

**NISTIR 8339**

**Workshop Report: Test Methods and  
Metrics for Effective HRI in  
Collaborative Human-Robot Teams,  
ACM/IEEE Human-Robot  
Interaction Conference, 2019**

Shelly Bagchi  
Murat Aksu  
Megan Zimmerman  
Jeremy A. Marvel  
Brian Antonishek  
Heni Ben Amor  
Terry Fong  
Ross Mead  
Yue Wang

This publication is available free of charge from:  
<https://doi.org/10.6028/NIST.IR.8339>

**NIST**  
National Institute of  
Standards and Technology  
U.S. Department of Commerce

**NISTIR 8339**

**Workshop Report: Test Methods and  
Metrics for Effective HRI in  
Collaborative Human-Robot Teams,  
ACM/IEEE Human-Robot  
Interaction Conference, 2019**

Shelly Bagchi, Murat Aksu, Megan Zimmerman, Jeremy A. Marvel, Brian Antonishek  
*Intelligent Systems Division  
Engineering Laboratory*

Heni Ben Amor  
*Arizona State University*

Terry Fong  
*NASA Ames Research Center*

Ross Mead  
*Semio AI, Inc.*

Yue Wang  
*Clemson University*

This publication is available free of charge from:  
<https://doi.org/10.6028/NIST.IR.8339>

December 2020



U.S. Department of Commerce  
*Wilbur L. Ross, Jr., Secretary*

National Institute of Standards and Technology  
*Walter Copan, NIST Director and Undersecretary of Commerce for Standards and Technology*

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

**National Institute of Standards and Technology  
Interagency or Internal Report 8339  
Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8339, 18 pages (December 2020)**

**This publication is available free of charge from:  
<https://doi.org/10.6028/NIST.IR.8339>**

## **Abstract**

Verified and validated test methods, being necessary to measure the performance of complex systems, are important tools for driving innovation, benchmarking and improving performance, and establishing trust in collaborative human-robot teams. This full-day workshop aims to explore the metrology necessary for repeatably and independently assessing the collaborative performance of robotic systems in real-world human-robot interaction (HRI) scenarios. This workshop aims to bridge the gaps between the theory and applications of HRI in industry, accelerating the adoption of cutting edge technologies as the industry state-of-practice. The workshop was held on March 11, 2019, at the ACM/IEEE International Conference on Human-Robot Interaction in Daegu, South Korea.

The interest in collaborative HRI is evident in the current market as well as standards efforts toward manufacturing, social, medical, and service robot solutions. Though these domains have been considered separate for many years, recent technological and scientific advancements show that, while their applications may differ, the underlying principles of HRI performance impact each identically. As such, this workshop seeks to identify test methods and metrics for the holistic assessment and assurance of collaborative HRI performance. The focus is on identifying the key performance indicators of these seemingly disparate sectors, and additionally to establish a community based on the principles of transparency, repeatability, and establishing trust in the assessment of collaborative HRI. The goal is to aid in the advancement of HRI technologies through the development of experimental scenarios, protocols, test methods, and metrics for the verification and validation of interaction solutions and interface designs.

## **Keywords**

Robotics; Human-Robot Interaction; Collaborative Robotics; Human-Robot Teaming; Metrics; Test Methods.

## Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Target Audience	2
1.2	Schedule and Format	2
1.3	Discussion Topics	2
<b>2</b>	<b>Invited Talks</b>	<b>4</b>
2.1	Keynote	4
2.2	Invited Talk 1: Dr. Tathagata Chakraborti, IBM Research	5
2.3	Invited Talk 2: Dr. José Lopes, Heriot-Watt University	6
<b>3</b>	<b>Accepted Presentations</b>	<b>7</b>
<b>4</b>	<b>Panel Discussion</b>	<b>9</b>
4.1	Discussion Questions	10
<b>5</b>	<b>Discussion of Standards</b>	<b>10</b>
<b>6</b>	<b>Outputs and Documentation</b>	<b>11</b>
<b>7</b>	<b>Future Plans</b>	<b>11</b>
	<b>References</b>	<b>11</b>

## List of Tables

Table 1	Schedule for the 2019 Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams workshop	3
---------	--	---

## **Disclaimer**

Certain trade names and company products are mentioned in the text or identified in certain illustrations. In no case does such an identification imply recommendation or endorsement by the NIST, nor does it imply that the products are necessarily the best available for the purpose.

The opinions expressed in this Workshop Report are those of the workshop participants and are not the official opinions of NIST. The summaries of the presentations have been reviewed by the speakers and the summaries reflect the speaker's main points.

## 1. Introduction

Despite large advances in interfaces and user-centric robot designs, the need for effective human-robot interaction (HRI) continues to present challenges for the field of robotics. As new technologies are integrated into human-robot teams in a myriad of application domains ranging from smart manufacturing to home automation, exposures to and expectations from robots are growing rapidly. A key limitation that is negatively impacting the achievement of effective human-robot teaming is that there are few consistent metrics for assessing how “effective” HRI actually is. The need for validated test methods and metrics is driven by the desire for repeatable and consistent evaluations of HRI methodologies. Such evaluations are critical for advancing the underlying theory of HRI, as well as establishing traceable mechanisms for vendors and consumers of HRI technologies to assess and assure functionality.

The full-day workshop began to address the issues surrounding the development of test methods and metrics for evaluating HRI performance across the multitude of application domains, including industrial, social, medical, and service robotics. The morning session focused on the diversity of approaches for addressing HRI metrology in different robotic domains via talks by invited speakers and contributing authors. The afternoon session addressed the underlying issues of traceability, objective repeatability, and transparency in HRI metrology by means of expert panel sessions, late-breaking research poster presentations, and open discussions of the HRI community.

The need for establishing consistent standards for evaluating HRI drove the creation of this workshop series. We additionally aimed to explore how the interfaces, technologies, and underlying theories impact the effective collaboration and cooperation of human-robot applications. Specific goals included the following:

- To develop and encourage the use of consistent test methods and metrics in evaluating HRI technologies, developing high quality data sets, and implementing and disseminating human subject studies for HRI;
- To establish benchmarks and baselines along a spectrum of key performance indicators for assessing and comparing novel HRI systems and applications;
- To support a discussion about best practices in metrology and what features should be measured as the underlying theory of HRI advances;
- To encourage the creation and sharing of high-quality, consistently-formatted datasets for HRI research.

The workshop was held on March 11, 2019, at the ACM/IEEE International Conference on Human-Robot Interaction in Daegu, South Korea, and was well-attended, with a peak audience of about 30 members.

## 1.1 Target Audience

The workshop served as a springboard for establishing a formalized and standardized HRI research community. Specific targeted interest groups include stakeholders in industrial, collaborative, medical, and service robotics:

- Researchers of novel HRI theories, applications, technologies, and systems;
- Researchers generating quality HRI datasets, or interested in consuming said datasets;
- Researchers studying the social impacts and acceptance of human-robot teaming;
- Researchers investigating HRI metrics, benchmarks, and other practices aimed at repeatable performance studies;
- End users and consumers of HRI technologies;
- Manufacturers and integrators of HRI or Human-Machine Interface (HMI) technologies;
- Standards communities interested in performance metrics for robotic systems.

## 1.2 Schedule and Format

Table 1 contains the workshop schedule. The first half of the day was dedicated to the technical aspects of metrology for collaborative HRI, and featured a keynote speaker and technical presentations of contributed papers. The papers were peer-reviewed prior to acceptance to assure quality and relevancy of the presentations. The second half of the day focused on the philosophy of collaborative HRI metrology, and featured both panel and open-forum discussions, and an informal poster session during an intermission break. The panel and open discussions in the second half of the workshop were intended to identify and document community approaches, challenges, and opportunities regarding test methods, metrics, artifacts, impacts, and data sets for HRI verification and validation.

## 1.3 Discussion Topics

Presentations by contributing authors focused on the documentation of the test methods, metrics, and data sets used in their respective studies. Keynote and invited speakers were selected from a targeted list of HRI researchers across a broad spectrum of application domains. Poster session participants were selected from contributors reporting late-breaking evaluations and their preliminary results.

Panel and open discussions were intended to highlight the various approaches, requirements, and opportunities of the research community toward assessing HRI performance, enabling advances in HRI research, and establishing trust in HRI technologies. Specific topics of discussion included the following:



**Table 1.** Schedule for the 2019 Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams workshop

Time	Topic	Presenter
09:00 - 09:15	Welcome and introduction	Megan Zimmerman
09:15 - 10:00	Keynote	Dr. Laurel Riek (UCSD)
10:00 - 10:30	Contributed Presentations	Dr. Malte Jung (Cornell) Kathleen Belhassein & Guilhem Buisan (CNRS, LAAS)
10:30 - 10:50	Break	
10:50 - 11:35	Invited Talk	Tathagata Chakraborty (IBM)
11:35 - 12:00	Contributed Presentation	Raquel Oliviera (ISCTE)
12:00 - 12:30	Invited Talk	José Lopes (HWU)
12:30 - 13:00	Recap and discussion	
13:00 - 14:00	Lunch	
14:00 - 14:45	Panelist Introductions	Dr. Chung-Hyuk Park (GWU) Dr. Yun Kyung Kim (NASA) Dr. Katrin Lohan (HWU)
14:45 - 15:30	Panel discussion	Moderator: Murat Aksu
15:30 - 15:50	Break & poster session	
15:50 - 16:05	Introduction to Standards	Shelly Bagchi
16:05 - 16:50	Standards Presentation	Dr. Seungbin Moon (Sejong U)
16:50 - 17:30	Recap and next steps	Organizers
17:30	Close	

- Human-robot collaboration and teaming test methods;
- Human data set content, formatting, metadata, transferability, and traceability;
- Human-robot interaction metrics (e.g., situation awareness, trustworthiness, and cultural influence);
- Human-machine interface metrics (e.g., timeliness, information quality, learnability, functionality, and security);
- Industry-specific metrology requirements (e.g., industrial, medical, and service robotics).

## 2. Invited Talks

The morning session of the workshop opened with a keynote presentation. There were also two additional invited talks, of a longer format than the contributed paper presentations.

### 2.1 Keynote

The opening keynote presentation, titled “Metrology and HRI”, was given by Dr. Laurel Riek, Associate Professor of Computer Science and Engineering at University of California, San Diego.

Dr. Riek began with the ongoing shift from robot-focused to human-focused HRI. In order to advance this goal, the concept of formative vs. summative research is key. Formative research can help identify problems that need to be solved by future work, while summative evaluates the performance of a completed project. Dr. Riek emphasized that rather than a binary relationship, this can be more of a scale. In her opinion, current HRI research is overly focused on summative work. In particular, referencing work by Andrew Gelman, the focus on hypothesis testing and p-values that are chosen to satisfy specific lab conditions leads to research that is hard to reproduce and difficult to compare between experiments [1].

The main issue discussed by Dr. Riek was that of how to select an evaluative paradigm for your HRI research. Relevant parameters included formative vs. summative, short vs. long term, whether the project is safety-critical, and other such factors. As an example case study, she described a joint project with Dr. Julie Shah (MIT) wherein they analyzed different aspects of a manufacturing environment with the aim of creating better workflow and upgrading to a ‘smart factory’ [2]. Some issues encountered in the factory included the dynamic environment, an aging workforce, a frequent turnover in employees and tasks, and a frequent reliance on manual workflow reducing throughput. Thus, as formative research, the first task involved comparing methods of modeling humans to more specifically identify areas of improvement or concern. For example, the industrial manufacturing environment necessitated greater safety concerns than a home environment might. Although the

work is ongoing (long-term research), a quantitative comparison was made between several Human Activity Recognition methods, including motion capture systems and wearable electromyography sensors.

A second case study involved robot caregivers for dementia patients [3]. The aim was to reduce the burden on their human caregivers, who have historically had poor health as a result of their physically and emotionally taxing roles. In order to determine what design features are important for robots in this role, formative research was conducted. Design ideas and objectives were collected from current caregivers, rather than only from patients and doctors as previous studies have done. From this investigation, the team was able to construct new guidelines for robot behavior in the space. In particular, an unexpected goal was to utilize robot caregivers to help bear the emotional load, and additionally to change robot behavior based on the stage of dementia a patient is experiencing.

Another topic of relevance was metrological decision making. Dr. Riek emphasized the importance of thinking beyond one paper or grant proposal when choosing metrics for a study; in other words, do not use what is most convenient, but rather what is most important for future work in the space. The Slow Research movement was mentioned as a community that can provide useful advice on this front. Additionally, rather than attempting to re-invent the wheel for every new study, one should investigate the metrics already used in related fields such as psychology, to leverage existing measures and provide a common basis.

After the talk, an audience member asked the question: how can students know what qualifies as research if it differs from the general standard of summative work? The answer given was that formative science or qualitative science is still science; it helps frame your problem, though it does not immediately yield decisive results. It is an important part of decision making in research and helps design future summative experiments.

Dr. Riek's talk was engaging and very well received. The workshop organizers appreciate her contribution and look forward to collaborating in this field in the future.

## 2.2 Invited Talk 1: Dr. Tathagata Chakraborti, IBM Research

Dr. Chakraborti presented a talk titled "Planning with Multiple Models: Plan Explanations and Explicability". In this case, the multiple models involved may include the human's conception of the robot's intentions, as well as the robot's plan which needs to take human actions into account. In addition, modeling the human's expectations to keep the human in the loop can be a challenge, and any of these models may diverge from the plan during robot execution.

Dr. Chakraborti presented an experiment where a robot's navigation plan is not possible in the human's knowledge of the world, and simultaneously the human's conception of the shortest path is not feasible due to the robot's internal map state. Now, the models must be reconciled through explanations and questioning by the human partner. The model update process was discussed in depth and the details can be found in Dr. Chakraborti's publications.

Various types of explanations were seen during the study. Two of note were Minimally

Complete Explanations, which were shortest in length, and Minimally Monotonic Explanations, which are short explanations that never cease to be valid even after later updates. However, in human-human explanation trials, when the participants were not restricted to giving the most concise explanation, they tended to give longer-form, more detailed explanations. Another factor mentioned was explicability and the process of evaluating the given explanations.

Next, on the robot's side of the problem, the robot must decide how to reconcile its model with the human's. This requires various levels of conforming, whether to conform to all the human's requirements or not, and how to factor in uncertainty about the human's model.

Finally, the talk concluded with a discussion of these topics applied to the authors' team in the DARPA Robotics Challenge <sup>1</sup>. The competition involved multiple agents (both human and human-robot teams) which had to reconcile their individual models over the course of the competition.

### **2.3 Invited Talk 2: Dr. José Lopes, Heriot-Watt University**

Dr. Lopes, a research associate at Heriot-Watt University in Scotland, presented about "Development of a Conversational Agent for Remote Autonomous Vehicles in a High Stakes Environment". The project is part of Heriot-Watt's ORCA Hub, which deals with human-robot interaction for naval disaster relief. The project in particular dealt with agents controlled by natural language (speech or text), with questions about how to generate data for test scenarios, how to determine if the agent is trusted, and how to track the human's state or reactions.

Data collection for the human study was done using Wizard of Oz (WoZ)-controlled robots of various models during an emergency scenario. Limited time was available to resolve the emergency, which only the robotic agent could investigate. Users were asked whether they wanted the agent to make decisions after the interaction (text-based dialogue) as a measure of trust. From this preliminary study it was determined that more situational awareness was needed in terms of visualizations to provide operator knowledge as well as urgency about the emergency scenario.

The remainder of the talk involved ways of determining cognitive load from natural language. A literature survey determined that factors including predictability of dialogue and number of utterances could be indicators of mental workload.

Future work indicates automation of the dialogue system rather than WoZ control. In addition, responses from more experienced participants indicated that an adjustable level of transparency would be desired for operators at different levels. This would help increase trust in the agent. Future questions include adaptive strategies to adjust cognitive load levels and investigation into metrics for better evaluation of the agent.

---

<sup>1</sup><https://www.darpa.mil/program/darpa-robotics-challenge>

### 3. Accepted Presentations

#### **Dr. Malte Jung, Cornell University**

Dr. Jung framed his talk, “Enabling Human-Robot Collaboration With Groups of People”, by contrasting the common research setup with a real-world situation. In HRI research, you are likely to see one human with one robot. However, in work environments, it is more common to find people working in collaborative teams. How do we design robots to leverage the unique situation of groups of people working together? Thus, Dr. Jung outlined his lab’s design criteria for group HRI research: it must enable data collection across group sizes and levels (e.g., individual as well as whole-group), employ a commonly used robot arm, and allow for different methods of robot control.

The task tackled for the present work was a group tower-building exercise where a collaborative robot arm delivered parts to two participants. In this case, who should the robot deliver the parts to? Where should its attention be? There are many more questions to be answered with the addition of more participants.

During the study, participants’ reactions were observed during a situation where the robot unevenly handed out the parts, or in other words, unexpectedly favored one participant over the other. In a larger study, the participants’ reactions could be collected, analyzed, and used to adjust the robot’s behavior. The setup is thought to be extensible to larger groups and could be used to examine the robot’s influence on the team dynamic. Additionally, various experimental metrics could be ‘auditioned’ in this simple setup as proof-of-concepts.

#### **Guilhem Buisan and Kathleen Belhassen, LAAS-CNRS, University of Toulouse**

The authors jointly presented their work, “Towards Methodological Principles for User Studies in HRI”. The authors, a multidisciplinary team from psychology and computer science, focused their work on the need to consider the human element during HRI user studies; i.e., human-in-the-loop research. First, for participant recruitment, the authors emphasized the need to select a representative population with sufficient sample size to draw conclusions. Additionally, one should consider the novelty effect when new users participate in a study.

Moving on to evaluation methods, the authors pulled in the existing concepts of Usability from human-computer interaction (HCI), specifically effectiveness, user satisfaction, and acceptance (in this case of the robot or interface). As with most subjective metrics, this would usually be administered in the form of a questionnaire. However, the authors point out that self-reporting often carries inherent problems, including that responses after the fact may not match actual feelings at the time. Also, often new questionnaires are not validated (especially if translated), while re-using questionnaires from other fields can be tricky for changes in application. Instead of relying solely on questionnaires, one should make sure to incorporate objective measures collected throughout the experiment: physiological measures and behavioral observation techniques, for example, in addition to performance

measures as appropriate for the task.

Another evaluation concept brought in from the psychology field was Standardization. This concept involves making sure that your results are only due to the experiment at hand, rather than any biases or uncontrolled factors. Thorough documentation is needed to try to avoid this issue, and this should also help with future study replication. Replication of studies was emphasized as well; the authors referenced a 2016 publication from Nature where 1500 researchers were asked to replicate a study. 70% failed to replicate due to insufficient documentation and other issues.

The authors felt this was of importance because if a piece of work is not replicable, there is likely a lack of detailed information about the methodology, leading to limited reusability of the work. Additionally, confirmation bias can come into play, and questionable research practices can be hidden by lack of detail. These issues apply to HRI studies as well as the psychology work specifically referenced, though with further complications such as unavailability of the hardware or software needed. Overuse of Wizard of Oz (WoZ) techniques also cause problems as those inputs are difficult to recreate.

Finally, the last recommendation of the authors was to be modest about one's results. Results are applicable to a specific context and may not necessarily carry over to a general case; i.e., one should critique one's own work in advance of others.

### **Raquel Oliveira, Instituto Universitario de Lisboa, Portugal**

Raquel Oliveira presented a talk titled "Methodologies for Human-Robot Interactions in Groups". The talk was inspired by the large increase in studies on HRI in groups in recent years. However, groups are difficult to study clearly due to the interplay of social interactions that can interact with the experiment at hand. It was pointed out that there are several other limitations, including a lack of validated instrumentation to measure group interactions, a lack of cross-cultural validation, and an over-reliance on measures from questionnaires. Finally, as in general HRI studies, there is often a bias towards significant results. These results are, at times, accepted without being looked at critically and determining whether they are valid.

One of the proposed solutions for future work was a switch towards longitudinal research studies. This could help eliminate anomalous results from one-time interactions with robots. Another suggestion was to examine the field of phenomenological research to perceive how specific groups interact with the robot. Group emotions through face tracking software can also be a useful avenue of study. Finally, consider alternative data analysis techniques, such as multi-level modeling or time-sequence sensitive techniques, that can help model the group as a whole rather than each member.

Some final conclusions included an appeal for more open scientific methods in the community. Additionally, the hope is for the community to be more accepting of the publication of non-significant studies and replication studies in the future. This generated a useful discussion on the feasibility of creating platforms for non-significant publications and how we can encourage the community to catch up and respect these types of research.

#### **4. Panel Discussion**

The panel took place in the afternoon session, with three panelists moderated by Murat Aksu from NIST. The panelists first gave a short overview of their research and then discussed both pre-prepared and audience questions.

##### **Katrin Solveig Lohan**

Katrin Lohan joined the school of Mathematical and Computer Sciences at Heriot-Watt University as an assistant Professor in 2013. She is deputy director of the Robotics Lab. She became SICSA team leader in the Cyber Physical Systems research theme in 2016. She was General Chair for the European Robotics Forum 2017. She is hired under the Global Platform Recruitment for Research Leaders and part of the Edinburgh Centre for robotics. Previously, she was working at the Italian Institute of Technology (IIT) as a Post Doc in the RobotDoc project funded by the Marie Curie Fellowship.

She obtained her Ph.D. in Engineering from Bielefeld University, Germany in 2012, where she was associated with the ITALK project. Her main research interests are in understanding the learning mechanisms between parents and infants, between adults and adults, and between humans and robots in order to create a natural interaction with a robot. Furthermore, she is interested in deep learning of semantic objects, both through vision and speech.

##### **Chung Hyuk Park**

Professor Chung Hyuk Park's Assistive Robotics and Tele-Medicine (ART-Med) Lab at The George Washington University studies the collaborative innovation between human intelligence and robotic technology, integrating machine learning, computer vision, haptics, and telepresence robotics. The current and future research topics are focused on the following three main themes: multi-modal human-robot interaction and robotic assistance for individuals with disabilities or special needs, robotic learning and humanized intelligence, and tele-medical robotic assistance. In the first research topic, his lab studies the impact of multi-modal feedback on the aspects of human-robot interaction and its application in assistive scenarios, such as telepresence for individuals with visual impairments or emotionally and socially interactive robotic systems for children with autism spectrum disorder (ASD). In the second line of research, they focus on computational methodologies of machine learning in the aspect of robotic learning of human behaviors and intelligence. Through the last research goal, they explore novel methodologies for utilizing robotic systems in biomedical applications, from simple care-giving tasks towards intelligent surgical assistance systems.

## **Yunkyung Kim**

Yunkyung Kim is a researcher and operations lead at NASA Johnson Space Center, previously of NASA Ames Research Center. At Ames, she worked on the Astrobe project, which she presented about at the workshop. Dr. Kim was the HRI lead on Astrobe, a flying robot for space applications, and developed its external design as well as a non-verbal interaction mechanism. Previously, Dr. Kim worked at Samsung, and received her doctorate from KAIST in Daejeon, South Korea.

### **4.1 Discussion Questions**

Among others, these are some of the questions discussed by the panelists and audience.

- What sorts of roles should robots take on in society (or avoid)?
- How can we model the interaction of a human being with a robot?
- Can robots lead interactions?
- How do you select populations who could benefit from service robots?
- How do you evaluate the impact of your work on the target user population and the broader society? How does the broader public respond to the robotics systems you develop?
- Are there social metrics applicable to HRI?
- What are the major challenges that the field of HRI aims at, what scientific or technological advances must be made to achieve these grand challenges, and what are the roadblocks to achieving these advances? What are the key research questions that must be addressed to make robots a practical solution for the needs of your target population?
- What are relevant applications and metrics for physical human-robot collaboration?

### **5. Discussion of Standards**

The remainder of the afternoon session was devoted to discussion of standards and how they could be applied to the human-robot interaction community. A brief overview of the concept was given by Shelly Bagchi from NIST. Topics included definitions of important terms, the standards process, a review of related robotics standards, and discussion of the gaps in the current HRI research landscape which could be assisted by standards.

A more in-depth discussion of relevant standards for robotics was given by Seungbin Moon from Sejong University. Dr. Moon has an extensive history working with industrial and service robotics, and participates on several standards bodies in the field. He gave an overview of the umbrella standards organizations in the field, such as the International



Organization for Standardization (ISO), ASTM International, and the Institute of Electrical & Electronics Engineers (IEEE).

Dr. Moon also gave an overview of several committees which he has participated on in the past. These working groups fall under the umbrella of ISO TC 299, originally "Robotics for Manufacturing Environments" and now just "Robotics", a group with 37 participating countries and 18 published standards, with 12 more under review (as of March 2019). An overview of the six working groups was presented, and several sub-groups were reviewed as well. Their topics include vocabulary and safety of robots in particular situations: personal care, industrial, service, and medical. Mention was also made of other robotics standards committees, including IEC TC 59, performance of household appliances, with its working group 16 on performance of household robots.

## 6. Outputs and Documentation

Peer-reviewed submissions by contributing authors were published via IEEE *Xplore*. Authors of high quality submissions have also been asked to contribute extended versions of their papers to an upcoming journal special issue. Late-breaking posters and invited speaker presentations were made publicly available via the workshop website per contributor consent.

Finally, this workshop is intended to be the first in a series of workshops leading toward formalized HRI performance standards. The IEEE Robotics and Automation Society has already been approached for hosting and supporting this standardization effort. Initial workshops will target community and consensus building. A follow-up workshop was held at the 2020 ACM/IEEE International Conference on Human Robot Interaction, and it built upon the discussion items generated from the 2019 workshop. A report detailing the 2020 workshop will follow.

## 7. Future Plans

A second workshop was held at the 2020 HRI conference, and a third is proposed for 2021. Additionally, a new conference track on "Reproducibility of Human-Robot Interaction" was introduced in 2020 and was well-received; future collaborations are intended.

Finally, a special issue of the Transactions on Human-Robot Interaction journal is currently in process and targeted to be published in 2021. The special issue will feature two years of papers from this workshop, as well as independent contributions.

## References

- [1] Gelman A, Loken E (2013) The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University* .

- [2] Kubota A, Iqbal T, Shah JA, Riek LD (2019) Activity recognition in manufacturing: The roles of motion capture and semg+ inertial wearables in detecting fine vs. gross motion. *ICRA IEEE* .
- [3] Moharana S, Panduro AE, Lee HR, Riek LD (2019) Robots for joy, robots for sorrow: Community based robot design for dementia caregivers. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, , pp 458–467. <https://doi.org/10.1109/HRI.2019.8673206>