Research Article

Zachary Henkel, Jesus Suarez, Vasant Srinivasan, and Robin R. Murphy

Medical Field Exercise With a Social Telepresence Robot

DOI 10.1515/pjbr-2016-0001
Received October 1, 2014; accepted May 9, 2016

Abstract: This article reports observations from a field study in which medical responders used a social telepresence robot to communicate with participants playing the role of a trapped victim in two search and rescue exercises. The interaction between the robot, victims, and responders suggests the coexistence of two distinct social identities for the robot. One which is a pure conduit for the remote medic, and another in which the robot is treated as an independent social actor. Participants acting as victims demonstrated fluidity in interacting with each identity. The social identify of a robot has important implications for the development of future telepresence systems, particularly in the healthcare domain. Since victims in the exercises gave attention to both the robot and the remote medic, it is possible that the robot's social actor role may divert attention from the remotely connected individual. The work provides a starting point for investigation of role conflict between a remote medical professional and the robot they are using to assist a patient.

Keywords: human-robot interaction; robot social identity; assistive robotics

1 Introduction

Robots are increasingly used in contexts where humans are highly dependent on a robot that serves as medium to a remote world or resource, such as eldercare [5], emergency response [1], and healthcare [15]. For example in eldercare, the robot acts as a link between an elderly patient and medical provider for interactions that may be intermittent or routine checkups. In these situations the human ("dependent") is connected to multiple other humans ("controllers") via the robot which acts as a proxy for the duration of the interaction. This scenario presents numerous challenges because attention may be split between the robot itself and the remote operator.

This work examines interactions between a social telepresence robot and participant playing the role of a trapped victim in a search and rescue scenario. Since victims trapped in the rubble of a disaster may be isolated for several hours while they await extrication, social telepresence robots are a promising medium for receiving support from the outside world. The study observed how two participants inserted into a simulated collapsed building at Disaster City ®responded to a robot controlled by a team of five medical responders from Texas Task Force 1. This setting was selected because its high fidelity allowed medical responders to explore using the robot in a scenario similar to actual response events. This setting also causes participants acting as victims to be highly dependent upon the robot, with results that are expected to transfer from search and rescue to healthcare.

The robot, Survivor Buddy, was tele-operated by the responders and exhibited different sets auditory and visual behaviors based upon responder commands and availability of the robot's subsystems. The medical responders primarily worked through Survivor Buddy to assess the dependent's condition and query the victim as to his state.

The observations collected during study suggest that the robot is implicitly recognized as a unique social entity and is given attention by the participant. The participant's decision to channel attention to the robot or remote operator is characterized by which entity the participant attributes an action to. The study also indicated that while participants demonstrated no evidence of confusion in shifting their attention between the robot and remote operator, the potential for a competition for attention between the two exists. Additionally, the study identified eight factors that assist in characterizing the participant's attentional focus.

The article is organized as follows: Section 2 provides a brief summary of prior work that is helpful for understanding the presented study. Section 3 summarizes the approach taken in the exploratory study. Section 4 details observations from the field study, while Section 5 de-

Zachary Henkel, Jesus Suarez, Vasant Srinivasan, Robin R. Murphy: Texas A&M University, 77843 College Station,TX USA, E-mail: zmhenkel@tamu.edu, jsuarez@cse.tamu.edu,, vasants@cse.tamu.edu, murphy@cse.tamu.edu

BY-NC-ND © 2016 Zachary Henkel et al., published by De Gruyter Open.
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 License.

scribes findings and relates them to prior work. Section 6 discusses implications and future work. The article concludes with Section 7 providing a short summary of the contributions.

2 Related Work

Three areas of prior research provide a helpful foundation for understanding the design, implementation, and interpretation of observations from this field study. This section briefly describes related work concerning telepresence robots, the use of social robots in search and rescue operations, and social identity theory.

2.1 Telepresence Robots

Telepresence robots offer the ability for remote users to connect to a distant location and acquire a sense of embodiment within the remote location. Telepresence robots have been extensively studied and utilized in a range of environments including offices, medical facilities, schools, and homes [12–14]. A wide variety of form factors and features are available in both commercially marketed systems and those developed primarily for research purposes [4].

Though telepresence robots are increasingly maturing, most research efforts focus on enhancing the user experience of the two humans communicating through the robot, considering the robot itself to be merely a medium or tool for facilitating this interaction. Research efforts which do focus on providing autonomous abilities to telepresence platforms often focus on abilities which facilitate the use of the robot as an appliance (e.g. automatic avoidance of low wifi coverage areas, automatic navigation to a charging station) or features which protect the platform (e.g. collision avoidance, stair detection) [4]. The overall purpose of telepresence robots is to enhance the communication between humans, carefully avoiding any ascription of agency to the robot itself. This goal is appreciable for a majority of telepresence applications, but scenarios like search and rescue operations may benefit from allowing the robot itself to take on a social role within telepresence scenarios.

2.2 Search and Rescue Robotics

During disaster scenarios those occupying the impacted area may become trapped, injured, and hidden from view, requiring emergency response experts to conduct search and rescue operations. Often it is advantageous to deploy mobile robots into areas which cannot be easily or safely searched by humans or canines [7]. A majority of prior work in this area has focused on examining the usability of robotic systems in disaster scenarios by quantifying the effectiveness of response teams using robots through measures of team performance and individual cognitive load [2, 6].

However research efforts have also expanded into the realm of using a search and rescue robot as a social communication platform to connect trapped victims and responders [3]. Upon discovering a trapped victim, people of many different roles (e.g. physicians, structural engineers, family members) will have an interest in communicating with the victim. Additionally, the victim may be isolated for several hours while waiting for extrication. This leads to the use of social telepresence robots in search and rescue domains becoming an area for further investigation.

2.3 Social Identity Theory

The Computers are Social Actors (CASA) model explains that humans often treat computers as social entities [8]. As human-robot interaction has become a field of study, the CASA model has been demonstrated to hold true when humans interact with robots. The model suggests that machines will either be regarded as a pure medium (e.g. a television set) or as a social actor (e.g. a social robot), depending on their specific qualities. However, the CASA model does not directly address the situation in which someone will be communicating with another human through a social robot.

Though many studies in social robotics have utilized a Wizard of Oz method (a remote human operator) for controlling social robots within lab studies, but this has the goal of participants believing the robot is a social actor with no remote human controlling it [9]. In a scenario somewhat similar to telepresence, Straub et al. used a public space setting to observe the control of and interactions with a tele-operated humanoid robot, with a focus on examining the identity ascribed to the robot [11]. The study examined whether participants, acting as operators, would create a unique identity for the robot or channel their own identity, while also examining if participants interacting with the robot would ascribe behavior to the robot or to the operator. Through video and audio coding, the study found that aside from two interactions with researchers, participants did not address the operator through the robot, rather they always directed their utterances to the robot as a unique social actor. The work presented in this article is distinct from prior studies, in that an explicit need for the operator to communicate through the robot in addition to the robot being its own social entity exists.

3 Exploratory Study

A field study, designed to be exploratory in nature, was employed as a vehicle for gathering responses to the robot's behavior. This approach was selected because it provides critical insight into the needs and behaviors of the actual users of search and rescue telepresence systems while also providing emergency responders an opportunity for further training. The unconstrained nature of a field study allows for the development of new research questions, which can be later examined and bolstered by rigorous controlled lab experiments. Thus, field studies can take on a methodological role by assisting in informing the design of lab experiments.

3.1 Scenario

Figure 1: The simulated disaster site used for the exploratory study.

In order to emulate actual disaster response conditions, the study used a physically simulated building collapse environment, located in the Disaster City ®training complex in College Station, TX (see Figure 1). The interaction was configured to be an unconstrained interaction between the trapped victim (dependent), rescue robot, and medical responders. Medical responders followed standard procedures for medical triage in an emergency event [10]. This type of content and communication was chosen to investigate the interaction patterns which would emerge as responders utilized a robot for conducting the triage process. The role of the robot was to facilitate interaction by acting as a proxy for medical responders, while also acting as a companion to the victim. Medical responders were asked to use the two-way communications capability of the Survivor Buddy robot to perform triage and gauge the severity of the injuries of the trapped victim.

Similar to actual deployment scenarios, responders worked from a large tent setup adjacent to the site of the simulated disaster. Responders did not have a clear line of sight to the simulated building collapse site and operated exclusively through the video feeds provided by the robotic platform. Researchers were able to monitor the responder and victim participants through a multi-camera surveillance system which was installed both in the responder tent and in the disaster site.

3.2 Robotic System

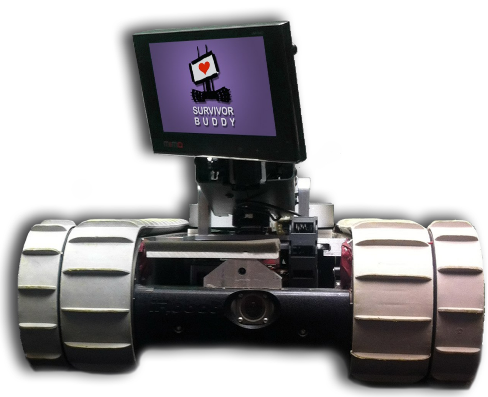

Figure 2: The robotic platform used to interact with participants.

The exercise used a iRobot SUGV robot, with a Survivor Buddy, affective robotic head to interact with victims (see Figure 2). The Survivor Buddy head is a four degree-of-freedom affective head which consists of a touch screen, webcam, speakers and microphone. The Survivor Buddy head is intended to aide in communicating with victims, as well as to increase their comfort levels while trapped. In the exercise, responders approached the trapped victim with the SUGV robot, activated the Survivor Buddy head and attempted to communicate with the victim.

For the purposes of this study, the base robot (SUGV) was controlled using a computer monitor and standard gamepad controller, while the Survivor Buddy head was controlled using a web interface on a standard laptop. The Survivor Buddy interface allowed responders to access:

two-way audio and video chat, robot head movement controls, synthetic speech, and screen sharing.

3.3 Participants

The exploratory study used two participants (Participant 1 and Participant 2), who were volunteers familiar with canine search trials to play the role of victims, and five medical specialists from Texas Task Force 1, acting as responders. All the participants were healthy adult males above the age of 18. This population was chosen because the volunteers had prior experience with confinement for long periods of time in a physically simulated disaster scenario and the medical personnel had extensive experience with triage and interacting with dependents who are trapped victims as a part of their duties with Texas Task Force 1.

3.4 Procedure

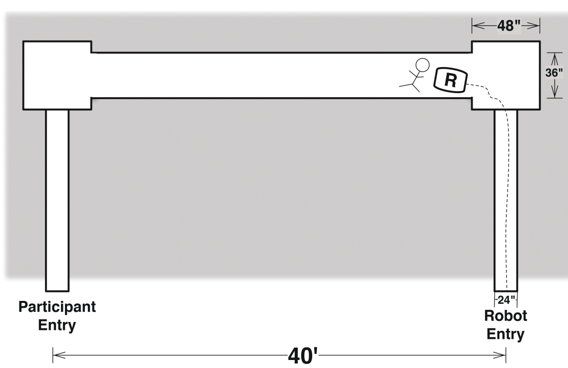

Figure 3: Layout for simulated disaster site.

The study was divided into three segments: 1) pre-interaction tasks, 2) interaction tasks, and 3) post-interaction tasks. The entire study took an average of 2 hours per participant, with participants spending nearly an hour with the robot.

During the pre-interaction phase, participants read and signed the consent forms, as required by the Texas A&M University Institutional Review Board and Disaster City ®. Participants playing the role of a victim were told that they would enter a confined space and remain there for up to 180 minutes. It was explained that during this time they would interact with a search and rescue robot, but no specific instructions or expectations were provided. The participants were given a radio which served as means

Figure 4: View of a participant interacting with the robot.

of communication in case of an emergency. Participants then entered into the rubble from a separate tunnel (left side of Figure 3) and made their way to a location marked with a glow stick. Once at the designated location, participants were asked to lay down on their right side. The study site was dark except for the natural light entering through cracks which was not substantial. Figure 4 shows a participant laying in the simulated disaster site and interacting with the robot. Responders participating in the study also reviewed and signed consent forms at a separate location from which the robot was operated. Responders were then briefed on the capabilities of the robotic platform and were provided answers to any specific questions about the robot.

At the beginning of the interaction phase, the Survivor Buddy robot mounted on the SUGV base entered the rubble through an opposing opening (right side of Figure 3) after the participant was in position for at least five minutes. The robot approached the victim, and the Survivor Buddy head opened, then medical responders attempted to introduce themselves to the victim. The medical personnel conducted triage and the interaction lasted for average of 39 minutes. Responders posed questions like: - *"Can you move your fingers?"*, and *"Do you see anyone else with you?"*. Though researchers were present to answer any questions responders might have about using the system, the interaction was guided primarily by the medical response team. Researchers were responsible for driving the robot base into position, but took instruction directly from responders. Once the interaction was completed, participants playing the victim role were informed through their radio that the interaction was complete and were asked to exit the rubble through the tunnel they entered from. The experiment was run in a dedicated area during mid-morning and afternoon hours.

In the post-interaction phase, participants were interviewed by two members of the research team. They were asked to describe their feeling during the interaction with the robot, whether they had positive or negative experiences, and what aspects need to be improved. The participants were then debriefed about the study after which their participation was complete.

Data was collected using video observations and unstructured audio-recorded interviews. Four infrared night vision cameras were utilized because the robot interactions occurred in low light conditions. Images and audio were obtained from four camera perspectives: participant front view, participant rear view, medical personnel left view, medical personnel right view.

4 Observations

Two disaster response simulations were performed using separate participants as victims, but the same medical response team. This section describes the qualitative narratives of each trial and the critical interaction events that were extracted from recordings of the trials. Since all events were extracted from video recordings of the exercise, some of the events captured were not available to those participating in the study (e.g. when robot audio systems failed participants could not hear each other, but this data is captured in video recordings).

4.1 Analysis Method

All reported observations are based upon video recordings of the two sessions along with audio recordings of interviews with participants playing the role of victims. The analysis began with two researchers familiar with the scenario reviewing video recordings of each session and creating independent listings of the robot's actions, the victim's actions, and the responder's actions linked to the video timecode. These reviews were first conducted with a focus on each single entity (e.g. only the robot's actions without regard for the victim or responders). Researchers then independently viewed each session as a whole and developed narratives along with notes of salient or interesting events. The independent listings of events were then compared and the two researchers resolved any discrepancies by reviewing the relevant portions of footage (using multiple camera angles).

Once a clear timeline of events was produced, the two researchers considered the audio recorded interviews with participants in conjunction with recorded video footage to better infer the perceptions participants had during the interaction. Though a majority of the events coded from reviewing video recordings are objective occurrences (e.g. the participant spoke, the robot moved), subjective judgement was required to interpret the meaning of these events in the context of social identity theory (e.g. sharp, loud, single-word responses were directed towards the robot itself, not remote responders). As a result, we emphasize that these observations are useful for provoking new areas of study, but at this point have limited viability in informing system design.

4.2 Interaction Narratives

This section offers a qualitative description of each exercise, while Figure 5 illustrates the occurrence of eight types of events over the interaction time for each participant's interaction. While reviewing recordings of the sessions, researchers found these eight event types to be useful in understanding interactions with the robot at any given time during the scenario and are further explained in Section 4.4.

Participant 1's interaction began with three minutes of the robot positioning itself and then opening the Survivor Buddy head. Immediately after opening the head, responders attempted to establish a two-way audio and video communication link with the victim. Due to subsystem failures, the audio stream of the two-way communications was not operational, leaving only a two-way video stream. During the first 15 minutes of the interaction, responders were unaware of the audio link failure. Since responders were not briefed on the instructions given to the participant, it was plausible that the participant was portraying an injury or cognitive confusion. Nearly 10 minutes into the interaction researchers were able to verify the instructions given to the participant and began to investigate potential audio system failures. Throughout the session Figure 5 indicates that the responders were disengaged at times. These are instances of responders conferring with each other to form a plan or of a researcher using the responder interface to diagnose system errors. After 15 minutes of the interaction, researchers verified that the participant could not hear the responders and informed responders of this limitation. During this time the participant fixated his gaze on the video feed, but did not attempt to communicate. Once responders were informed that the audio link was not working, they then began to display hand written signs over the video link to ask the participant standard triage questions. The participant responded both verbally and with gestures, de-

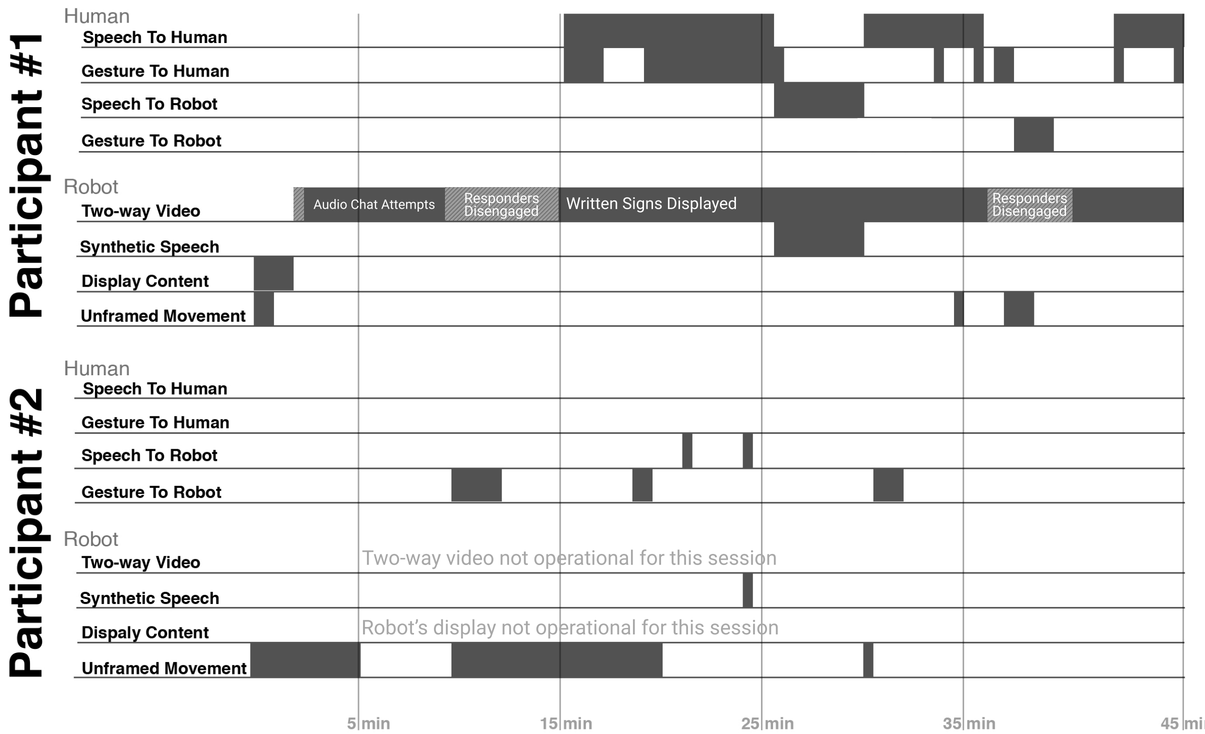

Figure 5: Timeline of the events occurring during each session.

spite being informed by responders that the audio link was not working. Approximately 25 minutes into the interaction, responders attempted to use synthetic speech to ask questions of the participant. The participant responded as if he was talking to a recognition system (i.e. carefully pronounced reduced vocabulary), but became frustrated when portions of the synthetic speech were not always understandable. The participant also responded to the synthetic speech using gestures to indicate frustration. After approximately 5 minutes of synthetic speech, responders switched back to using hand written signs for communication. At approximately 36 minutes into the interaction, responders disengaged from the video link (though the stream was left with no human appearing in the frame) to plan their next steps. During this time the robot engaged in unframed head movements. The participant responded to these head movements by using a gesture to re-indicate his injury to the robot, though there was no request to do so from responders. After 45 minutes of interaction responders ended the interaction and the participant was removed from the scenario.

Participant 2's interaction began with the robot approaching and opening the Survivor Buddy's head. The victim did not respond for several minutes, as unknown to responders and researchers, both the Survivor Buddy display and audio link were inactive. This resulted in responders being able to see the victim, but the victim was not able to see or hear the responders. Responders and researchers identified the audio link failure, but remained unaware of the failure of the robot's display for the duration of the interaction. To elicit a response from the participant, the responders cued the Survivor Buddy head to pan and tilt and make *Yes* gestures. The participant responded to these gestures by mimicking them back to the robot. Approximately 20 minutes into the interaction the participant briefly spoke to the robot, asking if anyone could hear him. At approximately 24 minutes into the interaction, synthetic speech was used, but was unintelligible, the participant responded verbally to the robot stating they did not understand. The participant continued to track the head movements of the Survivor Buddy robot for the remainder of the interaction. The participant was unable to communicate with any responders during the scenario. After 32 minutes the responders terminated the scenario and the participant exited.

4.3 Participant Interviews

Immediately following their interaction session and departure from the simulated disaster environment, both participants playing victim roles were interviewed by researchers about their experience.

Participant 1 indicated in his interview that he suspected responders were unable to hear him during the interaction. He also indicated that he was not certain where the camera was located on the robot and could not always see the content on the screen. The participant also indicated a preference for the robot being more physically proximate to him, and did not find the robot to be "creepy" or intimidating. Additionally, the participant confirmed that the synthetic voice was very difficult to understand within the environment. The participant also stated that he viewed the robot as "basically a person" assisting in rescuing him.

Participant 2 felt that the robot was physically close enough to him, but also expressed concerns about being unsure where the robot's sensors were located (i.e. camera location). Like Participant 1, he also indicated that the synthetic speech was not understandable. During the interview it was discovered that the robot's display had not been functional during the interaction session, however the participant assumed the robot's state to be typical and did not report the absence of screen content as a failure or malfunction. The participant also commented that the physical presence of the robot was comforting. The participant did not assume responders were controlling the robot. The participant attributed the robot's movements to the robot itself (not a remote operator), though he did express the belief that because the robot had located him people in the outside world were aware he was trapped. This is consistent with his verbalization 24 minutes into the interaction in which he asked if anyone could hear him.

4.4 Critical Interaction Events

Since the study was unconstrained and free of strict manipulations, analysis of the participant's interaction with the robot was conducted by examining particular events captured by video recordings of each session. Two researchers familiar with the study examined video recordings of the sessions and converged on eight event types of interest to be further examined from each session.

These eight event types were selected because they were easily observable attributes of both the human participant and the robot. Four of the event types were related to the robot's actions, while four were related to the participant's responses. The event types for the human's actions were:

1. *Speech To Human:* this event occurred if the participant was speaking to the human operator through the robot. An example of this behavior can be seen in Figure 5, on Participant 1's Speech To Human track, beginning at the 15 minute mark. The participant responds verbally to the responder's holding up a sign over the video link which is asking a question. In order to determine if the participant was speaking to the human, the context, content, and intonation of the speech were considered. For example, if the utterance was a direct response to a stimulus from the remote responder (like responding to a sign displayed over the video link), it was obviously intended as speech to the human.
2. *Gesture To Human:* this event occurred when the participant made a purposeful gesture which was intended to communicate information to the human responders working through the robot. An example of this occurs in Figure 5, on Participant 1's Gesture To Human track at approximately the 15 minute mark. In this example, responders asked the victim to point to their injury and the participant replied to the responder using a gesture to indicate the location of their injury (pointing). It is also notable that context and other cues are helpful in determining if a gesture meets the requirements for this event type. Some gestures existed without a targeted recipient (slight movements or adjustments in posture), while others like Participant 2 mimicking the robot's movements were considered as provoked by the robot itself.
3. *Speech To Robot:* this event includes any speech which was directed towards the robot itself or expected to be processed by the robot, rather than directly to the human responders working through the robot. An example of this is found in Figure 5 on Participant 1's Speech To Robot track at approximately the 25 minute mark. In this instance the participant was responding to synthetic speech which was asking a question. Speech to the robot was indicated by context, content, and intonation. For example, when prompted by the robot's synthetic voice, Participant 1 altered his responses to be loud, sharp, and short, creating the intonation commonly used with speech recognition systems. This alteration of speech clearly indicated that the participant was directing shaping his verbal responses for recognition by an automated system.

4. *Gesture To Robot:* this event occurred when the participant made a meaningful gesture which was intended for the robot's recognition. In Figure 5 on Participant 1's Gesture To Robot track at approximately 37 minutes this type of event occurs. In this case, the participant begins to indicate his injury by gesturing, not because of prompting by responders, but in response to the robot moving its head in a scanning manner. Though it is possible this was intended for responders, it is unlikely as responders had posed injury-related questions earlier in the dialog and were not actively engaged with the participant when the robot began to move its head.

In addition to the four event types considered for the human's actions, four event types for the robot's actions were also considered. These attributes require less observation and human judgement to be determined, and can be obtained directly from the robot's system logs. However, to ensure correctness these actions were also determined by examining video of the interactions. The four attributes for the robot's actions considered were:

1. *Two-way Video:* this event indicates that a video stream of the human responders was active on the robot and the participant could see the responders live. Due to subsystem failures, in this study this does not indicate a two-way audio signal, only a two-way video signal. Figure 5 indicates that for Participant 1, video conferencing was active for most of the interaction. However it is important to note that at times responders stepped out of view of the video feed, leading to the robot displaying the responder's environment without a responder in frame.
2. *Synthetic Speech:* this event occurred when the robot used synthetic speech to communicate with the participant. This occurs in Figure 5 on Participant 1's Synthetic Speech track at the 25 minute mark. In this case synthetic speech was used in an attempt to ease communications between responders and the participant, as the two-way audio subsystem had failed. The voice used was clearly synthetic and was unlikely to be confused with natural human voice.
3. *Display Content:* this event indicates that the robot's screen content was not a video stream, but rather a static image or logo which was not associated with human responders. As Figure 5 indicates, this only occurred at the beginning of Participant 1's interaction.
4. *Unframed Movement:* this event refers to the robot moving its head (pan or tilt) without the remote operator explicitly linking the action to their own intentions. In contrast, a framed movement is one which the remote operator clearly initiates. Unframed movements were most likely to occur in Participant 2's interaction with the robot, as remote responders were not able to communicate through video or audio with the victim. It is important to note that some participants may still attribute unframed movements to an operator.

5 Findings

Using observations from the two interactions combined with prior work detailing the social nature of computer systems and the social identity spectrum, five useful findings were distilled. First, it appears that participants experienced the robot as both a pure medium (i.e. communication device) and a social actor (i.e. the robot having intentions independent of remote responders), with Participant 1 possibly experiencing a middle ground, the robot as a social medium. Second, it was observed that the target of participant attention was affected by both the actions of the robot and of the remote responders. Third, despite a lack of framing the robot's abilities or purpose, an implicit social identity appeared to exist for the robot. Fourth, the analysis method used suggests that social identity may be a factor of who or what the participant attributes the robot's actions toward. Fifth, both participants exhibited intriguing responses during their sessions, with Participant 1 transitioning between social identities and Participant 2 mirroring the robot's movements.

5.1 The Robot's Social Identity

The Computer's are Social Actors (CASA) model by Nass et al. lays the foundation for a robot having a social identity[8]. When a remote operator is communicating through an appliance like a phone or computer to someone, the remote operator is regarded as a social actor and the appliance is typically regarded as a pure medium. However, if an appliance (or robot) is recognized as having social features itself and is being communicated with directly, it is regarded as a social actor. The scenario in which both the communication appliance (the robot) and remote operator have the potential to be recognized as social actors is unique to telepresence applications and has remained relatively unexplored.

While the Computers Are Social Actors paradigm highlights two extremes: the pure medium (e.g. a radio) and the social actor (e.g. a social robot), social telepresence devices which can both channel a remote social actor and themselves be a social actor present the potential for a third social identity: a social medium.

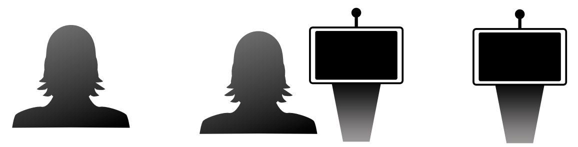

Figure 6: The social identity spectrum.

The social medium is posited to be a state in which the robot is not solely a pure medium and not solely a social actor (see Figure 6). For example, when Participant 1 shifted his attention to the remote operator, he was temporarily treating the robot as a pure medium. When remote responders moved out of video view to confer with each other and the robot began to move its head, Participant 1 responded with gestures to his injury, seemingly in response to the robot's own actions. In this case the participant was likely treating the robot as a social actor. When encountering a scenario of responders visible on screen, but the robot using synthetic voice to pose questions, Participant 1 adapted his speech responses for speech recognition, indicating he presumed the robot itself was attempting to understand his speech rather than remote human responders. This demonstrates both transitioning between social identities and the possible co-existence of the robot as a social actor and as a medium.

Such a scenario is not necessarily the equivalent of two humans and a robot interacting with each other. Since a telepresence robot channels the remote human social actor, it may have an advantage in gaining attention in comparison to the remote operator. It is also possible that a telepresence robot which does not take any actions to gain its own social agency, may remain as a pure medium. Framing of the scenario also influences the likelihood of a robot acquiring its own social actor identity. Though a full enumeration of specific actions that gain a telepresence robot social agency is not known, providing direct attribution of the robot's actions to the remote operator is likely to be a successful strategy for discouraging the development of the robot's own social agency.

5.2 The Participant's Attention

A review of the participants' interactions with the robot and remote operator revealed that the target of the participant's attention (as indicated by head gaze, speech, and gesture) could be the remote operator, the robot, or split between both. Using the eight interaction event types previously identified, the target of attention was estimated throughout the interaction.

Participant 2 appears to have primarily targeted the robotic platform as a social agent as there was a lack of explicit indicators of a remote operator's presence (recall the telepresence monitor was disabled). However, it is unclear if the use of speech by Participant 2 was directed at the robot platform or at an implicitly assumed remote operator (the participant was not questioned about this during the interview). In the post-interaction interview the participant expressed the belief that the robot was performing movement behaviors on its own, but that because the robot had located him people were aware he was trapped.

Participant 1 demonstrated several shifts of attention between the remote operator to the robot itself, and at times a seemingly split focus between the two. These shifts of attention were performed with fluidity, however these transitions may have been eased by the robot and remote operators common goals. The interview with Participant 1 following the interaction confirmed that he viewed the robot as having its own social role throughout the scenario rather than simply being a means to connecting with responders.

5.3 Implicit Nature of Identity

The social identity of the robot typically is either that of a pure medium (i.e. a portal to a remote social actor) or that of a social actor (i.e. the robot itself recognized as social). While previous studies have examined social identity in the context of explicitly manipulating and framing the robot's perceived social identity, the results from this study offer the first examination of social identity occurring implicitly [3]. This finding is particularly interesting, because observations demonstrate that the robot's actions can change the perceived identity of the robot during the interaction.

One compelling example of fluidity in identity is in observations of Participant 1's transitioning from a Pure Medium to a Social Actor. After communicating with responders through the robot for 26 minutes, the robot used synthetic speech to ask Participant 1 a question. Upon not understanding the question, Participant 1 responded with, *"repeat"* in the crisp tone and style commonly used by humans when the interact with speech recognition systems. This was in contrast to Participant 1's previous responses which had been casual and conversational in nature, like, *"yes, but for what?"*. Participant 1 continued to use this style of response, and engaged in altering synonymic phrases like, *"say again"* and *"repeat"*, until the robot stopped using synthetic speech. The participant's alteration of speech style likely indicates perception of the robot as an entity separate from the remote responders.

In a similar demonstration, 37 minutes into the interaction, while responders were disengaged, the robot began to pan and tilt its head in order to obtain a better viewing angle of Participant 1, which caused the participant to commence re-indicating his simulated injury. This response occurred despite no requests from responders to indicate injuries and despite an earlier 4 minute interaction with responders in which the participant clearly explained his injury. Participant 1 indicated his simulated leg injury by using his hand to gesture towards and massage the simulated injured area.

These two events transpired without any explicit framing of the robot as being a social entity, showing that social identity was actively perceived by the participant from primitive robot actions. This demonstrates that social identity is actively and continuously perceived by humans interacting with a robotic system and establishes the necessity of considering identity perceptions when designing social robots. Additionally, observations show that even with no explicit effort put forth to frame an identity or craft content, humans interacting with the robot resolve identities using any cue or nuance of the robot's behavior that is available to them. This serves as a reminder that humans engage in strong social inference in order to facilitate an interaction, even when that interaction is with a robot. As a consequence robot designers should be vigilant that artifacts of a robot's behavior do not cause an incorrect social inference.

5.4 Attribution of Intent

The two illustrations presented in the previous subsection lead us to posit that the identity perceptions of a robot can be characterized by the actions of the robot and who the participant attributes the action to. Succinctly: *social identity is determined by attribution of intent*. Attribution of intent refers to which entity the person interacting with the robot attributes the robot's actions or intentions to, the robot itself or the human controller of the robot. For example, if a responder informed a victim via two-way communications through the robot: *"I am going to move the robot forward."*, the victim would attribute any movements immediately following to be a manifestation of the intention of the responder, not the robot. Similarly, if a robot encounters a trapped victim and uses synthetic speech and autonomous head gaze acts to communicate: *"The responders will be with you soon."*, the victim attributes the speech and movements as intentions of the robot alone. The former example highlights a Pure Medium, while the latter is an instance of a Social Actor.

Attribution of intent can be suggested by examining auditory and visual behaviors of the robot. For example, if a human voice is channeled through a robot, it is probable that the intent of the voice is attributed to the human speaker, not the robot. In a similar fashion, if the robot is displaying animated eyes on its screen, it is likely that the intention conveyed by eye movements is attributed to the robot itself, not a human. For this study, auditory attributes included: synthetic speech or human speech; while visual attributes included: a video feed of responders, the robot's default screen, and the movements of the robot's head. By accounting for the presence of these attributes and cross referencing the participant's response, the robot's identity was estimated. Examining the stimuli presented by the robot as characterizations of attribution of intent and consequently social identity, enables third party observers to estimate the identity context in which a participant is working.

5.5 Intriguing Participant Reactions

The study revealed two intriguing behaviors related to the social identity of the robot. First, Participant 1 appeared to experience four transitions of social identity, and responded adaptively, despite the lack of any transitional cues. Second, Participant 2, who interacted only with the robot as a Social Actor, exhibited an empathetic mirroring of the robot's head movements.

Findings from prior research examining social identities in a search and rescue situation have indicated that in comparison to a social medium identity, a pure medium identity reduces the level of isolation participants feel and a social actor identity reduces fear in participants [3]. However, studies have been designed such that in the social medium condition the robot behaves as a social actor relaying information from a remote controller. In our study Participant 1 experienced transitions between identities as well as the co-existence of the remote responder and social robot. This is markedly different from previous instantiations of the social medium identity, as it allowed the remote human and social robot to be attended to simultaneously by the participant. This implies a more complex dynamic is possible and future efforts should explore the effects of a social medium which exceeds the role of a social relay.

Participant 2 engaged in the behavior of mirroring the robot's movements when encountering the robot as a social actor. During this interaction, the robot's display screen and audio system simultaneously encountered an unexpected failure. As a consequence, the participant encountered a speechless and blank-screened robot. Responders and operators were not aware of the display failure, but realized the audio failure. An effort was made to communicate using visual elements on the screen, but due to its failure the participant did not respond. Responders continued the triage assessment process, moving the robot's head to obtain different views of the participant, this resulted in the participant tracking the movements of the robot's head with their own head. Upon encountering this behavior, robot operators activated the robot head's *Yes* gesture. Participant 2 mimicked the yes gesture with his own head, this process mimicking occurred three times along with tracking of the robot's head movements. It is notable that even after 30 minutes of no verbal or visual communication with responders, the participant still smiled and tracked the robot's head. While this result primarily reflects on the flexibility and curiosity of humans interacting with novel technology, it also shows that the participant's responses were shaped by the robot's functionality. In this case the robot did not present capabilities outside of a Social Actor which could use head gestures, and the participant matched these capabilities in an attempt to convey responsiveness. This result suggests that constrained abilities or identities are not necessarily problematic, as humans adapt to the robot's abilities. Interestingly, the participant only inadvertently mentioned that the robot's display was turned off during the interaction at the end of their interview, not realizing this was unintended.

6 Implications

Observations from the field study also provide a discussion of insights and implications for human-robot interaction. This section describes the foundations of a coding scheme for robot social identity, specific insights for medical and assistive robotics, and issue of robot autonomy and competition for attention between the remote operator and the robot. In addition, the study's limitations and future research questions are described.

6.1 Towards a Classification Scheme

The study used eight factors from observational data to estimate the most likely recipient of the participant's attention. By examining in parallel four factors of the robot's actions with four factors of the participant's actions, the target of the participant's attention can be estimated. The stimulus-response nature of the robot's actions and the participants actions leads to the possibility of being able to classify which robot behaviors will direct attention away or towards the robot.

This scheme is important for robotic system designers as it enables characterization of how the robot is perceived. While the current scheme only accounts for a small number of auditory and visual cues, expansion to include other cues is likely to occur as robots with varying abilities are examined in terms of social identity. It is also necessary that the responses from the human towards the robot be considered in constructing a more complete model of social identity.

In order to form a useable classification scheme, lab studies which manipulate and carefully control different parameters of the robot, while receiving responses from the participant, will be necessary. After lab studies have established these parameters, verification of their influence and occurrence in the field can be performed to validate the scheme. An established classification scheme for social identity is important to the future of social and assistive robotics, as it enables the robotic system and remotely involved humans to understand the attentional focus of the person interacting with the system. A well defined scheme will clarify the design process of these robots and assist in shaping communication into most useful form for the application.

6.2 Insights for Medical & Assistive Robotics

While this work focuses on in situ medical care, the results can be applied to many other scenarios in the domain of medical and assistive robotics. The findings presented on social identity are likely to be observed in any natural interaction with a robot that serves as both a telepresence device and an assistive or social device simultaneously. These findings are useful in ensuring that the correct social identity of the robot is conveyed to patients using such robots as part of their care. A key expansion which may be necessary is the inclusion of bolder cues and robot state notifications, as some patients may be unable to devote significant cognitive resources towards interacting with the robot. In contrast to this study's participants, who were only playing the role of a victim, actual patients and victims encountered may be suffering from cognitive confusion and therefore may require more salient cues from the robot in order to correctly interpret its identity.

The results found in this study also provide insight into dealing with communications problems or system failures that might occur in assistive robots. For example, a robot which is channeling the identity of a remote physician may experience a network problem and loose connection to the physician. Rather than displaying a static failure message, the robot could switch identities into a Social Actor mode and engage with the patient until communications are restored with the physician. In a case like this, the robot's ability to attain multiple identities is beneficial to the user experience and works to prevent frustration.

6.3 Autonomy & Competition for Attention

Observations demonstrated that participants readily shifted their attention between the physical robot and the remote operator. Though the scenario examined did not create situations where the remote operator and robot were competing for attention or providing conflicting requests, there is the potential for such a competition to exist. If opposing demands for attention were to occur, it is unknown whether the remote operator could maintain their presence and authority in the situation. Since the robot is physically situated with the person it is interacting with, its attention demanding cues may be more powerful than those of the remote operator. In the case of this field study, by exhibiting small head movements and synthetic speech the robot was able to easily capture the attention of the participant. It is possible that a more socially engaging robot would merit even more attention from the participant.

This potential for divided attention can be remedied by providing the robot with behaviors that yield to the remote operator. For example, the idea of the robot tilting it's head back to indicate listening to the remote operator is one potential way of directing attention. Likewise, simple adjustments to the robot's social behavior, which take into account the remote operator's need for attention can be configured to help ensure the participant is not distracted by the social nature of the robot.

This aspect of the interaction also requires a careful balance between autonomous robot functions and those controlled by a remote operator. While traditional telepresence robots often have autonomous functions aimed at performing simple tasks, like avoiding collisions, a social telepresence robot has the potential to execute more complex autonomy. For example, in a disaster scenario trapped victims may benefit from the robot having higher level social autonomy in the event when the communication with responders fails. This also serves the practical function of limiting the amount of network communication that must occur between the robot and responders. Conversely, if the robot is unable to successfully manage a higher level social function (such as comforting a trapped victim) it may be inappropriate for the robot to assume a strong social actor role. Further studies are required to better understand how people shift their attention based on the robot and remote operator's actions and which function should be performable without the remote human.

6.4 Limitations and Future Questions

The scope of the findings is limited by: the domain being constrained to US&R, the participant's mental state, the robot being appearance constrained, and the small number of participants.

It is likely that the US&R environment assisted in creating focused interaction between the robot and the participant, which may not occur in other environments that contain distractions. Future studies should explore the occurrence of social identity in other scenarios. Furthermore, the chosen domain likely contributes some amount of stress and cognitive fatigue toward the participant, but this may not be as extreme as the cognitive confusion found in other users of assistive robots. This raises the possibility of modulating the intensity of transition cues and state cues as a function of the participant's cognitive state. This is also relevant to scenarios where users may interact with the robot, become distracted, and then re-engage with the robot. It is likely that providing clear robot context indicators will become more important in these situations.

The robot was an appearance constrained robot and had a form factor which enabled it to attach to a traditional search and rescue robot. This experience may be different from interacting with a robot that for example can be held in the participant's lap or from interacting with a robot with more degrees of freedom and affective expression abilities. Future work should clarify if an increase in affective capacity increases or decreases the ability to manipulate a robot's social identity. Additionally, other modalities like touch or tactile response may need to be examined as part of the model which defines social identity.

Most importantly, this study was exploratory in nature and only involved two participants interacting with the robot. Though a great deal of information regarding an unguided interaction was gained, it is anecdotal in nature and rigorous controlled studies which allow for reliable inference are necessary to further the development of social identity within robots. This will require both lab and field studies to be conducted in order to extract the relevant characteristics of each identity and then verify their occurrence in a natural interaction space.

7 Summary

This article has presented observations and interpretations gathered from a field exercise in a simulated in situ medical care scenario. The observations focus on specific events within the interaction and how the participants directed their attention between the robot and the remote operators. The findings indicated that a social telepresence robot with sufficient social cues may obtain its own social identity, separate from that of the remote operator. While participants demonstrated a fluidity in shifting their attention from the robot to the remote operator, the potential for divided attention and a loss of presence for the remote operator is a possible outcome of interactions where the robot and operator are competing for attention. The participants' ability to fluidly shift between the robot and the remote operator suggests that an identity spectrum may exist, in which the robotic platform exist in a role between a pure medium and social actor. This potential third identity represents the unique circumstances in which attention shifts between the social robot and the remote operator. Prior work has only observed social identities through explicit framing and content manipulation, while this study provided examples of implicit existence of social identity without framing. This work posited that ascriptions of social agency are the result of specific social cues from the robot, which are not attributed to the remote operator, but to the robot itself. The results presented in this study are relevant to assistive and social robots, as they illustrate the different perceptions humans may have of a robot as a result of the robot's behavior and presentation. Future work with an emphasis on building a more complete characterization of robot social identities through controlled studies which allow reliable inference and generalization is needed.

Acknowledgement: This work was sponsored in part by NSF Grants IIS-0905485 and Microsoft Research. The authors would like to thank Cliff Nass and David Sirkin for their helpful comments and TEEX for their assistance at Disaster City ®.

References

[1] Cindy L. Bethel and Robin R. Murphy. Non-facial/non-verbal methods of affective expression as applied to robot-assisted victim assessment. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 287–294, New York, NY, USA, 2007. ACM.

[2] Jennifer L Burke, Robin R Murphy, Michael D Coovert, and Dawn L Riddle. Moonlight in miami: Field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise. *Journal of Human–Computer Interaction*, 19(1-2):85–116, 2004.

[3] V. Groom, V. Srinivasan, C.L. Bethel, R. Murphy, L. Dole, and C. Nass. Responses to robot social roles and social role framing. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, pages 194–203, 2011.

[4] Annica Kristoffersson, Silvia Coradeschi, and Amy Loutfi. A review of mobile robotic telepresence. *Advances in Human-Computer Interaction*, page 3, 2013.

[5] Maja J Mataric. The role of embodiment in assistive interactive robotics for the elderly. In *AAAI fall symposium on caring machines: AI for the elderly, Arlington, VA*, 2005.

[6] Robin R Murphy and Jennifer L Burke. Up from the rubble: Lessons learned about hri from search and rescue. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 49, pages 437–441. SAGE Publications, 2005.

[7] Robin Roberson Murphy. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(2):138–153, 2004.

[8] B. Reeves and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press New York, NY, USA, 1996.

[9] Laurel D Riek. Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1), 2012.

[10] Scott M Sasser, Richard C Hunt, Ernest E Sullivent, Marlena M Wald, Jane Mitchko, Gregory J Jurkovich, Mark C Henry, Jeffrey P Salomone, Stewart C Wang, Robert L Galli, et al. Guidelines for field triage of injured patients. recommendations of the national expert panel on field triage. *Recommendations and reports/Centers for Disease Control: Morbidity and mortality weekly report.*, 58(RR-1):1–35, 2009.

[11] I. Straub, S. Nishio, and H. Ishiguro. Incorporated identity in interaction with a teleoperated android robot: A case study. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pages 119–124, 2010.

[12] Fumihide Tanaka, Toshimitsu Takahashi, Shizuko Matsuzoe, Nao Tazawa, and Masahiko Morita. Telepresence robot helps children in communicating with teachers who speak a different language. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 399–406. ACM, 2014.

[13] Tzung-Cheng Tsai, Yeh-Liang Hsu, An-I Ma, Trevor King, and Chang-Huei Wu. Developing a telepresence robot for interpersonal communication with the elderly in a home environment. *Jounral of Telemedicine and e-Health*, 13(4):407–424, 2007.
[14] Katherine M Tsui, Munjal Desai, Holly A Yanco, and Chris Uhlik. Exploring use cases for telepresence robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 11–18. IEEE, 2011.
[15] Antoine A. J. van de Ven, Anne-mie A. G. Sponselee, and Ben A. M. Schouten. Robo m.d.: a home care robot for monitoring and detection of critical situations. In *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, ECCE '10, pages 375–376, New York, NY, USA, 2010. ACM.