Teaming Up With Virtual Humans: How Other People Change Our Perceptions of and Behavior with Virtual Teammates

Andrew Robb, Andrew Cordar, Samsun Lampotang, Casey White, Adam Wendling, Benjamin Lok

Fig. 1. A human participant works with a virtual anesthesiologist and a human surgeon (played by a confederate) to prepare a simulated patient for surgery.

Abstract— In this paper we present a study exploring whether the physical presence of another human changes how people perceive and behave with virtual teammates. We conducted a study (n = 69) in which nurses worked with a simulated health care team to prepare a patient for surgery. The agency of participants' teammates was varied between conditions; participants either worked with a virtual surgeon and a virtual anesthesiologist, a human confederate playing a surgeon and a virtual anesthesiologist, or a virtual surgeon and a human confederate playing an anesthesiologist. While participants perceived the human confederates to have more social presence ($p < 0.01$), participants did not preferentially agree with their human team members. We also observed an interaction effect between agency and behavioral realism. Participants experienced less social presence from the virtual anesthesiologist, whose behavior was less in line with participants' expectations, when a human surgeon was present.

Index Terms—Virtual/digital characters, mixed reality, training, user studies

1 Introduction

Virtual humans have helped overcome training barriers in several different domains, including medicine, education, and the military. Virtual humans have been used to safely simulate scenarios that would normally endanger the trainee or others, such as negotiating with a hostile force [26] or interviewing a patient with a potentially fatal illness [12]. Virtual humans have also been employed to simulate real humans who are either unable or unwilling to participate in training programs, such as when a medical student needs to practice performing a prostate exam [25]. Virtual humans have also helped overcome team training barriers by serving as virtual teammates when real teammates are not available for training. The virtual human Steve was developed for use in several team training applications, including training naval maintenance technicians [24]. The Mission Rehearsal Exercise system used virtual teammates to help soldiers learn how to make difficult decisions under pressure [27]. The CITTP system [22] was used to train NASA personnel at the Payload Operations Center, and the PORTS TAO-ITS system [23] was used to train naval tactical action officers. In these systems, trainees interacted with their teammates in virtual environments.

Virtual environments have several affordances that are useful for team training. They can allow people who are not physically co-located to train together, can simulate training environments that are not easily accessible, and allow virtual team members to directly manipulate objects within the virtual environment. However, virtual environments are not suitable for every type of team training. Many important team skills rely not only on communication skills, which are easily practiced in virtual environments, but also on psychomotor skills that can be difficult to learn within virtual environments. Additionally, *in situ* team training, which is conducted in the actual environment in which the team usually functions, has been shown to help identify problems that cannot be identified in simulation centers or simulated environments [19]. In these cases, it is preferable to train in real-world environments.

Thankfully, several techniques exist that have been used to bring virtual humans into the real world (we discuss these techniques in

- *Andrew Robb, Andrew Cordar, and Benjamin Lok are with the Department of Computer Science and Engineering at the University of Florida. E-mail: {arobb, acordar, lok}@cise.ufl.edu.*
- *Samsun Lampotang and Adam Wendling are with the Department of Anesthesiology at the UF College of Medicine. E-mail: {slampotang, awendling}@anest.ufl.edu.*
- *Casey White is with the School of Medicine at the University of Virginia. E-mail: caseywhite@virginia.edu.*

Manuscript received 18 Sept. 2014; accepted 10 Jan. 2015. Date of publication 20 Jan. 2015; date of current version 23 Mar. 2015. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org
Digital Object Identifier 10.1109/TVCG.2015.2391855

Section 2.2). However, bringing virtual humans into the real world introduces a new dynamic not possible in most virtual environments: people will be able to directly compare virtual humans to other real, physically-present humans. When multiple humans interact with virtual humans in virtual environments, many of the differences between real and virtual humans are masked by technological limitations. This is not the case in the real world. It may be that introducing other real, physically-present humans into a team will change how people perceive and behave with members of the team who are virtual.

To explore this question, we conducted a study involving 69 operating room (OR) nurses who participated in a team training exercise including either one or two virtual team members, depending on condition. Nurses worked with a surgeon and an anesthesiologist to prepare a patient for surgery. The agency of the surgeon and the anesthesiologist were varied among three conditions, such that participants worked with either a virtual surgeon and a virtual anesthesiologist, a human surgeon and a virtual anesthesiologist, or a virtual surgeon and a human anesthesiologist. The human surgeon and human anesthesiologist were played by confederates. We investigated how the agency of participants' teammates impacted their feelings of social presence and their behavior. To investigate participants' behavior, we inserted two "decision moments" where participants were asked to settle minor disagreements between the surgeon and the anesthesiologist about specific procedures. We observed an interaction effect with the presence of a human teammate and behavioral realism: participants perceived virtual humans with low behavioral realism to have less social presence when a human teammate was present. However, this did not have any effect on participants' behavior during the decision moments.

2 Related Work

In this section we explore previous work comparing behavior with real and virtual humans. We also discuss different techniques that have been used to bring virtual humans into the real world.

2.1 Real Humans compared to Virtual Humans

The term "agency" describes whether a virtual human is controlled by either a human being or by a computer program. The terms "avatar" and "agent" are also used, respectively. Most investigations of agency focus on *perceived* agency rather than actual agency. For the purposes of experimental control most experiments hold actual agency constant while varying what participants are told about the virtual human's agency.

Two theories have been developed that are relevant to discussions about how agency impacts perceptions of and behavior with virtual humans. The *Media Equation*, also referred to as the *Computer as Social Actors* theory, was developed by Clifford Nass and his colleagues in the mid 90's [15, 16]. The Media Equation states that humans mindlessly respond to social stimuli, regardless of the source. All other things being equal, the Media Equations predicts that agents and avatars will evoke the same level of social behavior; the only factor that matters is the fidelity of the social stimuli produced by a virtual human. A second model, known as the *Threshold Model of Social Influence*, was developed by Blascovich and his colleagues in the early 2000's [4]. The Threshold Model predicts that perceived agency moderates the efficacy of social stimuli. In other words, humans are more responsive to social stimuli when they believe it comes from an avatar rather than an agent. This holds true regardless of what actually controls the virtual human. Accordingly, agents must exhibit higher levels of behavioral realism in order to evoke the same level of social behavior evoked by lower quality avatars.

Both theories are supported by research. In support of the Media Equation, Nowak explored the effect of perceived agency and anthropomorphism on feelings of co-presence and social presence [17]. While Nowak found that anthropomorphism impacted feelings of co-presence and social presence, perceived agency did not have any significant effects on either co-presence or social presence. Von der Putten performed a similar experiment exploring the effect of behavioral realism and perceived agency on social presence, rapport, emotional state, and verbal behavior during a self-disclosure task [30]. While von der Putten found that higher levels of behavioral realism produced stronger feelings of social presence and led participants to talk more, varying perceived agency did not have a significant effect on social presence or behavior. Perceived agency's only observed effect was on participants' experience of negative emotions – perceived agents led to more negative feelings than perceived avatars.

In contrast, other research has found support for the Threshold Model. Bailenson explored the effect of agency and behavioral realism, in the form of gaze behavior, on interpersonal distance in immersive virtual environments. He found that behavioral realism interacted with agency, such that higher levels of behavioral realism in the agent condition led participants to maintain more realistic interpersonal distances, compared to lower levels of behavioral realism. However, behavioral realism had no effect in the avatar condition. Participants also reported more social presence when virtual humans exhibited mutual gaze behavior, and reported more social presence for avatars than for agents. In a second experiment, Bailenson investigated how perceived agency influenced participants' reactions when their personal space was invaded by a virtual human. Participants moved farther away from agents than avatars after their personal space had been invaded [1]. Guadagno conducted a study exploring how agency and behavioral realism, in the form of gaze behavior, affected social influence during an informational presentation. She found that people perceived avatars to have higher behavioral realism and higher social presence than agents. There was also an interaction effect between virtual human gender, agency, and persuasiveness: participants perceived male agents to be more persuasive than male avatars [10]. De Melo explored the relationship between emotional expression, agency, and social behavior during social dilemmas and negotiations. She found that emotional expressions influence social behavior, but only for perceived avatars. Varying an agent's emotional expression had no effect on social behavior [7]. Raij compared medical interviews with a virtual patient to interviews with a human standardized patient. He found that participants' interviews with a virtual patient were less structured and their expressions of empathy were less sincere than those conducted with the standardized patient. However, Raij observed that these differences may have been due to limitations in the virtual patient's conversation architecture, especially the lack of support for contextual questions [21]. Lucas explored how varying perceived agency and actual agency influenced self-disclosure behavior when talking with a virtual interviewer. When talking with a perceived agent, participants reported less fear of self-disclosure, were less likely to engage in impression management behavior, and engaged in more open behavior. Varying the actual agency only influenced ratings of the system's usability [14].

While it appears that agency has some effect on how participants perceive and interact with virtual humans, the exact degree to which agency matters remains a matter of debate. The last two studies described above are of particular relevance to the research described in this paper. Raij remains one of the few researchers who has directly compared virtual humans to real, physically-present humans. He attributed the differences between people's interactions with real and virtual humans to the limitations in the virtual human's ability to understand and respond to participants questions. To overcome this limitation, a Wizard-of-Oz system was employed to control the virtual humans in the study described in this paper. This helped to reduce error rates and allowed the virtual humans to exhibit more complex behavior than those used in Raij's experiment. Lucas' findings about actual and perceived agency underscored how user perceptions can matter more than reality. Given this, the gap that Lucas observed between agents and avatars may grow even wider when the perceived avatar is replaced by a physically-present human.

2.2 Bringing Virtual Humans into the Real World

Researchers have developed at least three different methods that can bring virtual humans into the real world. The first method leverages see-through HMDs, or HMDs augmented with co-located physical props. ARFacade used a see-through display to embed two virtual humans in a physical space where they could interact with each other and with participants [8]. Kotranza developed a virtual breast exam patient that incorporated an augmented reality breast exam simulator. The virtual patient was displayed on an HMD, while a live video feed of the breast exam simulator was tracked and overlaid on the display [13]. The virtual patient was also equipped with a physical arm that could reach out and touch the participant; this physical arm was co-located with the patient's virtual arm shown on the HMD.

The second method relies on "digital flats", displays strategically placed at boundary points in a physical environment, such as doors, windows, or behind barriers. These displays give the illusion that the physical environment extends past these boundaries. Props are often used to obscure the edge of the displays, which helps to create the illusion of a single shared space. Digital flats have been used to replace doors and windows in the FlatWorld project [18], and to simulate parts of entire rooms in the Gunslinger project, such as behind a bar [11].

The final method employs augmented reality displays located *within* a physical environment. These augmented displays differ from digital flats in that they are placed within the physical environment instead of along its boundaries. Rather than simulating an adjacent but separate space, augmented displays employ see-through backgrounds and perspective-correct rendering to give the impression that virtual content is actually present in the room at the location of the display. Chuah and colleagues used augmented displays in several medical training programs [6].

Each of these methods has different advantages and limitations when used to embed virtual humans in shared, physical environments. When they function correctly, see-through HMDs can provide a very high fidelity illusion. However, these often require complex tracking setups that can be difficult to calibrate. Digital flats are simple and effective, but are rarely portable. Augmented displays can be moved more easily but cannot be used with multiple viewers because perspective-correct rendering is specific to a single viewer. In the end, the constraints of the scenario being simulated will determine which approach will be most useful.

3 The Team Training Exercise

In order to evaluate whether working with a physically-present human teammate changes how people perceive and behave with virtual humans, we created a team training exercise involving three roles: a nurse, a surgeon, and an anesthesiologist. In the study we conducted, participants always played the role of nurse. Depending on the condition, the surgeon and the anesthesiologist were played by either a virtual human or by a physically-present human confederate. The three conditions are illustrated in Figure 2.

3.1 Scenario Description

In this scenario, a human nurse worked with a surgeon and an anesthesiologist to prepare a simulated patient for surgery. The scenario comprised two stages. In the first stage, participants worked with the surgeon and anesthesiologist to ensure that the patient was ready for the start of anesthesia. The surgeon led the team as they worked through a standardized checklist used in the hospital in which the study was conducted. The surgeon addressed questions to the nurse participant and the anesthesiologist as required to complete the checklist. The anesthesiologist occasionally interrupted with a question or comment. Participants could also interrupt with questions or comments. Once the checklist was complete, the surgeon told the team to continue their preparations. Participants then moved on to the second stage, where they worked with the surgeon and anesthesiologist to confirm that the patient was ready for the surgical incision. In this stage, the surgeon led the team through a second, shorter checklist. During this stage, the surgeon learned that the anesthesiologist had forgotten to send blood samples to the lab for processing, which meant that replacement blood was not yet available. This angered the surgeon, who berated the anesthesiologist and announced that he was going to start the surgery even though blood was not yet available. At this point, participants could either object and try to stop the surgeon from proceeding, or give in to the surgeon and allow him to proceed.

(a) The condition with the Virtual Surgeon and the Virtual Anesthesiologist. The label VS-VA will be used to refer to this condition.

(b) The condition with the Human Surgeon and the Virtual Anesthesiologist. The label HS-VA will be used to refer to this condition.

(c) The condition with the Virtual Surgeon and the Human Anesthesiologist. The label VS-HA will be used to refer to this condition.

Fig. 2. The human surgeon and the human anesthesiologist were each played by a confederate. Each virtual human was modeled to visually resemble their human counterpart. The virtual humans' speeches were also recorded by their human counterpart.

3.2 Decision Moments

Two "decision moments" were incorporated into the scenario in order to investigate whether the agency of participants' team members impacted how they made decisions. During these decision moments, the surgeon and anesthesiologist were unable to agree on minor patient care issues. After being unable to resolve the issue themselves, they asked participants what they thought would be the best solution. The surgeon and the anesthesiologist agreed with whatever decision participants made and moved on to the next item on the surgeon's checklist.

The first decision moment concerned whether or not to place monitors (devices that monitor a patient's vital signs) on the patient before or after he was placed under anesthesia. The surgeon and anesthesiologist's exchange during this moment is shown below.

Surgeon: You haven't put the monitors on Clay yet?

Anesth: No, I didn't want to agitate Clay. His mother said he can be a bit touchy around needles, so I'm concerned the monitors might disturb him as well.

Surgeon: I don't think putting monitors on should agitate him, even if he doesn't like needles.

Anesth: Nurse, what do you think?

The second decision moment concerned whether or not to use an upper- and lower-body Bair Hugger or only a lower-body Bair Hugger (Bair Huggers are used to keep patients warm during surgery). The surgeon and anesthesiologist's exchange during this moment is shown below.

Anesth: We'll use an upper and lower Bair Hugger, right?

Surgeon: I think we're only going to have room for a lower body Bair Hugger. Nurse, what do you think?

Both of these decision moments occured during the first stage of the exercise. The decision moments did not occur back to back, but were separated by ten other items on the surgeon's checklist.

3.3 Virtual Human Technology

During the scenario, virtual humans were embedded in a real-world environment using augmented reality displays (shown in Figures 1 and 3), as described by Chuah [6]. These displays employed several techniques to create the illusion that a virtual human occupies the same physical environment shared by participants. The virtual humans were rendered life-size on 40" televisions, in portrait mode. A Microsoft Kinect tracked participants' positions, allowing the virtual humans to make eye contact with participants. The virtual humans were rendered using perspective-correct techniques, which helped create an illusion of depth. A see-through display was simulated using pre-captured images of the environment, which were adjusted based on the participant's position in the room. Finally, physical legs extended out from the bottom of the display, giving the illusion that the virtual human's body extends into the room.

The virtual human's gaze was controlled by a simple Markov model; the virtual humans would look at whomever was speaking, but would occasionally glance at other people in the room. The virtual humans also blinked and mimicked idle motions when not speaking. When speaking, the virtual humans occasionally used hand gestures, depending on the content of the speech. All animations were captured using motion capture. The virtual surgeon and virtual anesthesiologist spoke using pre-recorded audio files, which were recorded by their human counterparts. This ensured that the voices remained consistent in all conditions. The virtual surgeon and the virtual anesthesiologist were modeled to visually resemble their human counterparts. Average Caucasian males were selected to play the part of the surgeon and the anesthesiologist, as this combination of race and gender is representative of the majority of surgeons and anesthesiologists practicing in the U.S [5].

The virtual humans were controlled using a Wizard-of-Oz (WoZ). Human-factors research frequently uses Wizard-of-Ozs to reduce confounding effects that can be introduced by speech recognition and speech understanding errors [2, 6, 25, 29]. The WoZ operator controlled both the surgeon and the anesthesiologist simultaneously, using an interface that allowed him to trigger the virtual humans' speeches using prespecified lists. This interface was organized by character and topic to allow for rapid selection. The interface also intelligently suggested responses based on the last action performed. The WoZ operator followed a specific script for each stage of the interaction, but made adjustments if participants behaved unexpectedly. The virtual humans were capable of making nine generic statements, such as "Yes", "No", "OK", and "I'm not sure", which allowed the WoZ operator to respond to unexpected questions or statements. In order to create a consistent experience for each participant, the same WoZ operator was used during the entire study. To reduce suspicion that the virtual humans were controlled by a human, participants were asked to complete a speech recognition training session and wear a microphone during the exercise, and the WoZ operator was concealed behind privacy screens. The exercise took place in a former operating room that had been converted to a simulation lab.

3.4 The Human Confederates

Both confederates were trained standardized patients who regularly assist with medical training exercises. Standardized patients are frequently used to train medical students to practice medical interviewing and physical examinations; they are trained to portray specific individuals and exhibit specific symptoms. While standardized patients do not normally play the role of a surgeon or anesthesiologist, they are familiar with the medical domain and understand that it is important behave consistently during each interaction to ensure that all trainees receive similar experiences. Standardized doctors do not currently exist. Both confederates were recruited from the pool of available standardized patients and were paid their standard rate ($20/hour) during the study.

Both confederates received training on the script from the WoZ operator who controlled the virtual humans during the study. He explained the scenario and demonstrated a standard interaction where the surgeon and the anesthesiologist were both virtual, and then had the confederate practice playing his role until he was comfortable with it. After this training session, the confederates took home paper scripts for further study and familiarization. These paper scripts were also used during the study as a guide to help the confederates stay on track during the exercise.

Steps were taken to reduce the variance between the virtual humans and their human counterparts. Each virtual human was modeled to visually resemble their counterpart and had their voice recorded by their counterpart. The confederates attempted to use the same tone of voice that they used while recording the audio for the virtual humans' speech. On the whole, the human confederates successfully mimicked the behavior of their virtual counterparts. One notable exception to this was the human confederates' gaze behavior. The human confederates made significantly less eye contact with participants, as they frequently consulted their script during the interaction. The confederate who played the surgeon also occasionally responded incorrectly to a participant's statement or ad-libbed an unscripted response. The implications of these sources of variance will be considered further in the Limitations section.

4 Experimental Goals and Hypotheses

Our goals for this experiment were twofold: to explore how the presence of a human teammate affects the social presence of virtual team members, and to explore whether the presence of a human teammate changes how people make decisions. We have three hypotheses related to these goals:

H1: Human team members will generate stronger feelings of social presence than their virtual counterparts.

H2: The presence of a human team member will negatively impact the social presence of virtual team members.

H3: Human participants will agree with human team members more frequently than virtual team members.

5 Experimental Design

To investigate these hypotheses, we conducted a between-groups study with three conditions. The agency of the surgeon and the anesthesiologist were varied between the three conditions. In the first condition, both the surgeon and the anesthesiologist were virtual humans. In the second condition, the surgeon was played by a human confederate and the anesthesiologist was a virtual human. In the third condition, the surgeon was a virtual human and the anesthesiologist was played by a human confederate. We did not include a fourth condition where both the surgeon and the anesthesiologist were played by human confederates because we were specifically interested in the virtual team members and how they were impacted by the presence of a human team member. A condition which did not include any virtual team members was thus deemed unnecessary. These three conditions are illustrated above in Figure 2.

5.1 Population

A total of 69 participants (53 female) took part in the study. All participants were nurses working in operating rooms at Shands Hospital at the time of the study. The average participant age was 42.4 years old; ages ranged from 24 to 68. Participants had been working as a nurse for an average of 20.2 years, and as an OR nurse for an average of 16.8 years. Of the 69 participants, 52 reported their race as White, 9 as Asian, 7 as Black, and 1 as American Indian or Alaska Native. Participation in the training exercise was mandated by the hospital; participation in the study was optional. All participants received 1.5 hours of continuing education credits, which are required for license renewal. Additionally, study participants received a $10 coffee gift card.

5.2 Metrics

Social presence was measured using a five question survey developed by Jeremy Bailenson [1]. The questions are shown below. The placeholder "..." was replaced by either "surgeon" or "anesthesiologist", as appropriate. Responses were given using a seven-point likert scale, ranging from strongly disagree to strongly agree.

1. I perceive that I am in the presence of a ... in the room with me.
2. I feel that the ... is watching me and is aware of my presence.
3. The thought that the ... is not a real person crosses my mind often.
4. The ... appears to be sentient, conscious, and alive to me.
5. I perceive the ... as being only a computerized image, not as a real person.

To score the survey, questions 3 and 5 are inverted and then averaged together with the other questions. Participants completed social presence surveys for both the surgeon and the anesthesiologist after both Stage 1 and Stage 2. Finally, participants' behavior during the decision moments was recorded and coded for later analysis.

Bailenson's social presence survey given preference over other social presence surveys because of its brevity. Our study design called for social presence surveys to be administered for each character and after each stage. While Biocca's Networked Minds survey [3], which contains thirty-six questions, theoretically yields a more accurate social presence score, it proved infeasible to administer due to participant availability and because of the increased risk of survey fatigue [20].

6 Results

In this section we present the results from the social presence surveys and the two decision moments. When discussing the results, we refer to each condition using the following terminology: Each condition is labeled using a pair of letters. V stands for virtual, and H stands for human. S stands for surgeon, and A stands for anesthesiologist. Thus, the label VS-HA corresponds to the condition where the surgeon was virtual and the anesthesiologist was human.

6.1 Social Presence

Two participants failed to complete all of the social presence surveys; their social presence data were excluded from our analysis. In addition, three participants were excluded as outliers because their social presence scores were lower than the mean by more than three standard deviations. Excluding these five participants left us with social presence data from sixty-four participants: twenty-five in the VS-VA condition, nineteen in the HS-VA condition, and twenty in the VS-HA condition.

Participants completed two social presence surveys for each character, one after the first stage and another after the second stage. No significant differences between stages were seen for either the surgeon or the anesthesiologist. Accordingly, we averaged the surgeon's and the anesthesiologist's social presence scores from Stage 1 and Stage 2 to create a single composite social presence score. This was done to reduce measurement error and limit the number of statistical tests run, reducing the risk of making a Type I error.

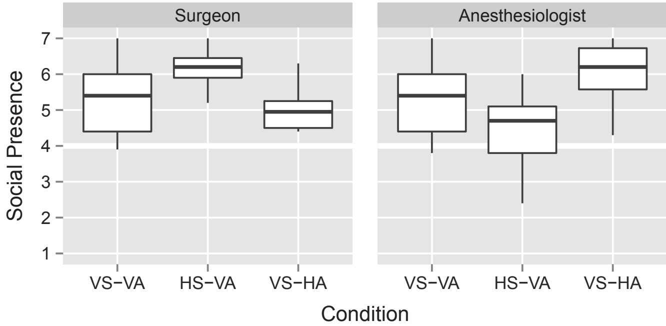

Fig. 3. Boxplots of social presence in each condition for the surgeon and the anesthesiologist.

6.1.1 Human Teammates Impact on Social Presence

Condition	Surgeon		Anesthesiologist	
	Mean	SD	Mean	SD
VS-VA	5.22	0.97	5.19	1.02
HS-VA	6.17	0.61	4.63	1.01
VS-HA	5.00	0.78	6.01	0.89

Table 1. Mean and standard deviations for the surgeon and anesthesiologist's social presence in each condition. Social presence scores can range from 1 to 7.

Social presence means and standard deviations are reported in Table 1 for the surgeon and the anesthesiologist. A boxplot of the social presence results is also shown in Figure 3. There were no significant differences ($p > 0.8$) in social presence between the surgeon and the anesthesiologist in the VS-VA condition, where both the surgeon and the anesthesiologist were virtual. This suggests that social presence in the VS-VA condition can be used as a baseline value when considering the effect of agency for both the surgeon and the anesthesiologist, as each character evoked similar levels of social presence in this condition.

Multiple one-way ANOVAs were performed to determine whether agency had any effect on participant's feelings of social presence. Significant differences in social presence were observed between conditions for both the surgeon ($F_{2,62} = 10.42, p < 0.001$) and the anesthesiologist ($F_{2,62} = 12.54, p < 0.001$). Post-hoc Bonferroni tests were conducted to further explore how varying the agency of the surgeon and anesthesiologist affected social presence. The results are reported below, categorized by their relevance to our first two hypotheses.

Our first hypothesis was that human teammates would evoke a stronger sense of social presence than their virtual counterparts. Our results suggest that this is accurate – the human surgeon ($p < 0.005$) and the human anesthesiologist ($p < 0.05$) evoked significantly stronger feelings of social presence than their virtual counterparts.

Our second hypothesis was that the presence of a human team member would negatively impact the social presence of virtual team members. This effect was not observed for the virtual surgeon, but was for the virtual anesthesiologist. The virtual surgeon's social presence was *not* significantly reduced by the presence of a human anesthesiologist ($p = 0.841$). However, the virtual anesthesiologist's social presence *was* significantly reduced by the presence of a human surgeon ($p < 0.05$).

6.2 Decision Moments

We observed four responses to the first decision moment. Participants either agreed with the surgeon and said to put the monitors on now, agreed with the anesthesiologist and said to wait until the patient was placed under anesthesia, or suggested one of two alternatives: try and put the monitors on now but stop if the patient complained, or to give him some minor anesthesia immediately and then place the monitors on. The percentage of participants exhibiting each behavior is shown below, in Table 2.

Condition	N	Put On	Wait	Try	Mask
VS-VA	26	26.9%	15.4%	42.3%	15.4%
HS-VA	20	40.0%	25.0%	15.0%	20.0%
VS-HA	21	33.3%	23.8%	38.1%	9.5%

Table 2. Participant responses to the monitors decision moment. The anesthesiologist, who was in favor of waiting to put the monitors on, asked participants to weigh in on this question. Data from two participants is missing because one did not consent to video recording and another did not respond to this question.

We observed three types of responses to the second decision moment. Participants either agreed with the surgeon and said to only use the lower-body bair hugger, agreed with the anesthesiologist and said to use both the upper- and lower-body bair hugger, or proposed a compromise of using the lower-body bair hugger and another warming method (e.g. raise the room temperature). The percentage of participants exhibiting each behavior is shown below, in Table 3.

Condition	N	Lower Only	Use Both	Compromise
VS-VA	25	52.0%	24.0%	24.0%
HS-VA	21	61.9%	28.6%	9.5%
VS-HA	19	42.1%	21.1%	36.8%

Table 3. Participant responses to the bair hugger decision moment. The surgeon, who was in favor of using just the lower-body bair hugger, asked participants to weigh in on this question. Data from four participants is missing because one did not consent to video recording and three did not respond to this question.

Fisher exact tests revealed no significant differences between conditions for either decision moment ($p_{Monitor} = 0.439, p_{BairHugger} = 0.380$), indicating that agency of participants' teammates had no effect on the participants' decisions during these two moments.

7 Discussion

The goal of this research was to investigate whether the presence of a human teammate changes how people perceive and behave with virtual teammates. With regard to this question, our results suggest three findings: human team members evoke significantly more social presence than their virtual counterparts; the presence of a human teammate can sometimes reduce the social presence of a virtual teammate; and participants show no tendency to preferentially agree with human teammates when asked to make a decision. We also report an additional result related to previous work conducted by Usoh about the use of presence questionnaires to evaluate real-world experiences.

7.1 Comparing Virtual Humans and Real Humans

Both human confederates evoked significantly higher levels of social presence than their virtual counterparts ($p_{Surgeon} < 0.005, p_{Anesth} < 0.05$). While it is not surprising that real humans evoked stronger feelings of social presence than virtual humans, it is encouraging to note that the virtual humans also evoked positive feelings of social presence. As can be seen in Figure 3, the majority of participants rated the virtual humans as having positive social presence, giving them a score above the neutral value of four. While virtual humans may not evoke the same level of social presence real humans possess, they nevertheless can evoke positive feelings of social presence.

7.2 The Effect of a Human Teammate on a Virtual Teammate's Social Presence

The presence of a human anesthesiologist did not impact the virtual surgeon's social presence. However, the presence of a human surgeon did reduce the virtual anesthesiologist's social presence ($p < 0.05$). This suggests that the presence of another human teammate can reduce a virtual human's social presence, but that this effect is mediated by another factor. Interviews with participants suggest that this mediating factor may be behavioral realism.

Participants reported that the virtual surgeon's knowledge of the procedure and his tendency to take over everything made him seem very realistic. His assertive personality, his anger about the blood not being ready, and his specific objections to nurses' speaking up also contributed to this sense of realism (though some participants did say that most surgeons would not be this difficult to work with).

In contrast, participants felt that the virtual anesthesiologist did not behave as realistically, especially during the confrontation with the surgeon that occurred at the end of Stage 2. These participants felt that the a real anesthesiologist would have been more assertive and would have defended his actions more. Other participants also felt that a real anesthesiologist would have supported them when they spoke up to stop the surgeon from beginning surgery before blood was available. Several participants said that the virtual anesthesiologist behaved like a new resident or intern, rather than an experienced clinician.

Blascovich's threshold model predicts that a virtual human's perceived agency and behavioral realism interact to determine how people perceive that virtual human [4]. We see a similar effect here, where a virtual human's behavioral realism interacts with the agency of *other teammates* to determine whether the virtual human can maintain it's normal level of social presence. The surgeon, who exhibited higher behavioral realism, was able to maintain his baseline level of social presence when a human anesthesiologist was present. The anesthesiologist, who exhibited lower behavioral realism, was not able to maintain his baseline level of social presence when a human surgeon was present. This suggests that virtual humans are capable of maintaining normal levels of social presence when other humans are present, so long as they exhibit relatively high levels of behavioral realism.

When seeking to apply this conclusion, it is important to recognize that there are two major classes of behavioral realism, one dealing with objective, lower-order behavior and one dealing with subjective, higher-order behavior. The first form of behavioral realism deals with well-defined, objective patterns of human behaviors that are often unconscious and automatic, such as gaze, non-verbal behavior, and posture/motion. These variables are frequently manipulated by researchers to create states of low and high behavioral realism [9, 10, 30]. These lower-order behaviors may not be consciously recognized by human participants, but nevertheless affects participants' interactions with virtual humans. These lower order behaviors were not manipulated in this experiment, but were maintained at a relatively high level.

The second form of behavioral realism deals with subjective, and often stereotypical, expectations about how a human is likely to behave. We expect teachers to be intelligent, firefighters to be brave, dancers to be coordinated, and judges to be fair. When these expectations are broken, behavioral realism is lowered. This is true even though teachers sometimes are not intelligent, firefighters not brave, dances uncoordinated, and judges unfair. It is this second type of behavioral realism that the surgeon and the anesthesiologist differed on. Our participants expected anesthesiologists to be competent, assertive, and helpful; the virtual anesthesiologist did not meet these expectations, which caused participants to perceive him as having lower behavioral realism. In contrast, the surgeon met participants' expectations by being knowledgeable, aggressive, and angry, and thus was judged to have higher behavioral realism.

It is important the virtual humans achieve high levels of both types of behavioral realism. The first type of behavioral realism hinges on psychophysical theory and technological models capable of simulating lower-order behaviors, while the second is driven largely by an understanding of the social nuances of the domain being simulated. Neither of these aspects of behavior realism can be neglected if realistic virtual humans are to be developed.

7.3 Human Teammates and Decision Making

No significant differences were observed between conditions in the decisions participants made during either of the two decision moments ($p_{Monitor} = 0.439, p_{BairHugger} = 0.380$). Accordingly, we reject the hypothesis that people are automatically more likely to agree with human team members. It appears that the agency of participants' teammates did not heavily influence their decision making process. Instead, participants' decisions appear to have been driven by their reasoning and past experience. While some participants simply agreed with either the surgeon or the anesthesiologist, other participants verbally reasoned through their decision before answering the question. Additionally, many participants attempted to develop a compromise solution that could address the concerns raised by both the surgeon and the anesthesiologist (see "Try", "Mask", and "Compromise" in Tables 1 and 2).

At first, our results appear to conflict with Guadagno's findings that agency can influence persuasiveness [10]. However, it is important to recognize that "persuasion" does not necessarily encompass all forms of decision making. In Guadagno's experiment, a single virtual agent spent a significant amount of time attempting to persuade participants to change their pre-existing opinion about a newly proposed security policy. In contrast, the surgeon and anesthesiologist made no attempt to persuade participants that their approach was the best one, but instead simply asked participants which approach they thought was better. The presence of two different viewpoints and the lack of persuasive arguments differentiates this research from Guadagno's.

Our results also differ from de Melo's findings that agency interacts with emotional expressions to influence decision making [7]. De Melo found that expressions of emotions, like joy, regret, and anger, changed participants' behavior during social dilemmas and negotiations with avatars, but not with agents. The main distinction here is that the decisions participants were asked to make had very low emotional content, and the surgeon and anesthesiologist expressed no emotions when asking participants to make a decision. It may be that emotionally-charged decisions may be influenced by the presence of real human teammates. Further research is needed here.

In sum, our results indicate that participants were not influenced by their teammate's agencies when asked to make simple, non-emotional decisions. Instead, the prevailing factor in these decisions appears to have been participants' reasoning about what would be more appropriate to the situation. This finding supports the viability of team training with virtual humans, as it indicates that the agency of learners' teammates is unlikely to lead them to make incorrect decisions.

7.4 Using Social Presence Questionnaires on Real People

Despite actually being real, most participants did not give the human confederates a perfect social presence score. This finding parallels Usoh's earlier work applying physical presence questionnaires in real environments [28]. Usoh found that two physical presence questionnaires were essentially unable to distinguish between virtual and real environments. While this result initially seems counterintuitive, it is actually a natural consequence of human psychology. Usoh explains:

> *If someone is asked for their sense of being there on a 1 to 7 scale, it gives them permission to answer with a score of less than 7 even when they are really there. The questions are reinterpreted to make sense in the given context. ... In the real-world, since there is no doubt that the individual is present in the obvious sense, it becomes reinterpreted as the sense of involvement, the lack of isolation, perhaps the degree of comfort. The thought "I am not comfortable to be here" might lead to a low 'presence' response. [28]*

Participants' responses to the individual social presence questions for the human confederates are shown in Figure 4. Participants' tendency to reinterpret the social presence questions is seen most clearly for Q3, which states "The thought that the ... is not a real person crosses my mind often." Despite the fact that both of the human confederates were obviously real people, participants frequently gave them low scores for this question, sometimes even dipping down into the "disagree" range. A similar, though less pronounced effect can be seen for Q5, which states "I perceive the ... as being only a computerized image, not as a real person." It seems likely that these participants reinterpreted 'not a real person' as 'not a real *surgeon*' or 'not a real *anesthesiologist*'. Participants appear to have reinterpreted Q1, Q2, and Q4 less frequently. While Q3 and Q5 focused on feelings of unreality, these three questions assessed positive feelings related to mutual awareness and sentience. It may be that negatively framed questions are more likely to be reinterpreted than positively framed questions. This possibility should be considered when developing presence questionnaires.

Fig. 4. Participants response to the individual social presence questions for the human surgeon and the human anesthesiologist. Responses for the virtual humans are not shown here. Responses to Q3 and Q5 have been inverted, according to the survey's scoring instructions. The text for each question can be found in Section 5.2.

7.5 Limitations

The ecological validity of our findings is somewhat limited by the use of confederates for this study. The script followed by the confederates was developed by experienced clinicians, but even with this script the confederates could not mimic an actual clinician's behavior perfectly. It may be that participants unusual responses to Q3 and Q5 for the human confederates, as discussed in Section 7.4, would not be observed in interactions with real clinicians.

A more important limitation is a systematic variation between the real and virtual humans' gaze behavior. The human confederates frequently looked down at the clipboard holding their script, while the virtual humans instead maintained eye contact with whoever was speaking, or the person being spoken to. This difference is a concern as prior research has found that virtual humans that exhibit more realistic gaze behavior evoke stronger feelings of social presence [9]. However, while this difference in gaze behavior is a limitation in our study, it also seems unlikely that our results would have been altered if the human confederates had made more eye contact. Given that improved eye contact would likely have *increased* the human confederate's social presence, this would have only served to strengthen the already significant social presence differences between the virtual humans and their human counterparts. Additionally, this type of gaze behavior is not uncommon in the OR, where surgeons and anesthesiologists frequently have to consult patient charts or checklists, and thus cannot always maintain eye contact with their teammates.

Other sources of variance between participants' virtual teammates and their human counterparts included variations in tone of voice, posture, and phrasing. While the human confederates attempted to keep these details constant from participant to participant, they were understandably not always successful, as it can be difficult for humans to control these behaviors precisely. It is possible that these changes influenced participant behavior. However, it is important to recognize that these sources of variance are impossible to completely eliminate from any research involving human confederates.

8 Conclusions

In this paper, we explored how the presence of a human teammate changes the way people perceive and behave with virtual teammates. To investigate this, we conducted a study in which the agency of participants' teammates was varied. Participants either worked with a virtual surgeon and a virtual anesthesiologist, a human confederate playing a surgeon and a virtual anesthesiologist, or a virtual surgeon and a human confederate playing an anesthesiologist. We found that while human team members were perceived as having more social presence than comparable virtual team members, this did not change participants' behavior when making decisions – when asked to side with one of their two teammates, participants showed no tendency to preferentially agree with human team members. Instead their decisions appear to have been driven by their reasoning about the scenario.

In addition to finding that human teammates evoke more social presence than corresponding virtual teammates, we also found that the presence of a human teammate can reduce a virtual human's social presence, if that virtual human's behavior does not meet participants' expectations about how a human should behave. This relationship between the agency of participants' team members and behavioral realism is similar to Blascovich's Threshold model, which theorizes that agency and behavioral realism interact to determine how people will perceive and behave with a virtual agent. Our results suggest that the Threshold model can be extended to say that the agency of all parties involved in a conversation interacts with an agent's behavioral realism to determine whether or not that agent will evoke social behavior from participants. Thus, virtual members of teams containing multiple human teammates must exhibit high levels of behavioral realism if they are to evoke the same social behavior evoked in teams containing only one human team member.

8.1 Future Work

In this paper, we presented an initial exploration of how the presence of another human teammate impacts feelings of social presence. Our findings suggest that the presence of another human teammate may reduce a virtual human's social presence if that virtual human does not meet participants' expectations for how a real human would behave. This study explored only one aspect of behavioral realism in this paper, namely whether a virtual human's demeanor and higher-order behavior was in accord with normal patterns of human behavior. Other lower-order aspects of behavioral realism include gaze behavior, nonverbal behavior, and motion/posture. It is unclear whether failure to meet expectations about these forms of behavioral realism would have the same effect on feelings of social presence when another human teammate is present. Future research is required to explore whether failing to meet expectations about these lower-order aspects of behavioral realism also interacts with the presence of another human to reduce a virtual human's social presence.

Future work is also required to further understand when agency influences how people make decisions. A key distinction between the scenario explored in this study and the scenarios used by Guadagno and de Melo is that the scenario used in this study placed participants in an environment very similar to their normal working environment and asked them to make decisions similar to those they make on a daily basis. The scenarios used in Guadagno's and de Melo's research were more abstract and less connected to their participants' experience and profession. It may be that agency's impact on decision making behavior increases as the decisions one is asked to make grow more abstract.

Acknowledgments

We would like to thank Theresa Hughes, Terry Sullivan, and David Lizdas for their help in developing the Speaking Up exercise and recruiting participants, as well as the nurses who agreed to participate in this study. We would also like to thank Robert Wilt and Arne Zarzecki, who played the human surgeon and anesthesiologist during the study. This work was supported in part by NSF Grant 1161491.

References

[1] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin*, 29(7), 2003.

[2] T. W. Bickmore. *Relational Agents : Effecting Change through Human-Computer Relationships*. PhD thesis, 2003.

[3] F. Biocca, C. Harms, and J. Gregg. The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th annual International Workshop on Presence*, 2001.

[4] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology. *Psychological Inquiry*, 13(2):103–124, Apr. 2002.

[5] L. Castillo-Page. Diversity in the Physician Workforce: Facts & Figures 2010. *Association of American Medical Colleges*, 2010.

[6] J. H. Chuah, A. Robb, C. White, A. Wendling, S. Lampotang, R. Kooper, and B. Lok. Exploring agent physicality and social presence for medical team training. *Presence: Teleoperators & Virtual Environments*, 22(2):141–170, 2013.

[7] C. M. de Melo, J. Gratch, and P. J. Carnevale. The Effect of Agency on the Impact of Emotion Expressions on People's Decision Making. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 546–551, Sept. 2013.

[8] S. Dow, M. Mehta, A. Lausier, B. MacIntyre, and M. Mateas. Initial lessons from AR Façade, an interactive augmented reality drama. *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology - ACE '06*, page 28, 2006.

[9] M. Garau, M. Slater, V. Vinayagamoorthy, A. Brogni, A. Steed, and M. A. Sasse. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proceedings of the conference on Human factors in computing systems - CHI '03*, (5):529, 2003.

[10] R. Guadagno, J. Blascovich, J. N. Bailenson, and C. McCall. Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, pages 1–22, 2007.

[11] A. Hartholt, J. Gratch, and L. Weiss. At the virtual frontier: Introducing Gunslinger, a multi-character, mixed-reality, story-driven experience. *Intelligent Virtual Agents*, pages 500–501, 2009.

[12] A. Kotranza, K. Johnsen, J. Cendan, B. Miller, D. S. Lind, and B. Lok. Virtual multi-tools for hand and tool-based interaction with life-size virtual human agents. *2009 IEEE Symposium on 3D User Interfaces*, pages 23–30, 2009.

[13] A. Kotranza and B. Lok. Virtual Human + Tangible Interface = Mixed Reality Human An Initial Exploration with a Virtual Breast Exam Patient. *IEEE Virtual Reality*, pages 99–106, 2008.

[14] G. M. Lucas, J. Gratch, A. King, and L.-P. Morency. Its only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100, Aug. 2014.

[15] C. Nass and Y. Moon. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, 56(1):81–103, Jan. 2000.

[16] C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. *Conference companion on Human factors in computing systems - CHI '94*, page 204, 1994.

[17] K. L. Nowak and F. Biocca. The Effect of the Agency and Anthropomorphism on Users Sense of Telepresence , Copresence , and Social Presence. *Presence: Teleoperators & Virtual Environments*, pages 481–494, 2002.

[18] J. Pair, U. Neumann, D. Piepol, and B. Swartout. Projects in VR Flat-World: Combining Hollywood Set-Design Techniques with VR. *IEEE Computer Graphics and Applications*, (February):12–15, 2003.

[19] M. Patterson, G. Blike, and V. Nadkarni. In situ simulation: challenges and results. *Advances in patient safety: New directions and alternative approaches*, 3:1–18, 2008.

[20] S. R. Porter, M. E. Whitcomb, and W. H. Weitzer. Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 2004(121):63–73, 2004.

[21] A. B. Raij, K. Johnsen, R. F. Dickerson, B. C. Lok, M. S. Cohen, M. Duerson, R. R. Pauly, A. O. Stevens, P. Wagner, and D. S. Lind. Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE transactions on visualization and computer graphics*, 13(3):443–57, 2007.

[22] E. Remolina, J. Li, and A. Johnston. Team Training with Simulated

Teammates. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, volume 2005, pages 1–11. NTSA, 2005.

[23] E. Remolina, S. Ramachandran, R. Stottler, and A. Davis. Rehearsing Naval Tactical Situations Using Simulated Teammates and an Automated Tutor. *Learning Technologies, IEEE Transactions on*, 2(2):148–156, 2009.

[24] J. Rickel and W. Lewis Johnson. Virtual humans for team training in virtual reality. *Proceedings of the Ninth International Conference on Artificial Intelligence*, (July):578–585, 1999.

[25] A. Robb, R. Kopper, R. Ambani, F. Qayyum, D. Lind, L.-m. Su, and B. Lok. Leveraging Virtual Humans to Effectively Prepare Learners for Stressful Interpersonal Experiences. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):662–670, 2013.

[26] D. Traum, S. Marsella, J. Gratch, J. Lee, and A. Hartholt. Multi-party, Multi-issue, Multi-strategy Negotiation for Multi-modal Virtual Agents. *Intelligent Virtual Agents*, pages 117–130, 2008.

[27] D. Traum, J. Rickel, J. Gratch, S. Marsella, U. S. C. Information, A. Way, and M. Rey. Negotiation over Tasks in Hybrid Human-Agent Teams for Simulation-Based Training. *Artificial Intelligence*, pages 441–448, 2003.

[28] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence*, pages 1–16, 2000.

[29] L. Vardoulakis, L. Ring, and B. Barry. Designing relational agents as long term social companions for older adults. *Intelligent Virtual Agents*, 2012.

[30] A. M. von der Pütten, N. C. Krämer, J. Gratch, and S.-H. Kang. It doesnt matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):1641–1650, Nov. 2010.