

Influence of Simulation and Interactivity on Human Perceptions of a Robot During Navigation Tasks

NATHAN TSOI, Yale University, USA

RACHEL STERNECK, Yale University, USA

XUAN ZHAO, Stanford University, USA

MARYNEL VÁZQUEZ, Yale University, USA

In Human-Robot Interaction, researchers typically utilize in-person studies to collect subjective perceptions of a robot. In addition, videos of interactions and interactive simulations (where participants control an avatar that interacts with a robot in a virtual world) have been used to quickly collect human feedback at scale. How would human perceptions of robots compare between these methodologies? To investigate this question, we conducted a 2x2 between-subjects study (N=160), which evaluated the effect of the interaction environment (Real vs. Simulated environment) and participants' interactivity during human-robot encounters (Interactive participation vs. Video observations) on perceptions about a robot (competence, discomfort, social presentation, and social information processing) for the task of navigating in concert with people. We also studied participants' workload across the experimental conditions. Our results revealed a significant difference in the perceptions of the robot between the real environment and the simulated environment. Furthermore, our results showed differences in human perceptions when people watched a video of an encounter versus taking part in the encounter. Finally, we found that simulated interactions and videos of the simulated encounter resulted in a higher workload than real-world encounters and videos thereof. Our results suggest that findings from video and simulation methodologies may not always translate to real-world human-robot interactions. In order to allow practitioners to leverage learnings from this study and future researchers to expand our knowledge in this area, we provide guidelines for weighing the tradeoffs between different methodologies.

CCS Concepts: • **Human-centered computing** → **Laboratory experiments**; • **Computer systems organization** → *Robotics*.

Additional Key Words and Phrases: Human-robot interaction; human perception; robot navigation

ACM Reference Format:

Nathan Tsoi, Rachel Sterneck, Xuan Zhao, and Marynel Vázquez. 2023. Influence of Simulation and Interactivity on Human Perceptions of a Robot During Navigation Tasks. 1, 1 (July 2023), 18 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

Different methodologies have been proposed to investigate human perceptions of robots in Human-Robot Interaction (HRI). Generally, the gold standard is to collect human perceptions through real-world, in-person studies [5]. However, in-person studies may carry with them administrative overhead, e.g., the recruiting of participants (perhaps through

Authors' addresses: Nathan Tsoi, Yale University, New Haven, CT, USA, nathan.tsoi@yale.edu; Rachel Sterneck, Yale University, New Haven, CT, USA, rachel.sterneck@yale.edu; Xuan Zhao, Stanford University, Stanford, CA, USA, xuanzhao@stanford.edu; Marynel Vázquez, Yale University, New Haven, CT, USA, marynel.vazquez@yale.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

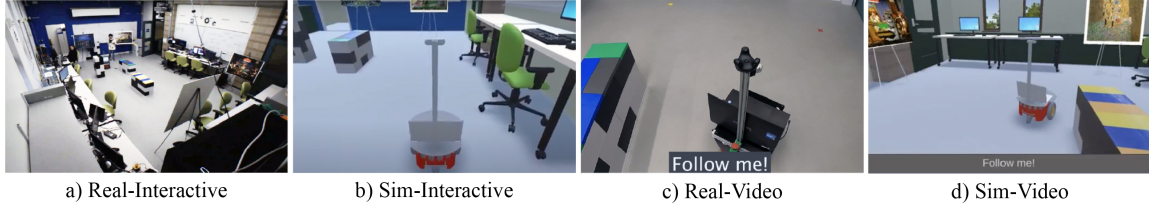


Fig. 1. Experimental conditions of our 2×2 between-subject study. Our independent variables were the interaction environment (Real vs. Simulated environment) and the level of interactivity of the research methodology (Interactive participation vs. Video observation).

flyers, social media or word-of-mouth) and scheduling. Moreover, each participant must travel in order to interact with a researcher in a set physical space. In practice, the need for in-person interaction and the associated administrative overhead could negatively impact the number of participants in an in-person study. Inadvertently, this could limit the sample size and statistical power a study may achieve [23].

An alternative to in-person studies is to record interactions between a human and a robot in videos and then gather human perceptions of the robot using a web survey that includes the recordings. Because of the online nature of the survey, participants can be recruited via online crowdsourcing platforms [26], allowing researchers to scale data collection and accelerate the pace of research. However, video studies are not without limitations. First, video interactions between a human and a robot can lack diversity compared to in-person studies due to the limited number of scenarios used to create videos. Second, participants who observe interactions through the recordings are one step removed from the human-robot interaction. In this case, participants providing the survey responses are not interacting with the robot but, instead, they passively view the robot interacting with another person. Information flow between the robot in the video and the person providing the label is unidirectional, as opposed to bidirectional, which characterizes interactive encounters with technology [6, 55].

Recently, simulations of human-robot interactions have been used instead of in-person or video-based studies in HRI [56, 65, 71]. Modern web infrastructure allows researchers to deploy simulations within online surveys so that online study participants can virtually interact with a robot in a simulator within their web browser and then provide their perceptions of social robots [65]. Due to the virtual nature of this process, simulations have the potential to improve the efficiency and scalability of data collection in HRI while offering a higher level of interactivity than video-based studies. Prior studies have explored how human perceptions of social navigation robots may differ between some methodologies, such as between videos and simulations [65]. Other studies have explored the potential benefits of in-person vs. virtual interactions [1]. Yet, open questions remain on how human perceptions of a mobile robot for social navigation might differ between such methodologies.

We conducted a study that utilized two navigation tasks to investigate human perceptions of a mobile robot along 4 dimensions (competence, discomfort, social presentation, and social information processing). As shown in Figure 1, the study considered two independent variables. One variable concerned the level of interactivity of the research methodology (Interactive participation vs. Video observation). The second variable was the interaction environment (Real vs. Simulated environment), because simulations used in HRI do not always fully mimic the visual appearance of the real world.

Our results suggest that there are subtle tradeoffs that must be considered when choosing the methodology with which one conducts a study. In particular, our results revealed that interaction environment and interactivity can influence human perceptions of robots in HRI studies. Moreover, the task can also influence perceptions of a robot's

performance. While simulations and video studies conducted online are pragmatic for HRI research, our results suggest that user perceptions of robots gathered with these methodologies may not always translate to perceptions from real-world human-robot interactions. In order to allow practitioners to leverage learnings from this study and future researchers to expand our knowledge in this area, we provide guidelines for weighing the tradeoffs between different methodologies in Section 7.

2 RELATED WORK

This section discusses related work in regards to the types of research methodologies considered in our study. First, we discuss video-based evaluations and simulation in Human-Robot Interaction. Then, we discuss related work on robot embodiment and physical presence, which are important aspects of in-person studies.

2.1 Video-Based Evaluation in HRI

Video studies have often been used in HRI to collect data on human perceptions of robots [24, 59], measure human understandability of robot behavior [13, 52], and gather preferences over robot behavior [31, 75]. Videos have also been used to portray recordings of human-robot interactions in a way that seems responsive to human actions [44] and for early robot prototyping [22].

Video recordings of human-robot interactions allow participants to provide feedback regarding their perception of a robot without directly interacting with it. Collecting feedback without in-person interaction is useful when it is infeasible to have a participant interact with the robot due to safety concerns [72] or when there are restrictions imposed by infectious disease outbreaks [15], which can limit access to research materials and robots.

While in-person studies require experimenters to find local participants (e.g., using flyers or word-of-mouth), online video studies can leverage crowdsourcing platforms (such as Prolific or Amazon Mechanical Turk) to reduce recruitment bottlenecks. Furthermore, crowdsourcing can enlarge the participant pool beyond a researcher's immediate geographic location, allowing for cross-cultural studies (e.g., [11, 27, 41]). Finally, once a study is posted online, crowdsourcing also allows the scaling of HRI research by enabling many participants to view videos of interactions and provide their feedback in parallel. However, because it is impossible to fully control the environment in which the video-based study is administered in these cases, there could be biases in the data collection. For example, bias could be introduced due to the screen size used by participants [67]. Nevertheless, because crowdsourcing has gained significant popularity in HRI (e.g., [3, 13, 29, 32, 49, 59, 62, 65]), we also used it in our study about human perceptions of a mobile robot.

2.2 Simulation in HRI

In HRI studies, simulations have been used to investigate interactions between participants and robots who engage in two-way flow of information, which is not present in videos. Early HRI simulators focused on providing graphical user interfaces for robot development and testing. For example, USARSim supported human-robot interaction research in the context of robot teleoperation [36]. Chernova et al. created an online multiplayer game that simulated human-robot interactions for learning interactive robot behavior [9]. Other robotics simulators allowed users to teleoperate human avatars to enable virtual interactions with robots. For instance, the Modular OpenRobots Simulation Engine (MORSE) [14] was integrated with human avatars to allow for virtual experimentation [35]. Also, the Social Environment for Autonomous Navigation 2.0 (SEAN 2.0) [66] integrated the Unity game engine with the Robot Operating System (ROS) to make it possible to train and evaluate social robot navigation policies.

A common limitation of simulation is the lack of visual realism. Rich-client simulations such as MORSE and SEAN 2.0 have partly addressed this limitation, but they typically require a powerful computer with a dedicated Graphics Processing Unit (GPU) to render the virtual world. Web technologies, such as SEAN-EP [65], have been used to increase accessibility to rich-client simulations by allowing a participant to interact with a robot in a simulated environment using a standard web browser. We used SEAN-EP in our study so that participants did not need to install simulation software locally or have a dedicated GPU.

One might naturally assume that more visual realism, via higher-fidelity simulations, is always better than less visual realism. Surprisingly, Truong et al. [63] found that lower fidelity simulations resulted in better sim-to-real transfer of robot navigation behavior. This result inspired us to compare human perceptions of a robot where visual realism can differ based on the interaction environment in which humans observe human-robot interactions. In our work, these observations were obtained in fully realistic environments (showing real-world interactions in a lab), or they were obtained in a simulation of the lab environment.

Close to our work, Tsoi et al. [65] examined differences in human perceptions of a Kuri robot in two setups: participants either interacted with the robot in SEAN [64], or they observed videos of human-robot interactions in the simulation. They found that, for navigation tasks, a robot viewed in a video was perceived as more competent than one experienced interactively in SEAN. Additionally, participants in the interactive simulation condition reported less mental demand than participants in the video condition. However, no comparison was made with respect to real-world interactions, as in our study.

2.3 Physical Robot Embodiment and Presence

One important difference between in-person studies and both video and simulation methods is robot embodiment and presence. These concepts are related but capture different aspects of the interaction [42]. Robot embodiment describes the morphology and visual characteristics of a robot, which can differ between the real world and virtual environments. Type of presence describes where a robot is located, and thereby can influence the medium over which the same robot is experienced (typically in-person, via teleconference, or in a one-way video). There has been much interest in how perceptions of robots are influenced by robot embodiment and presence, but results are inconsistent.

Robot embodiment can influence human perceptions of a robot and human-robot interactions [10, 12, 16, 37, 58, 68, 70]. Robot embodiment is not a binary concept, but exists on a spectrum [16] ranging from disembodied agents which communicate only over text or speech [10, 70], to agents simulated on a screen using a 2-dimensional interface or avatar [12], to agents modeled in a 3-dimensional simulation [36, 64, 66], to agents that exist with a physical presence in the real world. For example, Strait et al. [58] studied the effects of direct versus indirect speech on humans for an advice-giving robot where relevant factors in the study included robot appearance and robot presence. In another study, Wainer et al. [68] compared human perceptions of a co-located physical robot, a remotely located (telepresent) robot, and a simulated robot that explained and supervised a Towers of Hanoi puzzle. The study results suggested that physically embodied co-located interactions are more enjoyable than interactions with remote-located and simulated robots.

Research suggests that human behavior and human perception of robots can be influenced by robots' presence, although results vary in the literature. For example, Jung and Lee [28] and Lee et al. [34] found that the physical presence of a robot can influence its perceived social presence; however, Thellman et al. [61] found that the perceived social presence of a robot was not influenced significantly by its physical or virtual presence [61]. Other examples are found in Bainbridge et al. [1] and Salomons et al. [54], who compared physically present robots with a live video stream

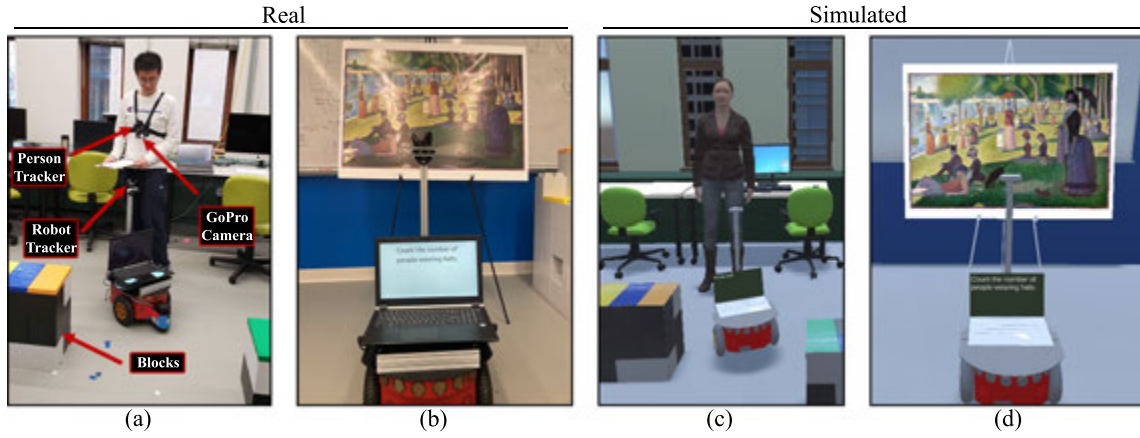


Fig. 2. Photos of the Real (a and b) and Simulated (c and d) environments. The Interactivity level manipulated how the participant interacted with each of the environments. A participant in the Real-Interactive condition (a) wore a chest harness with trackers for localization and a GoPro camera while interacting with the robot in the real world. A participant in the Sim-Interactive condition (c) used keyboard controls to control an avatar through the virtual lab. Participants in the Video conditions watched video recordings of the interactive participants. During the art task, the robot guided a participant to a poster and communicated with the participant using text on the real (b) or simulated (d) laptop screen.

of robots on a book-moving task and an exercise task, respectively. These studies found that people were more likely to fulfill an unusual request by the robot, afforded greater personal space to it, and made fewer exercise mistakes when it was physically present. But in social robot navigation, Woods et al. [72] found that perceptions of a robot approaching people were consistent between video and real-world settings. Our study further expands this line of work on the effects of presence on human perceptions of robots.

3 METHOD

Prior work on human perceptions of robots in video, simulation, and in-person studies has been largely fragmented by the research methodologies. To more comprehensively understand how human perceptions vary between these methodologies, we conducted a 2x2 between-subjects study with a mobile robot in a laboratory setting. The two independent factors of our study were: *Interaction Environment* (Real vs. Simulated environment), and the level of *Interactivity* of the research methodology (Interactive participation vs. Video observation). Photos of all experimental conditions are shown in Figure 1. The difference between Real and Simulated interactions is shown in Figure 2. To the best of our knowledge, our study, which utilized two navigation tasks, is the first to compare human perceptions of robots obtained in real-world interactions with perceptions obtained from interactive simulations, where humans control a virtual avatar. We compared these human perceptions of a robot in real-world interactions and interactive simulations with perceptions of the robot after viewing a video recording. Our study protocol was approved by our Institutional Review Board.

3.1 Hypotheses

As shown in Fig. 1, our two independent variables led to four conditions: Real-Interactive, Real-Video, Sim-Interactive, and Sim-Video. We studied whether these conditions had an effect on four aspects of human perceptions of the robot: Competence [17]; Discomfort [8]; Social Presentation, or “the robot’s ability to appear to be a desirable social partner” [4];

and Social Information Processing, which captures social intelligence [4]. We also studied the effect of interactivity on perceived workload [19]. These measures are common in the Human-Robot Interaction literature [18, 30, 33, 47, 57].

Our first set of hypotheses focused on the idea that human perceptions of a mobile robot in the Real environment would differ from perceptions of the robot in the Simulated environment. These hypotheses were motivated by prior work that suggests that peoples' perception of a robot can vary between simulation and real-world interactions (e.g., [38, 65, 69]). In particular, Tsoi et al. [65] provided evidence that human perceptions of robots collected via video studies and compared to those collected using interactive, online simulations could differ, but did not compare them to observations obtained in real-world human-robot interactions. More specifically:

H1. Human perceptions of the robot's competence (**H1a**), discomfort (**H1b**), social presentation (**H1c**), and social information processing (**H1d**) in the Real environment will differ from the Simulated environment.

Our second set of hypotheses tested the potential difference in human perception of a mobile robot between a participant interacting with a robot compared to a participant viewing an interaction with another person in a video. This hypothesis is motivated by the common use of videos in HRI studies, and the growing use of interactive simulations as a potential replacement [56, 65, 71]. Prior work suggests that people may perceive a robot more positively when physically present [37] and that people may be influenced by co-present robots (e.g., [1, 21]).

H2. Human perceptions of the robot's competence (**H2a**), discomfort (**H2b**), social presentation (**H2c**), and social information processing (**H2d**) will differ between interactive conditions (Sim-Interactive and Real-Interactive) and video-based conditions (Sim-Video and Real-Video).

Our third set of hypotheses considered data from the Real-Interactive condition as the gold standard for gathering human perceptions of robots. Then, because video observations lack interactivity in comparison to interactive simulations, we suspected that human perceptions collected with the Sim-Video and Real-Video conditions would be less similar to those obtained in the real-world than the perceptions obtained with the Sim-Interactive condition.

H3. Human perceptions of the robot's competence (**H3a**), discomfort (**H3b**), social presentation (**H3c**), and social information processing (**H3d**) in video-based conditions (Sim-Video and Real-Video) are more similar to the Sim-interactive condition than to the Real-Interactive condition.

Our fourth and final hypothesis is motivated by prior work that associates embodied and interactive experiences with low workload. For example, Wang et al. [70] found that robot agent embodiment resulted in lower perceived workloads during interaction with robotic agents compared to voice-only agents. Tsoi et al. [65] found partial support for lower perceived workload when completing an HRI survey that involved providing perceptions of a robot in interactive interactions compared to a survey that involved providing perceptions based on video observations

H4. The Interactive conditions will lead to a lower perceived workload by participants than the Video conditions.

3.2 Participants

In total, we recruited 213 participants for our study. For the Real-Interactive condition, participants were recruited via flyers and word of mouth. Participants for all other conditions were recruited online using the Prolific crowdsourcing platform.

All the participants were at least 18 years old, had normal or corrected-to-normal vision, and were fluent in English. The participants in the Real-Interactive condition were required to be able to walk comfortably and stand for the duration of the study (20-30 minutes). Participants in the online portion of the study were limited to those on non-mobile

devices, such as laptops and desktop computers to ensure a reasonable screen size on their device and the ability to control the virtual avatar in simulation using a physical keyboard.

We excluded 53 participants from analyses because 35 participants in an Interactive condition had incomplete video recordings due to technical issues or had incomplete surveys, 14 participants had other technical issues or did not follow directions, and 4 accidentally participated in the Sim-Video condition after participating in the Sim-Interactive condition.

Among the final 160 participants (40 per condition), 90 participants identified as male, 66 as female, 2 as non-binary, 1 as genderqueer, and 1 declined to state their gender. Additionally, 32 participants were between ages 25-34, 50 were between ages 35-44, 40 were between ages 45-54, 23 were between ages 55-64, 13 were between ages 65-74, and 2 were between ages 75-84. On average, the participants indicated neutral familiarity with robots on a 7-point scale ($M = 3.91, SE = 0.13$). The online participants had an average Internet speed of 163.46 Mbps ($SE = 15.86$), which was in line with prior use of SEAN-EP [65].

3.3 Setup

For the Real-Interactive condition, the experiment was conducted in a laboratory room on a university campus in the United States. The room contained physical obstacles consisting of EverBlock construction blocks, as shown in Figures 1(a) and 2(a). There were also four distinct pieces of artwork on easel stands positioned in the corners of the room. A close-up photo of one of the pieces of artwork in the real laboratory environment is shown in Fig. 2(b).

We designed our study such that a robot, controlled by the ROS Navigation Stack with Social Cost Layers [39], autonomously navigated near the participant to jointly complete two tasks: the *Follow Task* and the *Art Task*. The Follow Task was designed to place the participant's focus on the robot throughout the interaction. Follow tasks are typical for robots that serve as tour guides and have been investigated in the past in social navigation [7, 43, 45, 53]. Meanwhile, we designed the Art Task to allow participants to observe the robot's movement during a more dynamic and complex navigation task. These tasks are further described in the next section. Importantly, the robot that we used in the study was a Pioneer 3-DX on which we affixed a laptop, oriented with the screen pointing forward, to allow for robot communication with the participant. We also attached a depth sensor and localization beacon to the robot.

The participants in the Real-Interactive condition wore a GoPro camera on their chest (as in Figure 2(a)) to record videos from a first-person perspective while completing study activities. HTC Vive Trackers were used to localize the robot and the participants. Also, the participants used a custom web application on a mobile phone, which we provided, to do task-specific actions. This included pressing a button on the phone to begin each task and recording their answers to survey questions. The web application was also used to display text on the robot's laptop.

For the Sim-Interactive condition, we modeled the laboratory room used for the Real-Interactive condition as well as the Pioneer robot using the Unity game engine and SEAN 2.0 [66]. Figures 1(b), 1(d), 2(c) and 2(d) illustrate the virtual world that we created for the study. In addition, we used SEAN-EP [65] to embed our simulation in a Qualtrics web survey, which gathered participants' demographics data and all other relevant measures regarding their experience of virtual human-robot interactions. The participants used their keyboards to control a virtual avatar in the SEAN simulations and to complete the same activities as in the Real-Interactive condition.

For the Real-Video and Sim-Video conditions, we used recordings of participants' interactions with the robot in the real-world lab and the virtual re-creation, respectively. A GoPro camera worn by participants in the Real-Interactive condition (as in Fig. 2(a)) was used to record the interactions that were observed by participants in the Real-Video condition. For the Sim-Video condition, we used SEAN 2.0 to save video recordings of the human-robot interactions that

happened under the Sim-Interactive condition. The recordings were made from the perspective of the virtual avatar that was controlled by a human in SEAN. In order to ensure participants in the Video condition were able to understand what the robot was communicating, we added captions to all videos which displayed the same text that was shown on the robot’s laptop screen. We did not use audio in the simulation or the videos due to the difficulty of generating realistic audio. An example of the captions is provided in Figures 1(c) and 1(d). The videos were then embedded in a Qualtrics survey like the one used for the Real-Interactive condition.

3.4 Procedure

At the beginning of the study, the participant provided demographic information (as in Section 3.2). Then, the participant continued on to complete the study’s four phases: 1) Introduction, 2) Follow Task, 3) Art Task, and 4) Closing. In each task, the participant was specifically asked to pay attention to how the robot moved.

Phase 1: Introduction. In the Real-Interactive condition, the participant was introduced to the robot by an experimenter who told them that they would interact with the robot through a series of tasks. Then, the experimenter assisted the person as they put on the GoPro chest harness to record their activities during the study. In the Sim-Interactive condition, the participant completed a walk-through tutorial that showed them the virtual Pioneer robot and their randomly assigned avatar. The walk-through then explained how to navigate the simulated lab. In the Real-Video and Sim-Video conditions, the participant was given text instructions indicating that they would watch videos of a person or avatar interacting with a robot. The participant was also shown an image of the robot to familiarize the person with the Pioneer 3-DX platform.

Phase 2: Follow Task. In the Real-Interactive condition, the participant was instructed to move to a specific marker on the floor and then press a button on the mobile device to begin the follow task. Then, the participant followed the robot along a pre-defined path, which was composed of four segments.

The path involved navigating around EverBlock construction blocks placed throughout the room, as shown in Figures 2(a) and 2(c).

After following the robot along each of the four path segments, the participant answered survey questions about their impression of the robot. In the Sim-Interactive condition, the participant completed the same task but in a SEAN simulation.

For the Real-Video and Sim-Video conditions, we paired each participant with a study session that involved Real-Interactive and Sim-Interactive participation, respectively. Then, the videos of the Follow Task from the Interactive sessions were shown to the participants in the Video conditions. In this manner, a participant in Real-Video and Sim-Video conditions was able to watch recordings of the task and answer survey questions about their impression of the robot in the videos as in the Interactive conditions.

Phase 3: Art Task. In the Real-Interactive condition, the participant was told that there had been an art heist in the lab and some of the art had been replaced with fakes. The participant and the robot were tasked with collecting information about the four art pieces in the laboratory to help the experimenters figure out which were real and which were fake. Fig. 2(b) displays one of the art pieces in the real world and Fig. 2(d) shows it in simulation. For each of the four art pieces, a participant performed the following steps:

- (1) The participant was directed to find the robot.

- (2) Once the person found the robot, a text message was displayed on the robot’s computer screen which instructed them to follow it.
- (3) The robot then led the participant to a piece of artwork.
- (4) The participant was instructed via text on the robot’s computer screen to count the number of a given object shown in the art piece.
- (5) After instructing, the robot moved away to a different location and waited for the participant to complete the object counting.
- (6) The participant provided their answer to the counting request using the mobile device and was directed to find the robot again to repeat the process for the next art piece.

The Art Task was designed so that the person and the robot would engage in more dynamic interactions than in the Follow Task. In this case, while the person was counting objects in an art piece, the robot moved far from the participant and waited until they completed counting the objects in the picture. Only when the participant started moving away from the picture, did the robot start to move back towards the person. Then, both the robot and participant moved towards each other and soon thereafter engaged in face-to-face or side-by-side spatial formations (e.g., as in [25, 74]). In the Real-Video and Sim conditions, the description of the Art Task was provided in text before the participant began the task.

Also, in the Sim-Interactive condition, the participant used an interface which we implemented in the simulation to record their responses to the counting request by the robot. Meanwhile, in the Video conditions, the participant recorded their answers using the Qualtrics web survey. This survey included videos from Interactive conditions using the same participant-session pairing explained for the Follow Task.

Phase 4: Closing. Finally, the participant provided their impressions of their perceived workload for the tasks in the study.

In-person participants in the Real-Interactive condition were paid \$15.00 USD per hour rounded to the nearest 10-minute increment.

Participants in all other conditions completed the study online using Prolific. They were paid \$5.00 USD as we estimated the online study sessions to take 20 min.

3.5 Dependent Measures

We measured 2 aspects of participants’ experience during our study using widely adopted survey measures in HRI:

Human Perceptions of the Robot. We measured four aspects of human perceptions of the robot: 1) Competence, 2) Discomfort, 3) Social Presentation, and 4) Social Information Processing. The first two aspects were measured using the Robot Social Attributes Scale (RoSAS) [8], which includes robot Competence and Discomfort factors. The items were answered in relation to how the robot moved during the tasks. Ratings for the Competence and Discomfort scales were gathered on 7-point responding format ranging from 1 (Definitely Not Associated) with the robot to 7 (Definitely Associated), which was the same as the original RoSAS responding format.

Robot Social Presentation and Social Information Processing were measured using the short-form of the Perceived Social Intelligence (PSI) questionnaire [4]. The Social Presentation scale had a total of 7 items, all of which began with “This robot...” and ended with statements such as “enjoys meeting people,” and “cares about others.” The Social

Information Processing scale had a total of 13 items, which started with “This robot...” and ended with statements like “responds appropriately to human emotion” or “can figure out what people think.” Ratings for PSI statements were gathered on a 5-point responding format ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), which was the same as the original PSI responding format.

For each scale, we aggregated responses across items to calculate a composite measure after confirming high internal reliability. The Cronbach’s α values were 0.90 for Competence, 0.76 for Discomfort, 0.76 for Social Presentation, and 0.94 for Social Information Processing. The Cronbach’s α value for each aspect we measured was within the 0.7 to 0.95 acceptable value range [60].

Perceived Workload. We used items from the NASA Task Load Index (TLX) [19] to assess the perceived workload for the Follow and Art Tasks. Perceptions of Mental Demand, Physical Demand, Temporal Demand, Effort, and Frustration were gathered on a 7-point responding format from 1 (lowest) to 7 (highest). The 7-point responding format was used for consistency in the responding format with the other scales. The 7-point format was chosen over the 5-point format because responding formats with 6 or more categories have been shown to correlate better[51]. Example survey items included “How mentally demanding were the tasks?” (Mental Demand) and “How insecure, discouraged, irritated, stressed, and annoyed were you?” (Frustration). The Cronbach’s α for the NASA TLX survey items was 0.75, which is within the 0.7 to 0.95 range of acceptable values [60].

3.6 Analysis

We analyzed the results by task (Follow and Art) in two ways. First, we fitted linear mixed-effect models for all dependent measures with fixed effects for Interaction Environment (Real or Simulation) and Interactivity (Interactive participation or Video observation). We also assigned a unique identifier, Session ID, to each Interactive study session, which was added as a random effect in our linear model. A linear mixed-effect model was used due to the hierarchical nature of the data, i.e., Participant ID was nested within Session ID. This allowed us to associate the experience in the Interactive conditions, from which we made videos of human-robot interactions, with the corresponding data in the Video conditions. Unless otherwise noted, we used the Restricted Maximum Likelihood (REML) method for model estimation [48]. A linear mixed model was used for model estimation instead of ANOVA because of the nested nature of the data, i.e., Participant ID was nested within Session ID. Nesting was necessary because the video-condition stimuli was generated from a recording of the Interactive condition, which resulted in the interactive data and corresponding video recordings being paired. Note that within the paired data, the participant that interacted with the robot (either in the Real environment or simulation) was not the same as the participant who watched the video, so a unique Participant ID was used to identify all participants. Second, because H3 considered the Real-Interactive condition as the methodology that provides gold-standard results, we performed treatment contrasts between the Real-Interactive condition and all other conditions.

4 RESULTS

4.1 Perceptions of the Robot

4.1.1 Competence. The linear mixed model analysis per task revealed significant effects. In particular, for the Follow Task, we found Interaction Environment to have a significant effect on Competence, $F(1, 156) = 4.30, p = 0.04$. The effect size, as measured by Cohen’s d , was $d = 0.16$, indicating a very small effect. A post-hoc t -test showed that

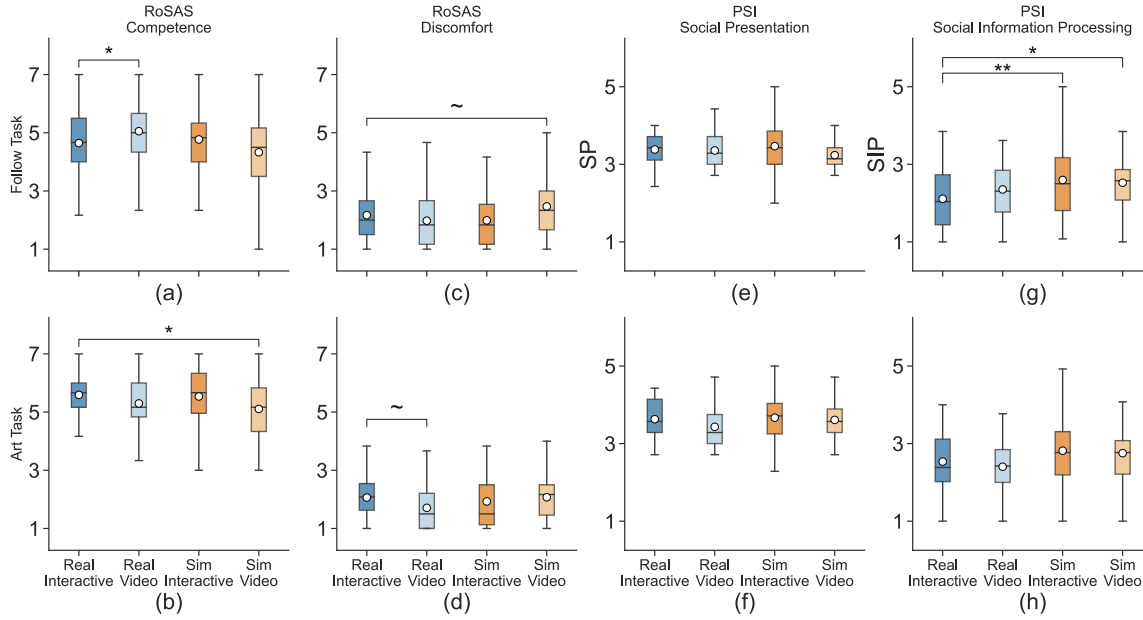


Fig. 3. Contrast results for RoSAS Competence (a,b), RoSAS Discomfort (c,d), PSI Social Presentation (e,f), and PSI Social Information Processing (g,h) by task. Box plots span the first to third quartile, a dark grey horizontal line through the box indicates the median, and a white circle indicates the mean. Box plot whiskers extend to ± 1.5 times the Interquartile Range. The \sim indicates $p < 0.10$, * indicates $p < 0.05$, and ** indicates $p < 0.001$.

people perceived the robot to be significantly more competent in the Real condition ($M = 4.85, SE = 0.06$) than in the Simulated condition ($M = 4.55, SE = 0.07$). The linear mixed model analysis on the Art Task showed that only Interactivity had a significant effect on Competence, $F(1, 156) = 5.39, p = 0.022$. The effect size, as measured by Cohen's d , was $d = 0.18$, indicating a very small effect. A post-hoc t-test indicated that competence ratings were significantly higher for Interactive participation ($M = 5.56, SE = 0.11$) than for Video observation ($M = 5.20, SE = 0.11$).

Comparing the Real-Interactive condition as the baseline condition against three other conditions with treatment contrasts revealed that the Real-Video condition significantly differed from the Real-Interactive condition in the Follow Task, $F(1, 156) = 3.94, p = 0.05$. The effect size, as measured by Cohen's d , was $d = 0.22$, indicating a small effect. Specifically, compared to interacting with the robot in the real world ($M = 4.65, SE = 0.09$), participants watching videos of the robot interacting with someone else in the real world perceived the robot to be even more competent ($M = 5.05, SE = 0.08$). For the Art Task, only the Sim-Video condition was significantly different from the Real-Interactive condition, $F(1, 156) = 4.79, p = 0.03$. The effect size, as measured by Cohen's d , was $d = 0.24$, indicating a small effect. This suggests that compared to watching a video of a person interacting with the robot in simulation ($M = 5.11, SE = 0.16$), participants that interacted with the robot in the real world viewed it to be even more competent ($M = 5.59, SE = 0.14$). These results are shown in Figures 3(a) and 3(b).

4.1.2 Discomfort. The linear mixed model analyses on both tasks resulted in no significant main effects on discomfort.

The contrast analyses for the Discomfort responses in the Follow and Art Tasks led to no significant differences. However, the discomfort ratings in the Sim-Video condition were marginally different from the Real-Interactive ratings

in the Follow Task, $F(1, 156) = 3.57, p = 0.06$. The effect size, as measured by Cohen's d , was $d = 0.21$, indicating a small effect. This indicates that compared to watching a video of a simulation ($M = 2.47, SE = 0.08$), participants that interact with a robot in the real world may view the robot as less discomforting ($M = 2.17, SE = 0.07$). Additionally, discomfort in the Real-Video condition was marginally different from the Real-Interactive condition in the Art Task, $F(1, 156) = 3.48, p = 0.06$. The effect size, as measured by Cohen's d , was $d = 0.21$, indicating a small effect. This indicates that compared to interacting with a robot in the real world ($M = 2.06, SE = 0.13$), participants that watch a video of the real-world robot interacting with another participant may view the robot as less discomforting ($M = 1.71, SE = 0.13$). These results are shown in Figures 3(c) and 3(d).

4.1.3 Social Presentation. The linear mixed model analyses and the treatment contrasts per task showed no significant effects on Social Presentation ratings. In general, most ratings were neutral in the Follow Task and slightly positive in the Art Task, as shown in Figures 3(e) and 3(f). The slight increase in Social Presentation perceptions for the Art Task was expected because the task involved more complex interactions than the Follow Task, as indicated in Sec. 3.4.

4.1.4 Social Information Processing. The linear mixed model analysis on Social Information Processing for the Follow Task revealed a significant main effect of Interaction Environment on the ratings, $F(1, 157) = 6.71, p = 0.01$. The effect size, as measured by Cohen's d , was $d = 0.41$, indicating a small effect. A post-hoc t -test indicated that people perceived the robot as better able to process social information in the Simulated condition ($M = 2.56, SE = 0.09$) than in the Real condition ($M = 2.23, SE = 0.09$). The linear mixed model analysis for the Art Task also indicated that Interaction Environment had a significant effect on Social Information Processing, $F(1, 157) = 5.02, p = 0.03$. The effect size, as measured by Cohen's d , was $d = 0.35$, indicating a small effect. The post-hoc test indicated that ratings were higher for the Simulated environment ($M = 2.79, SE = 0.10$) than for the Real environment ($M = 2.47, SE = 0.09$).

The contrast analyses on the Follow task indicated a significant difference in Social Information Processing ratings between the Sim-Interactive and Real-Interactive conditions, $F(1, 156) = 7.29, p = 0.008$, as well as between the Sim-Video and Real-Interactive conditions, $F(1, 156) = 5.31, p = 0.02$. The effect sizes, as measured by Cohen's d , were $d = 0.60$ and $d = 0.52$, respectively, indicating a medium effect for both contrasts. This suggests that compared to interacting with the robot in the real world ($M = 2.11, SE = 0.12$), participants viewed the robot as more capable of processing social information when interacting with it in simulation ($M = 2.60, SE = 0.15$) and when viewing it in a video in simulation ($M = 2.53, SE = 0.11$). These results are shown in Fig. 3(g). For the Art Task, the contrast analyses showed no significant differences on Social Information Processing with respect to Real-Interactive. The results for the Art Task are shown in Fig. 3(h).

4.2 Perceived Workload

We analyzed the perceived workload with linear mixed model analyses that included Interaction Environment (Real or Simulation), Interactivity (Interactive participation or Video observation) and their interaction as main effects. Also, we added Session ID as a random effect. In the case of workload, we did not perform contrast analyses as in Sec. 4.1 because H4 did not consider the Real-Interactive condition as a specific baseline for comparison.

The average ratings for Physical Demand and Temporal Demand were $1.48(SE = 0.07)$ and $1.76(SE = 0.08)$, respectively. We found no significant effects on these measures.

Interaction Environment had a significant effect on Mental Demand ($F(1, 156) = 8.60, p = 0.004$), Effort ($F(1, 156) = 6.94, p = 0.009$) and Frustration ($F(1, 156) = 5.77, p = 0.017$). The effect sizes, as measured by Cohen's d , were Mental Demand $d = 0.46$, Effort $d = 0.42$, and Frustration $d = 0.38$, indicating small effects. The post-hoc t -test on Mental

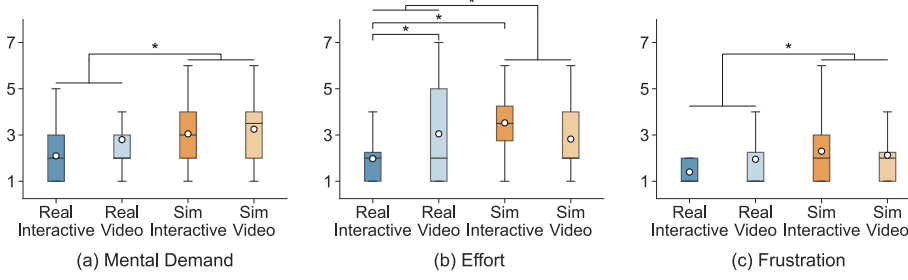


Fig. 4. Perceptions of Mental Demand, Effort, and Frustration by condition: Real-Interactive, Real-Video, Sim-Interactive, and Sim-Video. Box plots span the first to third quartile, a dark grey horizontal line through the box indicates the median, and a white circle indicates the mean. Box plot whiskers extend to ± 1.5 times the Interquartile Range. The * symbol indicates $p < 0.05$.

Demand indicated that participants provided higher ratings in the Simulated environment ($M = 3.15, SE = 0.16$) than in the Real environment ($M = 2.45, SE = 0.18$). The distribution of Mental Demand ratings is shown in Fig. 4(a). Likewise, in the case of Effort, the post-hoc test showed that the ratings in the Simulated environment ($M = 3.18, SE = 0.18$) were significantly higher than those in the Real environment ($M = 2.51, SE = 0.19$), as shown in Fig. 4(b). Finally, the post-hoc test for Frustration revealed that participants felt more “insecure, discouraged, irritated, stressed and annoyed” with the Simulated environment ($M = 2.21, SE = 0.17$) than with the Real environment ($M = 1.68, SE = 0.15$). Figure 4(c) shows the distribution of results for Frustration.

Interactivity had no significant effect on Mental Demand or Frustration; however, we found an interaction effect between Interaction Environment and Interactivity on Effort, $F(1, 156) = 12.45, p < 0.001, R^2_{Adjusted} = 0.10$. A post-hoc Tukey HSD test indicated that the Effort for the Real-Interactive condition ($M = 1.98, SE = 0.17$) was significantly lower than for Real-Video ($M = 3.05, SE = 0.32$) and Sim-Interactive ($M = 3.53, SE = 0.26$).

5 DISCUSSION

In our first set of hypotheses, our results indicated some support. Results showed a significant difference between perceptions of the robot in simulation compared to the real environment. In particular, we found higher Competence ratings (H1a) for the robot in the real laboratory environment than in simulation, although the effect was small. We suspect the difference was due to the greater level of visual realism exhibited by the real robot [69]. Also, we found that the real robot was perceived as less capable of processing social information than the simulated robot (H1d). Social information processing (SIP) refers to the robot’s ability to perceive the social behaviors, emotional states (including desires), and cognitions (including beliefs) of nearby people [4]. The effect for SIP was larger than the effect for Competence, but still small. It could be that human perceptions about the robot’s social information processing abilities were influenced by their virtual avatar in the simulations, which behaved in a much simpler way than people could in the real laboratory environment and looked less realistic as well.

We found evidence for some of our second set of hypotheses, which posited that human perceptions of the robot will differ between Interactive participation and Video observations. In particular, for the Art Task, participants viewed the robot as more competent with Interactive participation than when human-robot interactions were observed in Videos. Although the effect size was small, our results were surprising because they did not align with the results by Tsoi et al. [65], who compared human perceptions of the competence (H2a) of a Kuri robot in interactive SEAN simulations and in videos of the simulation. Beyond the fact that Tsoi et al. [65] did not consider real-world interactions, we believe that

the inconsistency in findings could be due to three reasons: 1) the laboratory environment used in our work had more obstacles and fewer people than the one used in [65]; 2) we used a Pioneer robot which could set different initial human expectations than the Kuri robot used in [65]; and 3) the Art Task was more complex than the Follow Task, and [65] only studied situations where participants followed the robot. Future work should investigate which factors specifically affect human perceptions of the competence of a robot between HRI studies involving Interactive participation and Video observation.

As to our third set of hypotheses, we obtained some evidence that human perceptions of the robot in the Video conditions are more dissimilar to the Real-Interactive condition than those in the Sim-Interactive condition. For example, contrast analyses indicated that robot competence (H3a) was significantly different between the Real-Interactive condition and the Real-Video conditions (for the Follow Task) and between the Real-Interactive and Sim-Video conditions (for the Art Task). No significant differences were found for competence between Real-Interactive and Sim-Interactive conditions. In terms of discomfort (H3b), we found trends that suggested similar differences but for the opposite task – compare Figures 3(a) with 3(c), and Figures 3(b) with 3(d). Again, no significant differences were found for discomfort between Real-Interactive and Sim-Interactive. However, for social information processing (H3d), Real-Interactive led to significantly different results than both Sim-Video and Sim-Interactive. This last result was unexpected and not in line with our hypothesis. Overall, the main take-away from these results is that perceptions of robots gathered through video observation and interactive simulation studies may not always translate to real-world interactions.

Finally, we found only a small amount of evidence in support of our last hypothesis, which stated that cognitive load would be lower for Interactive participation than Video observations. More specifically, only perceived effort was significantly lower for the Real-Interactive condition than for the Real-Video condition. Interestingly, most of our results in regard to workload were instead about differences between the Real and Simulated environments, including differences for mental demand, effort, and frustration. We thought that this result could be due to the fidelity of our SEAN 2.0 simulations. Although SEAN 2.0 generates the renderings through Unity and there is potential to make these simulations photo-realistic, our virtual laboratory environment looked much simpler than the real-world lab (as can be seen in Figure 1 and Figure 2). For example, while humans are adept at identifying coherent concepts from the visual clutter typically found in the real world [46], increased participant effort may be necessary to interpret and interact with the robot in the simulation environment which contains a distribution of visual clutter different from the real world. In the future, exploring how environmental clutter affects human perceptions of robots in HRI could be an interesting avenue of research, for example, by comparing with experiments in simulation that incorporate real-world clutter [73]. Another factor to consider is the usability and computing experience of the different systems implemented for each condition, which may have also had an impact on participant workload. Overall, this is a first step towards a better understanding of how different methodologies can influence the perceptions of mobile robots for social navigation. We hope future HRI studies can explore this direction on a larger scale.

6 LIMITATIONS

First, we conducted our study with only one simulation environment (SEAN 2.0 [65]). It would be interesting to verify in the future if our results hold with other types of simulators, e.g., built using other game engines like Unreal [40] or with lower-fidelity like Gazebo [50]. Second, as with all simulations, our simulated environment and the videos thereof were not perfect replicas of the real world. In the future, it would be interesting to investigate the impact of factors such as the lack of audio in simulation, which could have influenced perceptions of the robot in the Sim and Video conditions, the size and the resolution of the display or Head Mounted Display, and properties of the randomly assigned virtual avatars,

such as gender, which may not match that of the participant. Third, we focused on investigating people's perceptions of robots using subjective responses to well-established questionnaires. However, future research could benefit from including behavioral outcomes, like proxemics measures [20], when comparing research methodologies for social robot navigation. When evaluating results for other tasks, perhaps other behavioral measures like teamwork efficiency [2], could be used instead. Lastly, it would be interesting to investigate to what extent the crowdsourcing setup that we used to gather data in three experimental conditions affected our results. In particular, one could imagine replicating our study in the future with 100% in-person participants, such that no participant is subject to the distractions and technical challenges that often arise with remote participation through crowdsourcing [67].

7 GUIDELINES FOR METHODOLOGY SELECTION

The choice of methodology is one of the many considerations that a researcher must evaluate when approaching new experimental questions in HRI. The primary considerations are time and cost. Ideally, minimal time is required to set up and complete the study while minimizing the cost. Although in-person user studies are the gold standard, often video studies are used. Video studies can allow crowd-sourcing of user feedback which scales quickly, but the quality of responses can vary if participants are not engaged with or focused on the video. With recent technological advancements, interactive simulations can now scale with the use of crowd sourcing [65], they can encourage a participant to remain engaged with the task or detect if the person is not engaged. Other considerations include the availability of a real robot, the safety of the task experienced via different methodologies, and the quality of the simulation along the dimensions of importance. Perhaps in the future, we may have widely available, photo-realistic, real-time, interactive simulations that will decrease the gap between methodologies. However, until this is the case researchers should carefully consider the tradeoffs.

8 CONCLUSION

We investigated how people perceived the competence, discomfort, social presentation, and social information processing of a mobile robot during two navigation tasks. Our study compared methodologies with different Interaction Environments (Real vs. Simulated) and Interactivity (Interactive participation vs. Video observations). We found significant differences in human perceptions of a mobile robot when an interaction was experienced in the real world compared to simulation. In addition, we found significant differences in human perceptions when participants watched a video of a human-robot interaction compared to when they participated in the interaction, experiencing two-way flow of information.

Overall, our study suggests that results from user studies that rely on video observations and interactive simulations may not always mirror human perceptions of robots in real-world HRI. Importantly, we found trade-offs between Real-Video, Sim-Video and Sim-Interactive methodologies. First, our work provides initial evidence that suggests that human perceptions of a robot in video studies may be less similar to real-world in-person studies in comparison to interactive simulation studies. This suggests that an interactive simulation should be preferred over observing videos. Second, we found that participants perceived greater workload in simulated environments than in real-world environments. Lesser workload in the real-world may help explain why, in some prior work, humans preferred in-person human-robot interactions more than simulated or video interactions [1, 68]. Also, our results with respect to workload suggest that Real-Video may be preferred over Sim-Video and Sim-Interactive. Ultimately, it is important to consider whether human perceptions are likely to translate to the real-world, and human workload, when choosing a methodology other than in-person studies to investigate human-robot interactions.

ACKNOWLEDGMENTS

This work was partially funded by Amazon and the National Science Foundation (NSF) under Grant No. (IIS-1924802). The findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of Amazon or the NSF.

REFERENCES

- [1] Wilma A. Bainbridge, Justin W. Hart, Elizabeth S. Kim, and Brian Scassellati. 2010. The Benefits of Interactions with Physically Present Robots over Video-Displayed Agents. *Int. J. Soc. Robot.* 3, 1 (2010).
- [2] David P Baker and Eduardo Salas. 1992. Principles for measuring teamwork skills. *Human factors* 34, 4 (1992), 469–475.
- [3] Santosh Balajee Banisetty and Tom Williams. 2021. Implicit Communication Through Social Distancing: Can Social Navigation Communicate Social Norms?. In *HRI '21 Companion*.
- [4] Kimberly A. Barchard, Leisze Lapping-Carr, R. Shane Westfall, Andrea Fink-Armold, Santosh Balajee Banisetty, and David Feil-Seifer. 2020. Measuring the Perceived Social Intelligence of Robots. *THRI* 9, 4 (Sept. 2020).
- [5] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. 2020. *Human-robot interaction: An introduction*. Cambridge University Press.
- [6] Terry K Borsook and Nancy Higginbotham-Wheat. 1991. Interactivity: What is it and what can it do for computer-based instruction? *Educational Technology* (1991).
- [7] Wolfram Burgard, Armin B Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. 1999. Experiences with an interactive museum tour-guide robot. *Artificial intelligence* 114, 1-2 (1999), 3–55.
- [8] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): Development and Validation. In *HRI*.
- [9] Sonia Chernova, Jeff Orkin, and Cynthia Breazeal. 2010. Crowdsourcing hri through online multiplayer games. In *2010 AAAI Fall Symposium Series*.
- [10] Filipa Correia, Samuel Gomes, Samuel Mascarenhas, Francisco S. Melo, and Ana Paiva. 2020. The Dark Side of Embodiment Teaming Up With Robots VS Disembodied Agents. In *RSS*.
- [11] Kelly Cuccolo, Megan S Irgens, Martha S Zlokovich, Jon Grahe, and John E Edlund. 2021. What crowdsourcing can offer to cross-cultural psychological science. *Cross-Cultural Research* (2021).
- [12] Andrea Deublein and Birgit Lugin. 2020. (Expressive) Social Robot or Tablet? – On the Benefits of Embodiment and Non-verbal Expressivity of the Interface for a Smart Environment. In *International Conference on Persuasive Technology*.
- [13] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. 2013. Legibility and predictability of robot motion. In *HRI*.
- [14] Gilberto Echeverria, Séverin Lemaignan, Arnaud Degroote, Simon Lacroix, Michael Karg, Pierrick Koch, Charles Lesire, and Serge Stinckwich. 2012. Simulating complex robotic scenarios with MORSE. In *SIMPAR*. Springer.
- [15] David Feil-Seifer, Kerstin S Haring, Silvia Rossi, Alan R Wagner, and Tom Williams. 2020. Where to next? The impact of COVID-19 on human-robot interaction research. *THRI* (2020).
- [16] Kerstin Fischer, Katrin Lohan, and Kilian Foth. 2012. Levels of embodiment: Linguistic analyses of factors influencing HRI. In *HRI*.
- [17] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. *TiCS* (2007).
- [18] Yuxiang Gao and Chien-Ming Huang. 2022. Evaluation of socially-aware robot navigation. *Front. Robot. AI* (2022).
- [19] Sandra G.Hart and Lowell E.Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* (1988).
- [20] Edward Twitchell Hall. 1966. *The hidden dimension*. Vol. 609. Anchor.
- [21] Guy Hoffman, Jodi Forlizzi, Shahar Ayal, Aaron Steinfeld, John Antanitis, Guy Hochman, Eric Hochendoner, and Justin Finkenaur. 2015. Robot presence and human honesty: Experimental evidence. In *HRI*.
- [22] Guy Hoffman and Wendy Ju. 2014. Designing robots with movement in mind. *JHRI* (2014).
- [23] Guy Hoffman and Xuan Zhao. 2020. A Primer for Conducting Experiments in Human–Robot Interaction. *JHRI* (2020).
- [24] Yuhuan Hu and Guy Hoffman. 2019. Using skin texture change to design emotion expression in social robots. In *HRI*. IEEE.
- [25] Helge Hüttenrauch, Kerstin Severinson Eklundh, Anders Green, and Elin A Topp. 2006. Investigating spatial relationships in human-robot interaction. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE.
- [26] Patrik Jonell, Taras Kucherenko, Ilaria Torre, and Jonas Beskow. 2020. Can we trust online crowdworkers?. In *IVA*.
- [27] Michiel Joosse, Manja Lohse, and Vanessa Evers. 2015. Crowdsourcing culture in HRI: lessons learned from quantitative and qualitative data collections. In *3rd international workshop on culture aware robotics at ICSR*.
- [28] Younbo Jung and Kwan Min Lee. 2004. Effects of physical embodiment on social presence of social robots. In *Proceedings of Presence*.
- [29] Daphne E Karreman, Geke DS Ludden, and Vanessa Evers. 2019. Beyond R2D2: Designing multimodal interaction behavior for robot-specific morphology. *THRI* (2019).

- [30] Christian U Krägeloh, Jaishankar Bharatharaj, Senthil Kumar Sasthan Kutty, Praveen Regunathan Nirmala, and Loulin Huang. 2019. Questionnaires to measure acceptability of social robots: a critical review. *Robotics* (2019).
- [31] Minae Kwon, Erdem Biyik, Aditi Talati, Karan Bhasin, Dylan P Losey, and Dorsa Sadigh. 2020. When humans aren't optimal: Robots that collaborate with risk-aware humans. In *HRI*. IEEE.
- [32] Minae Kwon, Sandy H Huang, and Anca D Dragan. 2018. Expressing robot incapability. In *HRI*.
- [33] Alexis Lambert, Nahal Norouzi, Gerd Bruder, and Gregory Welch. 2020. A systematic review of ten years of research on human interaction with social robots. *JHCI* (2020).
- [34] Kwan Min Lee, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 2006. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *Int. J. Hum. Comput. Stud.* (2006).
- [35] Séverin Lemaignan, Marc Hanheide, Michael Karg, Harmish Kambhaita, Lars Kunze, Florian Lier, Ingo Lütkebohle, and Grégoire Milliez. 2014. Simulation and HRI recent perspectives with the MORSE simulator. In *SIMPAR*. Springer.
- [36] Michael Lewis, Jijun Wang, and Stephen Hughes. 2007. USARSim: Simulation for the study of human-robot interaction. *JCEDM* (2007).
- [37] Jamy Li. 2015. The benefit of being physically present. *Int. J. Hum. Comput. Stud.* (2015).
- [38] Rui Li, Marc van Almkerk, Sanne van Waveren, Elizabeth Carter, and Iolanda Leite. 2019. Comparing Human-Robot Proxemics Between Virtual Reality and the Real World. In *HRI*.
- [39] David V Lu, Dave Herschberger, and William D Smart. 2014. Layered costmaps for context-sensitive navigation. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 709–715.
- [40] Ratnesh Madaan, Nicholas Gyde, Sai Vempala, Matthew Brown, Keiko Nagami, Tim Taubner, Eric Cristofalo, Davide Scaramuzza, Mac Schwager, and Ashish Kapoor. 2020. Airsim drone racing lab. In *Neurips 2019 competition and demonstration track*. PMLR.
- [41] Maxim Makatchev, Reid Simmons, Majd Sakr, and Micheline Ziadee. 2013. Expressing ethnicity through behaviors of a robot character. In *HRI*.
- [42] Ali Mollahosseini, Hojjat Abdollahi, Timothy D. Sweeny, Ron Cole, and Mohammad H. Mahoor. 2018. Role of embodiment and presence in human perception of robots' facial cues. *Int. J. Hum. Comput. Stud.* (2018).
- [43] Amal Nanavati, Xiang Zhi Tan, Joe Connolly, and Aaron Steinfeld. 2019. Follow the robot: Modeling coupled human-robot dyads during navigation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3836–3843.
- [44] Stefanos Nikolaidis, Anton Kuznetsov, David Hsu, and Siddhartha Srinivasa. 2016. Formalizing human-robot mutual adaptation: A bounded memory model. In *HRI*.
- [45] Ali Noormohammadi-Asl, Kevin Fan, Stephen L Smith, and Kerstin Dautenhahn. 2024. Human Leading or Following Preferences: Effects on Human Perception of the Robot and the Human-Robot Collaboration. *arXiv preprint arXiv:2401.01466* (2024).
- [46] Aude Olivia, Michael L Mack, Mochan Shrestha, and Angela Peeper. 2004. Identifying the perceptual dimensions of visual complexity of scenes. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 26.
- [47] Valerio Ortenzi, Akansel Cosgun, Tommaso Pardi, Wesley P Chan, Elizabeth Croft, and Dana Kulić. 2021. Object handovers: a review for robotics. *T-RO* (2021).
- [48] HD Patterson. 1975. Maximum likelihood estimation of components of variance. In *IBC*. Biometric Soc.
- [49] Ashwini Pople, Roberto Martín-Martín, Patrick Goebel, Vincent Chow, Hans M. Ewald, Junwei Yang, Zhenkai Wang, Amir Sadeghian, Dorsa Sadigh, Silvio Savarese, and Marynel Vázquez. 2019. Deep Local Trajectory Replanning and Control for Robot Navigation. In *ICRA*.
- [50] Louise Poubel. [n. d.]. *Service Robot Simulator*. <https://github.com/osrf/servicesim>
- [51] Carolyn C Preston and Andrew M Colman. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica* 104, 1 (2000), 1–15.
- [52] Laurel D. Riek, Tal-Chen Rabinowitch, Paul Bremner, Anthony G. Pipe, Mike Fraser, and Peter Robinson. 2010. Cooperative gestures: Effective signaling for humanoid robots. In *HRI*.
- [53] Stephanie Rosenthal, Joydeep Biswas, and Manuela M Veloso. 2010. An effective personal mobile robot agent through symbiotic human-robot interaction.. In *AAMAS*, Vol. 10. 915–922.
- [54] Nicole Salomons, Tom Wallenstein, Debasmita Ghose, and Brian Scassellati. 2022. The Impact of an In-Home Co-Located Robotic Coach in Helping People Make Fewer Exercise Mistakes. In *RO-MAN*.
- [55] Gary W Selnow. 1988. Using interactive computer to communicate scientific information. *American Behavioral Scientist* (1988).
- [56] Stela H Seo, Denise Geiskovitch, Masayuki Nakane, Corey King, and James E Young. [n. d.]. Poor thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and a simulated robot. In *HRI*. IEEE.
- [57] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. 2006. Common metrics for human-robot interaction. In *HRI*.
- [58] Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 479–486.
- [59] Leila Takayama, Doug Dooley, and Wendy Ju. 2011. Expressing thought: improving robot readability with animation principles. In *HRI*.
- [60] Mohsen Tavakol and Reg Dennick. 2011. Making sense of Cronbach's alpha. *International journal of medical education* 2 (2011), 53.
- [61] Sam Thellman, Annika Silvervarg, Agneta Gulz, and Tom Ziemke. 2016. Physical vs. Virtual Agent Embodiment and Effects on Social Interaction. In *IIVA*.

- [62] Russell Toris, David Kent, and Sonia Chernova. 2014. The robot management system: A framework for conducting human-robot interaction studies through crowdsourcing. *JHRI* (2014).
- [63] Joanne Truong, Max Rudolph, Naoki Yokoyama, Sonia Chernova, Dhruv Batra, and Akshara Rai. 2022. Rethinking Sim2Real: Lower Fidelity Simulation Leads to Higher Sim2Real Transfer in Navigation. *arXiv preprint arXiv:2207.10821* (2022).
- [64] Nathan Tsoi, Mohamed Hussein, Jeacy Espinoza, Xavier Ruiz, and Marynel Vázquez. 2020. SEAN: Social Environment for Autonomous Navigation. In *HAI*.
- [65] Nathan Tsoi, Mohamed Hussein, Olivia Fugikawa, J. D. Zhao, and Marynel Vazquez. 2021. An Approach to Deploy Interactive Robotic Simulators on the Web for HRI Experiments: Results in Social Robot Navigation. In *IROS*.
- [66] Nathan Tsoi, Alec Xiang, Peter Yu, Samuel S. Sohn, Greg Schwartz, Subashri Ramesh, Mohamed Hussein, Anjali W. Gupta, Mubbasir Kapadia, and Marynel Vázquez. 2022. SEAN 2.0: Formalizing and Generating Social Situations for Robot Navigation. *RA-L* (2022).
- [67] Gentiane Venture and Dana Kulić. 2019. Robot expressive motions: a survey of generation and evaluation methods. *THRI* (2019).
- [68] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2006. The role of physical embodiment in human-robot interaction. In *RO-MAN*.
- [69] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2007. Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN*.
- [70] Manhua Wang, Seul Chan Lee, Harsh Kamallesh Sanghavi, Megan Eskew, Bo Zhou, and Myoungsoon Jeon. 2021. In-Vehicle Intelligent Agents in Fully Autonomous Driving: The Effects of Speech Style and Embodiment Together and Separately. In *AutomotiveUI*.
- [71] Ning Wang, David V Pynadath, and Susan G Hill. 2016. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *HRI*. IEEE.
- [72] Sarah N. Woods, Michael L. Walters, Kheng Lee Koay, and Kerstin Dautenhahn. 2006. Methodological Issues in HRI: A Comparison of Live and Video-Based Methods in Robot to Human Approach Direction Trials. In *RO-MAN*.
- [73] Fei Xia, William B Shen, Chengshu Li, Priya Kasimbeg, Micael Edmond Tchampi, Alexander Toshev, Roberto Martín-Martín, and Silvio Savarese. 2020. Interactive gibbon benchmark: A benchmark for interactive navigation in cluttered environments. *IEEE Robotics and Automation Letters* 5, 2 (2020), 713–720.
- [74] Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Akiko Yamazaki, and Keiichi Yamazaki. 2012. Development of a mobile museum guide robot that can configure spatial formation with visitors. In *ICIC*. Springer.
- [75] Jakub Zlotowski, Astrid Weiss, and Manfred Tscheligi. 2012. Navigating in public space: Participants’ evaluation of a robot’s approach behavior. In *HRI*.