Do Variations in Agency Indirectly Affect Behavior with Others? An Analysis of Gaze Behavior

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

Fig. 1: A participant works with a human actor playing a surgeon and a virtual anesthesiologist to prepare a simulated patient for surgery.

Abstract—In a group setting, it is possible for attributes of one group member to indirectly affect how other group members are perceived. In this paper, we explore whether one group member's agency (e.g. if they are real or virtual) can *indirectly* affect behavior with other group members. We also consider whether variations in the agency of a group member *directly* affects behavior with that group member. To do so, we examined gaze behavior during a team training exercise, in which sixty-nine nurses worked with a surgeon and an anesthesiologist to prepare a simulated patient for surgery. The agency of the surgeon and the anesthesiologist were varied between conditions. Nurses' gaze behavior was coded using videos of their interactions. Agency was observed to directly affect behavior, such that participants spent more time gazing at virtual teammates than human teammates. However, participants continued to obey polite gaze norms with virtual teammates. In contrast, agency was not observed to indirectly affect gaze behavior. The presence of a second human did not affect participants' gaze behavior with virtual teammates.

Index Terms—Virtual humans, Gaze, Agency

1 Introduction

Many social encounters involve groups of people. In group settings, it is possible for attributes of one group member to indirectly affect how other group members are perceived. For instance, Hebl and Mannix determined that participants who observed a job applicant sitting next to a socially stigmatized individual (in this case, one who was obese) rated the non-obese job applicant as less qualified and less likely to get the job than participants who saw the same applicant seated next to a non-obese individual [12]. Changes in the weight of the non-job applicant indirectly influenced how the job applicant was perceived. Walker and Vul observed a similar effect, though in the opposite direction [30]. They found that photos of individuals were rated as more attractive when those individuals were in the company of other people, compared to when all other people had been removed from the photo. In this case the presence of other group members indirectly affected the perceived attractiveness of an individual whose appearance had not changed.

As virtual agents become more commonplace in society, it is important to consider whether behavior with virtual group members is affected by the composition of the group as a whole. It is known that the agency of others, where agency refers to whether someone is a real human or a virtual character, can affect how people behave with them. For instance, Hoegen and colleagues found that, during the iterated prisoner's dilemma, people are more willing to violate social norms with virtual agents than with other humans [15], and Lucas and colleagues found that people are more willing to self-disclose personal information to virtual agents than human avatars [21]. Given that the agency of an individual can *directly* affect behavior with that individual, it is also possible that changes in the agency of one group member will *indirectly* affect behavior with other group members whose

- *Andrew Robb is with the School of Computing at Clemson University. E-mail: arobb@clemson.edu*
- *Andrea Kleinsmith is with the Deptartment of Information Systems at the University of Maryland, Baltimore County. E-mail: andreak@umbc.edu*
- *Andrew Cordar and Benjamin Lok are with the Department of Computer Science and Engineering at the University of Florida. E-mail: {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 21 Sept. 2015; accepted 10 Jan. 2016. Date of publication 20 Jan. 2016; date of current version 19 Mar. 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2016.2518405

agency has not changed. Despite its relevance to the use of virtual agents in group settings, this possibility that variations in agency can indirectly affect behavior with others has not yet been explored. This research is relevant to many uses of virtual humans, but is especially relevant within the domain of education and training. If, for instance, the presence of other humans indirectly causes virtual agents to be perceived as less capable or less relevant, then trainees could be subtly influenced to give less importance to the tasks or contributions of their virtual teammates, which could degrade their educational experience.

In this paper, we examine how the agency of one group member can *indirectly* affect behavior with other group members. We also consider how the agency of a group member can *directly* affect behavior with that group member. Specifically, we examine gaze behavior during a team training exercise involving a surgeon, an anesthesiologist, and a nurse. Direct and indirect effects of agency are illustrated visually in Figure 2. Gaze behavior was selected as a metric for two reasons. First, gaze behavior is primarily automatic and unconscious in nature. Indirect influences that occur within a group, such as the previous example of how stigmas concerning obesity can be contagious, typically manifest themselves at an unconscious level, rather than a conscious one. Thus, it seems likely that if variations in agency can indirectly affect behavior, this would manifest itself most readily in an unconscious behavior. Second, gaze behavior was selected as a metric because the *direct* effect agency has on gaze behavior has been explored in other research [26]. Rehms and André explored how variations in the agency of one group member affected gaze with that group member. This prior work can be used as a baseline to assess the validity of our findings concerning the *direct* effect of agency on behavior. Replicating their results concerning the direct effect of agency will strengthen our ability to draw conclusions about whether agency can indirectly affect behavior.

2 Related Work

2.1 Gaze with Virtual Humans

Rehm and André explored gaze behavior as a measure of engagement during interactions with real and virtual humans [26]. Two human participants and one virtual human played Mexicali, a simple dice game involving deception. Mexicali is played in pairs. Players take turns rolling the dice and telling their score to their partner. Players always have to say a score higher than the previously announced score, even if their actual score was lower. Players attempt to determine if someone is telling the truth, or lying. If a person is caught lying, they lose the round. In this experiment, the three players took turns playing with each other. This allowed data to be collected concerning interactions between a human and a virtual human, and between two humans. Players' interactions were video recorded, and these videos were later used to manually code each player's gaze behavior. Results showed that participants spent more time gazing at virtual partners while listening to them speak, but did not gaze at them more while participants were speaking. Additionally, participants stared at their virtual partners longer than their human partners, which suggests that participants violated polite gaze norms with their virtual partners.

While a large body of research pertaining to virtual human gaze exists, it has largely focused on developing gaze models that allow virtual agents to mimic human gaze behavior (e.g. [19, 20, 29, 31]). Pejsa et al. divide gaze models into two categories: data-driven models, which are based on either eye-tracking data or hand-annotated video, and procedural models, which are based on psychological and kinematic rules governing human gaze [24]. Gaze models have typically been evaluated in user studies, where participants interact with virtual agents possessing different gaze models. Though less common, other researchers have evaluated gaze models by comparing virtual agents to actual humans. The majority of these experiments have used recorded video of humans, rather than real humans who physically occupy the same space as participants (e.g. [24]). Various metrics have been used to assess the impact of different gaze models, including a range of self-report questionnaires, including presence [25], social presence [25], satisfaction [13], enjoyment [17], naturalness [24], and personality [13]. Non-verbal behavior [23], interpersonal distance [4], task performance [1, 13, 24] and eye tracking data [25] have also been used as metrics.

Rehm and André's study represents one of the few experiments that has explored whether gaze with virtual agents differs from gaze with humans. When other researchers have compared virtual humans to actual humans, they have rarely examined participants' gaze behavior. However, at least one study has compared gaze with humans to gaze with robots. Yu et al. explored gaze with real humans and robots and also found that participants spent more time looking at a robot's face than a human's face [32]. Taken together, these studies suggest that agency does have a direct influence on gaze behavior, though the exact reason for this effect remains unclear. However, it remains to be seen if variations in agency will be capable of indirectly affecting behavior with others.

2.2 The Function of Gaze in Conversation

Gaze and speech are closely linked [28]. During conversation, gaze can be used to emphasize speech, communicate emotions and relational information, gather information from listeners, communicate attention and comprehension to a speaker, regulate the flow of a conversation, and regulate visual information in order to reduce distraction [16, 3]. Gaze behavior differs significantly while listening compared to while speaking: people spend more time looking at someone while

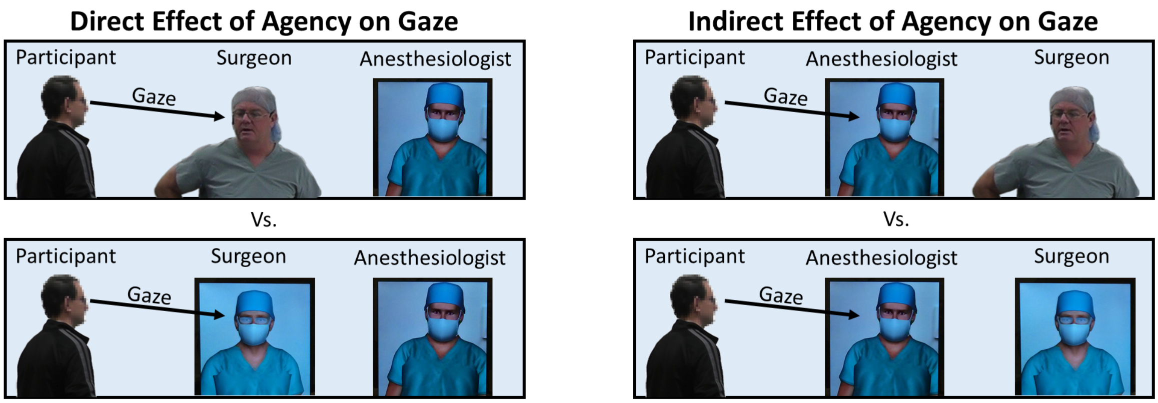


Fig. 2: In this paper, we examine whether agency directly and indirectly affects gaze with a teammate. To examine the direct effect of agency, we vary the agency of one teammate (e.g. the surgeon) and look at gaze with that teammate. To examine the indirect effect of agency, we vary the agency of one teammate and look at gaze with *the other teammate*, whose agency is held constant (e.g. the anesthesiologist). Direct and indirect comparisons were made for both the surgeon and the anesthesiologist.

listening than while speaking (75% vs 40% of the time) [2]. The number of people involved in a conversation can also affect gaze behavior: as the number of people present increases, speakers spend more time looking at their listeners [29].

Gaze aversions play an important role in communication. Though a great deal of information can by obtained by maintaining eye contact, people instead gaze at others in short, intermittent bursts, which are punctuated by averting gaze [3]. Gaze aversion while speaking has been associated with improved cognitive performance [22] and facilitates information recall [10]. Markson et al. found that cognitive performance only improved when gaze was averted from human faces, not other complex visual stimuli [22]. These findings have led researchers to speculate that gaze aversions manage the cognitive load associated with social interactions [7]. Doherty-Sneddon and Phelps explored the effect cognitive load had on children's gaze aversion during face-to-face interactions and video-mediated interactions. Questions of various difficulty (e.g. math, verbal reasoning, etc) were used to vary the current level of cognitive load. Gaze aversion increased when answering more difficult questions, though they found that gaze aversion was more common in face-to-face interactions than in video-mediated interactions. [8]. The above studies have specifically explored the effect of gaze aversions while thinking and while speaking. Fewer studies have explored the effect of gaze aversions while listening. Ehrlichman explored the effect of gaze aversions while looking at either a face or an oval on cognitive load while listening, thinking, and speaking, and found that gaze aversions had no effect on cognitive load, contrary to other research [9]. However, he did report that gaze aversions increased while listening when looking at an oval, compared to a face. Participants looked at the face for approximately 75% of the time, compared to 50% for the oval.

This research highlights the sensitivity of gaze to external stimuli. Gaze patterns are heavily influenced by visual, auditory, cognitive, and social stimuli. However, gaze patterns can also be remarkably consistent over time, between individuals, and even across cultures [3]. This combination of sensitivity and consistency makes gaze a highly useful metric for social experimentation.

3 Methods

We explored gaze in the context of a medical team training simulation. Nurses worked with a surgeon and an anesthesiologist to prepare a simulated patient for surgery, and the agency of the surgeon and the anesthesiologist was varied between conditions. This scenario required all teammates to interact with each other, which allowed us to investigate both direct and indirect effects of agency on gaze behavior.

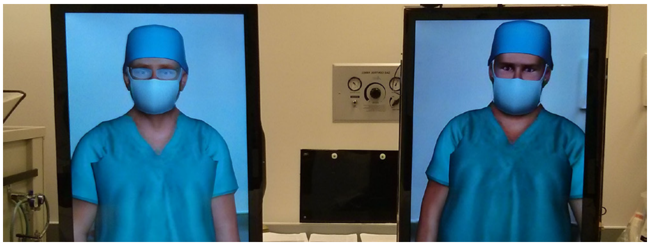
3.1 Participants

A total of 69 operating room (OR) nurses (53 female) participated in this study. All participants were employed by the UF Health Hospital at the time of the study. Participants were an average of 42.4 years old, and age ranged from 24 years to 68 years old with a standard deviation of 12.5 years. On average, participants had worked as a nurse for 20.2 years, and had worked in the OR 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.

Completing in the training exercise was mandated by the hospital, but participation in the study was optional. All nurses received 1.5 hours of continuing education credits, which are required for license renewal as a nurse. Study participants also received a $10 coffee gift card.

3.2 Conditions

Participants were divided into three conditions. The agency of the surgeon and the anesthesiologist was varied between conditions. Participants either worked with: 1) a virtual surgeon and a virtual anesthesiologist, 2) a human actor playing a surgeon and a virtual anesthesiologist, or 3) a virtual surgeon and a human actor playing an anesthesiologist. These conditions are shown in Figure 3. A fourth condition involving a human surgeon and a human anesthesiologist was not included due to a limited pool of participants and the limited availability of the confederates who played the surgeon and the anesthesiologist.



(a) The Virtual Surgeon and the Virtual Anesthesiologist

(b) The Human Surgeon and the Virtual Anesthesiologist

(c) The Virtual Surgeon and the Human Anesthesiologist

Fig. 3: The human surgeon and the anesthesiologist were both confederates. Their virtual counterparts were modeled to resemble their general appearance (e.g. skin tone, gender, size).

3.3 Training Exercise

The training exercise evaluated in this study was developed in conjunction with nursing management at the UF Health operating rooms and was based on an incident that had occurred previously at the hospital. The training exercise was conducted in a former OR that had been converted to a simulation room. During the training exercise, participants worked with the surgeon and the anesthesiologist to prepare a patient for surgery. The training exercise consisted of three stages, during which participants prepared a patient for surgery. Participants paused to complete surveys between each stage. In the first stage, participants spoke with the patient's mother and the anesthesiologist to gather information about the patient (the mother was always virtual). The primary purpose for this stage was to provide important information to participants about their patient and to give participants a chance to become more comfortable interacting with virtual humans. Data from this stage was not included in our analysis because the surgeon was not present in this stage. On average, the first stage lasted for 237 seconds, with a standard deviation of 62.3 seconds.

In the second and third stages, participants worked with the surgeon and the anesthesiologist to complete two checklists that prepared the patient for surgery. During these checklists, the surgeon read through each item and either confirmed that it was complete or asked the anesthesiologist or participant for a status update. The anesthesiologist occasionally interrupted with a question or a comment. Participants could also interrupt if they felt the need to do so. On average, the second stage lasted for 239 seconds, with a standard deviation of 28.3 seconds, and the third stage lasted for 114 seconds, with a standard deviation of 33.8 seconds.

During the final stage, the surgeon discovered that they were not ready to begin surgery, because the anesthesiologist had forgotten to send blood samples to the lab for processing. This angered the surgeon, who yelled at the anesthesiologist and then announced that he wanted to begin the surgery anyways. This moment was the focus of the training component of this exercise. Participants could either challenge the surgeon about the need to wait to begin the surgery until blood was available, or they could allow him to begin the surgery. If participants challenged the surgeon, he began arguing with them about how waiting was unnecessary. The argument could only be resolved by calling for a charge nurse to come in and speak with the surgeon. The exercise ended immediately after a charge nurse was called or after the nurse agreed to let the surgeon proceed. Participants were not primed to speak up to the surgeon before beginning the exercise. The agency of participants' teammates was not found to strongly impact their response to this training moment; many participants experienced difficulty speaking up to the surgeon, regardless of his agency[27].

3.4 Virtual Teammates

Participants could interact with their virtual teammates using speech and gesture. The virtual teammates were controlled by a wizard-of-oz who could hear participants directly and see them through a live video feed. A wizard-of-oz was used for this scenario to reduce speech recognition and speech understanding errors. The wizard-of-oz was able to select from a range of pre-recorded speeches that the virtual teammates could deliver. The interface was organized to facilitate rapid selection of speeches. Speeches were also suggested to the wizard based on what had been said recently. A single wizard was used for every participant. Participants were unaware that a wizard-of-oz was used to control the virtual teammates. To reduce suspicion that the virtual teammates were being controlled by a human, participants were asked to complete a speech recognition training session and to wear a microphone. Dragon Medial Practice Edition was used during the speech recognition training session, and the training session took approximately 10 minutes.

The virtual teammates were displayed on 40" 1080p television screens, making them approximately life-size (shown in Figures 1 and 3). Each virtual teammate was equipped with a Microsoft Kinect, used to track participants' location. This information was used to perform perspective correct rendering, and to allow the virtual teammates to make eye contact with participants. Pre-captured panoramic images were used to create the appearance of a see-through display. Different sections of the panoramic image were displayed depending on where participants were standing in the room.

The virtual teammates' gaze was controlled by a simple Markov model. When speaking, virtual teammates looked at whomever they were speaking to, with occasional glances at other teammates or the patient. When listening, 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 pre-recorded motion capture. The virtual surgeon and virtual anesthesiologist spoke using pre-recorded audio, 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 [6].

3.5 Human Teammates

Both human teammates were played by confederates. The confederates were selected from a pool of available standardized patients (e.g. trained actors who play the role of a patient during medical training exercise). 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 to behave consistently between different training sessions. The confederates were paid their standard rate ($20/hour) during the study.

The confederates were trained how to play their respective roles by the person responsible for controlling the wizard-of-oz. 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 available during the study as a guide to help the confederates stay on track during the exercise. The confederates were introduced to participants as Dr. Girard (for the surgeon) and Dr. Sanders (for the anesthesiologist), which is the name they used during the exercise. It was not explained to the participants that the confederates were actually actors.

On the whole, the human confederates successfully mimicked the behavior of their virtual counterparts. However, the human confederates' gaze behavior sometimes differed from the gaze behavior employed by the virtual teammates. The human confederates held clipboards which contained the script they were to follow. When necessary, the human confederates consulted the script to remind themselves what they were to say next. In contrast to this, the virtual teammates maintained eye contact with participants, except for occasional glances at other locations in the room. The implications of this source of variance are considered in the discussion.

3.6 Gaze Coding Procedure

Video data was recorded for 65 participants (four participants did not consent to video recording). Video was recorded at 30 FPS and a resolution of 640x480. A representative frame from the videos is shown in Figure 4.

Fig. 4: An example of the videos used for gaze analysis.

Video coding was performed using ANVIL [18]. The following gaze behaviors were coded using the recorded video: looking at the surgeon, the anesthesiologist, the patient, a clipboard, and looking up, down left, right, and straight ahead. The coding procedure used was adapted from the procedure used to create the Distress Analysis Corpus [11]. Three student annotators were hired to perform the coding. Annotators were trained how to use ANVIL and how to perform gaze coding. After receiving some basic instruction, annotators coded training videos until their inter-rater reliability (Krippendorff's alpha) exceeded 0.7. Annotators received feedback after each round of training. Once training was complete, each annotator was assigned a specific set of videos to code. Approximately 10% of the videos were assigned to multiple coders. These videos were used to monitor the annotators performance and ensure that they maintained inter-rater reliability through the coding process. Annotators were aware that their

Condition	N	Surgeon Speaking	Anesth. Speaking	Participant Speaking	Silence	*Total Duration*
Both Virtual	23	152.6 (±21.2)	35.6 (±4.5)	67.4 (±29.7)	130.9 (±31.2)	*386.5 (±61.7)*
Human Surgeon	20	155.8 (±17.2)	35.8 (±3.8)	65.2 (±23.7)	87.6 (±29.7)	*344.4 (±47.7)*
Human Anesth.	20	162.6 (±18.4)	31.5 (±2.7)	62.0 (±28.5)	113.3 (±23.2)	*369.4 (±58.6)*
Average	–	*157.0 (±19.5)*	*34.3 (±4.3)*	*64.9 (±26.4)*	*110.6 (±32.6)*	*366.8 (±66.3)*

Table 1: Average time (in seconds) that each teammate spoke, by condition. Standard deviations are shown in parentheses.

performance was being measured, but were not aware which videos were used to monitor their performance. Upon completion of the coding, inter-rater reliability was assessed for the ten shared videos. Gaze behavior coding yielded an average inter-rater reliability of $\kappa = 0.856$.

Gaze coders were not informed about the different conditions in the experiment, however this information could be deduced from the videos themselves. However, coders were not informed about the hypotheses of the study or the purpose of the gaze coding, making it unlikely that their coding was influenced by their awareness of the different conditions.

The times during which each teammate spoke was coded using Praat [5], a speech annotation system. Speech timings were coded by a second set of three coders. Inter-rater reliability was not calculated for speech timings due to the low subjectivity of speech.

4 Results and Discussion

In this section, we first consider how variations in the agency of one teammate *directly* affected gaze with that teammate. In the next section, we consider whether variations in the agency of one teammate *indirectly* affected gaze with other teammates. We use a similar analytical procedure to that employed by Rehm and André [26]. We consider three main data points: the total duration of gaze, the number of times participants gazed at an individual, and the average duration of each individual gaze fixation. To explore whether gaze was affected by who was speaking, we divided gaze into four periods based on who was currently speaking: the participant, the surgeon, the anesthesiologist, or no one.

Six participants were excluded from our analysis, either because they did not consent to video recording, or because of abnormal behavior during the training exercise that affected their gaze (e.g. having a conversation with the study proctor in the middle of the exercise). After these participants were removed, the following number of participants remained in each condition: 23 in the condition with two virtual teammates, 20 in the condition with the human surgeon, and 20 in the condition with the human anesthesiologist.

Gaze was analyzed during Stage 2 and 3 of the training exercise. Stage 1 was not analyzed because the surgeon was not present in that stage. Stage was included as a factor of each ANOVA considered in this section, but no significant interactions were observed between Stage and Agency. As such, stage-specific data is not reported in tables or graphs. Significant main effects of Stage were observed, most likely because of the differences in content between Stage 2 and 3. Specifically, participants spent more time speaking with the surgeon in Stage 3 due to the training component of the exercise.

4.1 Preparation of the data

Because participants' interactions could vary in length, we first performed a check to determine whether the duration of participants' interactions differed between conditions, and whether there was an interaction effect between condition and speaker. A two-way ANOVA did not reveal a main effect of condition ($F(2,483) = 1.656, p = 0.192$), but did reveal an interaction effect between condition and speaker ($F(6,483) = 3.741, p < 0.01$). Post-hoc comparisons revealed that there was significantly less silence in the Human Surgeon condition, compared with the Both Virtual condition ($p < 0.001$) and the Human Anesthesiologist condition ($p < 0.05$). No other significant differences were observed between conditions.

We then normalized participants' data to account for this significant effect. The following procedure was used to normalize participants' data: the average duration of each speaker period (e.g. when a specific teammate was speaking) was calculated (reported in Table 1). Each participants' data was then scaled such that the duration of their specific speaker periods matched the average duration for all participants. The amount of time spent gazing at the surgeon and the anesthesiologist during each speaker period was then calculated for each participant.

We also normalized the number of gaze fixation moments made by each participant to represent the number of gaze fixations that would have occurred during a 366.8 second interaction (the average interaction length). Normalization was performed by scaling the total number of gaze fixation moments by a factor equal to 366.8 divided by the actual length of the interaction.

4.2 Effect of Direct Variations in Agency

We now consider whether varying an individual's agency directly affected gaze with that individual. In this section, we perform the following comparisons: gaze behavior with the virtual surgeon is compared to gaze behavior with the human surgeon, and gaze behavior with the virtual anesthesiologist is compared to gaze behavior with the human anesthesiologist.

4.2.1 Total Gaze Duration

Figure 5 reports the proportion of an interaction that was spent gazing at the surgeon and the anesthesiologist while the various teammates were speaking, after normalization. Mixed factorial ANOVAs were conducted for gaze with the surgeon and the anesthesiologist, where direct variations in agency served as a between-subjects factor and speaker and stage served as repeated-measures factors. A significant main effect of direct variations in agency was found for the surgeon ($F(1,41) = 19.488, p < 0.001, d = 0.886$) and for the anesthesiologist ($F(1,40) = 16.914, p < 0.001, d = 0.674$). Significant interactions existed between agency and speaker for the surgeon ($F(3,123) = 9.963, p < 0.001, d = 0.587$) and the anesthesiologist ($F(3,120) = 4.186, p < 0.01, d = 0.344$).

Post-hoc tukey tests were conducted to explore the interaction effect between direct variations in agency and speaker. This revealed a significant difference for direct variations in agency during periods of silence for the surgeon ($p < 0.001$) and the anesthesiologist ($p < 0.001$). Over the course of the entire interaction, participants spent an average of 22.1 more seconds gazing at the virtual surgeon during periods of silence, an increase of 97.3%, and 14.9 more seconds gazing at the virtual anesthesiologist during periods of silence, an increase of 130.6%. A significant difference was observed for gaze with the surgeon when the surgeon was speaking ($p < 0.001$), where participants spent 25.3 more seconds gazing at the virtual surgeon while he was speaking, an increase of 45.1%. Similarly, a significant difference was also observed for gaze with the anesthesiologist while the anesthesiologist was speaking ($p < 0.05$), where participants spent 8.2 more seconds gazing at the virtual anesthesiologist while he was speaking, an increase of 46.4%. All other interactions were non-significant. The interactions effects between direct variations of agency and speaker are visualized in Figure 5.

Several important observations can be drawn from this data: first, we see that participants did not look at the virtual teammates more frequently when participants were speaking, but did look at the virtual

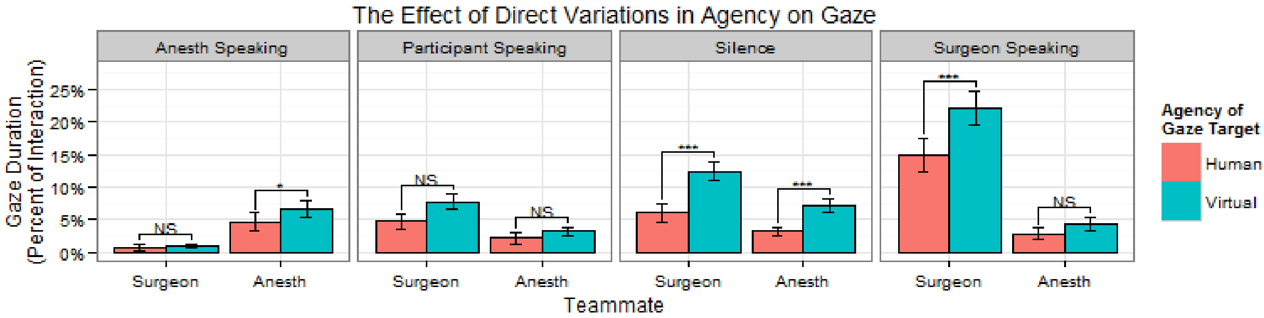


Fig. 5: The amount of time spent gazing at the surgeon and the anesthesiologist, while various teammates were speaking. Time is reported as a percentage of the entire interaction. Error bars represent 95% confidence intervals. The following labels are used for to report levels of significance: $NS = p \geq 0.05, * = p < 0.05, ** = p < 0.01, *** = p < 0.001$

teammates more frequently when the virtual teammates were speaking. Both of these findings mirror the results obtained by Rehm and André, who found that participants spent more time gazing at a virtual human when the virtual human was speaking, but not while the participants were speaking [26]. Additionally, we can also see that participants spent significantly more time looking at virtual teammates during periods of silence, compared to their human counterparts. Proportionally, this increase in time was two to three times larger than the increase in time observed while the virtual teammate was speaking. Finally, we also see that participants did not look at a virtual teammate more frequently when the other teammate was speaking (e.g. the agency of the anesthesiologist did not effect gaze with the anesthesiologist while the *surgeon* was speaking).

4.2.2 Gaze Fixation Moments

The increase in time spent looking at virtual teammates can be accounted for in two ways: either participants gazed at virtual teammates more frequently, or, each time participants gazed at them, they looked at them for longer. We consider both in this section.

Number of Gaze Fixation Moments Mixed factorial ANOVAs were conducted to explore whether participants looked at virtual teammates more frequently, where direct variations in agency served as a between-subjects factor and stage served as a repeated-measures factor. A significant main effect of direct variations in agency was observed for the surgeon ($F(1,43) = 11.051, p < 0.01, d = 0.934$) and for the anesthesiologist ($F(1,42) = 9.759, p < 0.01, d = 0.902$). A visualization of the direct effects of agency on the number of gaze fixation moments is shown in Figure 6. On average, participants produced 40.5% more gaze fixation moments with virtual teammates than human teammates. These results suggest that participants gazed at virtual teammates more frequently than at human teammates.

Duration of Gaze Fixation Moments No normalization was performed on the duration of gaze fixation moments, as they are examined here on average. Examining the average duration automatically compensates for variations in the length of an interaction, meaning normalization was not required.

To explore whether agency affected the average duration of gaze fixation moments, mixed factorial ANOVAs were conducted for the surgeon and the anesthesiologist, where direct variations in agency served as a between-subjects factor and stage served as a repeated-measures factor. A significant main effect of direct variations in agency was not found for the surgeon ($F(1,43) = 1.704, p = 0.199$) or the anesthesiologist ($F(1,42) = 0.037, p = 0.847$). A visualization showing the absence of an effect is shown in Figure 7a. These results suggest that, during a single gaze fixation moment, participants gazed at real and virtual teammates for similar lengths of time.

These results conflict with those obtained by Rehm and André, who found that participants made eye contact at similar rates with real and

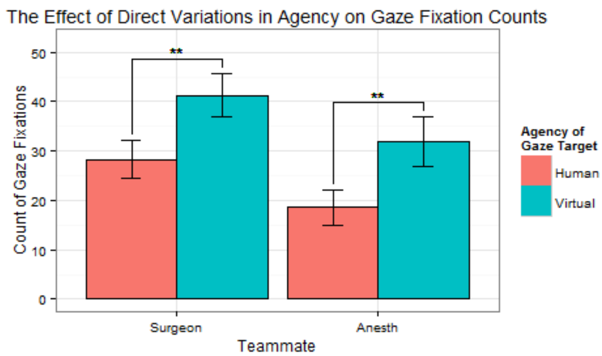


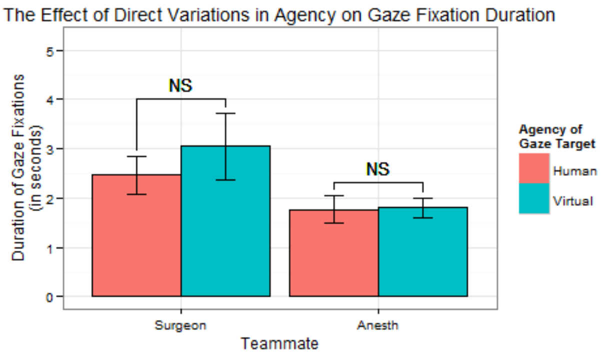
Fig. 6: Participants produced more gaze fixation moments with virtual teammates than their human counterparts. Error bars represent 95% confidence intervals. The following labels are used for to report levels of significance: $NS = p \geq 0.05, * = p < 0.05, ** = p < 0.01$

virtual teammates, but that they looked at virtual teammates longer once they had made eye contact. Instead, here we find participants making eye contact more frequently, but that they did not gaze at virtual teammates longer after making eye contact.

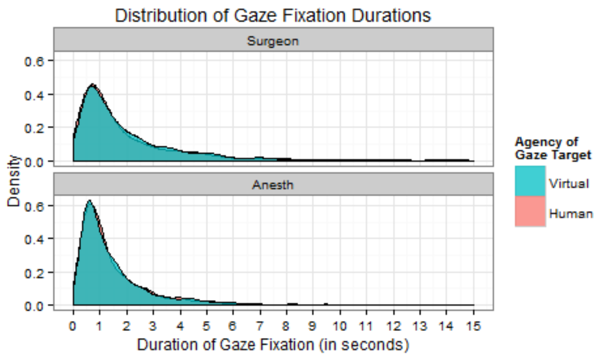
This difference is of particular importance when considering Rehm and André's conclusion that participants violated polite gaze norms with virtual teammates, as they were more willing to stare at them for prolonged periods of time. To further explore the question of politeness, we also examined the distributions governing the duration of gaze fixation moments with real and virtual teammates. If participants did violate politeness norms with virtual teammates, we would expect these distributions to be different. Figure 7b visualizes the distribution of gaze fixation durations using a density plot. This visualization is based on the duration of *every* gaze fixation moment, not just the averages for each participant that were examined previously.

Agency appears to have had very little effect on the duration of gaze fixation moments. The density plots for both virtual and human teammates almost completely overlap. All of the distributions are characterized by a sharp peak centered around 0.75 seconds, followed by a long and thin tail. The duration of gaze fixation moments with the anesthesiologist appears to be more strongly skewed towards shorter gaze fixation moments, which is likely reflective of his less prominent role in the exercise. He spoke less, which meant participants had fewer reasons to make prolonged eye contact. This data reinforces the conclusion that participants did not stare at virtual teammates for longer periods of time than human teammates.

As an aside, it is important to note that while the distribution of *all*



(a) The average duration of gaze fixation moments with the surgeon and the anesthesiologist. Error bars represent 95% confidence intervals. The following labels are used for to report levels of significance: $NS = p \geq 0.05$



(b) The distributions of each condition are overlaid on top of each other with 50% opacity, allowing differences between the conditions to be seen. The distributions are nearly identical, which may make it difficult to distinguish between the two.

Fig. 7: The direct effect of agency on the average duration of gaze fixation moments, and the distribution of the duration of gaze fixation moments.

gaze fixation durations is highly non-normal, the distribution of each participant's *average* gaze fixation duration, which was examined previously, was normally distributed, which made it possible to examine that data using an ANOVA.

4.2.3 Discussion

Agency was observed to directly affect gaze behavior with participants' teammates. When comparing gaze with virtual teammates to human teammates, participants gazed at virtual teammates for longer while the teammate was speaking ($p_{Surgeon} < 0.001, p_{Anesth} < 0.05$) and during periods of silence ($p < 0.001$). Agency did not affect how long participants looked at their teammates while participants were speaking, nor did it affect the average duration of gaze fixation moments. This last finding is important, as it suggests that participants continued to obey politeness norms when gazing at virtual teammates.

These results are largely consistent with those found by Rehm and André, who found that participants gazed at virtual humans longer while listening to them, but did not gaze at them longer while the participants were speaking. However, unlike Rehm and André, we found no indication that participants violated polite gaze norms with virtual teammates. The distributions governing how long participants maintained eye contact with real and virtual teammates were essentially identical. It is possible that our participants' familiarity with the task being performed contributed to our failure to identify violations of politeness norms. Our participants were very familiar with the process of preparing a patient for surgery, as this is an essential component of their daily activities. This familiarity with the task may have predisposed them to behave realistically with their virtual teammates. In contrast, the participants in Rehm and André's experiment were most likely unfamiliar with the task performed, and thus may have felt more free to violate polite gaze norms with their virtual teammates.

People's tendency to spend more time gazing at virtual faces (and robots, as found by Yu et al. [32]) than human faces could have several explanations. One potential explanation is *novelty*. Only two of our participants had interacted with a virtual human prior to this study. Rehm and André reported the same was true of their participants. However, if novelty was the primary reasons participants spent more time gazing at virtual teammates than real humans, we would expect this effect to be seen at all times, rather than mediated by speaker. Our observation that participants did not gaze at virtual teammates longer when participants were speaking and when the other teammate was speaking suggests that the increase in gaze time is at least partially linked to conversational factors. Novelty may explain part of why people spend more time gazing at virtual teammates, but novelty alone cannot explain the patterns observed in our data.

A second possible explanation is that gazing at virtual humans may produce a lighter cognitive load than gazing at real human faces. Cognitive load is often cited as an important factor of gaze aversion. Gaze aversion often increases as cognitive load increases (e.g. when being given questions of varying difficulty) [8]. Similarly, cognitive performance decreases if people are not allowed to avert their gaze from others [10]. Given the link between gaze aversion and cognitive load, if virtual humans do evoke a lighter cognitive load than real human faces, we would expect to see less aversion. This possibility is bolstered by prior work that showed that people averted their gaze less when looking at videos of human faces than actual human faces [8]. It may be that the absence of depth cues reduces the cognitive load required to interact with images shown on a screen. Similarly, virtual human faces are typically visually simpler than human faces, which could also reduce cognitive load. However, it is important to note that gaze aversion only decreases while listening to virtual humans, not while speaking to them. Cognitive load is typically considered a factor while speaking, not while listening. As such, if cognitive load were the primary factor related to the decrease in gaze aversion, we would expect to see a decrease in gaze aversion while speaking as well.

The observation that participants also gazed at virtual teammates more often during periods of silence suggests that the increase in time spent gazing at virtual humans may be related to information gathering. Participants spent additional time looking at virtual teammates while listening to them, and while waiting for someone else to speak. Gaze can be used during these periods to gather important information, such as non-verbal cues[3] or turn-taking cues [14]. It is possible that people find it more difficult to gather this information from virtual humans, which causes them to watch the faces of virtual humans more closely. If this is a factor, we would expect the amount of time spent gazing at virtual humans to decrease as virtual humans are developed that can communicate this information more clearly.

It seems unlikely that any of the above explanations are sufficient in and of themselves. The increase in time spent looking at virtual humans is likely caused by a combination of all three factors. Future research is required to understand the importance of each factor.

4.3 Effect of Indirect Variations in Agency

Having considered how variations in agency can directly impact gaze behavior, we now consider whether variations in the agency of one teammate indirectly affects gaze behavior with a second teammate. In this section, we perform the following comparisons: participants' gaze behavior with the virtual surgeon is compared when the agency of the anesthesiologist is changed, and participants' gaze behavior with the virtual anesthesiologist is compared when the agency of the surgeon is changed.

4.3.1 Total Gaze Duration

Figure 8 reports the proportion of an interaction that was spent gazing at the surgeon and the anesthesiologist while the various teammates were speaking, after normalization. Mixed factorial ANOVAs

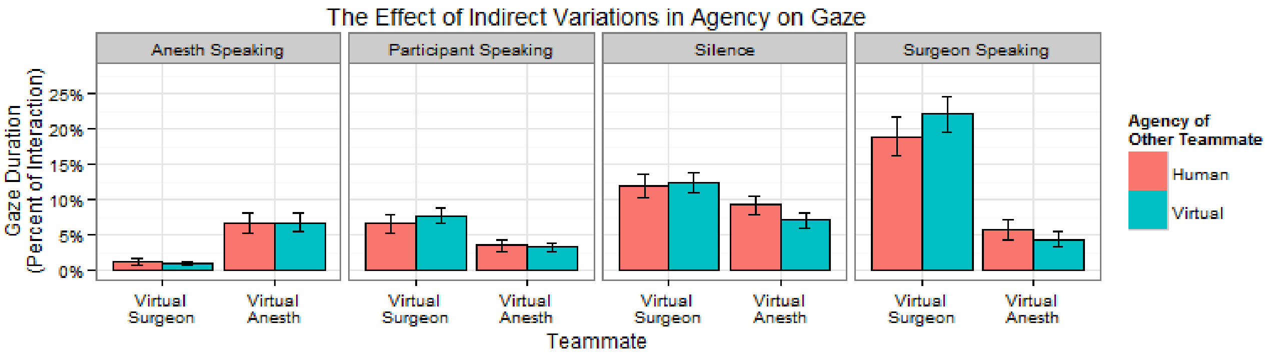


Fig. 8: Agency of other teammates did not indirectly affect gaze with virtual teammates. Error bars represent 95% confidence intervals.

were conducted for gaze with the surgeon and the anesthesiologist, where indirect variations in agency served as a between-subjects factor and speaker and stage served as repeated-measures factors. A significant main effect of indirect variations in agency was not found for the surgeon $(F(1,40) = 1.518, p = 0.225)$ or for the anesthesiologist $(F(1,41) = 2.072, p = 0.158)$. Significant interactions were not observed between agency and speaker for the surgeon $(F(3,120) = 2.564, p = 0.058)$ or for the anesthesiologist $(F(3,123) = 1.578, p = 0.195)$.

These results suggest that the agency of one teammate is unlikely to influence gaze with other teammates. A trend towards a significant interaction effect between agency and speaker was observed for the virtual surgeon, but no such trend was observed for the virtual anesthesiologist. A visual examination of Figure 8 reveals that the amount of time participants spent gazing at their teammates was remarkably consistent for all variations except for when gazing at the surgeon when the surgeon was speaking. However, this variation also had the widest confidence interval, suggesting that there was a large amount of variance in how often participants looked at the surgeon while he was speaking, which easily accounts for this difference.

4.3.2 Gaze Fixation Moments

We now consider where indirect variations in agency affected the number and duration of individual gaze fixation moments.

Number of Gaze Fixation Moments Mixed factorial ANOVAs were conducted to explore whether participants' gaze with virtual teammates was affected by changes in the agency of a second teammate, where indirect variations in agency served as a between-subjects factor and stage served as a repeated-measures factor. A significant main effect of indirect variations in agency was not found for the surgeon $(F(1,42) = 3.594, p = 0.064)$ or the anesthesiologist $(F(1,43) = 3.327, p = 0.075)$. A visualization showing the absence of an effect is shown in Figure 9.

Though the indirect effect of agency on the number of gaze fixation moments was not significant, trends were observed for both the surgeon and the anesthesiologist. These trends were of similar direction and magnitude, both representing a decrease in gaze fixation counts of approximately 20%.

Duration of Gaze Fixation Moments To explore whether agency affected the average duration of gaze fixation moments, mixed factorial ANOVAs were conducted for the surgeon and the anesthesiologist, where indirect variations in agency served as a between-subjects factor and stage served as a repeated-measures factor. A significant main effect of indirect variations in agency was not found for the surgeon $(F(1,42) = 0.784, p = 0.380)$, but was found for the anesthesiologist $(F(1,43) = 7.413, p < 0.01, d = 0.725)$. A visualization of the effect of indirect variations in agency on the duration of gaze fixation moments is shown in Figure 10a.

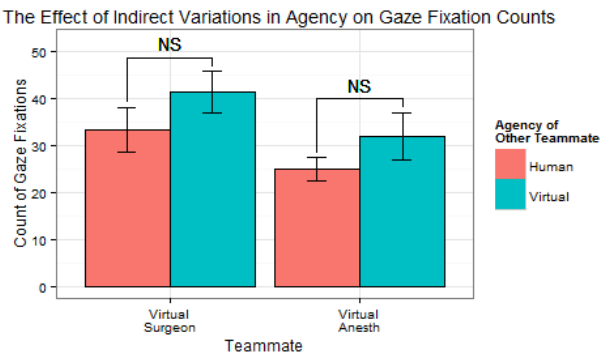


Fig. 9: Agency of other teammates did not indirectly affect gaze fixations with virtual teammates. Error bars represent 95% confidence intervals. The following labels are used for to report levels of significance: $NS = p \geq 0.05$

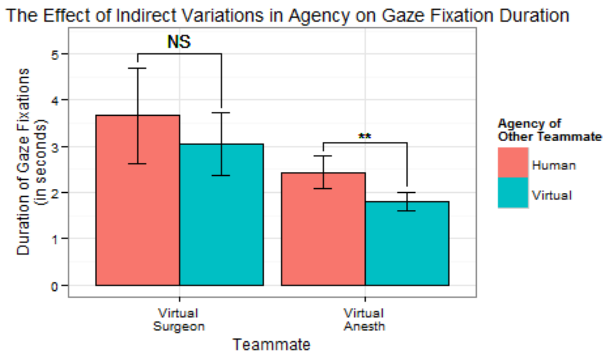
These results suggest that participants gazed at the virtual anesthesiologist for longer periods of time when a human teammate was present, but not at the virtual surgeon. This finding is also supported by the distributions governing the duration of gaze fixation moments, shown in Figure 10b.

Both of these distributions show a slight trend towards longer gaze fixation moments when another human teammate is present, though this trend is markedly larger for the anesthesiologist. Both distributions still peak near the 0.75 second mark, but the peak is less well defined. However, even though there is a shift towards longer gaze durations, it still does not appear that participants stared at virtual teammates in a manner that would violate polite gaze norms. To constitute impolite staring, the peak would need to be shifted several seconds further towards longer durations, which was not observed.

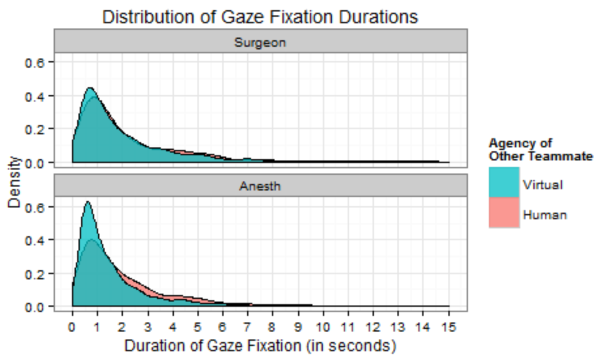
4.3.3 Discussion

Agency was not observed to have a consistently significant indirect effect on gaze behavior. Participants' gaze with virtual teammates was almost entirely unaffected by the agency of a second teammate. One significant effect was observed for the virtual anesthesiologist, such that participants spent slightly longer looking at him during specific gaze fixations.

Two observations can be made from this data: first, we see no indication that gaze with virtual humans can be indirectly influenced by the agency of other teammates. This suggests that behavior with virtual humans will not be subtly biased by the presence of other human group members. If virtual humans were perceived as "second-class citizens", we would expect to see that indicated in a reduction of the



(a) Agency of other teammates did not indirectly affect the duration of gaze fixations with the virtual surgeon, but did with the anesthesiologist. Error bars represent 95% confidence intervals. The following labels are used for to report levels of significance: $NS = p \geq 0.05, * = p < 0.05, ** = p < 0.01$



(b) The distribution of each condition is overlaid on top of each other with 50% opacity, allowing differences between the conditions to be seen.

Fig. 10: The indirect effect of agency on the average duration of gaze fixation moments, and the distribution of the duration of gaze fixation moments.

amount of gaze with virtual humans when other humans were present. Given that such a reduction was not observed, it appears that people will accept virtual humans as group members. This does not eliminate perceived differences between human and virtual group members, as gaze with virtual teammates remained consistent regardless of the agency of the other teammate, but it does allow a virtual human to stand or fall on its own merits, rather than on the agency of its teammates. This is an important finding, as it suggests that virtual humans can be used in group settings where multiple humans are present.

We can also draw inferences based on the consistency in the amount of gaze with virtual teammates. Participants spent more time looking at virtual teammates, regardless of the agency of a second teammate. This suggests that whatever effect causes people to spend more time looking at virtual humans than real humans operates without reference to the group. This is consistent with the theory that novelty and reduced cognitive load could be the causes of increased gaze with virtual humans. The presence of other humans will not make virtual humans any less novel, nor would it affect the cognitive load imposed by gaze with a virtual human. It may be inconsistent with the theory that participants gaze at virtual teammates more in order to gather additional information, given that information about turn taking could be gathered from both teammates. If participants felt that human teammates provide more accurate cues about turn taking, then it could be argued that people would look at virtual teammates less during periods of silence when a human teammate was present. However, this is a speculative argument, as it is not clear whether participants did consider humans more reliable sources of turn taking cues.

4.3.4 Limitations

One important limitation in this research is that the virtual teammates and human confederates exhibited different gaze behavior. The human confederates had copies of their script available on a clipboard during the interaction. When necessary, they referenced this script to ensure that they closely mimicked their virtual counterparts' behavior. This caused the human confederates to spend less time gazing at participants. Video data was not collected for the human confederates' gaze, which makes it impossible to quantify how much variation there was in gaze. Given that one person's gaze patterns can affect other people's gaze, it is possible that the variations in gaze between the virtual teammates and the human confederates biased our results. However, it is important to note that the human anesthesiologist spoke significantly less frequently that the surgeon, due to his role in the scenario. Accordingly, his gaze matched his virtual counterpart's gaze much more closely than the surgeon's. As such, if the human confederates' gaze had biased our results, we would not expect to see such consistent patterns emerging for gaze with the surgeon and the anesthesiologist. The observed level of consistently, along with the close correspondence between prior work and our results concerning direct variations in agency, makes it seem unlikely that this limitation created significant bias in our data.

5 Conclusion and Future Work

In this paper, we explored whether the agency of one group member can affect gaze behavior with other group members. We replicated previous findings that suggest agency does affect people's gaze while listening, but not while speaking. However, our results conflicted with existing research that suggests participants are willing to violate polite gaze norms with virtual humans. We found that agency had no effect on adherence to polite gaze norms with virtual teammates.

Unlike earlier studies that have only explored the direct effects of agency on gaze, we also examined whether the agency of one teammate can indirectly effect gaze with other teammates. We found that indirect variations of agency have very little effect on gaze with virtual teammates. No significant differences were observed, though some trended towards significance. Though virtual humans may be subject to contagious stigmas, such as those observed with obesity [12], the agency of one group member appears to have no effect on behavior with other virtual group members. This is an important finding, as it supports the use of virtual humans in group settings. This is particularly relevant to educational and training-centric uses of virtual humans, as the absence of indirect effects of agency supports the viability of using virtual humans for these tasks. Had agency produced indirect effects on others, then variations in agency could have produced subtle biases towards others that may have affected learning outcomes. This finding is also relevant for social scientists, who have begun to adopt virtual reality as an important tool for social research. Though direct effects of agency must be taken into consideration during this research, it seems unlikely that researchers will have to account for unwanted indirect effects during their experiments.

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] S. Andrist, B. Mutlu, and M. Gleicher. Conversational gaze aversion for virtual agents. *Intelligent Virtual Agents*, pages 249–262, 2013.
[2] M. Argyle. *Bodily communication.* Routledge, 2013.
[3] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, 1976.
[4] 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):819–833, 2003.

[5] P. Boersma and V. van Heuven. Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–347, 2001.

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

[7] G. Doherty-Sneddon, V. Bruce, L. Bonner, S. Longbotham, and C. Doyle. Development of gaze aversion as disengagement from visual information. *Developmental Psychology*, 38(3):438–455, 2002.

[8] G. Doherty-Sneddon and F. G. Phelps. Gaze aversion: a response to cognitive or social difficulty? *Memory & cognition*, 33(4):727–733, 2005.

[9] H. Ehrlichman. From gaze aversion to eye-movement suppression: An investigation of the cognitive interference explanation of gaze patterns during conversation, 1981.

[10] a. M. Glenberg, J. L. Schroeder, and D. a. Robertson. Averting the gaze disengages the environment and facilitates remembering. *Memory & cognition*, 26(4):651–658, 1998.

[11] J. Gratch, R. Artstein, G. Lucas, G. Stratou, S. Scherere, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D. Traum, S. Rizzo, and L.-P. Morency. The Distress Analysis Interview Corpus of human and computer interviews. *Proceedings of Language Resources and Evaluation Conference (LREC)*, 2014.

[12] M. R. Hebl and L. M. Mannix. The weight of obesity in evaluating others: a mere proximity effect. *Personality and social psychology bulletin*, 29(1):28–38, 2003.

[13] D. Heylen, I. van Es, a. Nijholt, and B. Van Dijk. Experimenting with the gaze of a conversational agent. *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, pages 93–101, 2002.

[14] L. Hirvenkari, J. Ruusuvuori, V. M. Saarinen, M. Kivioja, A. Peräkylä, and R. Hari. Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer. *PLoS ONE*, 8(8), 2013.

[15] R. Hoegen, G. Stratou, G. M. Lucas, and J. Gratch. Comparing behavior towards humans and virtual humans in a social dilemma. *Intelligent Virtual Agents*, 2015.

[16] A. Kendon. Some functions of gaze-direction in social interaction., 1967.

[17] K. Kim, A. Nagendran, J. Bailenson, and G. Welch. Expectancy Violations Related to a Virtual Human s Joint Gaze Behavior in Real-Virtual Human Interactions. pages 5–8.

[18] M. Kipp. Anvil - A Generic Annotation Tool for Multimodal Dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pages 1367–1370, 2001.

[19] J. Lee, S. Marsella, D. Traum, J. Gratch, and B. Lance. The Rickel Gaze Model: A Window on the Mind of a Virtual Human. *Intelligent Virtual Agents*, pages 296–303, 2007.

[20] S. P. Lee, J. B. Badler, and N. I. Badler. Eyes alive. *ACM Transactions on Graphics*, 21(3):637–644, 2002.

[21] G. M. Lucas, J. Gratch, A. King, and L. P. Morency. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37:94–100, aug 2014.

[22] L. Markson, K. B. Paterson, and K. Paterson. Gaze-aversion & visual-spatial imagination. *British Journal of Psychology*, pages 1–21.

[23] Y. I. Nakano, G. Reinstein, T. Stocky, and J. Cassell. Towards a Model of Face-to-Face Grounding. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, 2003.

[24] T. Pejsa, S. Andrist, M. Gleicher, and B. Mutlu. Gaze and Attention Management for Embodied Conversational Agents. *ACM Transactions on Interactive Intelligent Systems*, 5(1):1–34, 2015.

[25] N. Pfeiffer-Leßmann, T. Pfeiffer, and I. Wachsmuth. An Operational Model of Joint Attention - Timing of Gaze Patterns in Interactions between Humans and a Virtual Human. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, (2006):851–856, 2012.

[26] M. Rehm and E. Andre. Where do they look? Gaze Behaviors of Multiple Users Interacting with an Embodied Conversational Agent. *Intelligent Virtual Agents*, 2005.

[27] A. Robb, C. White, A. Cordar, A. Wendling, S. Lampotang, and B. Lok. A comparison of speaking up behavior during conflict with real and virtual humans. *Computers in Human Behavior*, 52:12–21, 2015.

[28] J. Streeck. Gesture as communication I: Its coordination with gaze and speech, 1993.

[29] R. Vertegaal, R. Slagter, G. Van Der Veer, and A. Nijholt. Eye Gaze Patterns in Conversations : There is More to Conversational Agents than Meets the Eyes. *Analysis*, pages 301–308, 2001.

[30] D. Walker and E. Vul. Hierarchical Encoding Makes Individuals in a Group Seem More Attractive. *Psychological Science*, 25(1):230–235, 2013.

[31] Z. Wang, J. Lee, and S. Marsella. Towards more comprehensive listening behavior: Beyond the bobble head. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6895 LNAI:216–227, 2011.

[32] C. Yu, P. Schermerhorn, and M. Scheutz. Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems*, 1(2):1–25, 2012.