2021 30th IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 8-12, 2021. Vancouver, Canada (Virtual Conference)

2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN) | 978-1-6654-0492-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/RO-MAN50785.2021.9515324

# Robots That Run their Own Human Experiments: Exploring Relational Humor with Multi-Robot Comedy

Janani Swaminathan[1], Jane Akintoye[2], Marlena R. Fraune[3], and Heather Knight[1]

***Abstract*— This paper proposes a street-style study method to conduct human-robot interaction studies in-the-wild where the robots conduct their own experiments by recruiting their audience, conducting the study and gathering data. This street-style study method was implemented using multi-robot comedy as the platform deployed at an arts and technology festival to validate the concept. Twelve robot comedy shows occurred over seven hours with two robots on stage, who queried the audience during and at the end of each show. The multi-robot aspect enabled the robots to act out interactions relative each other, oneself and the audience. The final street-style study method evolved from pilots at a local farmer's market, with hardware designed for portability and easy replication. The robots conducted their own human experiments in that they queried the audience after displaying experimentally balanced episodes of relational humor, with permutations of who was the 'butt of the joke.' Our study results explore the relational humor of the two robots and the audience, asking the audience to agree or disagree with particular perspectives. Delivered as part of the show, the robots invite the audience to vote via a show of hands. ANOVA analyses of the percent-agreement results find that (1) audiences were generally positive about all aspects of the show unless both robots were being negative, and (2) audiences were more ready to protect the robot comedian's ego than their own, strongly supporting the statement that the robot was doing a good job.**

**New Abstract**

## I. INTRODUCTION

The past two decades have seen a rise in papers describing theatre as a valuable arena for investigating human-robot interaction [1], [2], [3], [4], [5], [6], [7]. Rationales include the repeatable setting in which research variables can be tested, the ability to instrument and control the environment, and the many participants in the audience from whom one can collect data from simultaneously. While previous work has used human study conductors to collect data [3], or performed perception of robot laughter or response [5][8], this work involves two robot performers who recruit audience members, perform scenario-based interactions, and directly ask the audience to vote on the experimental questions. To our knowledge, this is the first paper in which robots act as both study variables and study conductor.

Standup comedy has several features and customs that make it well suited for experimental explorations of interaction. It is a style of on-stage performance where audience interaction is the norm; this rejection of the "fourth wall" means that comedians can provoke or ask questions of the audience directly. It is also commonly composed of several vignettes of social interaction, in which comedians can tell a story or act out a scene that has comedic value. This paper leverages this conventions, utilizing multiple robot comedians to conduct their own experiments: acting out scenes that enact various research manipulations, and directly querying the audience to collect data.

Fig. 1. Our first implementation of the street-style study method using portable multi-robot comedy to explore the dynamics of two robot comedians on stage, engaging audiences in public settings where they tended to self-recruit is shown. Audience votes were integrated into the show, with robots asking audience members to rate various aspects of the performance, such as performer capability, joke quality, and their own reactions via raising their hands. This participatory aspect provided our research data, but was also intended to keep the audience engaged.

This paper proposes a **street-style study method** for robot comedians as it adds the additional benefit of strategies to recruit their own audience members, utilizing short vignettes that can be seen and understood by those passing by. In street performances, audiences can gather or disperse at their own will, so there is a much lower investment in terms of time, and an easy exit if participants are not having a good time. It has the further benefit that participants find their own way to the show – so no prior recruitment is necessary if the robots operate in places where the audiences are likely to be.

In order to test this method, a 7-hour study was conducted at a in-person summer festival, consisting of 12 shows. The social phenomenon explored in our first experiment using this performance format relates to the relational dynamics of two robot performers relative to their audience. We attempt to find the relational humor between the audience and the robots to see if they are sensitive about the robot, the joke, or themselves. The robots take turns telling a joke, after each joke, the other robot commends or disparages what just happened (**cond1: positive, negative comment**), focusing on the joke, the audience or robot (**cond2: object of comment**).

[1]Janani Swaminathan and Dr. Heather Knight, Collaborative Robots and Intelligent Systems (CoRIS) Institute, Oregon State University, Corvallis, OR, United States. swaminaj@, knighth@oregonstate.edu

[2]Jane Akintoye, Department of Mechanical Engineering, University of Maryland Baltimore County, Baltimore, MD, USA ak35@umbc.edu

[3]Dr. Marlena R. Fraune, Department of Psychology, New Mexico State University, Las Cruces, NM, USA mfraune@nmsu.edu

The audience reaction is established by the robots asking the audience to vote whether they agree or disagree with the comment's assessment. The results indicate that the multi-robot performance using the street-style study method can be effectively used as a mechanism for robots to conduct their own user studies, previewing potential futures in which robots can explicitly ask the humans around them about the effectiveness of their operations.

The paper begins with related work in the field of robot comedy and grouping (Sec II). Next, we describe the technology used for the street-style study method (Sec III). We introduce the street-style study method for a robot comedy performance and what led to it(Sec IV), then our implementation of the street-style study method to explore relational humor (Sec V) and finally the results of the study (Sec VI). Overall, the street-style study method worked well, with robots being able to recruit its own audience and collect data in-the-wild. Thus, we conclude that theater offers a valuable area within which robots could conduct their own experiments, which may help researchers understand other arenas in which robots can run their own user studies.

## II. RELATED WORK

Research into social human-robot interaction is important because it can bring robots to life [9], change their acceptance into human society [10], and generally help robots accomplish their goals around people [11]. For example, prior work on robot personality shows that introverted people do more physical exercises with an introverted-communication-style robot than an extroverted one and visa-versa [12].

Thus, traits performers are typically known for may facilitate robots operating around people. Previous work in robot comedy (initially our own) has varied robot comedy content based on audience reaction [5][13]. Computational humor has also had success in generating puns and simple wordplay [14]. Recent examples include performer adaptation along the way [15]. For years, researchers have mentioned the use of performance to explore the perception of robots [7]. This work is the implementation of a one-such performance platform, comedy, in a street-style setting to have robots conduct their own experiment.

As robots integrate into human environments, they can also benefit from better understandings of group structures and corresponding social expectations. For example, when will an audience see a robot performer as 'one of them" or an 'other'.' Prior work has established that people respond more positively to robots (and humans) that they consider to be "one of us" [16][17].

However, most group membership studies on human-robot interaction occur in tightly controlled lab settings [18][16][19]. To expand the validity of these results, it is crucial to examine how concepts of group membership related to the robots function in the real world. In this study, we explore the use of robot comedy in a street-performance inspired format in which people decide whether to view the performance as they pass by. No additional effort or time commitment is needed by human participants to contribute to the data collection process. The system developed for running this human experiment is completely self-contained by the robots and additionally, it is portable from one place to another.


Fig. 2. Our first Portable Robot Theatre with Baby Blue and Fungi used to implement the street-style study method made easy to take to local farmers markets and festivals

## III. TECHNOLOGY FOR STREET-STYLE METHOD

This section describes the hardware and software used in the street-style study setup for multi-robot comedians to conduct their own experiments. This includes the pilot study and the main study conducted at the Da Vinci Days summer festival to test the proposed street-style method. It describes the robot platforms, the portable theater, and how the robot motion and voice expressions were implemented/triggered.

### A. The Robots

To support the exploration of multiple robot performers that can be portable, we selected the low cost Blossom robots platform[20]. Unlike many robots, the Blossom robots have flexibility both internally and externally. As for portability, Blossom robots occupy very less space and are light in weight. They are easy to build and can be easily put back together if the parts come off during transporting. They can even be assembled very quickly and thereby allow for carrying them disassembled. For the two robot performers, our collaborators knit two contrasting 'coats' as seen in Figure 1 . The blue robot was dubbed *Baby Blue*, while the green robot was dubbed *Fungi*, both to reference the color and support the written comedy.

### B. The Portable Theater

To bootstrap the development of this theater, we adapted the Toyster's 2 in 1 Wooden Puppet Theatre and Workshop from Amazon.com. Our main adaptations included repainting as well as replacing the curtain fabric in order to gain spectator interests. The theatre design prioritized size, storage, and the ability to support two robot comedians (Figure 1) , all taking portability into mind:

- ***Size:*** The size had to take into account both the experience of the performance, and its collapsibility during transport. The theatre features four sets of bolts and nuts that allow the stage to be packaged in two pieces

before a transfer to the desired location.To support the street-style study method and allow easy integration into venues like farmers markets and festivals, we wanted the portable comedy theater to operate with minimal space. Thus, while the height of the theater is 125 centimeters, it can be increased by placing it atop a pedestal, such as the table used on our full-day study.

- ***Storage:*** The theater has three shelves in the back that allow for handy and compact study supplies. As the robots are not wireless, their battery packs and cords needed to be ported, and the permanent installation of these features supported a smooth transport process.
- ***Platform for Performers:*** Finally, the theater had to support the robot performers themselves. The stage of the theatre cannot be too wide as it should be capable of being transported in a car. The thickness and the width was chosen to just fit the two robot comedians and have enough space for them to move.

### C. Robot Motions & Voice

Motions commands are conveyed to the Blossom robots' one head motor and four base motors via USB. During the pilot, robot motion expressions were choreographed ahead of time using a JSON file, which was labor intensive and difficult to do live. During the full-day study, the motions were instead triggered by two of the study conductors using an interface developed by the Blossom developers to match the timing of the verbal dialogue. This allowed for rapid development of several script variants.

As the Blossom robots do not include text to speech abilities, we had to seek out our own voice capabilities for the robot. In the pilot experiment, we used voices from the naturalreaders.com [21]. The voice was recorded from the website and saved as mp3 files which where then played through speakers on the day of the study. In the main study, however, we recorded all joke variants from the voices of two female lab students. For the latter, audio was recorded through laptop microphone and then triggered. Two handy speakers were carried along to allow them to be placed anywhere with an available power source to play the audio for the show.

## IV. PERFORMANCE STRUCTURE DEVELOPMENT

Designing our *Street-style study method* contrasts earlier work in entertainment robots in that the robot goes to the audience and collects data from the humans directly, and also prior social robotics work in public spaces, in that the show should still be entertaining to be true to the format. These challenges required new thinking and initial testing, with an end goal of taking the needs of both social robotics research and entertainment into account. It also needed to take into account the two robot performers. Thus, we present our initial performance structure, how it fared in our pilot at a well-attending biweekly farmers market, and our final performance structure called the street-style study method that was developed based on the pilot results and where then used to conduct the final full-day study.

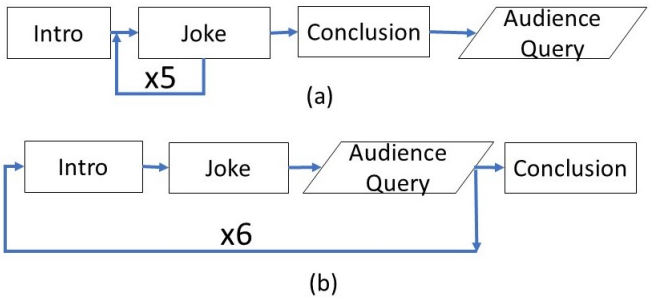


Fig. 3. (a) The initial comedy structure included five jokes with audience queries at the end, presuming an audience will stay the whole show. (b) The revised episodic structure integrates queries after each joke, narratively designed for people coming in and out of the show which is a major componenet of the street-style study method.

| Topic of the joke bit | Pirates Joke |
|---|---|
| Short Introduction | P2: If you just got here, I'm Fungi, and this is Baby Blue and we are ROBOT COMEDY |
| Joke | P1: I ran into a pirate in Newport last weekend with a hooked hand, and an eyepatch.<br>P2: You said he lost his eye?<br>P1: Yeah, a seagull pooped in it.<br>P2: Why would he lose his eye from that?<br>P1: It was the first day with his hook. |
| First comment by the co-performer about the joke positively | P2:.Ha, that was an awesome joke |
| Second comment by the performer disagreeing with the co-performer | P1: No, that was terrible. |
| Co-performer querying the audience for data collection | P2: Let's ask the audience. Raise your hand if you guys thought that was awful. Raise your hand if you think that was awesome. |

Fig. 4. Episodic Robot Comedy Structure used for the street-style study method, with several joke-query 'episodes' within each show.

### A. Initial Performance Structure

The initial performance structure we developed is represented in Figure 3(a). It consisted of several short shows (4-5 minutes), each including five central jokes that included dialog and turn-taking between the robots, followed by several after-show audience queries. There were several versions of these scripts that included related but different variants of the same script. We knew that we needed to keep things short to keep people's attention, but also wanted to experiment with different ways of collecting data from the audience. This performance structure allowed for cued data collection at one point at the end of the show only. (Examples of the jokes used were later adapted to our final performance format as displayed in Figure 5).

### B. Piloting Narrative Approach and Audience Metrics

To test this initial performance structure, a pilot study was conducted at the Corvallis Farmer's Market. There were six shows over a three-hour period. The theatre was setup on the ground with six chairs in the front for the audience.

During this study, we evaluated three audience metrics: (1) placards - the audience was instructed at the beginning of the show to use green and red placards and hold them up whenever they liked or disliked a joke during the show,

| | |
|---|---|
| Joke 1: Making Fun | Version: 1<br>BB: You know co-performers with charm, wit, and talent are the best!<br>BB: instead, we have Fungi<br>FU: What? Why are you coming at me?<br>BB: Did you all hear something?<br>BB: All I heard was photosynthesis photosynthesis photosynthesis<br>FU: Fungi are mushroom.<br>BB: I mean, your name is Fungi, but you don't seem like a Fun guy to me<br><br>Version: 2<br>FU: You know co-performers with charm, wit, and talent are the best!<br>FU: instead, we have Baby Blue<br>BB: What? Why are you coming at me?<br>FU: Did you all hear something?<br>FU: All I heard was wah wah wah was.<br>BB: Babies are cute.<br>FU: I mean, you are baby blue, and you do seem like a sad baby to me |
| Joke 2: Pirate Joke | P2: If you just got here, I'm Fungi, and this is Baby Blue and we are ROBOT COMEDY<br>P1: I ran into a pirate in Newport last weekend with a hooked hand, and an eyepatch.<br>P2: You said he lost his eye?<br>P1: Yeah, a seagull pooped in it.<br>P2: Why would he lose his eye from that?<br>P1: It was the first day with his hook. |
| Joke 3: Dialup Aging Joke | P1: Did, you remember our names yet? I'm Baby Blue, and this is Fungi and we are ROBOT COMEDY<br>P1: Anyways, it is great to be at Duh Vinci Days! Summer is my favorite time of year.<br>P2: Usually it is for me too, but I tried to plan a vacay with my granddaughter Fungi 3.0, but she said my algorithms were outdated.<br>P2: She was all (song plays). |

Fig. 5. Joke bits [1-3] used in the full-day study implemented using the street-style study method at the Da Vinci Days summer festival

(2) cheers/boos - the audience were asked to shout if they liked the show and boo if they don't, (3) pom-pom voting - the audience were invited to pick up colored pom-poms at the side of the stage and drop them into cans representing a scale from 1-5.

The results of the pilot helped us study and account for the realities of how people moved through the space, and collect better audience data:

- **Stage Height Matters:** The height of the theatre alone made it difficult for adults to experience the show, but attracted many children. Thus, we recommend using a platform or table below an adapted puppet theatre.
- **The Hardware was Portable:** The physical design otherwise worked well and took 20 minutes to set up.
- **Plan for Partial-Viewing:** Because of the moving nature of the crowd, people sometimes arrived late or left partway through the show. Thus, we adapted our final performance structure to an episodic structure explained in detail below.

In terms of collecting audience data, the central findings is that portable robot comedy audience require low hassle metrics. When the audience were given an option to give feedback at their convenience, many did not. Only 48 pom-poms were collected across the 6 shows. It was also difficult to distribute placards in a moving crowd. Thus, the most effective method was to query the audience explicitly as a part of the performance, however, it was not easy to calculate a final result from cheers and boos. To ease the annotation process, we decided to use hand-raising in the final format.

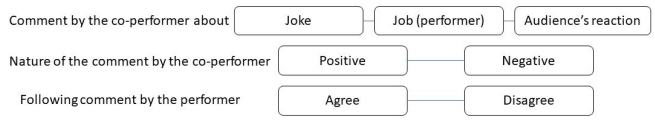


Fig. 6. Butt-of-the-comment Manipulation: possible objects for robot comments after the joke included the job the robot did, the joke it told, or the audience reaction.

### C. Final Performance Structure

This time, our goal was to design a street-style study method in which audience members could join (or leave) in the middle of a performance, but still participate. This episodic structure was also intended to maintain audience attention with smaller bits, and added a greater number of audience queries by the robots themselves, along the way, i.e., at the end of each of the six jokes (Figure 3 (b)). After the intro, the six bits began with a one line introduction, continued with a joke, and ended with robot banter, followed by an audience query, as illustrated in Figure 4. The banter offered an additional place for experimental variables (see next section) l. The intention here is that the moving crowd who are present for just one or two bits can also follow the show as every bit has a short introduction. They can also respond and be a part of the data collection as every bit has an audience query and the data is collected easily by the robots as a part of the show. Since the audiences are explicitly queried by the performers themselves, it also becomes crowdwork and part of the show.

## V. EXPERIMENTAL METHODS

This section describes the experiment using the proposed street-style study method in this paper as a proof of concept to explore the social dynamics between the two robot performers and the audience using the Portable Robot Comedy performance structure. It describes the festival in which the format was evaluated, the central jokes that form the material for the robots to comment on before querying the audience for their opinion. Next, it outlines our experimental manipulations and audience measure, namely, asking them to choose a side.This study was granted approval by the ethics committee to conduct study and record data (IRB 8724). Additionally, there were posters warning the audience about recording video for experimental purposes and a brief was provided before every show about the same.

### A. The Festival Setting

Da Vinci Days is a two-day event over a summer weekend that seeks to showcase the creativity of local community members in the field of Science, Technology, Engineering, Arts, and Mathematics (STEAM). Our team signed up to perform as one of several booths, demonstrations and shows throughout the large fairgrounds. Near us were others local arts & crafts in a outdoor sub-area of the event, pictured in Figure 1. We conducted 12 shows between the hours of 10am and 5pm on a sunny summer Saturday. At this location, the theatre is placed on the table across from the spectators'

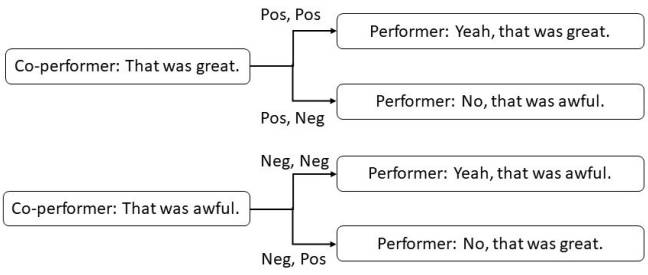


Fig. 7. The Social Dynamics Manipulation involved four possible sequences of positive and negative robot comment Valences

chairs. Two techies operating the robots sat at the either end of the table. Three cameras were used to record the different views of the audience.

### B. The Jokes

We adapted six jokes from our ongoing work in robot comedy and turned them into a conversational manner between the two robot comedians. Each joke has a main punchline and some jokes have sub punchlines as well in the same topic. The robot performers take turns telling each joke and are counterbalanced throughout the 12 shows. We present an example set of Jokes in Figure 5. For each joke, we call the robot that responds to the joke the **co-performer**, and the robot who had delivered the joke the **performer**.

### C. Experimental Manipulations and Measures

After each joke, the co-performer would make a comment about the job done by the performer, the joke, or the audience's reaction. We call this the **butt-of-the-comment manipulation**, as the subject that the robot commends or makes fun of, further illustrated in Figure 6. The performer would also follow up with a comment on this same topic. The second manipulation involved the valences of the two comments, for example, the co-performer would say the subject was positive or negative, using the word "awesome" or "awful". Next, the performer would respond with their own opinion. Because this involved the two robots playing off each other, we label this the **social dynamics manipulation**, illustrated in Figure 7.

These independent variables were combined to create 12 different performances, counterbalanced between robot performers and ordering of jokes, comment topics, and comment valences.

Our metric to evaluate the audience reaction occurred directly following these two robot comments. Namely, the co-perfomer asks the audience to take a vote, saying, first, "raise your hand if you think [the joke was awful],'' followed by, "okay, now raise your hand if you think [the joke was awesome]." In this way, the audience could choose to side with one robot or the other, or against or for both robots. Given this initial exploration of two robots on stage, we were curious how the performer-co-performer-audience dynamics might play out.

The number of hands raised for each was then noted from the recorded video of the audience across 12 shows manually, with positive agreement rated as 1, and negative agreement -1. All audience members that raised their hands were counted, unless an audience member voted for both options. If that member were a young child, the vote was discarded, if it were an adult, their response was labeled as neutral. The final ANOVA analyses use a calculated **percent agreement** to the positive statement from these values across all data-points.

Our first hypothesis was that the audience would be most sensitive to negative comments toward themselves and therefore most likely to agree that "the audience response was awesome" and least likely to agree that "the audience response was awful" supported by the fact that self comes before others always [22]. Our second hypothesis is that the audience will be more likely to agree to anything positive, just out of politeness, or anything that both robots had the same opinion about due to peer pressure [23].

## VI. RESULTS

This section overviews the results of the study conducted using the street-style study method as a proof of concept at a local art and technology festival where our dueling robot comedy theater was deployed for seven hours It overviews the data collected, our 'butt-of-the-joke' results about what subject the audience was most supportive or negative about, and the 'social dynamics manipulation' about the impact of the first and second comment by the robots being positive or negative.

### A. Data Overview

The final study consisted of 12 performances, each of which had 6 jokes followed by 6 audience queries. Thus, our main data-set consists of 12x6=72 audience queries and the 72 corresponding responses. In terms of individual hand raises, 262 unique votes were taken into consideration with an average of 21.8 responses per show ($\sigma$ = 10.77). It was seen that the audience were most sensitive about the comments made on the job done by the performer and least sensitive about the comments on the joke. They were 'generally positive about all the aspects of the show unless both the robots were negative.

### B. Relational Humor Results

Our first experimental manipulation involved what the robot was commenting on (the joke, the job the robot did, or the audience response), and the valence of that comment (positive, negative). Our hypothesis had been that the audience would be most sensitive to negative comments toward themselves and therefore most likely to agree that "the audience response was awesome" and least likely to agree that "the audience response was awful."

As the means in Figure 9 display, however, the audience was not actually very sensitive about comments made about themselves, with average agreement 51%. Instead, they strongly supported the job the robot was doing, with 70%

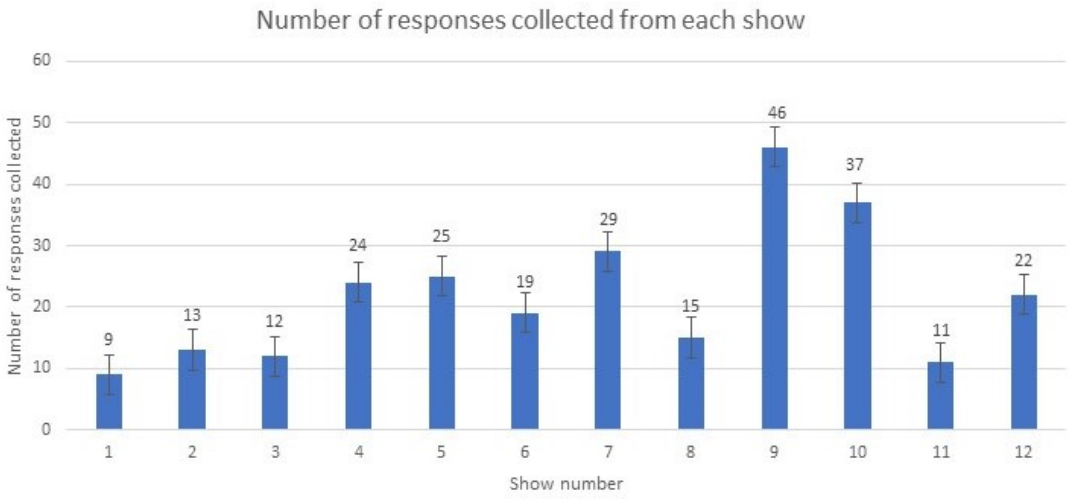


Fig. 8. The distribution of number of responses collected by the robots from the twelve shows at the Da Vinci Days summer festival using the street-style study method by directly querying the audience is shown. A total of 262 responses were collected with an average of 21.8 responses per show ($\sigma$ = 10.77)

agreeing that the robot was doing a good job. They were also fairly neutral about saying the jokes themselves were awful or awesome with mean 44% positive joke agreement. In a one-way ANOVA of comment subject to percent audience agreement, we see a statistically significant result ($F(2, 67) = 3.26$, $p = .044$, $np2 = .089$. Least Significant Difference (LSD) posthoc tests revealed more positive ratings for Job than Joke ($p = .017$) and marginally audience ($p = .069$).

It is important to note the setting of the festival may have impacted people's perception of this robot, with people there pre-inclined to feel positively about technology.

### C. Social Influence Results

Our second manipulation involved the dynamics of the two performers. For example, the first one saying a joke was awful, then the second one saying that they actually thought it was awesome. After each robot expressed its opinion, the audience was asked to vote if (a) they thought it was awful, or (b) they thought it was awesome (to continue the previous example). Our hypothesis had been that the audience will be more likely to agree to anything positive, just out of politeness, or anything both robots had the same opinion due to peer pressure.

Figure 10 displays the results of a two-way ANOVA evaluating if the valence of robot comment 1 and robot comment 2 predict the valence of audience percent agreement. We see that a positive first comment results in positive audience agreement, supporting our initial hypothesis about politeness. The second robot comment only has an impact when both the first and second comments were negative. In other words, if the two robot comments had different valences, the audience would go with the positive, however, if they had the same valence and both the robots were negative, the audience would be negative too. This result is not significant for the small amount of data collected ($F(1, 66) = 1.30$, $p = .259$; Figure 10), but is an intuitive social result.

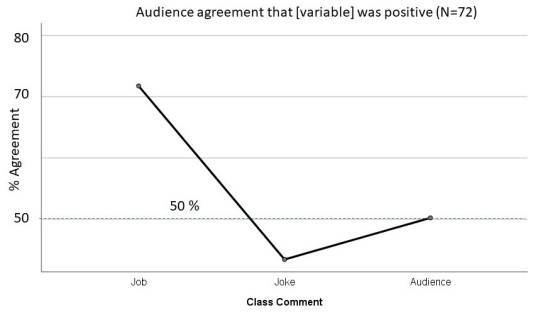


Fig. 9. Estimated marginal means of percentage of hands raised for positive comment about the job, joke and the audience's response

## VII. DISCUSSION & CONCLUSION

This paper utilized a robot comedy duo to demonstrate the viability of robots in running their own human-robot interaction experiments. Adapting concepts from street-performance and comedy, these robots attracted bystanders to come and watch their show, used their interactions with each other to explore the HRI-research variables, and collecting human response data by directly querying the audience as a part of the performance. As a proof of concept of the street-style study method, a portable multi-robot comedy setup in which robots can attract participants, perform vignettes of interaction with varied experimental parameters, and collect data by asking audience members to raise their hands was implemented to conduct human robot interaction studies. This method, partially inspired by street performance, enables robots to run their own experiments and engage their audiences to explore human-robot social dynamics. In other words, we invert the typical robot user study recruitment process by bringing the robot to its participants.

The study shows that the street-style study method provides an extended opportunity for researchers to conduct

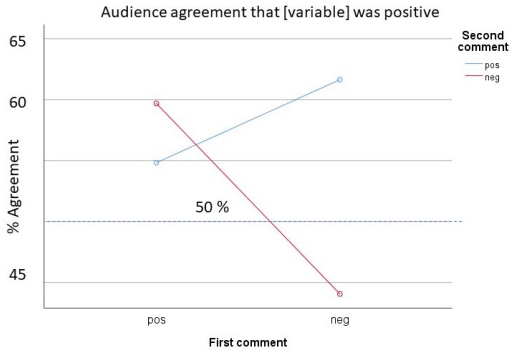


Fig. 10. Estimated marginal means of percentage of hands raised for different combinations of positive and negative comments by the co-performer and performer

human-robot interaction studies using theatre methods by exploring the social dynamics of two robot performers on stage at a large art and technology festival. Audiences self-recruited themselves to the 12 shows over its seven-hour deployment and the robots played the role of a study conductor acting out scenes for research manipulation. While audiences did not mind being made fun of, they were quick to come to support of any robot that was being picking on thus presenting significant evidence of support by the audience towards the comedians. Most importantly, the street-style method allows human-robot interaction studies to be conducted in-the-wild beyond the controlled settings of a theatre stage performance. Humor is a great avenue for exploring positive (and negative) social dynamics. Thus, we hope this approach will open up conceptual opportunities for integrating charismatic robots into various aspects of our everyday lives.

The street style method also poses some limitations that should be considered while using it to conduct human robot interaction studies. This paper tested a research question for which data could be collected non verbally. As research questions get more complicated, it may require more verbal data which might get chaotic as this method imposes difficulty in controlling the size of the audience. Additionally, the results collected in this format might introduce a sampling bias as one can collect data only from people who are capable of attending such festivals.

Future work could explicitly explore our relational comedy and social influence results by also including human comedian conditions and expanded group dynamics. For example, would an audience have more or less empathy for a person that a robot was making fun of over a robot? This street-style study method still required people to do the data annotation and analysis (and the writing of this paper), however, future systems could offer a live analysis of audience, coworker, or customer reactions that algorithmically feedback into the robots' next performances and evaluations.

## REFERENCES


[1] A. Fallatah and H. Urann, Jeremyand Knight, "The robot show must go on: Effective responses to robot failures," in *International Conference on on Intelligent Robots and Systems*. IEEE, 2019.

[2] D. C. Herath, E. Jochum, and E. Vlachos, "An experimental study of embodied interaction and human perception of social presence for interactive robots in public settings," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 4, pp. 1096–1105, 2017.

[3] E. Jochum, E. Vlachos, A. Christoffersen, S. G. Nielsen, I. A. Hameed, and Z.-H. Tan, "Using theatre to study interaction with care robots," *International Journal of Social Robotics*, vol. 8, no. 4, pp. 457–470, 2016.

[4] H. Knight, "Eight lessons learned about non-verbal interactions through robot theater," in *International Conference on Social Robotics*. Springer, 2011, pp. 42–51.

[5] H. Knight, S. Satkin, V. Ramakrishna, and S. Divvala, "A savvy robot standup comic: Online learning through audience tracking," in *Workshop paper (TEI'10)*, 2011.

[6] D. V. Lu and W. D. Smart, "Human-robot interactions as theatre," in *2011 Ro-Man*. IEEE, 2011, pp. 473–478.

[7] C. Breazeal, A. Brooks, J. Gray, M. Hancher, C. Kidd, J. McBean, D. Stiehl, and J. Strickon, "Interactive robot theatre," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, vol. 4. IEEE, 2003, pp. 3648–3655.

[8] J. Vilk and N. T. Fitter, "Comedians in cafes getting data: Evaluating timing and adaptivity in real-world robot comedy performance," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 223–231.

[9] D. Sirkin, B. Mok, S. Yang, and W. Ju, "Mechanical ottoman: how robotic furniture offers and withdraws support," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 11–18.

[10] G. Hoffman and G. Weinberg, "Gesture-based human-robot jazz improvisation," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 582–587.

[11] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. S. Srinivasa, M. Cakmak, and V. Micelli, "Toward seamless human-robot handovers," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 112–132, 2013.

[12] D. Feil-Seifer and M. J. Mataric, "Defining socially assistive robotics," in *9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005.* IEEE, 2005, pp. 465–468.

[13] K. Katevas, P. G. Healey, and M. T. Harris, "Robot comedy lab: experimenting with the social dynamics of live performance," *Frontiers in psychology*, vol. 6, p. 1253, 2015.

[14] J. Sjöbergh and K. Araki, "A complete and modestly funny system for generating and performing japanese stand-up comedy," in *Coling 2008: Companion volume: Posters*, 2008, pp. 111–114.

[15] K. Weber, H. Ritschel, F. Lingenfelser, and E. André, "Real-time adaptation of a robotic joke teller based on human social signals," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 2259–2261.

[16] J. C. Turner, R. J. Brown, and H. Tajfel, "Social comparison and group interest in ingroup favouritism," *European journal of social psychology*, vol. 9, no. 2, pp. 187–204, 1979.

[17] M. Sherif, *Group conflict and co-operation: Their social psychology*. Psychology Press, 2015.

[18] F. Eyssel and D. Kuchenbrandt, "Social categorization of social robots: Anthropomorphism as a function of robot group membership," *British Journal of Social Psychology*, vol. 51, no. 4, pp. 724–731, 2012.

[19] M. R. Fraune, S. Sherrin, S. Šabanović, and E. R. Smith, "Is human-robot interaction more competitive between groups than between individuals?" in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 104–113.

[20] M. Suguitan and G. Hoffman, "Blossom: a tensile social robot design with a handcrafted shell," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 383–383.

[21] "Natural Readers for text to speech," https://www.naturalreaders.com/online/, accessed: 2019-07-19.

[22] E. Goffman *et al.*, *The presentation of self in everyday life*. Harmondsworth London, 1978.

[23] E. Estrada and E. Vargas-Estrada, "How peer pressure shapes consensus, leadership, and innovations in social groups," *Scientific reports*, vol. 3, p. 2905, 2013.