Social Engagement in Public Places: A Tale of One Robot

Lilia Moshkina
NRC Post-Doctoral Fellow
Naval Research Laboratory
Washington, DC 20375 USA
lilia.tsvetik@gmail.com

Susan Trickett
ITT Exelis
Boulder, CO 80301 USA
+1-703-577-6035
sbtrickett@gmail.com

J. Gregory Trafton
Naval Research Laboratory
Washington, DC 20375 USA
+1-202-767-3479
greg.trafton@nrl.navy.mil

ABSTRACT
In this paper, we describe a large-scale (over 4000 participants) observational field study at a public venue, designed to explore how social a robot needs to be for people to engage with it. In this study we examined a prediction of Computers Are Social Actors (CASA) framework: the more machines present human-like characteristics in a consistent manner, the more likely they are to invoke a social response. Our humanoid robot's behavior varied in the amount of social cues, from no active social cues to increasing levels of social cues during story-telling to human-like game-playing interaction. We found several strong aspects of support for CASA: the robot that provides even minimal social cues (speech) is more engaging than a robot that does nothing, and the more human-like the robot behaved during story-telling, the more social engagement was observed. However, contrary to the prediction, the robot's game-playing did not elicit more engagement than other, less social behaviors.

Categories and Subject Descriptors
K.4.0 [**Computing Milieux**]: Computers and Society – *general.*

General Terms
Experimentation, Human Factors.

Keywords
Field experiment, CASA, engagement, human-robot interaction.

1. INTRODUCTION
Interest in social robot behavior has exploded in recent years, in part due to the Computers Are Social Actors (CASA) paradigm, advanced in Media Equation [1]. The CASA framework has been widely accepted in HCI and supported in a large set of experiments [2]. It claims that people will respond to a computer as a social partner, provided appropriate social cues are produced by the computer. This powerful concept extends to robotics as well, and it has been shown that human-like, social non-verbal behaviors can be advantageous in a robot. Breazeal et al. [3] found that implicit use of non-verbal behaviors, such as gaze shifts, shoulder shrugs, facial expressions, etc., was instrumental in human-robot team work. Sidner et al. [4] showed that a penguin robot that performed a few gestures was more engaging than its motionless version. Moshkina [5] reported that non-verbal expressions of anxiety and fear on a small humanoid robot Nao resulted in subjects' greater compliance with the robot's request to evacuate. Similarly, Chidambaram, Chiang, & Mutlu [6] discovered that presence of nonverbal bodily cues, such as gestures and gaze, increased a robot's persuasiveness. It has also been shown that robot form has an impact on human perception of robots: Groom et al.'s [7] self-extension experiment suggests that people are more likely to perceive a humanoid robot as a separate entity rather than an extension of themselves, as compared to their treatment of a robot car.

The CASA framework predicts that people will respond to robots in much the same way as people respond to other people as long as the robot presents human-like social cues (for a review, see Fong, Nourbakhsh, & Dautenhahn [8]). The CASA framework also predicts that as a system's social cues increase in number or fidelity in a technologically-consistent manner [9], people should find the system more socially appealing. To explore this issue, we focus on social engagement towards an autonomous robot.

Social engagement is a core social activity that refers to an individual's behavior within a social group [10]. In this study, we are interested in short term social engagement of individuals, and specifically what aspects of a robot's behavior will increase people's engagement. Because most previous studies of the CASA framework have been performed in the laboratory and we are interested in how to elicit social behavior from groups of people, we ran a large-scale observational study with over four thousand of participants in which we increased the social behavior of our robot. We then examined how different levels of the robot's social behavior impacted short term social engagement (listening to a robot tell a story).

2. STUDY DESIGN AND PROCEDURE
To examine the extent to which the presence of social cues in an anthropomorphic robotic platform influences social response in humans, an observational field study involving over 4000 of participants was conducted. In this study, attendees of a large public event had an opportunity to stop and listen to a humanoid robot recite a short story, as they were passing from one exhibit to another. For this study, we kept the verbal behavior the same, but changed non-verbal behavior across conditions, because previous studies found a strong impact of robot gestures on human social behavior [4, 6, 11], and paralinguistic cues had a much smaller or negligible impact on social behavior [6].

Our observational study followed a 2x4 between-subject design, where the first independent variable was Story Type (Informative vs. Humorous) and the second the level of social cues (Social Cues) produced by the robot, ranging from none (no movement) to full body movement. The level of Social Cues was increased incrementally between conditions, where the next level subsumed all the previous levels and added a new layer. The four conditions

Copyright © 2014 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the national government of United States. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

HRI'14, March 3–6, 2014, Bielefeld, Germany.
Copyright © 2014 ACM 978-1-4503-2658-2/14/03…$15.00.
http://dx.doi.org/10.1145/2559636.2559678

of Social Cues levels were as follows, in the order of increasing presence of social cues:

- Voice only – the robot produced no movement, just the narrative (Voice hereafter).
- Voice + lips – the robot's lips were making movements in sync with speech (Lips hereafter). Lip movements present a minimum level of motion expected from a talking creature.
- Voice + lips + facial expressions – in addition to lip-sync, the robot produced story-appropriate facial expressions (Face hereafter).
- Voice + lips + facial expressions + gestures – the previous condition was augmented with a variety of story-appropriate gestures and posture changes (Gestures hereafter).

The two Story Types were: Informative, in the form of a short, engaging lecture about the robot's capabilities, and Humorous, in the form of a short joke ending in a punch line. These two styles naturally elicit different social response in humans: nods during lectures [12], and smiles and laughter after jokes. Because of the differences in expected social behavior and content, they were placed in different conditions.

Two stand-alone conditions were also run. The first was a baseline condition where the robot was "in between acts" and was doing nothing – no movement, no talking, not being moved, etc.

In addition to our study, a robotics perception experiment was conducted using the same robot during the same public event [13]. In this experiment, volunteers interacted with the robot during a stylized game, in which the robot asked volunteers questions in an attempt to identify them. As game-playing is a very social activity, this experiment presented an additional comparison point to the original study by adding an interaction element. The robot's behavior here can be construed as exhibiting the most human-like, social behavior (as compared to robot story-telling), and was treated as the second stand-alone condition.

2.1 Stimuli and Robotic Implementation

Two vignettes of each Story Type were devised. All the stories were told in the first person singular; the robot talked about itself or a firefighting project it was involved in (informative stories), or told one of two inoffensive jokes: one about Sherlock Holmes and Dr Watson (relevant due to a recent movie), and another set in a zoo, which would appeal to any audience. All the stories were similar in length, within 10% both in the number of words and duration, and lasted, on average, 66 seconds. The last sentence in all the stories was always the same: "Thank you for listening to me!", and was preceded by a 3-second pause to signify the end of the narration. There were 16 variations total, with two vignettes per each condition. Appendix A contains the text of each vignette.

The Xitome Mobile-Dexterous-Social (MDS) humanoid robot platform was programmed to deliver the stories. The robot (referred to as "Octavia" in this study) was designed for human-robot interaction as an upper-body humanoid on a two-wheel Segway base, sized similarly to a larger adult. Octavia has a total of 41 DoFs (Degrees of Freedom), allowing a wide variety of human-like gestures and facial expressions. Figure 1 provides a close-up look at Octavia's head and a hand. The robotic implementation is described in the next subsections.

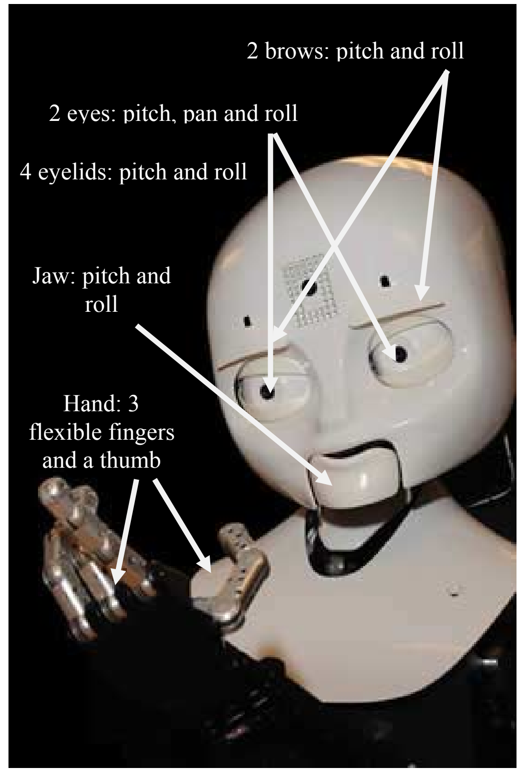

Figure 1: A close-up of Octavia's head and a hand.

2.1.1 Voice

Cepstral voice Allison was used to produce text-to-speech translation of the stories. Any mispronounced words by the TTS engine were manually adjusted to sound correct. The resulting speech was intelligible, though clearly computer-generated. The speech was transmitted through two speakers positioned on both sides of the robot. Figure 2 displays Octavia in a neutral position, which remained unchanged in the Voice condition: head up; eyes open and alert; arms and hands in a non-threatening position in front of the torso; body and head facing the audience.

Figure 2: Octavia in neutral position in Voice condition: head up, eyes open and alert, arms and hands in a non-threatening position, body and head facing the audience.

2.1.2 Lips

Octavia has a movable jaw, allowing its mouth to open or close. Octavia's jaw has two DoFs: pitch (up or down) and partial roll (where corners of the mouth can be made to appear at different heights, as in expressing a smirk). Pitch was used to vary how wide the robot's mouth was open, to correspond with more or less open sounds; the mouth was closed during pauses or between words. The jaw movements were manually synchronized for each vignette. Figure 3 (left) shows Octavia's mouth closed, and Figure 3 (right) open, with all other features in neutral position and unchanged throughout the story.

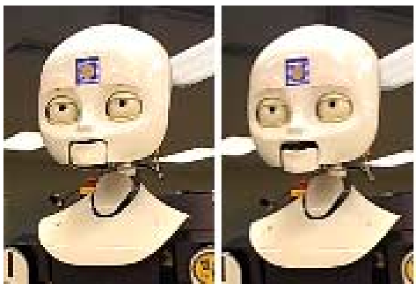

Figure 3: (Left) Mouth closed, as during a pause. (Right) Mouth half-open, pronouncing a phoneme (e.g. "ai").

2.1.3 Face

Octavia's face has nine movable parts which can be used to produce primitive facial expressions: two eyes, two eyebrows, two upper and two lower eyelids, and a jaw. Each eye has 3 DoFs: pitch for up or down, pan for side-to-side movement, and roll. Gaze shifts were used continuously throughout the story-telling to maintain the illusion of connecting with the audience; additionally, eye pitch and pan were used in producing gaze gestures accompanying the words indicating direction (e.g., up, high, there, etc.). Each eyebrow had two DoFs, pitch for up or down, and roll; both were used in production of facial expressions. The combination of upper and lower eyelids was instrumental in producing various degrees of eye closure, from fully closes as if sleeping, to fully open as in surprise; as well as periodic closing and opening of the eyelids was used to emulate blinking. Please note that head and neck positions were not changed in this condition, only facial features were varied. Figure 4 provides a few examples of Octavia's facial expressions used during the experiment: on the left, the eyes are shifted to imitate following individual attendees, in the middle, Octavia expresses surprise/anticipation, and the snapshot on the right shows skepticism.

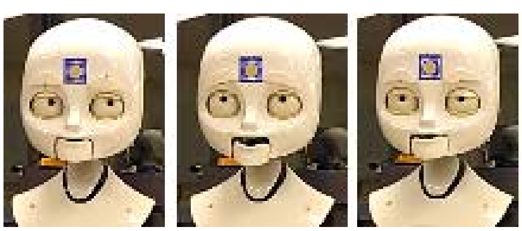

Figure 4: (Left) Eye shift to the right to connect with the audience; (Middle) Expression of surprise/anticipation; (Right) Expression of skepticism.

2.1.4 Gesture

The dexterous upper torso (two arms with shoulders, elbows, and hands with 3 fingers and a thumb) and head (head pitch, pan, roll, as well as neck pitch) allowed us to compose a variety of iconic, metaphoric and deictic gestures. For example, a pointing gesture would involve an extended arm, a turn of the head, and a turn of the torso; and a thumbs-up gesture had the fingers in a fist, a thumb up, and an arm half-extended in front of the torso. Each gesture was designed to correspond to the particular story being recited; in addition, head movements were used to accompany gaze shifts or arm gestures. Figure 5 displays examples of iconic (left), metaphoric (middle), and deictic (right) gestures performed by Octavia.

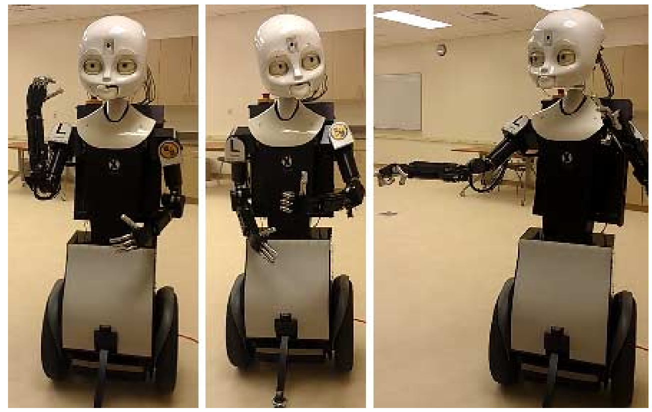

Figure 5: (Left) Octavia's arm gesture accompanies the phrase "when the fence was 40 feet high"; (Middle) Octavia shows a "thumbs up" gesture as she says "Mission accomplished!"; (Right) Octavia points to a hypothetical fire location.

2.1.5 Interactive Game

During the interactive game, 3 volunteers from the audience played a stylized version of a "shell" game with Octavia. In this game (with the goal of person identification) the robot asked each volunteer a question by which it could name them later (e.g., "What is your favorite ice-cream?"), then requested the participants to exchange places while it kept its eyes closed, and finally identified each of the participants by their answer-name (e.g., "You are vanilla, right?"), to the audience's general delight and cheering. The robot in this experiment was completely autonomous; it spoke with the participants and tracked them with its head, gaze and torso as it was addressing them.

2.1.6 Baseline control

In this control condition, the robot was positioned behind the stanchions, but performed no actions: neither speech nor movement. The experimenters were present at the exhibit, but were not manipulating the robot.

2.2 Hypotheses and Measures

The Computers Are Social Actors framework makes several clear predictions in this study. First, as a robot's behavior becomes more and more human-like, the more people should respond socially to it; in this case, we define social engagement as observing the robot for a minimum amount of time (described below). According to CASA, more people should observe or engage the robot as the social cues of the robot increase. Specifically, Idle < Voice < Lips < Face < Gesture < Game. The game should be the most engaging because the robot not only moved and spoke, but it was actively engaging with people and solving problems – very human-like behavior.

A weaker hypothesis concerns the two types of stories the robot told: jokes should be more engaging than informational talks, though people may perceive that robots should be informative over having a sense of humor.

Our primary measure in this study concerned the number of people that observed the robot for a specific length of time (15 seconds, or about ¼ of the length of the story) or that stayed for the entire story. Because the number of people in a public group can also impact whether new people will join the group (e.g., a small group may not look very interesting, but a large group may be too crowded), we also tracked the flow of people who walked by, but did not stay or observe the robot. These types of retention measures have been used successfully by other researchers [14].

2.3 Setting and Procedure

The study was conducted during Fleet Week 2012 in New York City, an annual event organized by the U.S. Navy to showcase its latest technology and allow civilians to tour their ships. A large number of exhibits were setup under a pier in Manhattan, where the general public could walk through and take a look at anything of interest. The exhibition area was located between the entrances to two modern US Navy ships, an assault ship USS Wasp which could be visited prior to entering the exhibits, and a destroyer USS Roosevelt which could be visited right after exiting the exhibits.

The attendees varied greatly in their age (both children and elderly were present), ethnic and language background, occupation (military vs. civilian) and gender from session to session. In addition to being diverse in their composition, the attendees also varied greatly in their current agenda: they could be leisurely strolling by or hurrying to climb onto USS Roosevelt or to the exit; interested in exhibits or just waiting for their companions; having a meal/snack, talking between themselves, or pointing out the robot to each other as it caught their interest.

Our exhibit occupied an approximately 20x15' area, the last one before the exit from the exhibit area, on the way to either USS Roosevelt or the exit from the entire Fleet Week area. Thus, Octavia had to vie for attention not only from other exhibits, many of which were interactive, but also with a tour of an impressive modern destroyer, currently in commission in the US Navy.

Within our exhibit, the robot was cordoned off from the public by stanchions, though it was fully visible. The area in front of the robot and to the left of it (as viewed on camera) was videotaped during the storytelling sessions, as well as a few seconds before and after; the view to the right, as the attendees were leaving the exhibit towards USS Roosevelt and the exit, was limited. Figure 6 provides a snapshot of a recorded session. Each session was started wirelessly; once started, each session ran autonomously.

The traffic (flow of people) through the exhibition area varied greatly throughout each day, from under 10 people passing in front of the robot over a 60-second period to over 70. Whenever there were at least 2 people present and the robot was not engaged in other activities, a session was initiated based on a pre-randomized order; there were at least 2 minutes between the end of one and the beginning of another session, to reduce the number of repeat participants.

Figure 6: View from the camera of attendees passing by the exhibit during a story session; the robot is positioned just out of the camera view, beyond the stanchions; another exhibit is located directly to the left, and the way to the exit is to the right.

As people were passing by the robot, they could choose to: ignore the robot completely (no gaze towards the robot), attend briefly (look at the robot for a few seconds), stay for a portion of a story, or stay for the entire story. Figure 7 shows the participants attending to robot, as opposed to simply passing by (Figure 6). The study was not announced as such, and no incentives were given for participation.[1] A total of 149 sessions were conducted and videotaped during a 6-day period.

Figure 7: Attendees watching Octavia present a story.

The interactive game sessions were videotaped as well from the same camera position. In order to make the interactive game clips comparable to the story sessions, the first 66 seconds (average story duration) of 23 game clips were extracted, as the games were longer in duration.

Finally, as the video recording was going on continuously for a large portion of the exhibit duration, it was possible to extract a number of clips during which the robot remained idle.

2.4 Video Coding

The measures that were used in the analysis were related to engagement: full engagement and attending to the robot for at least 15 seconds (partial engagement). The 15 second interval was deemed to sufficiently reflect observers' interest in the robot's performance. It took approximately 10 seconds to slowly traverse the length of our exhibit; therefore, a 10-second interval would be too small to judge engagement, and there were 2-4 gestures and

[1] IRB approval for this study was received by NRL.

facial expressions produced by the robot in each 15 second period to sufficiently differentiate between the conditions.

From the story sessions, a total of 93 clips (23 for each Social Cue condition, except for Facial Expressions, which had 24 sessions) have been video coded. In particular, the following measures were extracted:

- Full engagement: the number of people who were present and attending to the robot (looking at or actively listening to) for the entire story; this reflects engagement with the robot's storytelling;
- Partial engagement: the number of people attending for 15 seconds or more during the storytelling. People who were present, but not attending to the robot (e.g., engaged in a conversation) were not counted;
- Traffic: the total number of people who passed by the exhibit during the storytelling; included those present at the beginning of the story, and those who entered the exhibit area during the story.

In these 93 clips, 3314 was the total number of people who passed by the exhibit, and of those, 2165 passers-by at least looked at the robot. 15% of these clips were double-coded, and the inter-rater reliability, as expressed by Pearson's R was as follows: Traffic at 0.97, Partial engagement at 0.97, and Full engagement at 0.89.

To compare the story-telling with the interactive shell game, 23 excerpts from interactive games were coded in the same manner. Finally, we also coded 7 excerpts of 66 seconds each, where the robot was completely idle: not performing any task either autonomously or with the help of the experimenters. The combined total was 123 clips, in which 4222 observers passed by the robot.

3. ANALYSIS AND DISCUSSION

We present our results in two primary sections: analysis of the story-type x social cues data, and then compare those results with the two additional conditions. Recall that CASA predicts that as social cues increase, there should be more engagement with the robot. Additionally, CASA predicts that the idle robot with very little social activity beyond its anthropomorphism should be the least engaging and the interactive game with should be the most engaging.

3.1 Story-Telling

On average, 36 people passed by Octavia during a single vignette. Traffic did not differ across either story type, $F(1, 84) < 1$, MSE = 0.74, n.s., or social cues, $F(3, 84) < 1$, MSE = 135.14, n.s., nor was there an interaction, $F(3, 84) < 1$, MSE = 92.54, n.s..

These results show that one condition or another was not systematically run during a higher concentration of people. However, because traffic did differ greatly across sessions, we used traffic as a covariate in all future analyses.

3.1.1 Full Engagement

As suggested by Figure 8, the type of story did not have an impact on the number of people who stayed for the entire story, $F(1, 76) < 1$, MSE = 0.19, n.s., nor did it interact with social cues, $F(3, 76) < 1$, MSE = 2.47, n.s., or Traffic $F(3, 76) < 1$, MSE = 6.05, n.s..

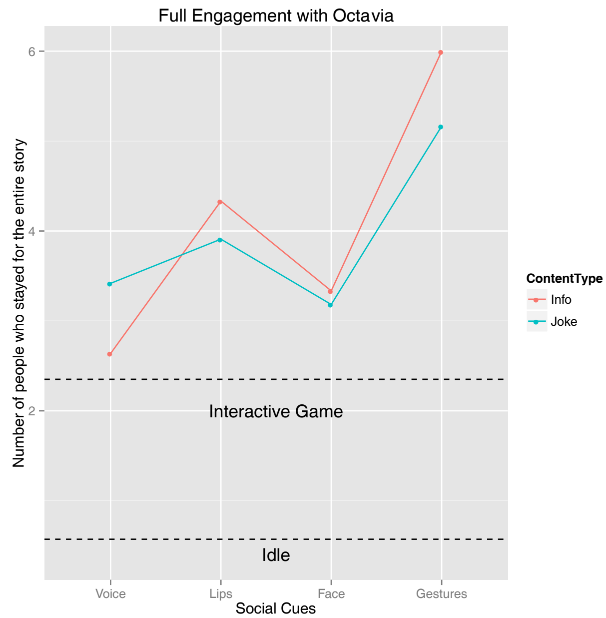

Figure 8: Full Engagement. The number of people who listened to the entire story increases with higher levels of Social Cues; no difference between Story Types is observed.

Not surprisingly, there was an effect of traffic: as traffic increased, more people stayed to watch the robot, $F(1, 76) = 29.40$, MSE = 198.82, $p < 0.05$, partial eta squared = 0.28. Traffic did not interact with the level of social cues, $F(1, 76) < 1$, MSE = 6.05, n.s.. Our explanation for traffic is straightforward: as more people walked by, more of them were likely to stay to watch the robot.

As predicted by CASA, the level of social cues did have an overall impact on the number of people who stayed for the entire story, omnibus $F(3, 76) = 4.04$, MSE = 27.34, $p < 0.05$, partial eta squared = 0.1. Because CASA predicts a specific pattern (an increasing trend), we used a contrast. As predicted by CASA, the contrast showed that as the level of social cues increased, the number of people also increased, $p < 0.05$.

3.1.2 Partial Engagement

Full engagement showed support for CASA. However, the full engagement measure required people who happened to be near the robot at the beginning of the story to stay for the entire story. It could be that a better or stronger measure of engagement would be to look at partial engagement (staying for at least 15 s). This partial engagement measure may show a stronger trend than the full engagement measure.

As suggested by Figure 9 (Partial Engagement) and similar to full engagement, the type of story did not have an impact on the number of people who stayed for at least 15s, $F(1, 76) < 1$, MSE = 5.8, n.s., nor did it interact with social cues, $F(3, 76) < 1$, MSE = 8.2, n.s., or traffic $F(3, 76) < 1$, MSE = 2.4, n.s..

Also similar to the full engagement analysis, there was an effect of traffic: as traffic increased, more people stayed to watch the robot, $F(1, 76) = 65.72$, MSE = 645.7.82, $p < 0.05$, partial eta squared = .46. Traffic did not interact with the level of social cues, $F(1, 76) = 1.32$, MSE = 12.9, $p > 0.05$. We interpret this finding in a straightforward manner: as more people walked by, more of them were likely to stay to watch the robot.

Similar to the full engagement analysis and as predicted by CASA, the level of social cues did have a strong impact on the number of people who watched for at least 15s, omnibus $F(3, 76) = 8.87$, MSE = 87.1, $p < 0.05$, partial eta squared = .18. We also performed a trend analysis for partial engagement and again found that as the level of social cues increased, the number of people who watched the robot also increased, $p < 0.05$.

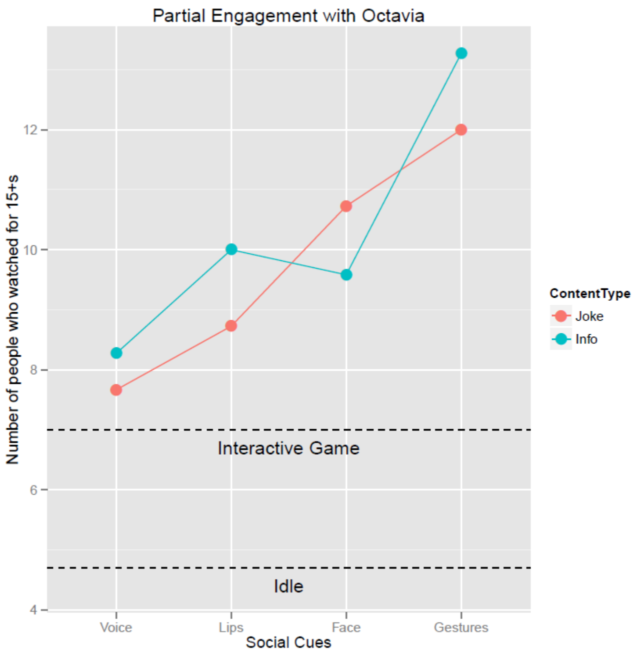

Figure 9: Partial Engagement. The number of people who attended to Octavia's story-telling for 15+ seconds increases with higher levels of Social Cues; no difference between Story Types is observed.

3.1.3 Idle Control

As described above, we also coded video excerpts when the robot was completely idle and there was no one interacting with the robot. The robot had almost no social appeal at all. In fact, the only interesting aspect about it was that it was a robot. CASA predicts that, because this robot was behaving in the least social manner, people should be the least engaged toward it compared to other conditions.

As both Figures 8 and 9 suggest, the idle robot was far less engaging than any of the other more interactive conditions. To test this statistically, we performed two different tests. We first compared the idle condition to all the story-telling conditions, collapsed across content type (since the idle condition obviously did not have that factor). We found that for the full engagement measure, fewer people watched the robot for a full minute when it was idle than when it was telling a joke or being informative, $F(1, 95) = 20.1$, MSE = 137.2, $p < 0.05$, partial eta squared = 0.16. Similarly, there was a strong effect for partial engagement: the idle robot was less engaging than when it was telling a story, $F(1, 95) = 15.9$, MSE =183.2, partial eta squared = 0.12. As in previous analyses, traffic did not differ between conditions ($p > 0.05$) and it did have a positive effect on the number of people who engaged with the robot, $F(1, 95) = 34.9$, MSE = 238.3, $p < 0.05$, partial eta squared = 0.27 for full engagement and $F(1, 95) = 68.0$, MSE = 783.7, $p < 0.05$, partial eta squared = 0.41 for partial engagement.

We also performed a more conservative test where we compared the idle condition to the least interactive story-telling condition, the Voice condition. We found that for the full engagement measure, fewer people watched the robot for a full minute when it was idle than when it was telling a story with its voice only, $F(1, 26) = 26.7$, MSE = 71.34, $p < 0.05$, partial eta squared = 0.52. Similarly, there was a strong effect for partial engagement: the idle robot was less engaging than when it was telling a story with its voice alone, $F(1, 26) = 7.0$, MSE =56.4, partial eta squared = 0.22. As in previous analyses, traffic did not differ between conditions ($p > 0.05$) and traffic did have a positive effect on the number of people who engaged with the robot, $F(1, 26) = 11.2$, MSE = 30.0, $p < 0.05$, partial eta squared = 0.30 for full engagement and $F(1, 26) = 15.1$, MSE = 121.08, $p < 0.05$, partial eta squared = 0.37 for partial engagement.

As predicted by CASA, the idle robot was less engaging socially than a story-telling robot.

3.2 Interactive Game

While the idle condition was predicted by CASA to be the least engaging, the interactive game was predicted to be the most engaging. In this condition, the robot spoke to people playing a game, telling people to move around, and of all the conditions discussed so far, was the most human-like. CASA predicts that, because it is the most human-like, people should be engaged with it more than any other condition.

However, as Figures 8 and 9 suggest, the interactive game was actually **less** engaging than when the robot told stories. The statistical tests in this case will be limited because the direction of the prediction is opposite to what the data shows. We found that for the full engagement measure, people were less engaged with the interactive game robot than a robot telling a story, $F(1, 111) = 9.3$, MSE = 72.8, $p < 0.05$, partial eta square = 0.03. Similarly, for partial engagement people were less likely to watch the interactive game than the robot telling a story, $F(1, 111) = 11.9$, MSE = 163.2, $p < 0.05$, partial eta squared = 0.03.

Please note again that these results are in the opposite direction to that predicted by CASA.

3.3 Engaged vs. Unengaged

While there was general support for CASA as the robot told stories, there is another measure that should be looked at: unengaged people. As Figures 8 and 9 suggest, and earlier analyses confirm, the number of people who engaged with the robot showed a reliable and robust increasing trend across levels of social cues. However, it is also possible to look at the number of people who were not engaged, or who left without paying attention to the robot. This analysis used the full engagement analysis: people who stayed for the entire story or people who left mid-way through. We collapsed across story type because there was no statistical difference between joke or information in any of our analyses.

As Figure 10 suggests, twice as many people were unengaged and left ($M = 8.8$) than were engaged and stayed ($M = 4.0$), $F(1, 176) = 61.8$, MSE = 1037.9, $p < 0.05$, partial eta squared = 0.26. A similar finding occurs for the partially engaged measure.

It is surprising that more people decided to leave than stay, and somewhat counter to the general CASA framework.

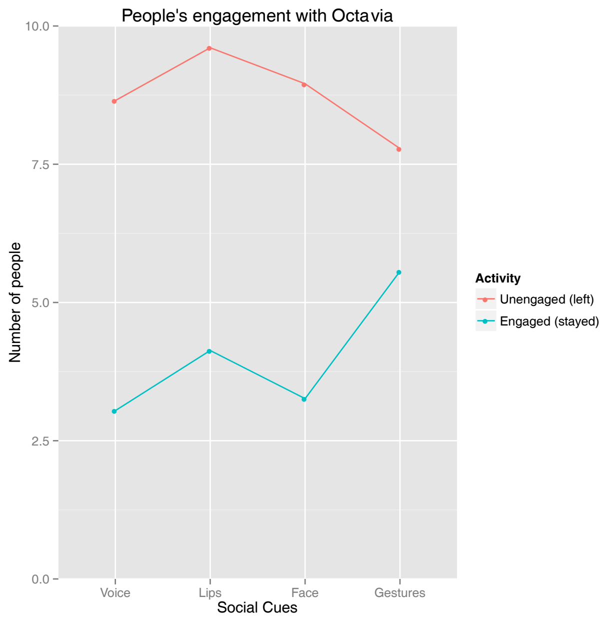

Figure 10: Number of the attendees who stayed for the entire story (fully engaged) vs. those who left (unengaged).

4. GENERAL DISCUSSION

In this paper, we explored one of the fundamental questions of human robot interaction: how to encourage people to engage with a robot. We addressed this question in a very theoretical manner. We used the Computers Are Social Actors (CASA) framework to create robotic interactions that varied in their social cues, from absolutely no active social cues (our idle condition) to increasing levels of social cues (Voice, Lips, Face, Gesture) to full human-like game-playing interaction (interactive game).

We collected data in a naturalistic setting and had the opportunity to examine how "normal" people would engage with a novel robotic platform. Because data collection occurred in a public venue, we were able to collect data on over 4000 individuals as they made a simple choice: should they stay and engage with our robot.

We found several strong aspects of support for CASA. First, we found that a robot that provides even minimal social cues (e.g., talking) is more engaging than a robot that does nothing. While seemingly obvious, it should be noted that the robot that was used in this study was an actual physical robot, and most people in the US have not seen or interacted with a robot before, so the novelty was quite high.

Second, we found that as the robot's social cues increased, people's engagement also increased. Specifically, we found that, while the robot told a story, people were more engaged and interested in the robot if it acted more human-like – if it gestured while making faces and moved its lips as it spoke. People became progressively less engaged with the robot as each of those features was removed.

However, we also found some reasons to question whether CASA is the best or only framework to use in order to increase social engagement with robots. The biggest concern we found was that the robot that had the most human-like social behaviors – conversational talking, movement, game-playing – did not elicit more social engagement from people than other conditions. If anything, the game playing robot engendered **less** social engagement than other, less social interactions. This finding is in direct opposition to CASA.

Another concern is that, even when the robot was telling a story, more people left than stayed to watch the robot. Perhaps this finding is not completely surprising: people at this event had varied agendas and may not have been interested in engaging or watching a robot. However, if one of the goals of the human-robot interaction field is to understand how and why people and robots interact the way that they do, it is sobering to think that a lab, using a state-of-the-art robotic platform and the best current theory on how to elicit social engagement, was able to capture, at most, less than 33% of a naïve population's attention.

Finally, it should be noted that this study did not specifically test whether the social cues accompanying the story-telling were the primary explanation of the engagement findings, or whether just random motion by the robot would produce a similar effect. Given the duration of the engagement, we believe the latter is unlikely, although further studies would be needed to disambiguate this notion.

5. CONCLUSION

This paper used a strong methodology in a naturalistic setting to examine social engagement between people and a robot. We found mixed support for the Computers Are Social Actors framework. We believe that CASA is the best current theory about how and why people will socially engage with robots. However, we also believe that more theoretical and applied work needs to be done to improve or replace the current framework.

6. ACKNOWLEDGMENTS

Our thanks go to those who conducted the experiment during the Fleetweek and made sure the robot behaved properly, namely Magda Bugajska, Eric Martinson, Ed Lawson, Bill Adams, Anthony Harrison, Laura Hiatt and Sunny Khemlani. This research was supported by a grant to JG Trafton from the Office of Naval Research.

7. REFERENCES

[1] Reeves, B., & Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press.

[2] Nass, C., & Moon, Y. (2000). Machies and Mindlessness: Social Responses to Computers. Journal of Social Issues (56), pp. 81-103.

[3] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin. (2005) Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In Intelligent Robots and Systems, IEEE/RSJ International Conference on, pages 708–713.

[4] Sidner, C., Kidd, C., Lee, C., & Lesh, N. (2004) Where to Look: a Study of Human-Robot Engagement. In Proceedings of the 9th International Conference on Intelligent User Interfaces.

[5] Moshkina, L. (2012). Improving Request Compliance through Robot Affect. In Proceedings of Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, Canada.

[6] Chidambaram, V., Chiang, Y.-H., & Mutlu B. (2012). Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Nonverbal Cues. In Proceedings of the 7th ACM/IEEE Conference on Human- Robot Interaction (HRI 2012)

[7] Groom, V., Takayama, L., Ochi, P., & Nass, C. (2009) I am my robot: The impact of robot-building and robot form on operators. In Proceedings of the Human-Robot Interaction Conference: HRI 2009, San Diego, CA.

[8] Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003) A survey of socially interactive robots. Robotics and Autonomous Systems, 42(3-4), pp. 143-166.

[9] Nass, C., & Gong, L. (1999). Maximized Modality or Constrained Consistency? In Proceedings of the International Conference on Audio-Visual Speech Processing (AVSP'99), Santa Cruz, CA

[10] Avison , W.R., McLeod, J.D., & Pescosolido, B.A. (Eds.) (2007) Mental Health, Social Mirror. New York: Springer.

[11] Huang, C.-M., & Mutlu, B. (2013). Repertoire of Robot Behavior: Enabling Robots to Achieve Interaction Goals through Social Behavior. *Journal of Human-Robot Interaction, 2(2), 80–102*.

[12] Helweg-Larsen, M., Cunningham, S. J., Carrico, A., Pergram, A. M. (2004) To Nod or Not to Nod: An Observational Study of Nonverbal Communication and Status in Female and Male College Students, Psychology of Women Quarterly, 28(4), pp. 358-361.

[13] Martinson, E., Lawson, W.E., & Trafton, J.G. (2013). Identifying People with Soft Biometrics at Fleet Week. In Proceedings of the 8th ACM/IEEE Conference on Human-Robot Interaction (HRI 2013).

[14] Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic Interactions with a Robot and Robot-Like Agent. Social Cognition (26:2, pp. 169-181).

8. Appendix A: Text of the Vignettes Used in the Study

8.1 MDS Vignette (Informative)

Hello! My name is Octavia. I work at the Navy Center for Applied Research in Artificial Intelligence. I am an MDS robot. M is for mobile. D is for dexterous. S is for social. I was designed so that it would be easy to work with me, just like with people. I have two video cameras in my eyes; and a special infrared camera in my forehead that let me see shapes. I have four microphone ears that allow me to hear you and help me figure out where sounds are coming from. I have a laser range finder on my base. It helps me avoid obstacles. I have a wide range of motion in my arms and hands. My face is also very expressive. Thank you for listening to me!

8.2 Firefighting Vignette (Informative)

Hello! My name is Octavia. I work at the Navy Center for Applied Research in Artificial Intelligence. I am an MDS robot. M is for mobile. D is for dexterous. S is for social. Our latest project was developing robots that can fight fires on board navy vessels. A real fire was set up in our lab. First, I found my team leader with the cameras in my eyes. Next, he showed me where the fire was. A special camera, in my forehead, helped me recognize a few gestures, like pointing, and come here. Then, I found the fire using two infrared cameras. Finally, using a hose, attached to my left arm, I sprayed the fire with a stream of water. The fire was extinguished - mission accomplished! Thank you for listening to me!

8.3 Kangaroo Vignette (Joke)

Hello! My name is Octavia. I work at the Navy Center for Applied Research in Artificial Intelligence. I heard a funny joke yesterday at fleet week - I hope you like it! Here it is. A kangaroo kept getting out of his enclosure at the zoo. Knowing that he could jump high, the zoo officials put up a ten-foot fence. He was out the next morning, just roaming about the zoo. A twenty-foot fence was put up. He got out, again. When the fence was forty feet high, a camel in the next enclosure asked the kangaroo: How high do you think they'll go? The kangaroo said: About a thousand feet, unless somebody shuts the gate at night! Thank you for listening to me!

8.4 Sherlock Holmes Vignette (Joke)

Hello! My name is Octavia. I work at the Navy Center for Applied Research in Artificial Intelligence. I heard a funny joke yesterday at fleet week - I hope you like it! Here it is. Sherlock Holmes and Dr. Watson are going camping. They pitch their tent under the stars and go to sleep. Sometime in the middle of the night, Holmes wakes Watson up: Watson, look up at the stars, and tell me what you deduce! Watson says: I see millions of stars, and even if a few of those have planets, it's quite likely there are some planets like Earth, and if there are a few planets like Earth out there, there might be life. Holmes replies: Watson, you idiot, somebody stole our tent! Thank you for listening to me!