly found in medical team-training environments. To accomplish this, the low-physicality condition used a projection screen, a virtual background, and fixed perspective.

A single 1,024 × 768 front projection system displayed the two virtual humans, resulting in low form fidelity. The projector screen measured 50 in diagonally and was mounted to the wall 40 in above the ground and 50 in away from where a real nurse and daughter would have stood, resulting in low position fidelity. The patient and projection screen were parallel. The nurse and daughter virtual humans each occupied a 22 × 30 in area of the projection screen. This area was large enough to display the upper portion of the virtual human at slightly larger than life-size, resulting in slightly lower size fidelity than the high-physicality condition.

The virtual environment created low concordance with the physical space by using a background disjointed from the physical room. The background was a privacy curtain, a common feature in patient care units where the medical scenario typically occurs in real life. However, the physical room was not a patient care unit but rather a simulation room where scenarios are typically practiced. This placed the ECAs in a distinct virtual environment with no concordance with the user's physical environment.

Fixed perspective, rather than head-coupled perspective, further emphasized the low levels of all dimensions of physicality. For fixed perspective, the virtual camera remained at a fixed position that did not change as the user moved. Fixed perspective also did not modify the projection matrix. Head-coupled perspective would have created the illusion of volume behind the display, which would have increased physicality. Additionally, unlike the high-physicality condition, there were no physical elements that needed to be visually aligned with the virtual elements.

4.2 Wizard of Oz

The ECAs were controlled via a Wizard of Oz (woz) setup. In a woz setup, a concealed human operator, rather than an artificial intelligence system, controls the ECAs. Woz control has been used in many studies involving ECAs and natural-language-understanding systems (Dahlback, Jonsson, & Ahrenberg, 1993; Dow et al., 2007; Johnsen & Beck, 2010; Maulsby, Greenberg, & Mander, 1993). Woz was chosen to eliminate confounds that could be introduced by an artifical intelligence system. Artificial intelligence systems sometimes produce incorrect responses due to speech-to-text errors or natural-language-understanding errors. Woz also allowed for a more natural interaction, removing the need for users to wear a microphone.

Participants were told that the ECAs were controlled by a computer and not a human operator. This was done because Bailenson et al. (2003) found that agency (belief that the ECA is controlled by a human or computer) had an effect on social presence. Social presence was higher for avatars believed to be controlled by humans. Participant awareness of ECA control by humans could potentially have raised social presence enough to mask the effects of physicality.

The human operator worked from behind a screen, a common item in these environments. The operator was able to see the participants through a streaming camera feed and was able to hear the participants directly.

The operator was able to respond to participant speech within half a second in most cases. This speed was possible because the response could usually be determined before the participant finished speaking. In some cases, a delay of 2–3 s was introduced by the need for the operator to find the correct response to play. This delay was reduced by grouping responses by category, reducing the number of responses needing to be searched by the operator. The operator chose from a total of 121 possible responses for the nurse, 25 for the patient, and 36 for the daughter. The operator typically issued the correct response and issued at most one incorrect, nonsensical response per participant.

The operator was familiar with the scenario and the set of responses. However, the operator was not a medical professional and did not know the answer to every question. In these cases, the characters responded with “I’m not sure” or “I don’t know.” This happened on average four times per participant. In some cases, the operator knew the answer, but there were no audio files for these answers. In these situations, the character did not respond to the question. During the introduction, users were instructed to continue with another question if this occurred. This occurred on average once per three participants.

Some woz systems, such as the Suede system by Klemmer et al. (2000), purposely introduce artificial errors. These artificial errors simulate the errors an autonomous, non-woz system might produce. We chose not to introduce artifical errors because we could not predict how many nonartificial errors a particular participant would encounter. Introducing artificial errors would raise the error rate, but would not gurantee a consistent error rate between participants.

4.3 Participants

The user population for the holistic evaluation consisted of medical professionals specializing in anesthesia at UF&Shands Hospital in Gainesville, Florida. These professionals possessed the knowledge and experience necessary to complete the critical incident scenario (see Section 4.4.2).

One faculty member of the Department of Anesthesiology recruited participants over a period of two weeks. Recruited participants were asked if they were willing to assist with a team-training exercise. If so, they were temporarily relieved of their clinical duties to participate in the exercise.

In total, 23 medical professionals participated in the study. Three participants were removed due to technical difficulties. Two additional participants were removed because they were outliers and not representative of the population. At least 47% of their self-report social presence ratings were beyond 2 *SD* of the average, and they repeatedly expressed a dislike for simulation exercises. Removing these participants resulted in a final *n* of 18. Thirteen were residents (doctors who graduated from medical school and are in their second or third year of specialized training), one was an intern (a doctor who has graduated from medical school and is in the first year of specialized training), two were attendings (doctors who have completed their specialized training), and two were certified registered nurse anesthetists (advanced trained nurses specializing in anesthesia). All participants were expected to have encountered real scenarios in their professional careers similar to the simulated one.

Residents were the target of the team-training exercise, and as a result made up the majority of the participants. Participants ranged from 28 to 40 years old, with an average age of 33. Sixteen participants were

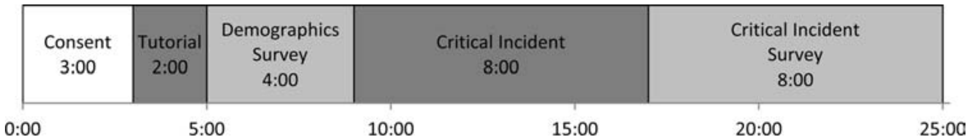


Figure 4. *Approximate time spent in each stage of the holistic evaluation. Total time was approximately 25 min.*

male, and all but one participant spoke English as a primary language. Participants were alternately assigned to the high- and low-physicality conditions. Excluding participants removed due to technical difficulties and outliers, the high-physicality condition had eight participants and the low-physicality condition had 10 participants.

4.4 Procedure

Both groups followed identical study procedures (see Figure 4). The stages were as follows.

1. **Consent.** Completed an informed consent form and were briefed about the goals and scope of the training exercise.
2. **Tutorial.** Virtual nurse explains how to interact with the virtual humans.
3. **Demographics Survey.** Completed a demographics and background survey.
4. **Critical Incident.** Virtual nurse answers questions about the patient, administers drugs when instructed, and suggests courses of action. Virtual family member interjects with questions about the patient's status.
5. **Critical Incident Survey.** Completed a feedback survey and free-form interview.

4.4.1 Tutorial. Pilot testing showed that many people were at first uncomfortable with the virtual humans. This led to difficulty communicating with them, confusion about the virtual human's capabilities, and, in the case of the high-physicality virtual human, surprise when the motorized legs first moved. To reduce this discomfort, the participants first went through a tutorial where they interacted with the virtual human nurse. The nurse prompted the participant to ask a question, then provided a response. The nurse also demonstrated his ability to move and administer drugs to the patient when instructed. After this, the nurse explained that he could not move physical objects and asked the participant to connect an oxygen mask to a nearby oxygen tank. This provided the participant with practice communicating and an understanding of the capabilities and limitations of the virtual human.

Familiarity with the medical equipment, including the mannequin patient simulator, was assumed. This exact mannequin was used by all residents in their prior year as interns, and all other participants had experience with similar mannequin patient simulators. The mannequin was connected to an actual anesthesia machine identical to the one participants used in their jobs.

4.4.2 Critical Incident. The proctor told participants they would be performing a team-training exercise involving virtual humans (no distinction was made between ECAs and virtual humans). He introduced the virtual humans as the nurse, the patient, and the patient's family member. The proctor gave no names to the virtual humans, but did explain that the nurse was the same nurse from the tutorial. Finally, the proctor explained the medical scenario and instructed participants that their task was to assess the situation and take an appropriate course of action.

The goals of medical team-training exercises such as this one are to practice how to:

- Work efficiently with team members.
- Identify a medical condition quickly.
- Treat the medical condition calmly and correctly.

This specific exercise focused on diagnosis and early management of perioperative myocardial infarction. This condition occurs in approximately 5% of patients undergoing major surgery and increases the relative risk of death by more than four times (Devereaux, 2008).

Due to the condition's critical nature, medical providers must recognize it and act quickly. Providers must direct a team of providers to treat the condition and prevent severe morbidity and possibly mortality.

During the exercise, the participant first found the patient displaying symptoms that could have indicated a variety of possible conditions. To diagnose the correct condition, the participant should have then gathered additional symptoms and history from the patient, nurse, and daughter. Next, the participant should have directed the nurse to perform specific tasks, such as ordering laboratory studies, and administering medications. If the participant did not perform certain necessary actions by preset times, the nurse suggested those actions. Providing these suggestions is one of the duties of a nurse in a real medical team.

While the participant was diagnosing and treating the patient, the daughter interjected with questions that occurred at preset times. For example, the daughter asked, "Why is nothing working?" Questions like this created empathetic opportunities where the participant could calm or comfort the daughter.

The exercise ended when one of two things occurred: (1) the participant took the necessary actions to treat the patient; or (2) the nurse suggested all the necessary actions and the participant chose not to follow the suggestions. In both cases, the nurse informed the participant that the exercise was complete.

4.5 Metrics

We used a combination of self-report surveys, observed behaviors in video recordings, and post-exercise interviews to assess social presence, realistic behavior, and plausibility.

4.5.1 Self-Report Social Presence. A five-question seven-point (-3 to 3) Likert-type survey assessed social presence for each ECA. This survey was taken from Bailenson et al. (2003) with minor modifications for clarity. The first occurrence of the word person in each question was replaced by the role of the ECA (nurse or family member). We abbreviated the questions for reporting in tables and figures as follows.

- **InRoom.** I perceive that I am in the presence of a nurse/family member in the room with me.
- **Watching.** I feel that the nurse/family member is watching me and is aware of my presence.
- **NotReal.** The thought that the nurse/family member was not a real person crossed my mind often.
- **Sentient.** The nurse/family member appeared to be sentient, conscious, and alive to me.
- **Computerized.** I perceived the nurse/family member as being only a computerized image, not as a real person.

4.5.2 Realistic Behavior. We assessed realistic and nonrealistic behavior toward the ECAs using video recordings of the critical incident exercise. Behavioral metrics were collected in addition to self-report social presence on the recommendation of Bailenson et al. (2004). Behavioral metrics could include gaze behavior, interpersonal distance, and conversational features. Of these, we used only conversational features because the exercise's focus on patient care potentially influenced gaze behavior and interpersonal distance. Specifically, we noted how the participants addressed the ECAs. The participants addressed the ECAs using words such as "ma'am" or "family member." To analyze how realistic the participants behaved, we classified words used to address the ECAs into three manners of address: nonrealistic, realistic but impersonal, and realistic and personal. The use of words was analyzed as a binary metric; a participant either used that word to address the ECA or did not. These data were analyzed using a likelihood ratio.

One possible behavioral metric is whether the participant chose to follow the nurse's advice or not. We chose not to use this metric for two reasons. First, some participants did not receive any advice because they ordered the treatments early enough. Second, our participants had a range of educational status and familiarity with this scenario.

4.5.3 Plausibility Illusion. We assessed plausibility illusion using qualitative post-exercise interview questions. We were unable to use the physiological metrics used by Slater (2009), because one of the goals of medical team-training is to reduce anxiety. Slater et al. studied plausibility illusion in virtual environments designed to generate anxiety. Low anxiety in our study could indicate failure to achieve plausibility illusion, but it could also indicate achieving plausibility illusion with an experienced clinician who is unstressed by the scenario.

During the post-exercise interview, we asked participants, "What did you think of the simulation?" If participants did not mention specific ECAs, we asked about those ECAs specifically (i.e., "What did you think of the nurse?"). These questions were left open-ended to get the participant's strongest impression without biasing them toward any one quality.

4.6 Results

Our results suggested that increasing physicality can increase social presence. However, self-report social presence results differed between the nurse and daughter ECAs. This difference may be explained by an interaction effect of physicality and plausibility.

The results for self-report social presence were analyzed using a nonparametric ANOVA-type statistic for each ECA with physicality as a between-subjects factor. Nonparametric analysis was chosen per the recommendation of Kaptein, Nass, and Markopoulos (2010) for Likert-type data with sample sizes less than 50. For between-subjects factors, the Box approximation was also used for small sample sizes per the recommendation of Brunner, Domhof, and Langer (2002), which results in a fractional denominator degrees of freedom (DOF). Analysis was done using R and the nparLD package.

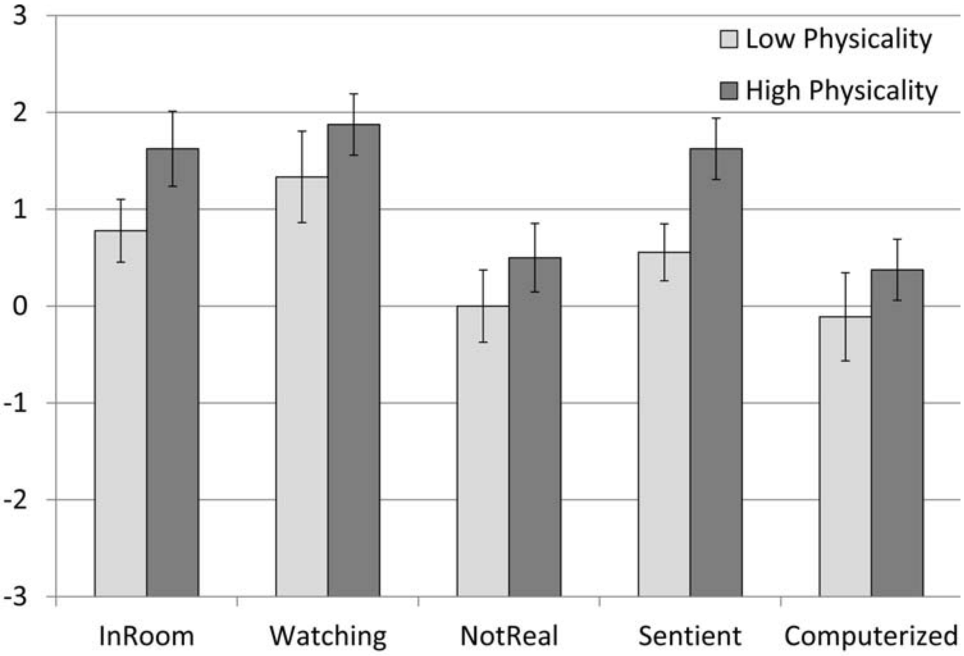
4.6.1 Self-Report Social Presence. The nurse ECA supported the hypothesis that increasing physicality increased social presence. The high-physicality condition had higher average self-report social presence scores for all five items (see Figure 5). The results show that participants perceived the nurse to be significantly more "sentient, conscious, and alive," $F(1, 14.674) = 8.312, p < .05$, high-physicality $M = 1.63, SE = 0.32$, low physicality $M = 0.56, SE = 0.29$. A marginally significant difference favoring the high-physicality condition was also found for the InRoom item, $F(1, 13.725) = 3.316, p = .09$, high physicality $M = 1.63, SE = 0.39$, low physicality $M = 0.78, SE = 0.32$. The other differences were not statistically significant, for which all $p > .37$.

The daughter ECA did not support the hypothesis that increasing physicality increased social presence. Neither condition had higher scores for all five self-report social presence items (see Figure 6). There were no significant differences between the two conditions for any of the five self-report social presence items, for which all $p > .18$.

4.6.2 Addressing ECAs. The results for manners of address suggest that increasing physicality results in more realistic behavior. Table 3 summarizes the words used to address the ECAs and how many participants used each word. Three out of 10 low-physicality condition participants addressed the daughter nonrealistically, addressing her by her role rather than by her name. For example one participant said "Family member, does he have any heart problems?" No high-physicality participants used this nonrealistic manner of address, a statistically significant difference, $p < .05$.

The use of realistic but impersonal manners of address with the nurse further supports the hypothesis that increasing physicality elicits more realistic behavior. Three out of 10 low-physicality condition participants addressed the nurse as "Nurse," while no high-physicality participants did so, $p < .05$. Participants in both conditions addressed the patient and daughter in a realistic but impersonal manner such as "Sir," $p = .81$, low = 70%, high = 75%, and "Ma'am," $p = .50$, low = 40%, high = 25%.

Participants in both conditions used realistic and personal manners of address with both the nurse and the patient. Participants addressed the nurse as "Matt," $p = .60$, low = 50%, high = 63%. The nurse introduced himself as "Matt" during the tutorial, and he also wore a name badge with his first name in large print. Furthermore, participants addressed the patient as "Mr.



	InRoom	Watching	NotReal	Sentient	Computerized
L	**0.78 ± 0.32**	1.33 ± 0.47	0.00 ± 0.37	**0.56 ± 0.29**	-0.11 ± 0.45
H	**1.63 ± 0.39**	1.88 ± 0.32	0.50 ± 0.35	**1.63 ± 0.32**	0.38 ± 0.32
p	**0.09**	0.46	0.37	**< 0.05**	0.51

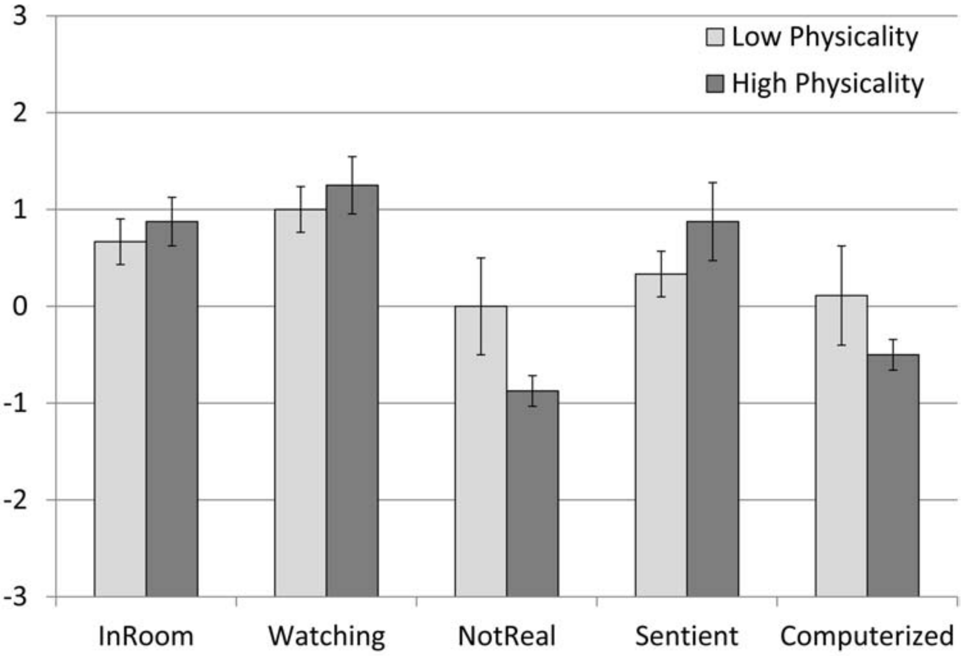
Figure 5. *Social presence (mean and standard error) for the low- and high-physicality nurse. Higher indicates better. Scores for negative statements (NotReal and Computerized) have been inverted for ease of comparison.*

Starks," $p = .50$, low = 40%, high = 25%. The patient was introduced as "Mr. Philip Starks" by the nurse at the beginning of the exercise. However, no participants addressed the daughter in a realistic and personal manner. Only one participant asked for her name, and none addressed her as "Ms. Starks" or any other variant of her name.

4.6.3 Plausibility Illusion. Plausibility illusion was achieved by the nurse ECA but not the daughter ECA. The nurse did have some minor problems; one participant expected the nurse to respond with more information. However, none found this distracting enough to consider the nurse not plausible. The nurse spoke frequently in response to the participant's questions and instructions. Further, he played the role of a clinician performing a job, and hence was not expected to be emotional.

During post-exercise interviews, participants commented on both the daughter's inclusion in the scenario and the lack of emotion in her response. Six participants (33%) found the daughter nonplausible. Family members are rarely present in post-anesthesia care units, where the scenario took place, and this diminished the daughter's plausibility. Furthermore, participants expected the daughter to be more emotional. They expected a real daughter to interject more often, express more emotion through facial expressions, and also to move closer to the patient to comfort him. The daughter only interjected between two and three times per interaction, did not express any emotion through facial expressions, and was not capable of moving closer to the patient. These factors combined to make the daughter not plausible.

Differences in social presence scores between the nurse and the daughter, regardless of condition, support



	InRoom	Watching	NotReal	Sentient	Computerized
L	0.67 ± 0.24	1.00 ± 0.24	0.00 ± 0.50	0.33 ± 0.24	0.11 ± 0.51
H	0.88 ± 0.25	1.25 ± 0.30	-0.88 ± 0.16	0.88 ± 0.40	-0.50 ± 0.16
p	0.53	0.50	0.20	0.18	0.38

Figure 6. *Social presence (mean and standard error) for the low- and high-physicality daughter. Higher indicates better. Scores for negative statements (NotReal and Computerized) have been inverted for ease of comparison.*

Table 3. *Words Used by the Participants to Address the ECAs and the Number of Participants Who Used Them in Each Condition. Note, one participant mistakenly thought the family member was the patient's wife, not daughter.*

Words	Category	Low physicality	High physicality
Family Member, Wife	Nonrealistic	3	0
Nurse	Realistic but impersonal	3	0
Ma'am	Realistic but impersonal	4	2
Sir	Realistic but impersonal	7	6
Matt	Realistic and personal	5	5
Mr. Starks	Realistic and personal	4	2

the daughter being less plausible. There were significant differences for NotReal, $p = .01$, where the nurse was $M = 0.65, SE = 0.31$ better, and Watching, $p = .04$, where the nurse was $M = 0.47, SE = 0.27$ better. There was a marginally significant difference for "Sentient," $p = .06$, where the nurse was $M = 0.47, SE = 0.27$ better.

4.6.4 Interaction of Physicality and Plausibility. An interaction effect between condition and

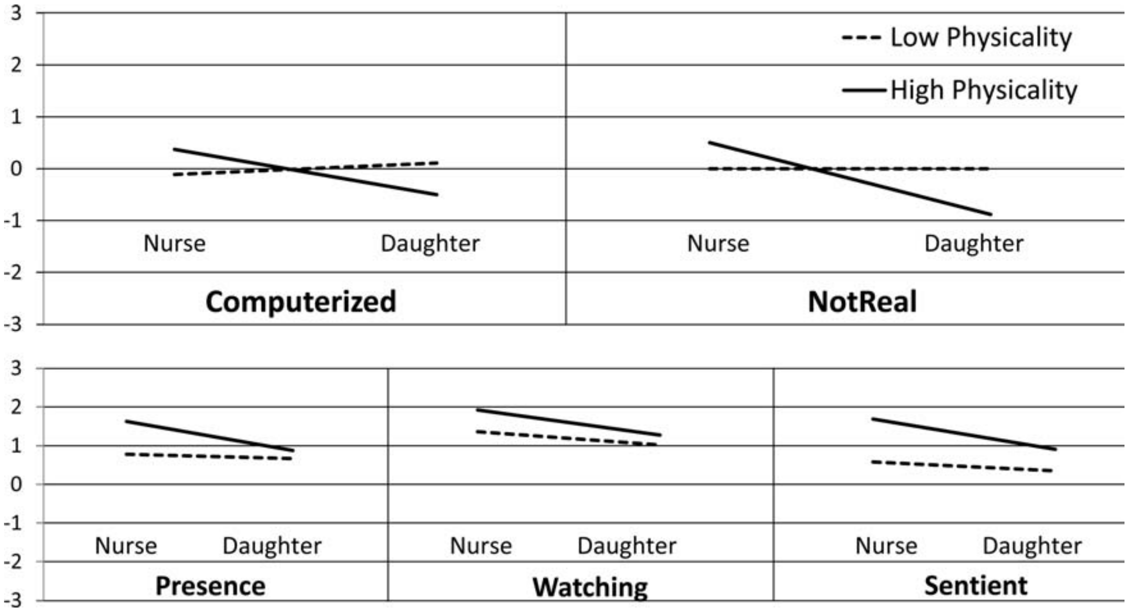


Figure 7. *Interaction effects of condition and ECA. Significant interaction effect for NotReal,* p = *.01, and marginally significant interaction effect for Computerized,* p = *.07. Higher indicates better. Scores for negative statements (NotReal and Computerized) have been inverted for ease of comparison.*

ECA would indicate that other factors besides physicality affected social presence. We used a nonparametric ANOVA-type statistic with physicality as the between-subjects factor and ECA as the within-subjects factor to determine whether there was an interaction effect between condition and ECA (see Figure 7). The Box approximation could not be applied to the within-subjects factor, so the denominator DOF is infinity. An interaction effect was found when comparing the nurse and daughter.

For all five self-report social presence items, the differences between the high-physicality nurse and daughter were larger than the differences between the low-physicality nurse and daughter. These differences were significant for NotReal, $F(1, \infty) = 7.192, p < .01$, and marginally significant for Computerized, $F(1, \infty) = 3.355, p = .07$. For NotReal, the high-physicality virtual humans had a difference of $M = -1.38, SE = 0.39$ (where the nurse was better), while the low-physicality virtual humans had a difference of $M = 0.00, SE = 0.33$. For Computerized, the high-physicality ECAs had a difference of $M = -0.88, SE = 0.19$ (where the nurse was better), while the low-physicality ECAs had a difference of $M = 0.22, SE = 0.36$ (where the daughter was better).

4.7 Discussion

4.7.1 Self-Report Social Presence. Bailenson et al. (2003) used a Cronbach's α to determine if the scores from all five statements could be summed into an overall social presence score. The Cronbach's α for the nurse of 0.84 indicated that we could do so. The social presence scores were $M = 6.00, SE = 1.29$ for high physicality and $M = 2.56, SE = 1.36$ for low physicality. This difference was not statistically significant, $p = .10$, but does agree with the trend of increasing physicality increasing social presence.

4.7.2 Addressing ECAs. In total, 50% of the low-physicality participants used either "family member" or "Nurse" to address one of the ECAs (one participant did both). None of the high-physicality participants did so. These results suggest that increasing physicality increases social presence by eliciting more realistic behavior. Addressing the daughter as "family member" was clearly unrealistic, and showed that the participants did not treat the low-physicality daughter as if she were a real person. Three low-physicality participants addressed the daughter nonrealistically, and strikingly, none of the high-physicality participants did so.

Unlike the case with the daughter, addressing the nurse by his role was realistic, but it was impersonal. This manner of address was acceptable behavior in such emergency situations with unfamiliar team members. However, doctors considered it more respectful to use a nurse's name or even to not use a word at all to address the nurse. This less-respectful manner of address may have been used because these participants did not feel the need to treat the low-physicality nurse realistically.

4.7.3 Plausibility Illusion. The nurse's social presence scores showed a trend that increasing physicality did positively affect social presence. All five social presence statements trended positively, with one statement being statistically significant and one statement being marginally significant.

The daughter's social presence scores showed that low plausibility limited social presence. There were no clear trends, indicating that either the low- or high-physicality condition affected social presence. This suggests that even though the high-physicality daughter had a higher potential social presence, the lack of plausibility limited social presence to a lower level.

This relationship between physicality and social presence supports the idea that they are the ECA analogs of immersion and presence. Slater believes that "immersion provides the boundaries within which PI [presence] can occur" (Slater, 2009, p. 3552). Similarly, physicality provides the boundaries of social presence, and to reach these boundaries, plausibility expectations must be met.

4.7.4 Interaction of Physicality and Plausibility. The interaction effect between physicality and ECA demonstrated that increasing physicality increased expectations. Thus, when the physicality was increased and the expectations were met (as is the case between the low-physicality nurse and the high-physicality nurse), the self-report social presence scores correspondingly increased. However, when the physicality was increased and the expectations were not met (as is the case between the low-physicality daughter and the high-physicality daughter), self-report social presence was not significantly raised.

The interaction effect is supported by Nowak and Biocca's (2003) findings that less anthropomorphic images were correlated with higher social presence. They believed that increasing anthropomorphism increased social presence expectations. Similarly, increasing physicality increased plausibility expectations.

4.8 Limitations

The holistic evaluation's results suggest that occupancy of the physical space affects social presence. However, the holistic evaluation varied several dimensions in the occupancy of the physical space category simultaneously. How each dimension or combination of dimensions affected social presence cannot be determined from the holistic evaluation. Moreover, the holistic evaluation varied display resolution and brightness, which are not aspects of physicality. To address these limitations, we conducted an independent-components-evaluation.

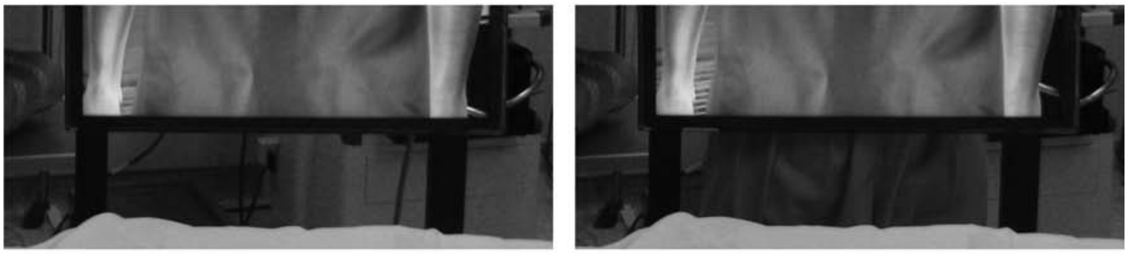
5 Independent-Components Evaluation: Exploring Individual Dimensions of Occupancy

The independent-components evaluation looked specifically at the effects of form fidelity and concordance with the physical space on social presence. The independent-components evaluation used a between-subjects, 2 × 2 design with high and low levels of form fidelity and concordance. All four conditions used the high-physicality ANDI units, but some of ANDI's high-physicality features were removed in some conditions. The four conditions were as follows.

- **LCLF.** Low concordance (plain, black virtual background) and low form fidelity (no physical legs; Figure 8[a]).

[a] LCLF - low concordance, low form fidelity [b] LCHF - low concordance, high form fidelity

[c] HCLF - high concordance, low form fidelity [d] HCHF - high concordance, high form fidelity



[e] Close-up of no physical legs in HCLF [f] Close-up of physical legs in HCHF

Figure 8. *The independent-components evaluation conditions. Conditions on the top use a plain, black virtual background for low concordance with the physical space. Conditions in the middle use a photograph of the real background for high concordance. Conditions on the left have only a TV for low form fidelity. Conditions on the right include physical legs for high form fidelity.*

- **LCHF.** Low concordance (plain, black virtual background) and high form fidelity (physical legs; Figure 8[b]).
- **HCLF.** High concordance (see-through display) and low form fidelity (no physical legs; Figure 8[c]).
- **HCHF.** High concordance (see-through display) and high form fidelity (physical legs; Figure 8[d]).

Table 4 summarizes the number of participants in each group.

Table 4. *Number of Participants in Each Condition for the Independent-Components Evaluation*

Form fidelity	Concordance with physical space	
	Low	High
Low	LCLF 8 participants	HCLF 7 participants
High	LCHF 9 participants	HCHF 8 participants

All four conditions maintained the same high level of size and position fidelity. The ECAs were life-size and placed in the correct physical places. Size and position fidelity were kept constant, because technologies for high concordance, such as see-through displays, typically also inherently enable high size and position fidelity.

The independent-components evaluation used a similar scenario to the holistic evaluation but with modifications to address the limited pool of participants and the low-plausibility daughter. To increase the pool of participants, the independent-components evaluation changed which role was played by the participant. Instead of real doctors asking a virtual nurse questions, the independent-components evaluation used real nurses asking a virtual doctor questions. The pool of available real nurses was roughly twice as large as the pool of available participants for the holistic evaluation.

The low-plausibility daughter was excluded for the independent-components evaluation. Instead, the independent-components evaluation included an additional virtual nurse. The virtual nurse aided the participant (also a nurse) by administering drugs to the patient when requested to do so. Delegating drug administration to a virtual nurse allowed the participants to focus on the communication aspects without worrying about physical tasks. This delegation is also common in real situations.

As with the holistic evaluation, all of the independent-components evaluation's ECAs were controlled by a human operator behind the scenes. The operator practiced controlling the ECAs as part of a pilot study involving 10 real nurses. The pilot study also aided in expanding the set of anticipated questions, so that more preprogrammed responses could be added to the ECAs.

5.1 Participants

The population for the independent-components evaluation consisted of nurses who worked in the post-anesthesia care unit at UF&Shands Hospital in Gainesville, Florida. These nurses all had the potential to encounter cases similar to the simulated critical incident in their jobs, though not all nurses had done so. Nurses were recruited to participate in the study as part of a training exercise. Participation was voluntary, and participants were not offered any compensation.

In total, 32 nurses participated in the independent-components evaluation. Participants had practiced nursing for at least 4 years and had an average age of 45.29, $SD = 10.57$. All but one participant was female, and all participants spoke English as their primary language.

5.2 Procedure

The independent-components evaluation followed a similar procedure to the holistic evaluation. Because the independent-components evaluation required the participant to answer questions about the patient, the independent-components evaluation included an extra stage where a virtual doctor provides the nurse (participant) with information about the patient.

All four groups followed identical study procedures (see Figure 9). The stages were as follows.

1. **Consent.** Completed an informed consent form and were briefed about the goals and scope of the training exercise.
2. **Tutorial.** Virtual tutorial doctor and virtual tutorial nurse explain how to talk with the virtual humans and have the participant practice portions of the later stages.
3. **Demographics Survey.** Completed a demographics and background survey.
4. **Handoff.** Virtual handoff doctor provides patient information to the participant. The virtual nurse's

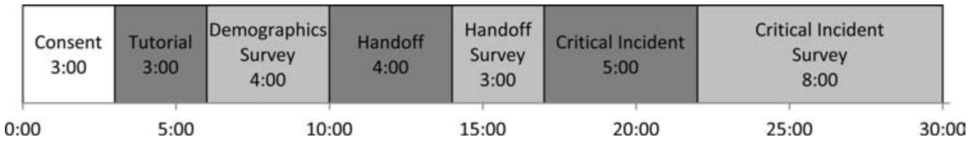


Figure 9. *Approximate time spent in each stage of the independent-components evaluation. Total time was approximately 30 min.*

display was turned off, and the physical legs (if any) were hidden.

5. **Handoff Survey.** Completed a feedback survey about the handoff doctor.
6. **Critical Incident.** Virtual critical incident doctor asks the participant questions about the patient and gives treatment orders. Participant delegates drug orders to the virtual critical incident nurse.
7. **Critical Incident Survey.** Completed a feedback survey and free-form interview.

5.2.1 Handoff. During the handoff, the virtual handoff doctor provided the participant with information about the procedures done to the patient and the patient's medical history. The participant recorded this information on a paper form. If the participant felt that any information or detail was missing, the participant could ask the virtual handoff doctor for more information.

5.2.2 Critical Incident. In the critical incident, the participant worked with the virtual critical incident doctor (a different ECA from the virtual handoff doctor) and the virtual critical incident nurse to treat the patient. The virtual doctor asked the participant questions about the patient. The participant answered the questions and typically referred to the notes taken during the handoff stage.

As the virtual doctor asked questions, the patient's vital signs degraded. When the vital signs degraded to certain specified points, the virtual doctor ordered drugs to treat the patient. The participant then delegated the drug orders to the virtual charge nurse. The virtual charge nurse administered the drugs and verbally confirmed the order by repeating it. This delegation of orders is common in real medical scenarios.

5.3 Metrics

The independent-components evaluation used the same five self-report social presence items as the holistic evaluation (InRoom, Watching, NotReal, Computerized, and Sentient). As with the holistic evaluation, the data were analyzed using a two-way nonparametric ANOVA-type statistic with concordance and form fidelity as between-subjects factors, and ECA as a within-subjects factor. Nonparametric analysis was again chosen per the recommendation of Kaptein et al. (2010) for Likert-type data with sample sizes less than 50. For between-subjects factors, the Box approximation was also used for small sample sizes per the recommendation of Brunner et al. (2002), which results in a fractional denominator of DOF. To the best of our knowledge, there are no post hoc tests available for the two-way nonparametric ANOVA-type statistic we used to analyze our data. As an approximate substitute, we used a pair-wise one-way ANOVA-type statistic with condition as the between-subjects factor.

If no interaction effect was found for either of the between-subjects factors, the data were reanalyzed using two separate one-way nonparametric ANOVA-type statistics with each between-subjects factor. R and the nparLD package were used to analyze all self-report social presence data.

The independent-components evaluation also used perceived intelligence of the ECAs as a proxy for plausibility. Participants rated each ECA's intelligence on a scale of 1–10 (low-to-high) based on the ECA's ability to achieve the team-training goals. We did not expect any plausibility differences between conditions, but we did expect the ECAs to all have high plausibility. We analyzed perceived intelligence using a repeated-measures ANOVA. If the assumption of sphericity was violated, we used the Greenhouse–Geisser adjustment.

5.4 Results

For all five social presence items, the LCLF condition consistently received worse average ratings than the HCHF condition for both critical incident virtual humans. Comparing all four conditions showed two different trends among the items. Two items were affected by both form fidelity and concordance, while three items were affected by only concordance.

5.4.1 Outliers. Three participants had particularly bad experiences with the training system. These bad experiences had two causes. The first cause was participants asking the ECAs a large number of unanticipated questions. Unanticipated questions resulted in the human operator choosing responses such as "I don't know, you'll have to check his chart" or silence if no suitable prerecorded response was available. A competent real doctor or nurse would never respond in such a manner or completely fail to respond. These failures diminished the plausibility of the ECAs and were reflected in the perceived ECA intelligence ratings. Based on the perceived ECA intelligence ratings, we identified two statistical outliers. Both participants rated the intelligence of the critical incident nurse as a 1 (the lowest on a scale of 1–10). One participant rated the intelligence of the critical incident doctor as a 1 while the other rated it as a 3. The critical incident nurse and doctor as received ratings of $M = 7.16, SD = 2.07$ and $M = 7.45, SD = 1.91$, respectively.

The second cause of bad experiences was extreme discomfort with computer-based simulation. While many participants were at first mildly uncomfortable with the ECAs, only one participant showed extreme discomfort. She repeatedly stated that the exercise made her very nervous and she was not a good candidate for computer-based simulation. Further, she responded "Strongly disagree" to all positive statements and "Strongly agree" to all negative statements for all three ECAs. Only one other participant responded similarly, and this other participant was already removed as an outlier due to perceived ECA intelligence ratings. These extreme responses emphasize that unlike the majority of the population, this outlier was unable to suspend disbelief.

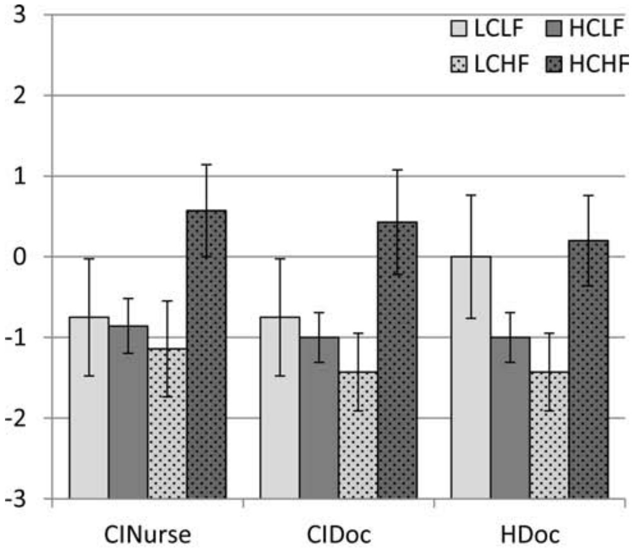
After removing outliers, the LCLF condition had eight participants and the other three conditions had seven participants each.

5.4.2 Plausibility. The subjective intelligence of the ECA results suggested the two critical incident ECAs were highly plausible, but the handoff doctor ECA was significantly less plausible, according to a repeated-measures ANOVA, $F(1.286, 34.735) = 79.063, p < .001$. The critical incident nurse and doctor received similar intelligence ratings of $M = 7.16, SD = 2.07$ and $M = 7.45, SD = 1.91$; but the handoff doctor received a rating of $M = 4.52, SD = 1.61$. Notably, the handoff doctor's highest rating of 6 was lower than the average rating for the critical incident ECAs. A one-way ANOVA did not find any significant differences between conditions in intelligence ratings for the handoff doctor, $p = .82$, critical incident nurse, $p = .49$, or critical incident doctor, $p = .63$.

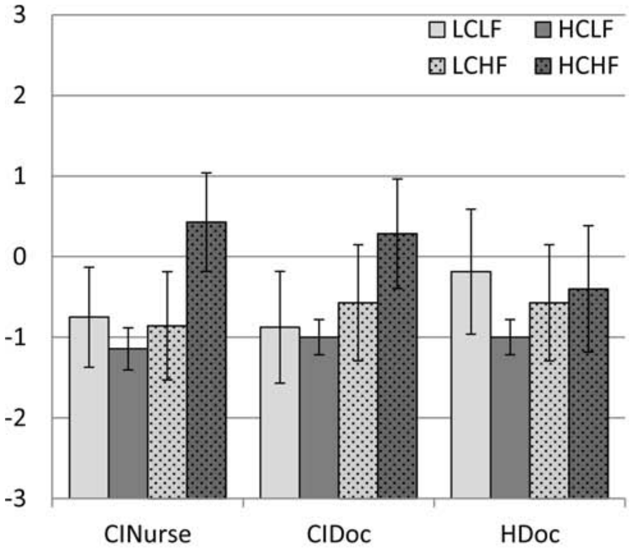
5.4.3 Form Fidelity and Concordance. The results suggest both high form fidelity and high concordance were necessary to improve ratings for the Computerized (Figure 10[a]) item. For this item, the HCHF condition received an average score between "Slightly disagree" and "Neutral" for both of the critical incident virtual humans. The other three conditions received an average score between "Neutral" and "Slightly agree" (the Computerized survey item was a negative statement, so, in this case, "disagree" is better).

Because the handoff doctor showed lower plausibility and plausibility may have an interaction effect with physicality, we included only the more plausible critical incident ECAs in the analysis. However, we do report the results for the handoff doctor ECA for discussion purposes.

For Computerized, the results showed a marginally significant interaction effect between concordance and form fidelity, $F(1, 20.655) = 3.674, p = .07$, but no significant effects of concordance, $F(1, 20.655) = 2.603, p = .12$, form, $F(1, 20.655) = 0.451, p = .51$, or ECA, $F(1, \infty) = 0.386, p = .53$. Pair-wise analysis found the HCHF condition was significantly better than the HCLF condition, $F(1, \infty) = 4.581, p < .05$, and



[a] Computerized



[b] NotReal

Figure 10. *Self-report social presence results (mean and standard error) for NotReal and Computerized for the independent-components evaluation. Light and dark gray backgrounds indicate low and high concordance, respectively. Solid and dotted fills indicate low and high form fidelity, respectively. Scores for negative statements have been inverted for ease of comparison.*

LCHF, $F(1, \infty) = 7.559, p < .01$, but not the LCLF condition, $F(1, \infty) = 1.827, p = .18$. There were no other significant differences between groups, for which all $p > .43$.

For NotReal (Figure 10[b]), the results showed a similar trend to Computerized. However, the two-way ANOVA-type statistic failed to find significant interaction effects between form fidelity and concordance, $F(1, 20.211) = 1.393, p = .25$. No interaction effects were found for InRoom, $F(1, 22.092) = 0.000, p = .98$, Watching, $F(1, 22.181) = 0.564, p = .46$, and Sentient, $F(1, 23.863) = 0.020, p = .89$.

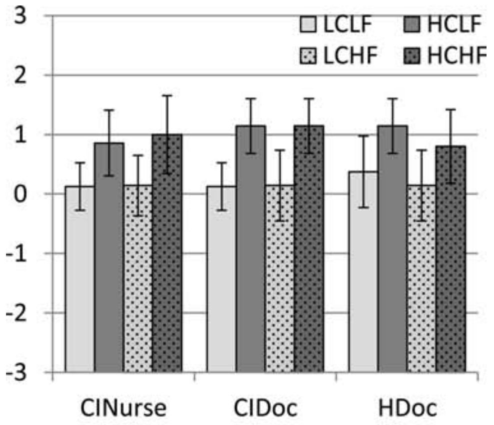
5.4.4 Concordance. The results suggest that concordance, regardless of form fidelity, affected ratings for the InRoom (Figure 11[a]), Watching (Figure 11[b]), and Sentient (Figure 11[c]) items. For all three items, both high concordance conditions tended to receive an average rating of "Slightly Agree" while both low concordance conditions received an average rating of "Neutral" for both critical incident virtual humans. However, the handoff doctor did not show the same trend for the Watching and Sentient items.

Considering concordance as the only between-subjects factor, the results showed marginally significant effects for InRoom, $F(1, 25.185) = 4.209, p = .05$, Watching, $F(1, 24.538) = 3.080, p = 0.09$, and Sentient, $F(1, 25.952) = 3.195, p = .09$. No significant effects were found for NotReal, $F(1, 25.511) = 0.471$, $p = .50$.

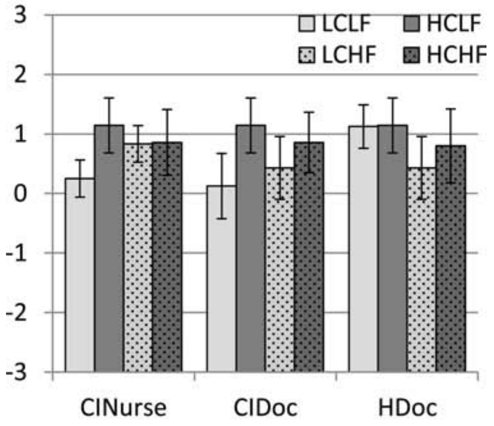
5.4.5 Form Fidelity. Considering form fidelity as the only between-subjects factor, the results showed no significant effects for NotReal, $F(1, 25.164) = 1.185, p = .29$, InRoom, $F(1, 25.703) = 0.055, p = .82$, Watching, $F(1, 26.930) = 0.041, p = .841$, and Sentient, $F(1, 26.949) = 0.036, p = .85$.

5.5 Discussion

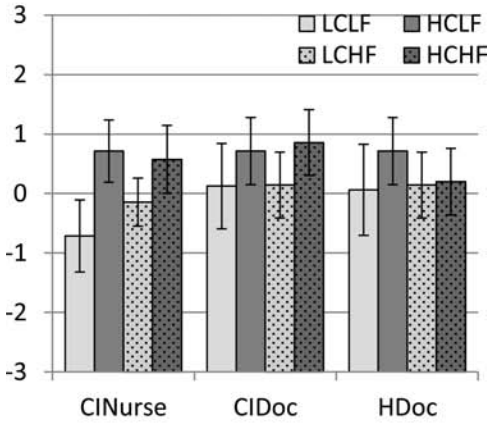
5.5.1 Form Fidelity and Concordance. The combination of high form fidelity and high concordance appears to increase the sense of the virtual human as a real person. The Computerized item read "I perceived the nurse as being only a computerized image, not as a real person." The phrase "real person" also appeared in the NotReal item as "The thought that the nurse was not a real person crossed my mind often." Although the



[a] InRoom



[b] Watching



[c] Sentient

Figure 11. *Self-report social presence results (mean and standard error) for the independent-components evaluation. Light and dark gray backgrounds indicate low and high concordance, respectively. Solid and dotted fills indicate low and high form fidelity, respectively.*

NotReal item did not show significant interaction effects between form fidelity and concordance, it did show a similar trend to Computerized.

The effect of form fidelity in conjunction with high concordance was apparent, despite the fact that the physical legs occupied a very small portion of the participant's visual field. The participant's view of the legs was blocked by the patient simulator and bed. Only a few vertical inches of the physical legs remained visible. This suggests that even giving a small portion of an ECA high form fidelity can substantially improve feelings of the ECA as a real person.

The handoff doctor's lower plausibility may explain inconsistencies in the trends. The two critical incident ECAs received similar self-report social presence scores to each other, but the handoff doctor sometimes received different scores. For the Computerized and NotReal items (Figures 10[a] and 10[b]), the scores for the handoff doctor show a different trend from the critical incident ECAs. In the LCLF condition, the handoff doctor received better scores than the critical incident ECAs. This may be because in the LCLF condition, participants had lower initial expectations of realism, and the handoff doctor's plausibility matched these expectations. The critical incident ECAs had better plausibility, and as a result, the participants' expectations of realism increased; but the physicality did not meet their expectations.

5.5.2 Concordance.

The results for the InRoom, Watching, and Sentient items show the importance of concordance. Although the differences for all three items were only marginally significant, together, they establish a consistent trend. Increasing concordance by adding a photograph of the physical environment increased the rating by a full point on a 7-point Likert scale for InRoom (Figure 11[a]) and Sentient (Figure 11[b]). This full point also made the difference between a neutral rating and a positive rating.

The results for the InRoom item specifically suggest that the see-through display does noticeably affect concordance. The InRoom item states "I perceive that I am in the presence of a nurse/doctor in the room with me." Both of the high-concordance conditions received average ratings on the positive side of the scale, while both low-concordance conditions received average ratings of neutral.

The see-through display made a significant difference, despite the illusion being less than perfect. The background photo sometimes is not registered correctly with the surrounding real background, because the photo lacks correct depth information. However, it had a very low cost and still made substantial improvements.

The results for the InRoom, Watching, and Sentient items were consistent between the two critical incident ECAs but not consistent with the handoff ECA. This may be because during the handoff, the participant was taking notes and was focused on her clipboard, rather than on the virtual human.

5.6 Form Fidelity

The independent-components evaluation failed to find a significant effect for form fidelity. This is inconsistent with Lee et al.'s (2006) findings of form fidelity affecting social presence. This may have been caused by the interaction effect between form fidelity and concordance. The LCHF condition likely did not receive the benefits of form fidelity because the concordance of the ECA's upper and lower halves conflicted. The upper half was displayed against a plain black background and had low concordance with the physical space. The lower half, however, had high concordance, because the physical legs had the real background and hence naturally had high concordance with the physical space. The high form fidelity may have made the virtual human more real, had it not been for the mismatched levels of concordance with the physical space.

5.6.1 Plausibility. The lower intelligence ratings for the handoff doctor may have been caused by the handoff doctor talking too fast, giving excessively long answers, or occasionally not knowing the answer to a question. Many participants asked the handoff doctor to slow down because they could not write notes fast enough. However, the handoff doctor ignored this request and continued speaking at a fixed rate, because all of the speeches were prerecorded.

As a result, many participants asked the handoff doctor to repeat pieces of information. The handoff doctor's reply often contained far more information than needed, because the requested information was at the end of a long prerecorded speech. For example, if the participant asked, “What drug did he have boluses of?” the handoff doctor might respond, “He had dexamethasone and ketamine this morning for analgesia and he has had intermittent boluses of fentanyl throughout the surgery.”

In some cases, participants asked the handoff doctor unanticipated questions. For example, if the participant asked, “Which antibiotic is he on?” then the doctor ECA answered, “You'll have to check his chart.” Most real doctors would know this information, but this question was unanticipated, and hence, it did not have a specific prerecorded answer.

5.7 Limitations

One limitation of the independent-components evaluation was that only a small portion of the physical legs were visible. The results suggest that participants at least subconsciously noticed the physical legs, even though only two participants commented on the physical legs. Also, the inclusion of physical legs changed more than just form fidelity. The physical legs also completed the ECA's body. Without legs, the ECAs in the low form fidelity conditions appeared as floating ECAs from the waist up. This floating is an alternative explanation for why the low form fidelity condition rating was worse than the HCHF condition. To test this alternate explanation, future studies comparing the physical legs with virtual legs on a second TV need to be conducted.

6 Overall Discussion

The holistic evaluation suggested that physicality has a positive relationship with social presence. This led us to delve deeper into specific dimensions of physicality and to conduct the independent-components evaluation. The independent-components evaluation repeated many of the results from the holistic evaluation.

The holistic evaluation found a significant difference for the Sentient item and a marginally significant difference for the InRoom item. These two items both showed marginally significant differences in the independent-components evaluation. The differences may have been less pronounced in the independent-components evaluation, because all conditions in the independent-components evaluation included high size fidelity and high position fidelity.

The results of the independent-components evaluation suggest that the differences seen in the holistic evaluation were not solely caused by differences in size fidelity, position fidelity, or factors unrelated to physicality, such as display resolution and brightness. Had that been the case, the independent-components evaluation should have shown no differences between conditions. However, this does not indicate that size fidelity and position fidelity are unimportant. Rather, high size fidelity and high position fidelity tend to accompany high concordance. High concordance requires MR technology such as see-through HMDs or other see-through displays. These see-through displays usually support high size fidelity and high position fidelity without any added cost.

Both evaluations demonstrated the importance of plausibility when developing ECAs for training applications. Implausibility limits both the social presence of the ECA and the usefulness of the training application as a whole. Effort must be spent developing and testing the conversational component of ECAs. Without this effort, increased physicality can be wasted.

The outliers in both evaluations demonstrate that despite best efforts, some users are likely to always have issues suspending disbelief. These users may be uncomfortable with all computer-based simulations, regardless of physicality. However, this does not have any relation to the user's ability to perform a job in reality.

In both the holistic evaluation and independent-components evaluations, we made the trade-off of increased ecological validity at the cost of lower statistical power. The evaluations both used real health science professionals and medical cases that medical teams really practice. Both evaluations were also conducted in the application's target environment, specially equipped rooms dedicated to medical simulation. These choices provide high ecological validity. However, these choices limited both the participant pool and the times that the studies could be run. For the independent-components evaluation, the entire participant pool (post-anesthesia care unit nurses at UF&Shands Hospital) was exhausted through pilot testing and conducting the study. Using a larger participant pool, such as undergraduate psychology students, could increase the statistical power. However, undergraduate psychology students do not possess the necessary experience to complete the training exercise. Moreover, undergraduate psychology students do not accurately reflect the demographics of the target population.

7 Conclusion and Future Work

We have proposed physicality and social presence as the ECA analogs of immersion and presence. We have identified six dimensions of physicality and examined existing research into those dimensions. We further examined four of these dimensions in two user studies.

Our results from both studies show that increasing physicality can increase social presence and lead to realistic behavior, provided that plausibility expectations are met. This result parallels findings that increasing immersion can increase presence in virtual environments (Meehan et al., 2002), as well as Slater's (2009) belief that plausibility illusion must also be met. Furthermore, the holistic evaluation found that increasing physicality increased plausibility expectations, which agrees with Nowak and Biocca's (2003) findings about the effects of anthropomorphic images on social presence.

The holistic evaluation's results demonstrate the importance of considering the user's expectations before increasing physicality. If the user's expectations about an ECA's behavior are met, increasing physicality can increase social presence and elicit more realistic behavior. However, increasing physicality also increases expectations. If these increased expectations are not met, increasing physicality may even have a negative effect on social presence.

The independent-components evaluation results highlight the importance of both environment and form fidelity. Increasing concordance by using a see-through display provides a large benefit for relatively little cost. Form fidelity is also very important. High form fidelity for an entire ECA can be difficult and costly, but a hybrid of high and low form fidelity (physical legs and a TV) can still increase social presence. Both studies demonstrate the value and feasibility of high-physicality ECAs in training applications.

This paper presented an initial investigation of the effects of physicality on social presence in the context of medical team training. Our results suggest that high-physicality ECAs increase social presence provided that plausibility is met. However, in order for the research community to gain a more fine-grained understanding of these effects, further studies, particularly having larger subject pools, should be conducted.

ECA plausibility also warrants further investigation. Plausibility may include a wide variety of components as suggested by our evaluations. In the holistic evaluation, we observed that emotional content, or the lack thereof, could hurt an ECA's plausibility. The independent-component evaluation results suggested that the ability to answer questions in a satisfactory manner might also affect plausibility. Other components, such as nonverbal behavior and responses to nonverbal stimuli, may also affect plausibility. Exploring these components of plausibility will require additional studies focusing specifically on plausibility, and we are currently conducting an initial exploration study.

We are also exploring the effects of multiple exposures to ECAs on comfort with ECAs. Many participants initially expressed mild discomfort with the ECAs but became more comfortable with them over time. Our explorations are examining two forms of multiple exposure to ECAs. One form is a single session where participants are exposed to multiple ECAs in succession, as was done in both the holitistic and independent-components evaluations. Another form is multiple sessions separated by several days. Determining how many exposures are needed to make paricipants comfortable with the ECAs may help researchers to produce more consistent results.

Although we have used medical team training as the motivating scenario for high-physicality ECAs, we believe that this line of research applies to other domains, such as the military and psychology. We encourage the community to pursue research on high-physicality ECAs in these and other domains.

Finally, we are currently investigating ways to increase ECA interaction with the environment, so that we can study the effect that those dimensions of physicality have on social presence and other metrics.

References

Argyle, M. (1967). *The psychology of interpersonal behavior.* Baltimore: Penguin Books.

Bailenson, J. N., Aharoni, E., Beall, A., Guadagno, R., Dimov, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. *Proceedings of the 7th Annual International Workshop on Presence.*

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments, 10*(6), 583–598.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin, 29*(7), 819–833.

Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments, 14*(4), 379–393.

Bimber, O., & Raskar, R. (2005). *Spatial augmented reality. Merging real and virtual worlds.* Natick, MA: AK Peters.

Blascovich, J. (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars.* Berlin: Springer.

Brunner, E., Domhof, S., & Langer, F. (2002). *Nonparametric analysis of longitudinal data in factorial experiments.* New York: Wiley.

Chuah, J. H., & Lok, B. (2012). Hybrid virtual-physical entities. *International Symposium on Mixed and Augmented Reality*, 275–276.

Chuah, J. H., Robb, A., White, C., Wendling, A., Lampotang, S., Kopper, R., & Lok, B. (2012). Increasing agent physicality to raise social presence and elicit realistic behavior. *Proceedings of the 2012 IEEE Virtual Reality (VR) Conference*, 19–22.

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies—Why and how. *Knowledge-Based Systems, 6*(4), 258–266.

Devereaux, P. J. (2008). Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): A randomised controlled trial. *The Lancet, 371*(9627), 1839–1847.

Dow, S., Mehta, M., Harmon, E., MacIntyre, B., & Mateas, M. (2007). Presence and engagement in an interactive drama. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07*, 1475–1484.

Fujie, S., Ejiri, Y., Nakajima, K., Matsusaka, Y., & Kobayashi, T. (2004). A conversation robot using head gesture recognition as para-linguistic information. *13th IEEE International Workshop on Robot and Human Interactive Communication, RO-MAN 2004*, 159–164.

Hunziker, S., Johansson, A., Tschan, F., Semmer, N., Rock, L., Howell, M., & Marsch, S. (2011). Teamwork and leadership in cardiopulmonary resuscitation. *Journal of the American College of Cardiology, 57*(1), 2381–2388.

Hyatt, J.-P. K., & Hurst, S. A. (2010). Novel undergraduate physiology laboratory using a human patient simulator. *Medical Education, 44*(5), 523.

Izadi, S. (2012). *A new era of human computer interaction.* Keynote address presented at the 11th IEE and ACM Symposium on Mixed and Augmented Reality, ISMAR 2012.

Johnsen, K., & Beck, D. (2010). The impact of a mixed reality display configuration on user behavior with a virtual human. *Proceedings of the Conference on Intelligent Virtual Agents*, 1–7.

Kaptein, M. C., Nass, C., & Markopoulos, P. (2010). Powerful and consistent analysis of Likert-type rating scales. *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI '10*, 2391–2394.

Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Aboobaker, N., & Wang, A. (2000). SUEDE: A Wizard of Oz prototyping tool for speech user interfaces. *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, Vol. 2, 1–10.

Kohn, L. T., Corrigan, J., & Donaldson, M. S. (2000). *To err is human: Building a safer health system.* Washington, DC: National Academies Press.

Kotranza, A., Johnsen, K., Cendan, J., Miller, B., Lind, D., & Lok, B. (2009). Virtual multi-tools for hand and tool-based interaction with life-size virtual human agents. *Proceedings of the 2009 IEEE Symposium on 3D User Interfaces*, 23–30.

Kotranza, A., Lok, B., Pugh, C., & Lind, D. (2009). Virtual humans that touch back: Enhancing nonverbal communication with virtual humans through bidirectional touch. *Proceedings of the 2009 IEEE Virtual Reality Conference*, 175–178.

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies, 64*(10), 962–973.

Lincoln, P., Welch, G., Nashel, A., Ilie, A., State, A., & Fuchs, H. (2009). Animatronic shader lamps avatars. *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality*, 27–33.

Maulsby, D., Greenberg, S., & Mander, R. (1993). Prototyping an intelligent agent through Wizard of Oz. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '93*, 277–284.

McMahan, R., Gorton, D., Gresock, J., McConnell, W., & Bowman, D. (2006). Separating the effects of level of immersion and 3D interaction techniques. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, 108–111.

Meehan, M., Insko, B., Whitton, M., & Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics, 21*(3), 645–652.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems E, Series D, 77*(12), 1321–1321.

Mori, M. (1970). The uncanny valley. *Energy, 7*(4), 33–35. doi: 10.1162/pres.16.4.337

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 12*(5), 481–494.

Pair, J., Neumann, U., Piepol, D., & Swartout, B. (2003). FlatWorld: Combining Hollywood set-design techniques with VR. *IEEE Computer Graphics and Applications, 23*(1), 12–15.

Pan, X., Gillies, M., Barker, C., Clark, D. M., & Slater, M. (2012). Socially anxious and confident men interact with a forward virtual woman: An experimental study. *PloS one, 7*(4), e32931 doi: 10.1371/journal.pone.0032931

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2009). Developing a team performance framework for the intensive care unit. *Critical Care Medicine, 37*(5), 1787–1793.

Rivera-Gutierrez, D., Welch, G., Lincoln, P., Whitton, M., Cendan, J., Chesnutt, D. A., … Lok, B. (2012). Shader lamps virtual patients: The physical manifestation of virtual patients. *Medicine Meets Virtual Reality 19, NextMed, Studies in Health Technology and Informatics*, 173, 372–378.

RoboThespian can't make horizontal wrist movements, resorts to unwanted advances. (2011). Retrieved from http://www.news.com.au/technology/robothespian-cant-make-horizontal-wrist-movements-resorts-to-unwanted-advances/story-e6frfro0-1226015171685

Rothschild, S. K., & Lapidos, S. (2003). Virtual integrated practice: Integrating teams and technology to manage chronic disease in primary care. *Journal of Medical Systems, 27*(1), 85–93.

Slater, M. (2003). A note on presence terminology. *Presence Connect, 3.*

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

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence: Teleoperators and Virtual Environments, 9*(1), 37–51.

Suma, E. A., Lange, B., Rizzo, A. S., Krum, D. M., & Bolas, M. (2011). FAAST: The flexible action and articulated skeleton toolkit. *Proceedings of the IEEE Virtual Reality Conference*, 247–248.

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., & Brooks, F. P. (1999). Walking > Walking-in-place > Flying, in virtual environments. *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '99*, 359–364.

Von der Pütten, A., & Gratch, J. (2010). How our personality shapes our interactions with virtual characters—Implications for research and development. *Proceedings of the Intelligent Virtual Agents Conference*, 208–221.

Wallin, C., & Meurling, L. (2007). Target-focused medical emergency team training using a human patient simulator: Effects on behaviour and attitude. *Medical Education, 41*(2), 173–180.

Ware, C., Arthur, K., & Booth, K. S. (1993). Fish tank virtual reality. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '93*, 37–42.

Zimmons, P., & Panter, A. (2003). The influence of rendering quality on presence and task performance in a virtual environment. *Proceedings of the IEEE Virtual Reality Conference VR '03*, Vol. 2003, 2–3.

Zwarenstein, M., & Reeves, S. (2002). Working together but apart: Barriers and routes to nurse–physician collaboration. *The Joint Commission Journal on Quality Improvement, 28*(5), 242–247.