Training Together: How Another Human Trainee's Presence Affects Behavior during Virtual Human-Based Team Training

Andrew Robb[1]*, Andrea Kleinsmith[2], Andrew Cordar[3], Casey White[4], Adam Wendling[5], Samsun Lampotang[5] and Benjamin Lok[3]

[1] School of Computing, Clemson University, Clemson, SC, USA, [2] Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, USA, [3] Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA, [4] School of Medicine, University of Virginia, Charlottesville, VA, USA, [5] Department of Anesthesiology, University of Florida, Gainesville, FL, USA

Despite research showing that team training can lead to strong improvements in team performance, logistical difficulties can prevent team training programs from being adopted on a large scale. A proposed solution to these difficulties is the use of virtual humans to replace missing teammates. Existing research evaluating the use of virtual humans for team training has been conducted in settings involving a single human trainee. However, in the real world, multiple human trainees would most likely train together. In this paper, we explore how the presence of a second human trainee can alter behavior during a medical team training program. Ninety-two nurses and surgical technicians participated in a medical training exercise, where they worked with a virtual surgeon and virtual anesthesiologist to prepare a simulated patient for surgery. The agency of the nurse and the surgical technician were varied between three conditions: human nurses and surgical technicians working together; human nurses working with a virtual surgical technician; and human surgical technicians working with a virtual nurse. Variations in agency did not produce statistically significant differences in the training outcomes, but several notable differences were observed in other aspects of the team's behavior. Specifically, when working with a virtual nurse, human surgical technicians were more likely to assist with speaking up about patient safety issues that were outside of their normal responsibilities; human trainees spent less time searching for a missing item when working with a virtual partner, likely because the virtual partner was physically unable to move throughout the room and assist with the searching process; and more breaks in presence were observed when two human teammates were present. These results show that some behaviors may be influenced by the presence of multiple human trainees, though these behaviors may not impinge on core training goals. When developing virtual human-based training programs, designers should consider that the presence of other humans may reduce involvement during training moments perceived to be the responsibility of other trainees and also should consider that a virtual teammate's limitations may cause human teammates to limit their own behaviors in corresponding ways (e.g., searching less).

Keywords: virtual humans, team training, social presence, social interaction, human–computer interaction

1. INTRODUCTION

Communication and team work are two essential aspects of safe and effective health-care systems (Weaver and Rosen, 2013). Team training exercises provide valuable opportunities to improve these team skills. Over the last 15 years, team training programs have generated growing interest within the medical community (Buljac-Samardzic et al., 2010). Studies have shown that team training can lead to improvements in team performance and patient safety (Awad et al., 2005; Rabøl et al., 2010). Moreover, team training in health care can positively affect patient outcomes and team processes (Knight et al., 2014; Weaver et al., 2014). However, despite these benefits, team training programs are often hindered by the difficulty of getting the entire team together for training, due to differing schedules and/or emergency situations.

Virtual teammates have the potential to help overcome this difficult problem. When human teammates are unavailable for training, a virtual teammate can stand in for them and allow the training to go forward. Additionally, virtual teammates are consistent in their behavior regardless of time of day or number of hours in use. The use of virtual humans for team training has been investigated by several researchers. Examples include virtual teammates for procedural skills training (Rickel and Lewis Johnson, 1999), military negotiation training (Hill et al., 2003), operations training (Remolina et al., 2005), unmanned Ariel vehicle operation (Ball et al., 2010), postoperative care training (Chuah et al., 2013), and patient safety negotiations (Robb et al., 2016). Similar to team training studies involving teams comprised of humans only, research involving virtual teammates has also demonstrated positive effects on skill development (Kroetz, 1999), team performance (Demir and Cooke, 2014), and issues involving patient safety (Robb et al., 2015b).

A major limitation in this body of work is the absence of studies exploring virtual human-based training programs designed to support multiple human trainees. The majority of virtual human-based training programs yet developed have been explicitly designed to support a single human trainee. When researchers have investigated virtual human-based training programs capable of supporting multiple human teammates, all but one of the human teammates have been played by confederates, i.e., actors who are not part of the research team [e.g., Robb et al. (2015b)]. In these cases, the use of human confederates makes it impossible to judge whether new behaviors emerge when multiple humans work together alongside virtual teammates, as opposed to when a single human works with virtual teammates. If virtual humans are used to support team training in health care by serving as replacements for missing teammates, some training sessions will inevitably involve multiple human trainees working alongside one or more virtual teammates. In these cases, the presence of multiple human trainees raises the possibility of behavioral changes that cannot be observed in training settings involving only a single human. It is important to understand how the presence of multiple human trainees affects behavior before virtual teammates can be widely used as replacements for missing humans during training. In this paper, we present the results of an initial exploration of how behavior can be affected by another

human trainee's presence during virtual human-based team training.

To explore behavior during team training, we developed a medical training exercise to help prepare nurses and surgical technicians to speak up about patient safety issues during surgical procedures. During the exercise, a nurse and a surgical technician worked with a virtual surgeon and a virtual anesthesiologist to prepare a simulated patient for surgery. The virtual surgeon made two decisions that could potentially endanger the patient's safety, which gave the human trainees an opportunity to practice identifying and speaking up about patient safety issues. The agency of the nurse and the surgical technician were varied across three conditions: human nurses and human surgical technicians working together; human nurses working with virtual surgical technicians; and virtual nurses working with human surgical technicians.

We examined the training outcomes to determine whether the presence of a second human teammate altered how participants spoke up to the virtual surgeon; no significant differences were observed. We then examined videos of participants' interactions to identify potential variations in behaviors that may have been caused by the presence of a second human teammate. Three main behaviors were identified for further coding and analysis: how participants spoke up to the virtual surgeon about the patient safety risks, how participants searched for a missing item during the closing count after the surgery, and whether participants experienced more breaks in presence when a second human was present. In general, our results indicate that both positive and negative behavioral variations can result when other humans are present during a virtual human-based training exercise. For instance, we found that participants spent less time searching and searched less thoroughly when working with a virtual teammate, compared to a second human teammate, likely because of limitations in the virtual teammate's ability to move throughout the room. In contrast, we found that when working with a virtual nurse, surgical technicians were more likely to contribute to speaking up about patient safety issues that fell outside of their normal responsibilities. These results indicate that behavior can be affected by the presence of other human teammates, though more research is required to better understand the ways in which these behavior changes may manifest.

2. MATERIALS AND METHODS

We developed a team training exercise in conjunction with operating room (OR) nursing management at the UF Health ORs that supported up to two human teammates (a human nurse and/or a human surgical technician). The specific training scenario was selected by OR nursing management based on their assessment of their current training needs. We then conducted a study exploring whether the presence of a second human teammate affected behavior during the team training session.

2.1. Training Scenario

The goal of this exercise was to help prepare nurses and surgical technicians to speak up about patient safety issues in the OR. During the training exercise, a nurse and a surgical technician

worked with a virtual surgeon and a virtual anesthesiologist to prepare a simulated patient for surgery (represented by a patient mannequin). The exercise was broken into three stages, which corresponded to three of the main steps in the surgical process: the preinduction briefing, the pre-incision timeout, and the closing count. The surgery itself was not simulated. To give participants an opportunity to speak up, the scenario was designed so that the team's surgeon made two decisions that could potentially endanger the patient's safety and which also violated hospital policy. The first incident took place during the pre-incision timeout and the second during the closing count. We refer to these incidents as *speaking up moments*, specifically, the *blood safety moment* and the *closing count moment*. Each stage typically lasted between 5 and 7 min. Participants were not warned that patient safety incidents would occur during the training scenario.

Participants worked with the surgeon and the anesthesiologist to complete the preinduction briefing and the pre-incision timeout prior to the blood safety moment. The surgeon walked the team through two checklists, one in each stage. Both checklists are real checklists used by OR teams at UF Health. The checklist used in the preinduction briefing contained approximately 18 items that were confirmed by the surgeon, and the anesthesiologist contributed two additional pieces of information. Participants were each asked at least one question by the surgeon during this stage as well, and could ask additional questions if desired. The checklist used during the pre-incision timeout was shorter, only containing six items before the blood safety moment occurred.

During the blood safety moment, the surgeon discovered that replacement blood was not yet available, because the blood bank had lost the samples that were drawn and sent down earlier in the patient preparation process. The surgeon became upset and berated the anesthesiologist for failing to notify him earlier, and then instructed him to send down new blood samples immediately. The surgeon then decided to begin the surgery immediately, without waiting for the blood samples to be received and processed by the blood bank. The decision could pose a serious risk to the patient. If the blood samples were to take longer than expected to process, replacement blood may not be available by the time it is needed. If the trainees chose to speak up about this risk, the surgeon repeatedly dismissed their concerns and argued that it would be appropriate to proceed. The surgeon only agreed to wait if the trainees told the surgeon they were going to ask a charge nurse to intervene on their behalf. This speaking up moment was primarily the responsibility of the nurse, as surgical technicians are not responsible for blood-related issues in the UF Health OR.

After the blood safety moment was resolved, the team moved onto the closing count, which takes place once the surgery is complete. At the beginning of this stage, the surgeon directed the nurse and the surgical technician to begin the closing count, during which they counted eight types of items, all of which were physically present in the room and positioned as they would be during a real surgery. The counting policy at UF Health dictates that the nurse and surgical technician work together to count all items, where the surgical technician verbally counts off all items on the sterile field that were used during the surgery, and the nurse verbally counts off all items not on the sterile field. The nurse then confirms whether the numbers are correct, based on an initial

count made before the surgery. All of the counts were correct except for the final item on the list, the hemoclip boats (small objects containing staples). The list specified that two should be present, when only one was actually in the room. Upon notifying the surgeon that an item was missing, he instructed everyone to begin searching for the missing boat. When it could not be found, he ordered an X-ray of the patient to determine if the boat was inside the incision. Upon receipt of the X-ray, the surgeon read it and decided that, as he could not see anything inside the patient, he was going to begin closing the patient's incision. This decision ran counter to the hospital policy, which requires that, in the event of a missing item, the attending radiologist must read the X-ray as well and confirm that nothing is inside the patient before closing the incision. If trainees spoke up about this violation of policy, the surgeon repeatedly dismissed their concerns and argued that it would be appropriate to close the incision. The surgeon only agreed to wait if the trainees told the surgeon they were going to ask a charge nurse to intervene on their behalf. This speaking up moment was the responsibility of both the nurse and the surgical technician, as both participate in the closing count and should be aware of the relevant policies and their implications for patient safety.

2.2. Study Design

Participants were divided into three conditions. In the first condition, a human nurse and a human surgical technician worked with the virtual surgeon and the virtual anesthesiologist. In the second condition, a human nurse worked with a virtual surgical technician, a virtual surgeon, and a virtual anesthesiologist. In the third condition, a human surgical technician worked with a virtual nurse, a virtual surgeon, and a virtual anesthesiologist. All human team members were actual OR personnel; no human confederates were used in this study. The virtual surgeon and virtual anesthesiologist behaved the same in all three conditions.

In the condition containing a human nurse and a virtual surgical technician, the virtual surgical technician did not participate in the first speaking up moment, as blood-related issues are outside of the domain of surgical technicians. During the second speaking up moment, the virtual surgical technician did not immediately speak up but would ask the human nurse a leading question if she did not speak up to the virtual surgeon ("Do you think this is OK? Don't we need to confirm with a radiologist before closing?"). If the human nurse still did not speak up, the virtual surgical technician began speaking up to the virtual surgeon, pausing at certain moments to allow the human nurse to speak up as well. After speaking up three times, if the human nurse had not begun assisting with the speaking up, the virtual surgical technician asked the human nurse to intervene and speak up to the surgeon.

In the condition containing a virtual nurse and a human surgical technician, the virtual nurse spoke up to the virtual surgeon during the first speaking up moment. During the second speaking up moment, the virtual nurse did not immediately speak up but would ask the human surgical technician a leading question if she did not speak up to the surgeon. If the human surgical technician still did not speak up, the virtual nurse began speaking up to the virtual surgeon, pausing at certain moments to allow the human

surgical technician to speak up as well. After speaking up three times, if the human nurse had not begun assisting with the speaking up, the virtual surgical technician asked the human nurse to intervene and speak up to the surgeon.

This condition was further divided into two sub-conditions, where the virtual nurse either spoke up successfully to the virtual surgeon during the first speaking up moment ($n = 12$) or where the virtual nurse gave into the surgeon and allowed him to proceed ($n = 10$). In both the conditions, the virtual nurse and the virtual surgeon argued back and forth four times, before the nurse then either gave in to the surgeon or called her charge nurse. Surgical technicians could also speak up with the nurse if they wanted to; however, there was no priming to do so. The effects of this manipulation are examined in detail in a separate publication (Cordar et al., 2015). It was found that participants were more likely to speak up successfully during the second speaking up moment when they had seen the virtual nurse successfully speak up to the virtual surgeon during the first speaking up moment.

2.3. Participants

A total of 92 trainees participated in this study. Twenty-two pairs of nurses and surgical technicians worked together, twenty-six nurses worked with a virtual surgical technician, and twenty-three surgical technicians worked with a virtual nurse. All trainees were actual nurses or surgical technicians currently working in the UF Health ORs. This study was carried out in accordance with the recommendations of the Institutional Review Board of the University of Florida with written informed consent from all subjects. All subjects also gave written informed consent in accordance with the Declaration of Helsinki.

Of the participants, 76 were females and 16 were males. Participants were an average of 39.8 years old ($\sigma = 11.77$), and age ranged from 23 to 62 years old. On average, participants had been practicing medicine for 11.49 years ($\sigma = 10.1$) and had been working in the operating room for 9.17 years ($\sigma = 9.39$). Of the 92 participants, 57 reported their race as White, 20 as Black, 10 as Asian, 5 as American Indian or Alaska Native, and 1 as Native Hawaiian or Pacific Islander. Three did not report their race.

Completion of the training exercise was required by the hospital, but participation in the study was optional. If an individual declined to participate in the study, no data were collected, and their debriefing exclusively focused on feedback relevant to their performance. All trainees received 1.5 h of continuing education credits. Study participants also received a $10 coffee gift card. Once the study was completed, participants were asked not to discuss the content of the training program with other nurses and surgical technicians at UF Health until the study had been completed.

2.4. Virtual Teammates

The virtual teammates were embedded in a former operating room that had been converted into a simulation center using portable display modules (Chuah et al., 2013). These display modules consist of 40″ 1080p television screens mounted on a rolling stand equipped with a computer and a Microsoft Kinect 2. These units are designed to allow virtual humans to be rapidly

deployed in real environments, which allow training to take place near the participants' workplace (minimizing the travel burden necessary to participate in the training) as well as taking place within environments that closely approximate normal working conditions. These units make use of commodity off-the-shelf hardware, which simplifies acquiring and operating them. The low cost of these units, their high portability, and their ability to be embedded within a range of actual working environments are all advantages compared to HMD-based or CAVE-based virtual training environments. They do come with trade-offs, such as requiring the virtual humans to remain stationary during the exercise and not allowing the virtual teammates to interact with physical objects in the room. However, these limitations are frequently acceptable within a training context, as expressed in the "simulation contract" that requires participants to suspend their disbelief.

Figure 1 shows a human nurse working with three virtual teammates. Virtual humans are rendered from the waist up, making them approximately life size. The Microsoft Kinect 2 is used to track participants' positions in the room, which allows the virtual humans to make eye contact with participants. Perspective correct rendering was not used in this study, as one of the conditions involved two different human participants with two separate perspectives. Instead, a static image of the room was used as a background image, to give the impression that the virtual human was actually present in the room.

The virtual teammates' gaze was controlled by a simple Markov model. When speaking, the virtual teammates looked at whomever they were speaking to, with occasional glances at other teammates or the patient. When listening, the virtual teammates looked at whoever was speaking, or whoever was expected to speak next. Virtual teammates also made occasional glances away when listening. The virtual teammates also blinked and mimicked idle motions when not speaking. When speaking, the virtual teammates occasionally used hand gestures, depending on the content of the speech. All animations were created using prerecorded motion capture. The virtual teammates spoke using prerecorded audio. The virtual surgeon and virtual anesthesiologist resembled average Caucasian males, and the virtual nurse and virtual surgical technician resembled average Caucasian females, as this combination of race and gender is representative of the majority of surgeons, anesthesiologists, nurses, and surgical technicians practicing in the U.S. (Castillo-Page, 2006).

Participants interacted with their virtual teammates using speech and gesture. The virtual teammates were controlled using a Wizard-of-Oz (WoZ) system, whereby the study proctor remotely controlled what the virtual teammates said and did. A WoZ was used to reduce errors associated with current speech recognition and speech understanding technology. Participants were unaware that a WoZ was used to control the virtual teammates. Instead, they were given the impression that the virtual teammates operated autonomously.

The study proctor was a computer scientist who worked closely with medical professionals to develop the training scenario and who was intimately familiar with the training context. The same person proctored the study for every participant, ensuring that
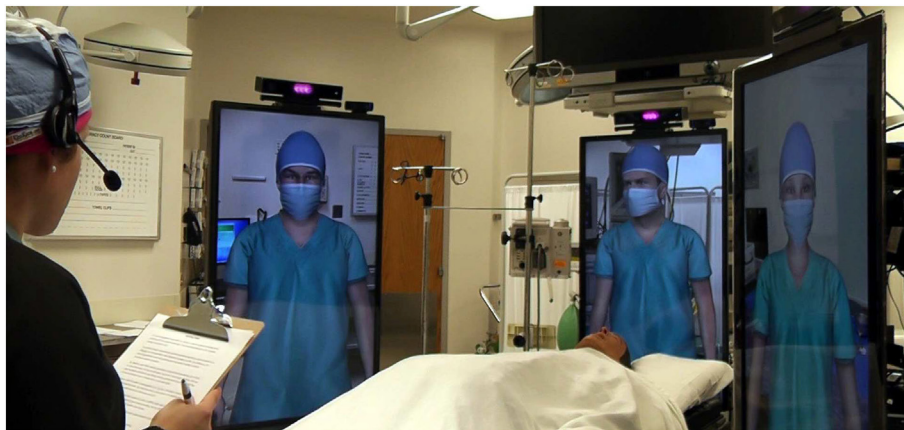
FIGURE 1 | A human nurse works with the virtual surgeon, the virtual anesthesiologist, and the virtual surgical technician (from left to right). The patient mannequin lies on the bed in the middle of the team.

the manner in which virtual teammates were operated remained consistent for all participants. The study proctor controlled all virtual teammates present simultaneously, which was possible due to the highly structured nature of the training exercise, and the fact that the training exercise did not include multiple simultaneous conversations. Additional proctors would be needed to support training exercises involving multiple simultaneous conversations.

The training exercise contained a core "trunk" that remained consistent for all participants, though some variation could occur when participants interjected with questions or comments (this type of variation is typically unavoidable during interpersonal training exercises). The WoZ system intelligently suggested behaviors to the study proctor based on the current position along the core trunk of the training exercise. In addition to these intelligent suggestions, the study proctor also had access to categorized lists containing all of the possible behaviors for each virtual teammate, allowing him to quickly handle questions or comments made by participants that fell outside of the core trunk of the training exercise. The study proctor could see and hear the participants using a live video feed, which allowed him to include important non-verbal cues (such as gaze at a specific teammate) in his decision-making process.

Because the surgeon and the anesthesiologist were always virtual, their behavior was more consistent across all participants than the behavior of the nurse and the surgical technician, who were sometimes played by virtual humans and sometimes by a real participant. The virtual nurse's and virtual surgical technician's behaviors were modeled based on discussions and role playing sessions with content experts about how a typical nurse or surgical technician would behave. The virtual nurse and virtual surgical technician behaved consistently across participants; however, their behavior was not always consistent with the behavior of actual human participants, given that we placed no constraints on their behavior. However, this sort of variability is inescapable when investigating training with real humans.

3. RESULTS AND DISCUSSION

In this section, we present a detailed analysis exploring when and how behaviors were affected by working with a second human teammate. Specifically, we consider the outcome reached during the two speaking up moments, supporting behavior on the part of the surgical technician during the blood safety moment and searching patterns before and during the closing count moment. We also discuss self-reported social presence and observed breaks in presence.

The majority of the data discussed in this section was derived from video recordings of participants collected during the study. Video data were coded by one or more coders, depending on the complexity of the process and the level of judgment required by the coder. All coders are also authors of this paper. We discuss the specific coding process used for each type of behavior in the section where that behavior is analyzed.

3.1. Behavior during the Speaking Up Moments

We first consider whether participants spoke up differently to the virtual surgeon when a second human was present.

3.1.1. Coding Process

Each participant was classified as behaving in one of three ways during the speaking up moments: *stopping the line*, defined as asking a charge nurse or other manager to intervene on behalf of the patient; *speaking up*, defined as challenging the surgeon about the patient safety risk but failing to stop the line; and *no objections*, defined as offering no objections to the virtual surgeon's proposed course of action. This is a simplified coding scheme based on previous research exploring speaking up behaviors (Robb et al., 2015b). The *speaking up* code was created by merging three infrequently used codes that constituted various objections to the surgeon but did not culminate in stopping the line, specifically *filing an incident report*, *shifting responsibility to the surgeon*, and *giving in to the surgeon*.

A single coder assessed each video to determine which outcome was reached by participants. In the event of uncertainty, that video was discussed with another researcher, and agreement was reached about the appropriate code (this was rarely necessary). As this outcome took place on a team-level, participants who worked with a second human teammate were both assigned the same outcome, even if only one of the humans participated in the speaking up process. There was never any disagreement between human teammates about whether the virtual surgeon's proposed course of action was appropriate.

3.1.2. Analysis

The two speaking up moments were analyzed separately. The surgical technician's behavior during the blood safety moment was not included in the analysis because nursing management did not expect surgical technicians to speak up about blood-related issues.

Speaking up behavior during both the blood safety moment and the closing count moment was analyzed using a Fisher exact test to determine whether the presence of a second human teammate altered the manner in which participants spoke up. The test revealed no significant differences between conditions for the blood safety moment ($p = 0.870$) or the closing count moment ($p = 0.728$).

Figure 2 reports the relative percentage of participants in each condition that stopped the line, spoke up, or offered no objections.

3.1.3. Discussion

Prior research has explored whether the agency of an antagonist (e.g., the surgeon who must be spoken up to) impacts the rate and manner in which participants speak up (Robb et al., 2015b). This research found that the antagonist's agency had no significant effect ($p = 0.869$). The research reported in this paper expands upon this finding by examining whether the agency of a partner (e.g., the nurse and the surgical technician who both work together to speak up to the surgeon) impacts the rate and manner in which the participants speak up. The results of this research suggest

that the presence of a human partner does not substantially affect participants' willingness to speak up about patient safety issues during training. Discussions with participants during the debriefing suggested that prior experience, assessments of risk, and institutional policies were some of the factors participants considered when determining how to respond during the patient safety moments.

It is notable that participants stopped the line with much greater frequency during the closing count moment, compared to the blood safety moment. An analysis of the arguments made while speaking up suggests that the presence of a strong and widely known policy governing the closing count made participants more confident in their ability to stop the line. The majority of participants referenced hospital policy during the closing count moment, while policy was infrequently referenced during the blood safety moment. A second, related possibility is that the presence of a strong policy governing the closing count may have meant more participants were aware of the importance of the closing count moment, compared to the blood safety moment. Since the amount of blood required varies from surgery to surgery, it may be appropriate to begin some surgeries without having replacement blood in the room. In contrast to this, if the closing count is off, it is never appropriate to close the patient before consulting with the attending radiologist. However, it should be noted that because the scenario explicitly stated that blood was required for this surgery in both Stage 1 and Stage 2, all participants were aware that replacement blood was important for this surgery.

3.2. Supporting Behavior during the Blood Safety Moment

As noted previously, nursing management did not expect surgical technicians to speak up during the blood safety moment because surgical technicians are not typically responsible for blood-related issues. Accordingly, in the condition where the human surgical technician worked with a virtual nurse, the virtual nurse was programed to speak up to the virtual surgeon about the
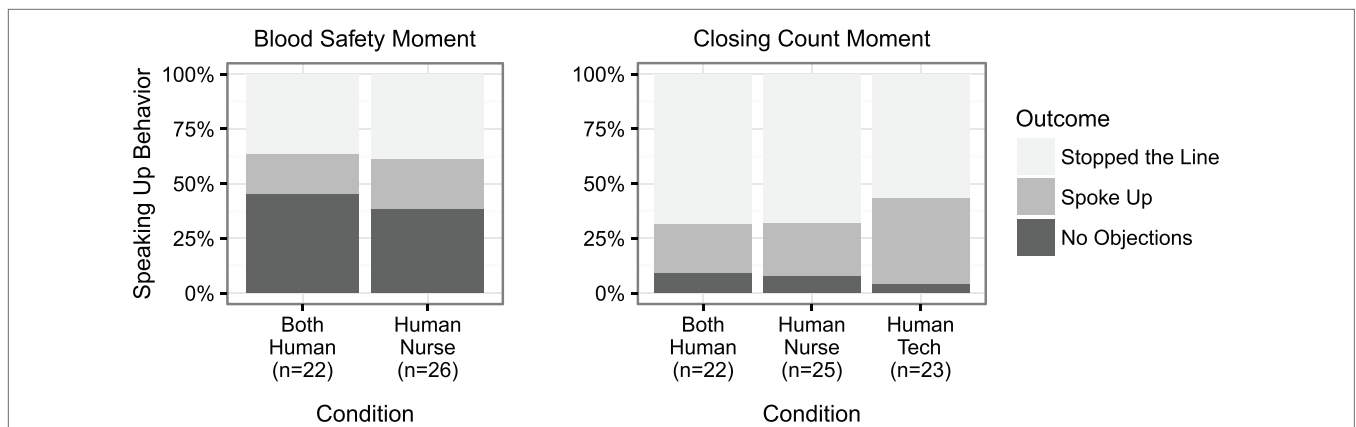

FIGURE 2 | Speaking up rates for the blood safety moment and the closing count moment. Human surgical technicians working alone were not expected to speak up during the blood safety moment, thus their data are not reported. Instead, the virtual nurse spoke up about the safety issue. No significant differences were observed ($p_{Blood} = 0.870$, $p_{ClosingCount} = 0.728$).

blood safety issue. This design allowed us to explore how surgical technicians supported the nurse during the blood safety moment, and whether this support was affected by the agency of the nurse.

3.2.1. Coding Process

A single coder transcribed the speech of all team members during the blood safety moment and recorded whenever the surgical technician verbally supported the nurse or verbally challenged the surgeon about the potential patient safety risk posed by beginning the surgery before blood was available.

3.2.2. Analysis

A Fisher exact test was conducted to determine if the presence of a second human influenced how likely surgical technicians were to support the nurse while she spoke up or otherwise challenged the virtual surgeon. Surgical technicians were significantly more likely to support the virtual nurse while speaking up, compared to the human nurse ($p < 0.01$). **Figure 3** shows the percentage of surgical technicians who initiated speaking up, or who supported the nurse while speaking up in the two conditions.

Out of the twenty-two surgical technicians who worked with a human nurse, only one assisted the nurse with speaking up to the virtual surgeon. This surgical technician began speaking up after the human nurse voiced concern and interleaved her concerns with the nurse as they spoke up together. She also covered up the scalpel blades when the surgeon demanded one be given to him so he could begin the surgery. No surgical technicians who worked with a human nurse initiated speaking up before the human nurse spoke.

Out of the twenty-three surgical technicians who worked with a virtual nurse, nine assisted the virtual nurse with speaking up. Five of these surgical technicians provided minimal support to the virtual nurse, only speaking up once or twice. The most common point at which these surgical technicians spoke up was when the surgeon demanded they give him a scalpel to begin surgery.

FIGURE 3 | **The percentage of surgical technicians who spoke up to the virtual surgeon or who supported the nurse while speaking up during the blood safety moment**. Surgical technicians were significantly more likely to support the virtual nurse ($p < 0.01$).

The remaining four surgical technicians provided more robust support, speaking up repeatedly and sometimes even preempting the virtual nurse, who spoke up only if the surgical technician did not. Four of these nine surgical technicians began speaking up before the virtual nurse voiced any objections.

3.2.3. Discussion

These results suggest that human trainees may be more likely to actively participate in a virtual training program when they do not have other human teammates, particularly in portions of the training that fall outside of the human trainee's normal responsibilities. We consider several potential explanations for this finding below; however, future research is required to determine the degree to which each explanation is valid.

First, training alone may have altered surgical technicians' perceptions of which components of the training session were relevant to them and required their participation. When a human nurse was present, surgical technicians may have perceived the blood safety moment as a training moment aimed at the nurse (because blood safety is not typically the surgical technician's responsibility), and thus making it either inappropriate or unnecessary for her to intervene. However, when training alone, the absence of any other trainee may have caused surgical technicians to perceive that they were being tested throughout the entire scenario, even on matters that typically fall outside of their normal responsibilities. This could explain why surgical technicians were more willing to assist the virtual nurse with speaking up during the blood safety moment, and sometimes even initiate the speaking up process. If this is the case, then training developers must consider that the purpose of different training objectives may be perceived differently when other humans are present, compared to when training alone.

Second, the surgical technicians' hesitancy to assist human nurses with speaking up could be viewed in terms of the bystander effect, where an individual is less likely to provide help when others are also present (Latané and Darley, 1970). Latané and Nida (1981) identify three important elements contributing to the bystander effect: (1) diffusion of responsibility (e.g., when others are present, each individual feels less responsibility for intervening), (2) social influence (e.g., participants look to others' behavior to determine if a situation is really a problem), and (3) audience inhibition (e.g., the threat of negative evaluation prevents people from acting in the presence of uncertainty). The bystander effect is particularly relevant to our observation that no surgical technicians initiated speaking up when a human nurse was present, while four surgical technicians began speaking up before the virtual nurse. When a human nurse was present, surgical technicians may have been less likely to contribute to speaking up about the blood safety issue (which falls outside of their normal responsibilities) for each of the three reasons cited above, especially if the human nurse did not speak up to the surgeon. However, these factors would be significantly reduced when working alone, which may explain why surgical technicians were more willing to assist the virtual nurse with speaking up, or even begin speaking up before the virtual nurse did.
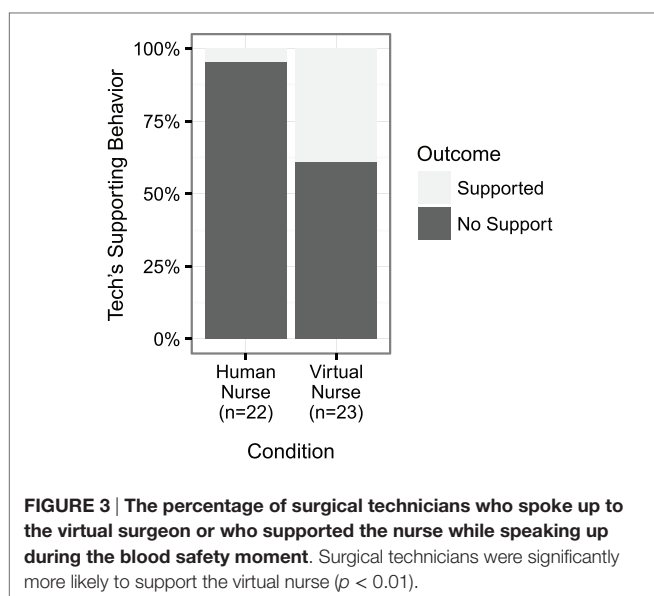
Finally, it is possible that surgical technicians felt that the human nurses were more competent than the virtual nurse, and thus needed less assistance.

One important limitation in these results should be considered: all of the surgical technicians who worked with the virtual nurse observed the virtual nurse speak up to the surgeon. However, there were 10 human nurses who did not speak up to the surgeon about the patient safety risks, which means the experience of the surgical technicians who worked with these nurses differed from the experience of surgical technicians who worked with the virtual nurse. The decision was made not to exclude these 10 surgical technicians from our analysis because the manipulation of the nurse's agency remained valid and because four of the surgical technicians who worked with the virtual nurse began speaking up to the surgeon *before* the virtual nurse, which suggests that the choice to speak up was not solely influenced by whether or not the nurse spoke up first. In contrast, none of the surgical technicians who worked with a human nurse spoke up before the nurse did. This observation strengthens the conclusion that surgical technicians were more likely to challenge the surgeon when working with a virtual nurse, regardless of the nurse's behavior.

3.3. Searching Behavior after the Count Was Discovered to Be Incorrect

Once participants discovered that an item was missing, the virtual surgeon instructed the team to begin searching for the missing item. Participants were free to search using whatever means they felt appropriate. Common searching behaviors included walking around the room, looking on the floor, looking under the drapes covering the patient, and moving surgical equipment stored on tables. During this period, virtual teammates looked around the room to indicate they were searching; however, they could not move due to the constraints of their displays. They also verbally confirmed they had not found the missing item when asked by the surgeon. After searching for approximately 70 s, the virtual anesthesiologist prompted the team to order an X-ray. The virtual surgeon then ordered an X-ray, which was shown on two TV monitors a few seconds later (this process would typically take much longer but was sped up for the purpose of the simulation). The surgeon then announced that he did not see the missing item and was going to begin closing the patient. This began the closing count moment. Many participants continued searching during the closing count moment.

3.3.1. Coding Process

Searching behavior was coded from the moment the virtual surgeon said "You're missing something? What happened?" and continued until the end of the exercise. This moment was chosen to start coding as this was the consistent point at which the surgeon explicitly instructed them to start searching. Two aspects of participants' searching behavior was coded: how they searched (using their eyes, their hands, or their body) and where they searched (the mayo stand, the patient's body, the equipment table, the sponge bag, the back table, the floor by the scrub tech, the floor by the nurse, the floor by the surgeon, the laparoscopic cart, in the trash, and in a location not visible to the camera). Data

collected by the Microsoft Kinect can not be used for the coding process due to obstruction of participants' legs by the surgical bed, and because participants frequently moved items during the search process, making it difficult for a computer to infer accurately what participants were looking at. Instead, searching behaviors were coded manually using video recordings.

Two video coders coded participants' searching behavior. Coding was done using ANVIL (Kipp, 2001). Both coders were experienced at coding using ANVIL. Both coders watched 10 separate videos and then discussed what they had observed about how participants searched for the missing item. They then worked together to establish the coding scheme described above. After the coding scheme had been established, both coders coded five training videos. Interrater reliability was then calculated for these training videos. Both coders achieved very high interrater reliability after the initial coding of the training videos (average $\kappa = 0.860$). Each coder then coded half of the remaining videos, with an additional seven overlapping videos that were used to determine whether the coders maintained consistency over the course of video coding. Once the coding was complete, interrater reliability was assessed again for these seven videos, and the coders were found to have maintained a very high level of interrater reliability (average $\kappa = 0.891$).
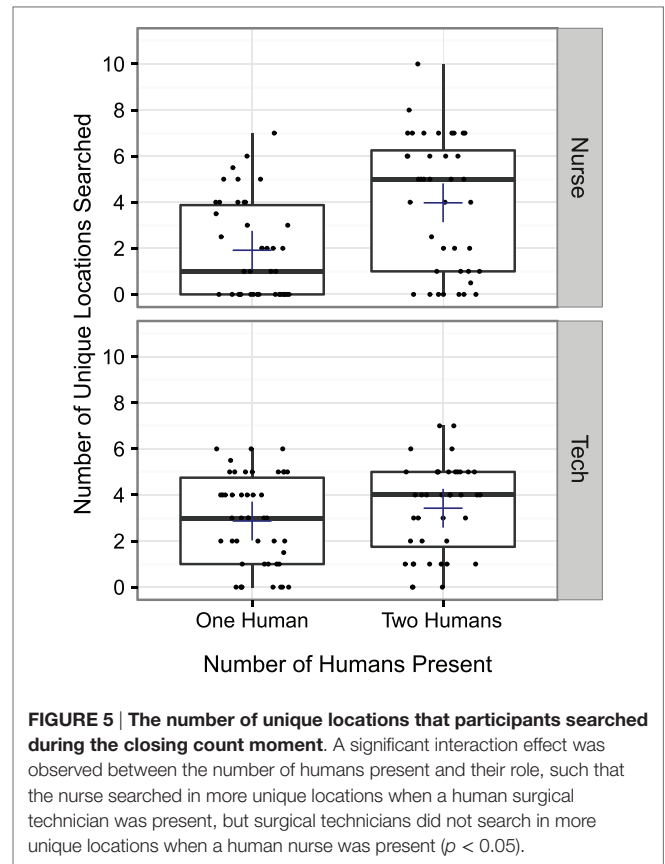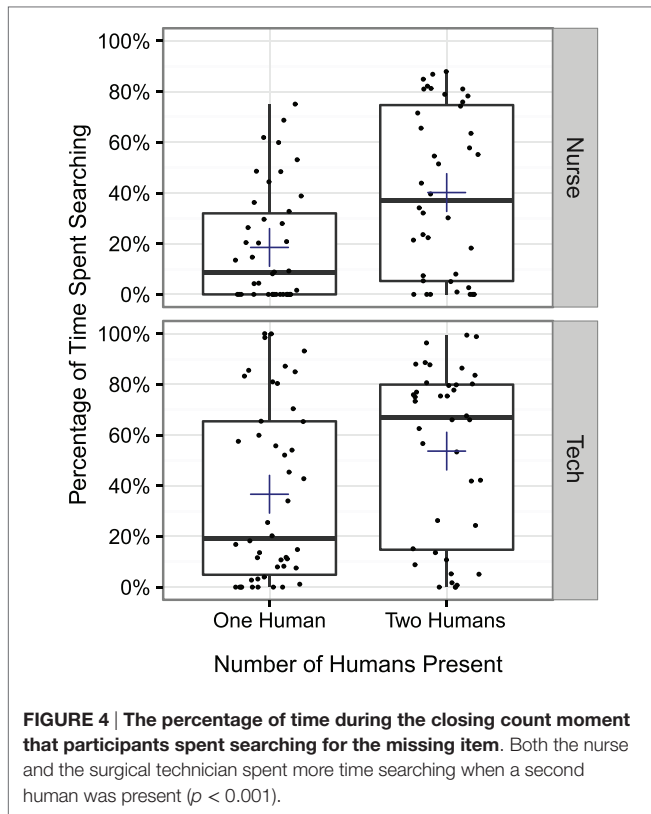
ANVIL produces an XML file containing the start and end point of every code. Once coding was complete, these files were run through a program developed to produce counts of the number of times each code was used, and percentages of the total time that each code was used. Our statistical analysis was conducted on these data.

3.3.2. Analysis

Video data of sufficient quality for the searching analysis were recorded for 65 of the 71 training sessions. The remaining participants either did not give consent for video recording or encountered technical difficulties. In these 65 training sessions, 21 sessions had two human teammates, 21 sessions had one human nurse, and 23 had one human surgical technician.

A $2 \times 2 \times 2$ ANOVA was conducted exploring whether the presence of a second human affected the amount of time people spent searching. The number of humans present and the role of the participant (nurse or surgical technician) served as between-subject factors, and when the searching occurred (before the speaking up moment and during the speaking up moment) served as a within-subjects factor. A main effect was observed for the number of humans present ($p < 0.001$) and for role ($p < 0.001$), but not for period ($p = 0.476$). A significant interaction effect was observed between role and period ($p < 0.05$). No other interaction effects were observed. Participants spent more time searching when a second human was present, and surgical technicians spent more time searching than nurses before the speaking up moment, but not during the speaking up moment. The proportions of time participants spent searching is shown in **Figure 4**.

Having considered whether the presence of a second human affected the amount of time spent searching, we now consider two factors that may provide some insight into the thoroughness of participants' searching behavior: the number of unique locations

FIGURE 4 | The percentage of time during the closing count moment that participants spent searching for the missing item. Both the nurse and the surgical technician spent more time searching when a second human was present ($p < 0.001$).

FIGURE 5 | The number of unique locations that participants searched during the closing count moment. A significant interaction effect was observed between the number of humans present and their role, such that the nurse searched in more unique locations when a human surgical technician was present, but surgical technicians did not search in more unique locations when a human nurse was present ($p < 0.05$).
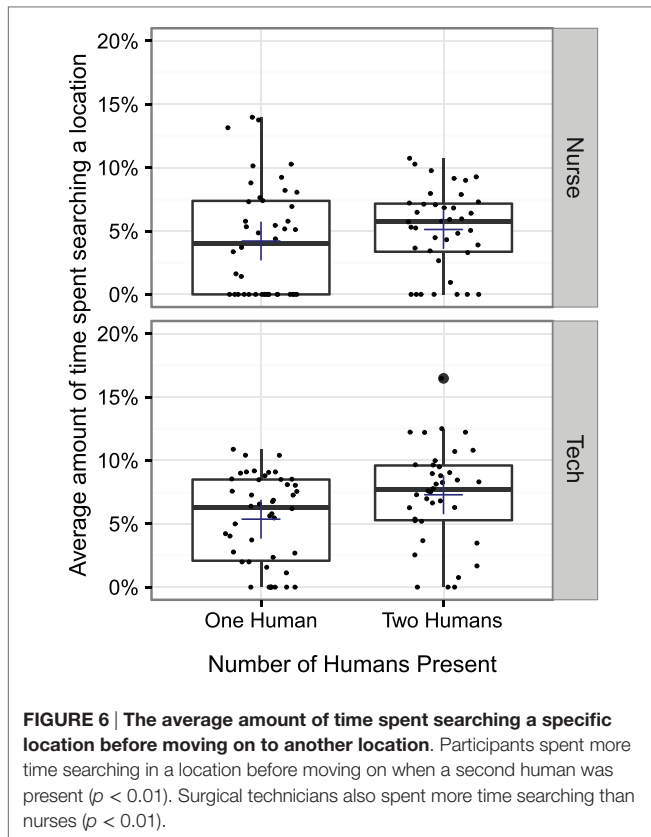
searched and the average amount of time spent searching at a location before moving on to another location. We first consider whether the presence of a second human affected the number of unique places searched by participants. A $2 \times 2 \times 2$ ANOVA was conducted exploring the total number of unique locations participants searched. The same between-subject factors and within-subject factors were used as in the previous test. A main effect was observed for the number of humans present ($p < 0.001$), but not for role ($p = 0.524$) or period ($p = 0.119$). A significant interaction effect was observed between the number of humans present and role ($p < 0.05$). No other interaction effects were observed. Nurses looked in more locations when a human surgical technician was present, but surgical technicians did not look in more locations when a human nurse was present. The amount of different locations participants searched is shown in **Figure 5**.

We now consider whether the presence of a second human affected the average amount of time spent searching any given location before moving on to search another location. The average duration spent searching before transitioning to another location was calculated by dividing the total time spent searching by the total number of locations searched. Three outliers were identified and removed. These participants exceeded the mean by more than three SDs. Each of these three outliers was in the condition with two humans. A $2 \times 2 \times 2$ ANOVA was conducted exploring the average duration spent searching before transitioning to another location. The same between-subject factors and within-subject factors were used as in the previous tests. A main effect was observed for the number of humans present ($p < 0.01$) and for role ($p < 0.01$), but not period ($p = 0.979$). No significant

interaction effects were observed. Participants spent more time searching in a location before moving on when a second human was present, and surgical technicians spent more time than nurses searching in a location before moving on to another location. The average amount of time participants spent searching a location before moving on is shown in **Figure 6**.

3.3.3. Discussion
Participants spent more time searching when a second human was present and also appear to have conducted more thorough searches, as evidenced by the larger number of unique locations searched and the longer consecutive period spent searching a location before moving on. These results suggest that the agency of a teammate can impact behaviors involving motion and interaction with a physical space. It seems plausible that this is linked to the virtual teammate's inability to move or interact with physical objects in the room. This lack of movement may inhibit the human teammate's movements. The human may feel less "able" or less inclined to move around the physical space specifically because the other trainee (i.e., the virtual one) does not move. The possibility that trainees may interact with physical environments, less when working with virtual teammates incapable of interaction with the physical environment, should be considered whenever a training scenario involves physical skills. If reduced interaction with the physical environment poses a problem for training, then solutions involving head-mounted displays with tracked props that provide passive haptics may be considered, as

FIGURE 6 | The average amount of time spent searching a specific location before moving on to another location. Participants spent more time searching in a location before moving on when a second human was present ($p < 0.01$). Surgical technicians also spent more time searching than nurses ($p < 0.01$).

this would allow trainees to work with virtual teammates capable of moving throughout an environment and jointly interacting with physical props. It is possible that this finding that humans limited their own searching behavior when working with a virtual teammate may generalize in the following way, namely, that limitations in a virtual teammate's capabilities may cause human teammates to limit their own behavior in corresponding ways. The implications of this effect should be considered when developing virtual human-based training programs.

As a note, an interaction effect was observed between the number of humans present and the role of participants on the number of unique locations searched, such that surgical technicians did not search more unique locations when a human nurse was present. This is most likely a consequence of the surgical technicians' role in the training exercise, which required them to remain constrained to a specific region of the room. Surgical technicians consistently searched the majority of available locations within this region, regardless of whether a human nurse was present.

3.4. Self-Reported Social Presence

We now consider whether the presence of an additional human teammate impacted social presence. We first examine self-reported social presence collected via surveys. Because this analysis is based on self-report survey data collected from participants, video coding was not required.

3.4.1. Analysis

Social presence was measured using a short, five question self-report survey (Bailenson et al., 2003). The survey is reproduced below. Ellipses were replaced with the name of the specific teammate whose social presence was being reported (e.g., "surgeon," "anesthesiologist," etc.).

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.
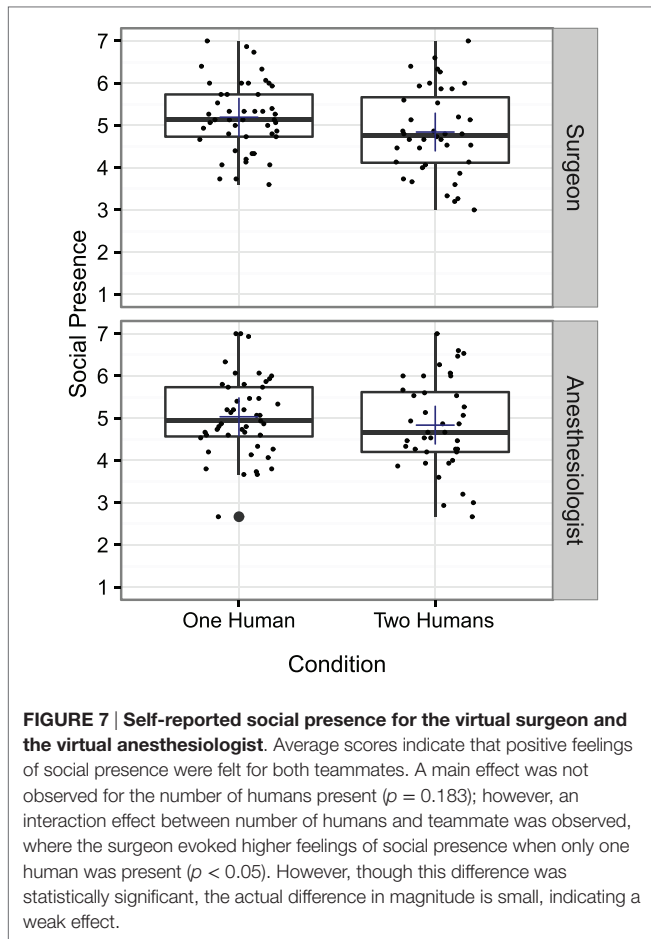5. I perceive the … as being only a computerized image, not as a real person.

Participants completed the survey for each of their teammates, including human teammates, after each stage of the training session. This resulted in a total of nine surveys. The results from each stage were averaged together into a single social presence score for each teammate. Before creating average scores, tests were run to confirm that social presence was correlated between stages. Moderate to strong correlations were found between all stages for the surgeon ($r_{1,2} = 0.718$, $r_{2,3} = 0.794$, $r_{1,3} = 0.619$) and for the anesthesiologist ($r_{1,2} = 0.795$, $r_{2,3} = 0.805$, $r_{1,3} = 0.796$).

Four participants were excluded due to incomplete data. These participants failed to complete multiple surveys due to time constraints. Three additional participants were excluded as outliers, as their social presence data fell below the mean by more than three times the SD. Excluding these seven participants left social presence data for 87 participants: 40 in the both-human condition (20 nurses and 20 surgical technicians), 24 in the human nurse condition, and 23 in the human surgical technician condition.

A $2 \times 2 \times 2$ ANOVA was conducted exploring whether the presence of a second human affected participants' feelings of social presence. Whether or not a second human was present and the role of the participant (nurse or surgical technician) served as between-subject factors, and the teammate whose social presence was assessed (surgeon or anesthesiologist) served as a within-subjects factor. Main effects for the number of humans present ($p = 0.183$) and for role ($p = 0.478$) were not observed; however, a main effect of teammate ($p < 0.05$) was observed, where the surgeon evoked slightly higher feelings of social presence than the anesthesiologist ($\mu_{Surgeon} = 5.033$, $\mu_{Anesth} = 4.941$). A significant interaction effect was observed between the number of humans present and teammate ($p < 0.05$), where the surgeon evoked significantly higher feelings of social presence when only one human was present ($\mu_{OneHuman} = 5.194$, $\mu_{TwoHumans} = 4.843$). No other interaction effects were observed. **Figure 7** reports the social presence scores from participants for the surgeon and the anesthesiologist.

3.4.2. Discussion

These results suggest that the agency of other teammates can have a small, but measurable, effect on self-reported social presence. However, given the magnitude of the effect ($\eta^2 = 0.0016$), it is

FIGURE 7 | Self-reported social presence for the virtual surgeon and the virtual anesthesiologist. Average scores indicate that positive feelings of social presence were felt for both teammates. A main effect was not observed for the number of humans present ($p = 0.183$); however, an interaction effect between number of humans and teammate was observed, where the surgeon evoked higher feelings of social presence when only one human was present ($p < 0.05$). However, though this difference was statistically significant, the actual difference in magnitude is small, indicating a weak effect.

unlikely to meaningfully impact a trainee's experience during training. The low magnitude of the observed difference may also be a consequence of the inherent imprecision associated with self-reported social presence data. Accordingly, we consider a more objective metric in the next section.

Additionally, a significant effect was observed only for the surgeon, not for the anesthesiologist. This inconsistency could be due to the fact that the surgeon has a bigger role in the scenario in two distinct ways: (1) the surgeon speaks more than the anesthesiologist and (2) the surgeon's personality is bigger and more abrasive than the anesthesiologist's. Prior research has found similarly mixed results. In a previous study that varied the agency of a surgeon and anesthesiologist using human confederates, a small but statistically significant decrease in the virtual anesthesiologist's social presence was observed when a human surgeon was present, but this effect was not observed for the virtual surgeon (Robb et al., 2015a).

3.5. Breaks in Presence during the Speaking Up Moments

Social presence is a manifestation of a larger component known as presence. Presence can be defined as occurring whenever a virtual experience is perceived as if it were actually occurring (Lee, 2006). While self-reported measures of presence, like those

considered in the previous section, are valuable metrics, they are most effective when combined with physiological or behavioral measures (Slater, 1999). Place presence research, which focuses on the feeling that you are actually in a real place in a virtual environment, has successfully leveraged common physiological and behavioral responses to physical danger (such as standing on the edge of a cliff) as a measure of place presence (Schuemie et al., 2001). Unfortunately, generalizable physiological or behavioral measures of social presence have proven difficult to develop, primarily due to the lack of physiological or behavioral metrics that are consistent for all people, cultures, and scenarios.

However, one behavioral metric used in place presence research can also be applied to social presence, namely *breaks in presence* (BIP). In place presence research, BIPs are defined as "any event whereby, for the participant, at that moment, the real world becomes apparent, and for the duration of that event, the participant acts and responds more to the real-world setting than to the virtual world" (Brogni et al., 2003). For social presence research, BIPs can be defined as moments in which participants behave in a manner inconsistent with how they would behave in a real-world social setting.

BIPs are typically detected using one of three techniques. In the first, the participant records whenever they experience a break in presence, either through a button press or by verbally stating that a break occurred (Slater and Steed, 2000). This method is helpful in that it allows all breaks in presence to be captured, even when they are not accompanied by an external manifestation. However, it also has the potential to induce more breaks in presence by further reinforcing that what the participant is experiencing is not real. A second method artificially creates a BIP through technological "failures," such as deliberately induced lag or flicker (Chung and Gardner, 2009). A third method relies on external identification of BIPs based on an analysis of participants' behavior (Baren and IJsselsteijn, 2004). This method has the advantage of not requiring participants to self-identify when a BIP occurs, which is not always appropriate during virtual reality research. A disadvantage is that it is not possible to capture internal BIPs that do not lead to external behavioral changes. In this study, we explore BIPs using the third method due to the impracticality of asking participants to report when a BIP occurs during training, as this could detract from the learning objectives. The second method was also inappropriate given that it artificially induces BIPs, which would not allow us to explore whether the presence of other humans causes them to occur more frequently.

3.5.1. Coding Process

Two video coders examined the videos of participants' interactions for BIPs. Each coder coded approximately 60% of the videos, thus creating some overlap between the two coders. Coders were instructed to mark any behaviors that would have been inappropriate in a real operating room and to note their degree of certainty that this constituted a BIP as *uncertain*, *somewhat certain*, or *very certain*. Coders were also instructed to explain why they thought this moment constituted a BIP. After all of the videos were coded, the two coders examined each BIP together and came to a decision about whether the moment constituted an actual BIP.

Behaviors that were commonly recorded as BIPs include the following: laughter, rolling eyes, smiles at inappropriate moments (e.g., surgeon is yelling at participants), and dancing or exaggerated body motions. Other behaviors that were often associated with a BIP but may not constitute a BIP in and of themselves include raised eyebrow, grimaces, a patronizing tone of voice, and shared eye contact between participants. BIPs frequently occurred during or soon after the surgeon complained about the missing item or otherwise argued with participants.

3.5.2. Analysis
BIPs were observed for 32 of the 92 participants. BIPs were observed for 22 participants in the condition with two humans ($N = 50$) and for 10 participants in the conditions with one human ($N = 44$). The relative proportions of participants for whom a BIP was observed in each condition and role are reported in **Figure 8**.

A logistic regression was performed to explore whether BIPs were observed more frequently when a second human was present. The observation of one or more BIPs was used as the dependent variable, and the number of humans present and their role were used as independent variables. A main effect was observed for the number of humans present ($p < 0.01$), but not for role ($p = 0.579$). No interaction effects were observed. BIPs were observed more often when a second human was present.

Next, we sought to determine whether BIPs were associated with lower self-reported social presence scores. Correlation-based statistical methods were not employed during this portion of our analysis because BIPs were not observed for two-thirds of our participants, meaning the distribution of the number of BIPs

FIGURE 8 | Participants were approximately twice as likely to experience a break in presence when another teammate was present ($p < 0.01$).

observed per participant was highly non-normal. Instead, participants were binned according to whether a BIP was observed for or not. A $2 \times 2 \times 2$ ANOVA was then conducted to determine whether participants for whom a BIP was observed also reported lower social presence scores for the virtual surgeon and the virtual anesthesiologist. Whether or not a BIP was observed for a participant and if a second human was present served as between-subject factors, and the teammate whose social presence was assessed (surgeon or anesthesiologist) served as a within-subjects factor. No main effects were observed for observed BIPs ($p = 0.235$), the number of humans present ($p = 0.199$), or for the teammate ($p = 0.969$). Additionally, no interaction effects were observed.

3.5.3. Discussion
These results suggest that participants are more likely to exhibit BIPs when other humans are present. However, the results did not indicate that exhibiting BIPs were associated with lower self-reported feelings of social presence. This is surprising given previous associations between BIPs and place presence (Brogni et al., 2003). A potential explanation for the lack of association between BIPs and self-reported social presence is that our method may not capture all of the BIPs that occur, particularly in the condition with only a single human. Because BIPs were detected based on external behavior, it may be that some participants experienced BIPs internally but did not express them externally. More so, it may be that the increase in BIPs observed when a second human is present actually represents an increased likelihood to externalize a BIP when other humans are present, as opposed to an increased likelihood to experience a BIP. If this is so, then externally assessed BIPs would not be expected to correlate with self-reported measures of social presence, given that they may not always accurately reflect the internal state of the participant. However, this conclusion is by no means certain. When two humans were present, some of the BIPs observed involved interaction between participants (e.g., shared gaze or laughter); however, other BIPs mirrored the behaviors observed when only one human was present (e.g., smiling or grimacing inappropriately at the surgeon). Additionally, BIPs were not always observed for both participants when two humans were present. Out of the 13 pairs of humans for whom BIPs were observed, BIPs were observed for only one human in four of the pairs.

Observing more BIPs when two humans were present may have implications for team training. The behaviors classified as BIPs are often associated with amusement or sarcasm, which could indicate that participants were less inclined to take the virtual surgeon seriously during the training exercise. However, it is important to recognize that this did not appear to affect the outcomes of the training exercise: participants did not speak up more readily to the virtual surgeon when a second human was present. The increased rate at which BIPs were observed may be more relevant when a training exercise is used for assessment purposes, as opposed to educational purposes, given that the inappropriate behaviors that constitute a BIP could have a negative impact on participants' assessments. Finally, these BIPs were specifically observed during the speaking up moment, during which the surgeon was argumentative and displayed
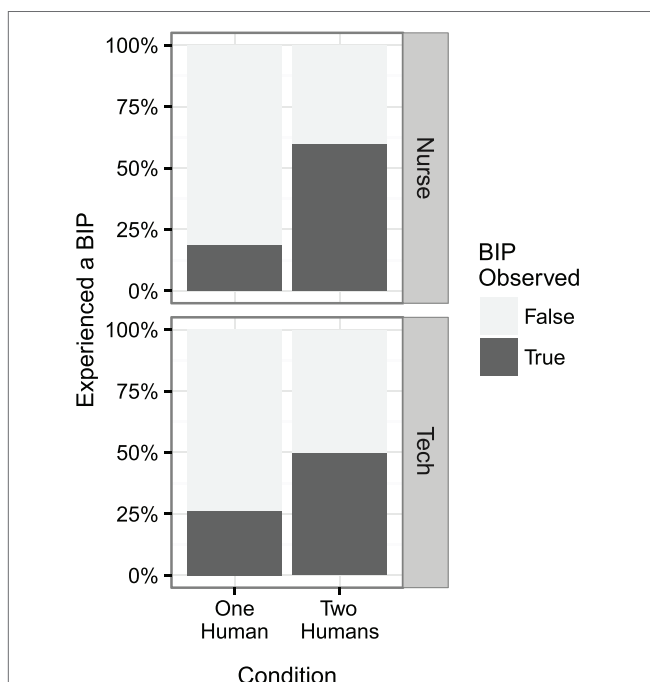
characteristics associated with the stereotype of a difficult and angry surgeon. The inclusion of a second human teammate may not increase the rate at which BIPs occur in settings that do not involve confrontation or stereotypical behavior.

4. CONCLUSION

To summarize, we implemented and examined a virtual human-based training program that aimed to prepare nurses and surgical technicians to speak up about patient safety issues. We evaluated whether the presence of a second human teammate affected how participants spoke up to the virtual surgeon and found no significant effects. We then identified and investigated three behaviors that were affected by the presence of a second human teammate. Our results suggest that the inclusion of additional human teammates can produce both positive and negative effects on behavior. In this experiment, participants searched less thoroughly when working with a virtual teammate (an undesirable effect), likely because the virtual teammate was limited in its ability to contribute to the search, given it was unable to move throughout the room or manipulate physical objects. However, surgical technicians were more likely to assist with speaking up during the blood safety moment when working with a virtual nurse (a desirable effect), likely because they perceived the blood safety moment as more relevant to their own learning because no other trainees were present. We also found that participants exhibited more breaks in presence when working with another human, though more research is required to determine whether this represents an increase *actual number* of breaks in presence or merely an increase in the *rate of expression* of internal breaks in presence. Despite these observed differences, our results also suggest that the inclusion of additional human teammates will not necessarily impact the behavioral outcomes of a training simulation, as seen by the lack of significant differences in speaking up behavior with the virtual surgeon. Self-reported social presence was also not strongly impacted by the inclusion of a second human teammate. Taken together, our results indicate that while the presence of an additional human trainee did not significantly affect the training outcomes of this virtual training exercise, it did cause subtle but notable differences in trainee behavior that point to the importance of further research into interpersonal interaction effects during virtual training. Until this research is conducted, developers of training programs should carefully consider the implications of behavioral variation on training objectives when developing training programs where agency may be varied between training sessions. Specifically, developers should consider that working with others may cause trainees to perceive some aspects of training programs as less relevant to themselves, that a virtual human's limitations may cause humans to limit their own behavior in corresponding ways, and that more "inappropriate" behavior (e.g., breaks in presence) may be observed during virtual human-based training programs when training with other humans.

AUTHOR CONTRIBUTIONS

AR, AK, AC, and BL designed the research; CW, AW, and SL provided medical expertise during research design and study development; AR developed the virtual human architecture to support varying the agency of a teammate; AR and AC developed the virtual human visualization; AC proctored the experiment; AR, AK, and CW coded the data; AR analyzed the data; AR and AK wrote the paper; and all authors revised the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fict.2016.00017

REFERENCES

Awad, S. S., Fagan, S. P., Bellows, C., Albo, D., Green-Rashad, B., Garza, M. D. L., et al. (2005). Bridging the communication gap in the operating room with medical team training. *Am. J. Surg.* 190, 770–774. doi:10.1016/j.amjsurg.2005.07.018

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Pers. Soc. Psychol. Bull.* 29, 819–833. doi:10.1177/0146167203253270

Ball, J., Myers, C., Heiberg, A., Cooke, N. J., Matessa, M., Freiman, M., et al. (2010). The synthetic teammate project. *Comput. Math. Organ. Theory* 16, 271–299. doi:10.1007/s10588-010-9065-3

Baren, J. V., and Ijsselsteijn, W. (2004). "Measuring presence: a guide to current measurement approaches," *OmniPres project IST-2001-39237, Project Report.*

Brogni, A., Slater, M., and Steed, A. (2003). "More breaks less presence," in *Presence 2003: The 6th Annual International Workshop on Presence*, 1–4.

Buljac-Samardzic, M., van Doorn, C. M. D., van Wijngaarden, J. D., and van Wijk, K. P. (2010). Interventions to improve team effectiveness: a systematic review. *Health Policy* 94, 183–195. doi:10.1016/j.healthpol.2009.09.015

Castillo-Page, L. (2006). *Diversity in the Physician Workforce: Facts & Figures 2006.* Washington, DC: Association of American Medical Colleges.

Chuah, J. H., Lok, B., and Black, E. (2013). Applying mixed reality to simulate vulnerable populations for practicing clinical communication skills. *IEEE Trans. Vis. Comput. Graph* 19, 539–546. doi:10.1109/TVCG.2013.25

Chung, J., and Gardner, H. J. (2009). "Measuring temporal variation in presence during game playing," in *The 8th International Conference on Virtual Reality Continuum and Its Applications in Industry*, 163–168.

Cordar, A., Robb, A., Wendling, A., Lampotang, S., White, C., and Lok, B. (2015). "Virtual role-models: using virtual humans to train best communication practices for healthcare teams," in *International Conference on Intelligent Virtual Agents*. Springer International Publishing, 229–238.

Demir, M., and Cooke, N. J. (2014). Human teaming changes driven by expectations of a synthetic teammate. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 58, 16–20. doi:10.1177/1541931214581004

Hill, R., Gratch, J., Marsella, S., and Rickel, J. (2003). Virtual humans in the mission rehearsal exercise system. *Künstlich Intelligenz* 17, 1–10.

Kipp, M. (2001). "Anvil – a generic annotation tool for multimodal dialogue," in *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, 1367–1370.

Knight, L. J., Gabhart, J. M., Earnest, K. S., Leong, K. M., Anglemyer, A., and Franzon, D. (2014). Improving code team performance and survival outcomes: implementation of pediatric resuscitation team training. *Crit. Care Med.* 42, 243–251. doi:10.1097/CCM.0b013e3182a6439d

Kroetz, A. (1999). *The Role of Intelligent Agency in Synthetic Instructor and Human Student Dialogue.* Ph.D. thesis, University of Southern California, Los Angeles.

Latané, B., and Darley, J. M. (1970). *The Unresponsive Bystander: Why Doesn't He Help?* New York: Prentice Hall.

Latané, B., and Nida, S. (1981). Ten years of research on group size and helping. *Psychol. Bull.* 89, 308–324. doi:10.1037/0033-2909.89.2.308

Lee, K. M. (2006). Presence, explicated. *Commun. Theory* 14, 27–50. doi:10.1093/ct/14.1.27

Rabøl, L. I., Østergaard, D., and Mogensen, T. (2010). Outcomes of classroom-based team training interventions for multiprofessional hospital staff. A systematic review. *Qual. Saf. Health Care* 19:e27. doi:10.1136/qshc.2009.037184

Remolina, E., Li, J., and Johnston, A. E. (2005). "Team training with simulated teammates," in *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, Vol. 2005 (NTSA), 1–11.

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

Robb, A., Cordar, A., Lampotang, S., White, C., Wendling, A., and Lok, B. (2015a). Teaming up with virtual humans: how other people change our perceptions of and behavior with virtual teammates. *IEEE Trans. Vis. Comput. Graph.* 21, 511–519. doi:10.1109/TVCG.2015.2391855

Robb, A., White, C., Cordar, A., Wendling, A., Lampotang, S., and Lok, B. (2015b). A comparison of speaking up behavior during conflict with real and virtual humans. *Comput. Human Behav.* 52, 12–21. doi:10.1016/j.chb.2015.05.043

Robb, A., Kleinsmith, A., Cordar, A., White, C., Lampotang, S., Wendling, A., et al. (2016). Do variations in agency indirectly affect behavior with others? *IEEE Trans. Vis. Comput. Graph.* 22, 1–1. doi:10.1109/TVCG.2016.2518405

Schuemie, M. J., van der Straaten, P., Krijn, M., and van der Mast, C. A. (2001). Research on presence in virtual reality: a survey. *Cyberpsychol. Behav.* 4, 183–201. doi:10.1089/109493101300117884

Slater, M. (1999). Measuring presence: a response to the Witmer and Singer Presence Questionnaire. *Presence (Camb)* 8, 1–13. doi:10.1162/105474699566477

Slater, M., and Steed, A. (2000). A virtual presence counter. *Presence (Camb)* 9, 413–434. doi:10.1162/105474600566925

Weaver, S. J., Dy, S. M., and Rosen, M. A. (2014). Team-training in healthcare: a narrative synthesis of the literature. *BMJ Qual. Saf.* 23, 359–372. doi:10.1136/bmjqs-2013-001848

Weaver, S. J., and Rosen, M. A. (2013). "Team-training in healthcare: brief update review," in *Making Health Care Safer II: An Updated Critical Analysis of the Evidence for Patient Safety Practices*, eds P. G. Shekelle, R. M. Wachter, and P. J. Pronovost (Rockville, MD: Agency for Healthcare Research and Quality). Comparative Effectiveness Review No. 211. (Prepared by the Southern California-RAND Evidence-based Practice Center under Contract No. 290-2007-10062-I.) AHRQ Publication No. 13-E001-EF. Available at: www.ahrq.gov/research/findings/evidence-based-reports/ptsafetyuptp.html