Seeing is Comforting: Effects of Teleoperator Visibility in Robot-Mediated Health Care

Kory Kraft, William D. Smart

Abstract—Teleoperated robots can be used to provide medical care to patients in infectious disease outbreaks, alleviating workers from being in dangerous infectious zones longer than absolutely needed. Nevertheless, patients’ reactions to this technology have not been tested. We test three hypotheses related to patients’ comfort and trust of the operator and robot in a simulated Ebola Treatment Unit. Our findings suggest patients trust the robot teleoperator more when they can see the teleoperator.

I. INTRODUCTION

Providing medical care to people with highly infectious diseases exposes health care work to the risk of becoming infected themselves. This is especially significant when the diseases are either highly contagious or extremely lethal. Both of which were true with the 2014 outbreak of Ebola Virus Disease (EVD) in West Africa. Health care professionals had to wear high levels of personal protective equipment (PPE) in order to safeguard against infection [1], [2]. This PPE, when coupled with the high heat and humidity of the region, meant that health care workers could only work for 40-minute shifts before being at risk for heatstroke. This dramatically decreased the quality of care that they were able to provide and contributed to the high mortality rate of the outbreak.

We are investigating the use of teleoperated robots in this setting to allow health care workers to perform some of their duties at a safe distance from infected patients. However, these robots can be scary things, especially when hovering over a medical cot with a patient in it. In this paper, we investigate whether being able to directly observe the human teleoperator makes the patient feel more trusting and comfortable, when being attended to by a robot.

Specifically, we address the following hypotheses:

• H1: A patient’s increased visibility of the operator leads to higher levels of patient trust in the operator.
• H2: A patient’s increased visibility of the operator leads to higher levels of patient trust in the robot system.
• H3: A patient’s increased visibility of the operator leads to higher levels of overall patient comfort.

We test these hypotheses under two conditions where the operator was either visible or not visible to the patient. The patient lay in a simulated Ebola Treatment Unit (ETU) while a human-sized mobile robot performed various tasks via teleoperation. Questionnaires and psychophysiological responses were used to evaluate the hypotheses.

Our major contribution suggests that greater visibility of the operator gives patients higher levels of trust in the operator. Our results raise important questions regarding the current adoption and style of telemedicine systems, as companies currently sell remote medical telepresence devices [3] and pursue remote telesurgery platforms [4]. These results also have immediate practical implications for the design of a teleoperation control unit to be deployed in the field.

While the latest Ebola outbreak was the stimulus for our work, the principles and practices learned here generalize to other types of infectious disease outbreaks, especially those requiring workers to wear PPE.

II. BACKGROUND AND RELATED WORK

We give background information and related work pertaining to Ebola and infectious diseases and robotics, the importance and measures of trust and comfort in medicine and human-robot interaction.

A. Ebola and Infectious Diseases

The costs of infectious disease outbreaks can be far-reaching. They can claim lives, disrupt economies, and set back a generation’s development [5], [6].

The recent Ebola outbreak is a good example of the potential destruction caused by infectious disease. The EVD outbreak, declared in March 2014, took 11,261 lives out of 27,642 cases as of the time of writing [7]. These deaths took place in a region already beset by medical fragility. EVD is both highly infectious and can be highly virulent (the mortality rate ranges from 20-90%) [2], [8]. Despite
ample fear and political pressure it still took a year to begin extensive human trials of a promising EVD vaccine [9].

B. Medical and Health Care Robotics

Okamura et al. segment the Medical and Health Care Robotics field into six application areas: surgical and interventional robotics, robotic replacement of diminished/lost function, robot-assisted recovery and rehabilitation, behavioral therapy, personalized care for special-needs populations, and wellness/health promotion [10].

This partitioning, while helpful in its specificity, misses a large part of what happens in many hospitals and doctors’ offices: general medical diagnosis and treatment involving complex patient-doctor interactions. They also miss the behind-the-scenes organizational and logistical work needed for stable medical infrastructures.

The commercial robotics sector is addressing some of these issues. Aethon Inc’s TUG robot, used for transportation and delivery of goods in hospitals and warehouses, has “travelled over 1,000,000 miles” [11]. Xenex sells a xenon ultraviolet disinfection robot that irradiates rooms to prevent hospital acquired infections. It is set up via a touch screen and remains stationary during irradiation [12].

There has been research on general nursing robots that involve physical interactions with patients [13] [14]. Despite some advances, most commercial robotic systems are either 1) stationary manipulator platforms for executing precision tasks or 2) mobile platforms without manipulators meant to connect people together, move goods around, or perform static ultraviolet decontamination. In this work, however, we are interested in exploring a more general robotic platform that is mobile and manipulator-capable in a domain that has been largely ignored – responding to infectious disease outbreaks.

C. Trust in Medicine

Trust is an important component of medical care, yet it is multifaceted and culturally dependent. Trust in medical systems can be influenced by a variety of factors (e.g., payment method) and is linked with a variety of outcomes (e.g., patient retention and likelihood to seek medical services in the future) [15]–[18].

One important factor in our study is the link between trust of physicians and the willingness of patients to seek future medical treatment. This is especially important in outbreak conditions. PPE obscured the human appearance enough that it distanced potential patients from health care providers. *Médecins Sans Frontières* (MSF) doctors value physician-patient trust so highly that they have learned to go into new communities, regardless of contamination risk, without wearing their PPE in order to gain communities’ trust [19].

The Wake Forest Physician Trust Scale (PTS) is one standard measure for patient-physician trust and has undergone verification and testing [20]. We use the PTS in our pre-experiment survey.

In our post experiment survey, we examine both the patient’s trust of the robot system and the operator of the robot using similar language and style to the PTS. This is discussed more in Section IV-E.

D. Trust in HRI

Trust has been examined in a number of contexts in HRI. For example, Freedy et al. developed a task-specific, objective measurement of trust for a collaborative human-robot task [21]. Desai et al. found that reliability impacts user trust in shared-autonomy robots as evidenced by increased switching to manual mode in driving a robot [22]. Bainbridge et al. examine the trust of a robot versus a video-displayed agent both by using questionnaires and examining the likelihood of participants to follow through with certain actions [23]. Similarly Salem et al. examine participants likelihood to trust a “faulty” robot versus a properly functioning robot [24]. All of these studies include quantitative measures for trust using post-hoc questionnaires.

There is a growing trend to quantitatively evaluate trust by counting the number of times participants allow a robot to do an action autonomously or, as in [21] and [24], follow through with a request from the robot. In each of these studies the participants played a much more active role in interacting with the robot than a hospitalized patient would. The passive role of the patient in our experimental setting requires the use of subjective, self-reported data instead of more objective, observational data.

E. Comfort in Medicine

Comfort is seen as “central” to medical practice amongst practitioners [25]. MSF doctors in the Ebola outbreak went through the trouble of putting a picture of themselves with their handwritten name next to it on the front of their PPE. This offset the unease brought on by the obfuscating PPE and made patients feel more comfortable with doctors [19].

The nursing literature provides a rich set of theory and tradition for measuring and valuing patient comfort. The literature includes studies that assess comfort for both patients and caregivers using questionnaires and ethnographic interviews [26], [27].

The health care and medical literature sometimes defines comfort as the absence of pain and use measures such as the Visual Analog Scale for Pain [28], [29]. However, this is not relevant to our study because our study did not induce physical pain.

F. Comfort in HRI

Comfort has been studied in the context of socially-aware robots that optimize for people’s comfort when navigating around them [30] [31], in human-robot handovers [32], and in response to robot touch in a nursing context [14]. [30] uses a 5-point Likert scale questionnaire to evaluate comfortable approach paths, speeds, and distances by robots. [33] developed a unique hand-held device that allows users to input current comfort ratings for the duration of the experiment; this method has the obvious drawback of requiring the participant to continually assess and report their comfort level. [31] defined and modelled pedestrians’ walking comfort as
the “subjective impression of one’s easiness of traversing an environment.” They recorded pedestrians’ traversal using laser range finders and then averaged three 7-point Likert items to calculate comfort.

Robots like the Paro and MEDI (built on the NAO robot) are being trialled in health care settings to provide comfort to patients [34]–[37].

We define comfort as the relative absence of stress and arousal and test this using questionnaires and physiological methods. Following Chen et al. we use a Galvanic Skin Response (GSR) system to measure skin conductance which is linearly correlated with valence and arousal [14]. Our methods are described more in Sections IV-F and IV-E.

III. IMPLEMENTATION

This section outlines the robot, robotic control, and environment used in the experiment.

A. Robot Description

For this study, we used a PR2 humanoid robot [38]. The PR2 has two 7 degree-of-freedom arms. The mobile base allows for semi-holonomic motion. In addition, the PR2 has a telescoping spine and ranges in height from 1.33-1.64m.

The PR2 is equipped with a variety of sensors: two laser range-finding sensors (one on the base and one on the head), an Asus Xtion RGB-D camera mounted on top of the robot’s head, a wrist camera for each arm, and stereo cameras. Each of these sensors, except for the stereo cameras, was used to provide situational awareness to the robot operator.

The robot ran on battery power, but was tethered via Ethernet for data transmission. To simplify networking, a remote desktop was connected via Ethernet to the service port of the robot.

B. Robot Control

An operator teleoperated the robot using the Robot Operating System (ROS) PR2 Surrogate package [39], [40]. This allows the teleoperator to control the position of the robot’s arms via Razer Hydra motion sensing controllers [41].

The operator holds the controllers, moves his/her hands freely, and presses the enable button when he/she desires the robot’s end effectors to match that of the operators. An open gripper, close gripper, and killswitch button are mapped to the controller. The killswitch immediately disengages the robot’s joint actuator forces.

The operator relied on an Robot Visualization (RVIZ) graphical user interface (GUI) to perform the teleoperation. The operator’s GUI gives access to a variety of information including a live feed from the robot’s sensors, the three ETU environment cameras, and a model of the robot as it appears in the world. The operator relied solely on the visualization found on the monitors to move the robot and complete each task.

The operator was trained by repeatedly completing the tasks in the ETU without any human subjects. The training time took roughly 2 hours. The primary author of this paper acted as the operator.

C. Environment

The robot performed tasks in a simulated Ebola treatment unit (ETU) built according to information found from MSF [42]. Our scaled-down ETU is resemblant of a high-risk zone within a larger Ebola treatment center.

The ETU is an enclosed space approximately 4m x 5m x 2.5m. It is covered in anti-static, fire-retardant white plastic for a ceiling and walls and has a concrete floor. There is only one entryway measuring approximately 2m H x 1m W. The space includes: 2 single cots (195cm L x 66cm W x 40.6cm H), an IV pole and bags, 2 bedpans, a medical tray, and sundry first aid supplies. The full-size simulated patient’s cot is located in the middle of the room, while the human subject’s cot is next to the walls.

The ETU is equipped with 1 regular webcam and 2 ultra wide angle webcams. These provide greater situational awareness for the teleoperator and the safety attendant. The sensors were controlled via ROS drivers on a desktop machine on the outside of the ETU.

Fig. 2: Patient’s perspective of the room, including view of simulated patient and medical supplies.

D. Safety

The PR2 robot is equipped with both onboard and wireless emergency stops. A human safety attendant continually monitored the scene via a remote screen with live feed from 3 different cameras in case intervention was needed. The robot’s maximal gripper effort was set to 30N. Participants were given an orange flag with instructions to raise it in case they desired to stop the experiment.

A thorough explanation of the experiment and the inherent risks was given to each potential participant before the experiment started. Consent was required, and each participant was free to stop or leave the experiment at any time, for any reason. The university’s IRB approved the study.

IV. METHODS

A. Experiment Design Overview

To test our 3 hypotheses we conduct a 2 condition, between-subjects experiment wherein we vary the participant’s visibility of the operator. Participant visibility of the operator is broken into 2 conditions described below.

The study is broken into 3 stages and takes roughly 35 minutes. The initial stage consists of informing the participant of the study via a brief overview, obtaining written consent to participate, and filling out an initial questionnaire. The participant is then taken to the simulated ETU and
shown how the E-stop works on the robot. Following that, the participant puts on an exercise heart rate monitor and the patient gown. This stage takes approximately 10 minutes.

In the second stage the participant is led into the ETU. The GSR is attached to the participant’s left hand. The subject is then instructed to lay down. Recording of the psychometric signals begins, as well as video and audio recording if the subject gives consent. A 2 minute baseline of the signals is collected, with the operator in the ETU standing beside the robot for roughly 1 minute. The operator then leaves the ETU and closes the door based on the visibility condition. The robot is activated and executes 5 tasks in close proximity to the participant via teleoperation. Execution of the tasks takes approximately 8 minutes.

The participant is then led out of the ETU and instructed to take off the gown. The post-hoc survey is administered. The participant takes off the exercise heart rate monitor, is paid $20 for their time, and allowed to ask questions.

B. Participant Visibility Conditions

In both conditions, the teleoperator is located outside the ETU approximately 5m away from the the participant’s cot. The only thing changed between conditions is the material of the entryway door covering. Participants are randomly assigned a condition.

In the no visibility condition an opaque sheet of plastic, of the same material as the rest of the ETU, is drawn over the entryway to the room. Thus, the operator is not visible to the participant at all during the robot’s task execution. In the visibility condition the participant and operator are physically separated by a sheet of 4mm clear plastic drawn over the entryway to the room, allowing the participant full visibility of the operator, constrained by the participant’s requirement to stay on the cot. The participant’s view of both conditions is shown in Figure 3.

C. Tasks

The robot performed five tasks in the ETU, in close proximity to the participant:

1) Deliver wrapped gauze to the medical tray
2) Pick up a no-touch thermometer and hold it over participants torso
3) Move IV pole closer to the participant’s cot
4) Remove the wash cloth hanging on the wall next to the participant and place it in a bucket
5) Move the bucket from the end of the cot to the other side of the ETU entrance

The primary criteria for task selection is that each had to be feasible and consistently repeatable. This was a two-fold decision. First, we were interested in what robotic technology could actually do in the near-future. Second, we wanted to standardize robot performance between participants.

We also wanted realistic tasks that would lighten the workload of the health care staff. The selected tasks were deemed helpful based on conversations with trained medical personnel (nurses and doctors), some of whom were principal leaders in the latest Ebola response [19].

Lastly, we wanted tasks that are generalizable to other tasks in this domain. The selected tasks involve the robot picking up objects of various sizes, stiffness, textures and colors that requires the robot to use its full range of height and to traverse much of the ETU space.

D. Participant Arrangement

Participants were recruited from the surrounding community and campus. They were required to proficiently speak and read English and be at least 18 years old. 23 people participated in the study (8 females and 15 males, ages 18 to 55, median of 24). All but one had at least some college experience. 19 participants gave consent to allow us to record audio and video.

We wanted the participant to feel like he/she was in a medical environment as much as possible. Not only was our ETU realistically modelled, we also sprayed a small amount of bleach under the participant’s bed before entering to give the room a sterilized smell.

In the initial briefing for the experiment, the participant was told that we were interested “in human-robot interactions” and “how to use robots to help in the fight against infectious diseases." We framed the robot to the participant as “a robot, controlled by a human operator” twice in the study overview and once in the consent form. With the exception of specifying that the robot would be “moving the IV pole from one cot to another,” the actions of the robot were not specified in advance. We characterized the robot’s actions as “general health care tasks in the ETU.” It was not disclosed to the participants that we were studying their levels of comfort and trust. The initial briefing and surveys were conducted outside of the ETU, with partitions blocking the subject’s view of the ETU.

The participant was shown how the emergency stop functions on the robot in the ETU before putting on a heart
rate monitor in another room. Then the participant put on a standard hospital gown over his/her clothes. He/she was led into the ETU where the GSR was hooked up. Further questions by the participants were deferred until the end of the study unless the questions regarded safety practices.

E. Questionnaires

An initial questionnaire was administered to each participant following the consent process. Basic demographic information was collected, including 4 items relating to age, gender, level of education and current employment status. Each participant then took the Negative Attitude Towards Robots (NARS) [43] survey and the PTS.

Following the robot task execution stage, the participants returned to the initial study area to take a post-experiment survey. The post-experiment survey is broken into three categories: operator trust, robot trust, and comfort (see Table 1). Each question is on a 5-point Likert scale, with 5 being “Strongly Agree.” The operator and robot trust questions were designed to follow the language and spirit of the trust questions found in the PTS and the work by Chen et. al [14], [44]. Since the sensor is highly sensitive to physical movement, the participant was instructed to lie on his/her back with his/her instrumented hand on the cot for the entire robot task execution stage. The GSR outputs 0-5v and was sampled at a rate of 3.3 Hz.

The second minute of recording was used to calculate a baseline for the individual participant using Equation [1]. A baseline is needed since the gain has to be hand-adjusted in the initial reading period and since the initial reading period varies per individual.

\[ \text{Baseline}(gsr_p) = \frac{\sum_{t=200}^{400} \text{volt}_t}{200} \] (1)

This is used to calculate a proportional GSR value:

\[ \text{PropGSR}_t(gsr_p, baseline_p) = \frac{\text{volt}_t}{\text{baseline}_p} \] (2)

Task-phase proportional means (TPPM) for each participant are calculated using Equation [3]. Start and end times for the task-phases were calculated by taking the mean of hand-coded times from the recorded sessions.

\[ \text{TPPM}(g_p, b_p) = \frac{\sum_{t=\text{phase}_{\text{start}}}^{\text{phase}_{\text{end}}} \text{PropGSR}_t(g_p, b_p)}{\text{phase}_{\text{end}} - \text{phase}_{\text{start}}} \] (3)

This is used to calculate the proportional mean during a task-phase for an entire conditional group using Equation [4]

\[ \text{CondTPPM}([p], [g], [b]) = \frac{\sum_{i=1}^{n} \text{PPM}(p_i, g_i, b_i)}{n} \] (4)

One-sided t-tests are performed on TPPMs between significance to check for significant differences.

V. Results

A. Questionnaire Results

For the survey responses, n=11 for the no visibility condition and n=12 for the full visibility condition.

Figure 4 is a boxplot for trust in the operator between conditions. A higher y-value represents a higher level of trust in the human operator. The median and mean are above the neutral line for operator trust for both conditions. However, the full visibility condition had a higher mean than the no visibility condition (M=4.38 vs. M=3.95). Thus, the full visibility group expressed higher levels of trust in the human operator than the no visibility group. This difference is significantly explained by the condition (p<0.05, Cohen’s d=0.81) and supports H1.

Figure 5 shows the boxplot for trust in the robot by each condition. Again, a higher y-value indicates higher levels of trust. The medians are equal whereas the mean in the full visibility condition is higher than the no visibility condition.
mean (M=3.77 vs M=3.56). While interesting, this result is not statistically significant enough ($p>0.1$, Cohen’s $d=0.30$) to support H2.

Figure 6 shows the self-reported level of comfort for each condition. Comfort levels determined by the questionnaire show a lower mean and median in the full visibility group. While in contrast to our initial inclination, these results are not significant enough to reject H3 ($p>0.1$, Cohen’s $d=0.34$).

**B. GSR Results**

Three of the participants’ data had to be excluded from analysis because the signal was saturated immediately following the baseline time period, indicating a poor gain setting. Thus, for the GSR data, n=9 for the no visibility condition and n=11 for the visibility condition.

Figure 7 shows the average proportional GSR reading for both conditions (note again that minute 2 was used as a baseline and the robot did not start moving until after minute 2). A higher proportional GSR reading indicates greater levels of arousal and excitement. The no visibility group had a higher level of arousal compared to the full visibility group throughout the trial. Both conditions experienced an increase between minutes 3 and 4, but the rise for the no visibility condition was much higher. The full visibility condition’s signal looks much smoother, indicating fewer changes in arousal across time.

Figure 8 shows the proportional task-phase mean GSR values for each condition. The no visibility condition’s mean is higher than the full visibility condition at each stage and across all time. The difference in the proportional mean across all time yields a marginal degree of significance ($p<0.1$). This suggests that the no visibility condition had higher levels of arousal/valence compared to the full visibility group during the robot’s task execution, and only
marginally supports H3.

In agreement with our intuition, Figure 8 suggests that the “Thermometer” phase, where the robot held a no touch thermometer over the torso of the participant, is the most arousing/exciting task-phase for both groups, aside from the baseline reading.

VI. DISCUSSION

In support of H1, our results suggest that people are less trusting of the human operator when the operator is unseen. The reason for this effect is still unknown. The patients may tend to project more autonomy on the robot when the operator remains unseen during task execution, and this might lead them to doubt the operator’s ability to command and control the robot. Contrast this to the full visibility condition where the patient is continually reminded of the operator’s role in the control and manipulation of the robot. This might bolster the patient’s trust of the operator because the patient can constantly link the operator’s capabilities and power to the robot’s actions.

In this experimental design, the operator was introduced to the patient before the robot even began to move and had at least 5 minutes of interaction with the patient during the consent process and safety explanation. The patient even walks by the operator’s control unit. We hypothesize that the effect size may have been much greater if the patient did not meet the operator beforehand or see the control station. A less-rational state due to medication or symptoms may also increase the effect, causing patients to project more autonomy to the robot, further decreasing their perceived ability of the operator to control the robot.

From the support of H1 it follows that a patient-centered robotic ETU would feature a mobile teleoperator control unit that could be temporarily stationed outside of patients’ rooms. This would allow each patient to see the operator while the operator controls the robot. This would in contrast to a central teleoperation control unit which would not allow patients to see the person manipulating the robot.

The deeper implication of H1 is to question the current and imagined-future form of telemedicine. Obscuring the operator behind the robot could be disconcerting and off-putting for many people in regards to their trust of the operator unless, as with most current robot surgery, the patient is unconscious. A loss of operator trust could mean a decrease in long-term physician-patient relations and a lower likelihood to seek future medical care. More work needs to be done to evaluate how things such as screens, speech communication, or even greater familiarity with robots can help overcome this specific barrier to operator trust in telemedicine applications involving physical action around the patient.

Our results do not support H2; nevertheless, it is very interesting to find that the patient’s trust of the operator varied significantly while the trust in the robot did not. Participants may have been assessing different attributes of trust when evaluating the operator versus the robot. The trust attributes assessed for the operator (e.g., level of control) may have varied greatly between conditions. The same measured trust attributes might not have varied as much for the robot. This might help explain why there was not a significant difference in robot trust.

Our results are somewhat mixed in regards to H3. The survey responses do not support H3, and if anything, point the other way. The GSR data however shows that the no visibility condition experienced significantly more arousal across time compared to the full visibility condition in support of H3. The higher levels of excitement in the no visibility condition coincide with their lower levels of operator trust. The lower levels of arousal in the full visibility condition coincide with their higher levels of operator trust. Our results then fit an inverse pattern between trust levels (in the operator) and GSR response. This seems to make intuitive sense that, all else being equal, the more trust a person has in the human operator, the calmer he/she would be. Thus, the GSR data not only supports H3, it also fits with H1.

Participants may have responded to the comfort questions with more regards to the physical attributes of the ETU rather than the overall situation of the robot moving around in the space. This could explain variation in the results.

There appears to be a general downward trend in the GSR results for both conditions. This may suggest that people adjust to the robot over time. The downward trend could also be explained by the variation in the tasks. People were more excited when the robot hovered over them with a no-touch thermometer than when the robot dropped the towel in the bucket. To better understand the overall downward trend the GSR baseline should not start until the person’s GSR voltage stabilizes and tasks should be varied. Also, these results are drawn from a limited interaction time between the robot and test subject; it is unclear how longer interaction times, on the order of days or hours, would impact the levels.

There are limitations to the current study. While the support of H1 is statistically significant with a large effect size, the result is limited in statistical power. Ideally, at least double the number of participants is needed to increase statistical power. Lastly, interviews may have aided in the interpretation of the results.

VII. CONCLUSION

In this work we explored how patients’ visibility of the teleoperator affects their own levels of trust in the operator, robot, and their overall comfort. We tested our three respective hypotheses by having a human-sized robot complete general health care tasks around participants under two conditions of operator visibility. The experiment took place in a simulated Ebola treatment unit.

Our major contribution suggests that a patient’s increased visibility of the operator leads to higher levels of patient trust in the operator. More needs to be done to assess the reasons for this effect and ways to potentially overcome this barrier to trust. The mixed results regarding H3 invite more research and clarification into how visibility impacts patient comfort.

Robotic technology in infectious disease outbreaks appears to offer tremendous benefits to health care worker safety, but
more work needs to be done to explore the particular challenges of deploying these systems around human patients.

VIII. ACKNOWLEDGEMENTS

This project was funded by the National Science Foundation under award IIS 1518652 and IIS 1540483.

REFERENCES


