"What's That Robot Doing Here?": Perceptions Of Incidental Encounters With Autonomous Quadruped Robots

Elliott Hauser*
The University of Texas at Austin
School of Information
Austin, Texas, United States
elliott@utexas.edu

Yao-Cheng Chan*
The University of Texas at Austin
School of Information
Austin, Texas, United States
ycchan@utexas.edu

Geethika Hemkumar
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
geethika.hemkumar@utexas.edu

Parth Chonkar
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
parthchonkar@gmail.com

Daksh Dua
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
dakshdua@utexas.edu

Efren Mendoza Enriquez
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
emefren@utexas.edu

Tiffany Kao
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
tiffanyiris1004@gmail.com

Shikhar Gupta
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
shikhar.gupta.tx@gmail.com

Huihai Wang
The University of Texas at Austin
School of Architecture
Austin, Texas, United States
hw9998@utexas.edu

Junfeng Jiao
The University of Texas at Austin
School of Architecture
Austin, Texas, United States
jjiao@austin.utexas.edu

Reuth Mirsky
Bar Ilan University
Department of Computer Science
Ramat Gan, Israel
reuthde@gmail.com

Justin Hart
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
hart@cs.utexas.edu

Joydeep Biswas
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
joydeepb@cs.utexas.edu

Peter Stone
The University of Texas at Austin
Department of Computer Science
Austin, Texas, United States
pstone@cs.utexas.edu

ABSTRACT

Autonomous service robots in a public setting will generate hundreds of incidental human-robot encounters, yet researchers have only recently addressed this important topic in earnest. In this study, we hypothesized that visual indicators of human control, such as a leash on a robot, would impact humans' perceptions of robots in the context of human-robot encounters. A pilot study (n = 26) and a revised study (n = 22) including semi-structured interviews (n = 21) were conducted. The interview data suggested that the presence of another human during the encounter elicited positive reactions from the participants. Counter to these interview findings, the Godspeed-based survey data yielded largely statistically insignificant results between the conditions. We interpret this as evidence that traditional HRI survey instruments focused on the perception of robot characteristics may not be suitable for incidental human-robot encounters research. We suggest that human-robot encounters can be meaningfully characterized by participants' ability or inability to answer implicit questions such as, "what is that robot doing here?". We conclude with recommendations for human-robot encounters research methods and call for research on the intelligibility and acceptability of perceived robot purpose during human-robot encounters.

*Authors contributed equally to this research.

TAS '23, July 11–12, 2023, Edinburgh, United Kingdom

ACM ISBN 979-8-4007-0734-6/23/07.
https://doi.org/10.1145/3597512.3599707

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; **User studies**; *Laboratory experiments*; • **Computer systems organization → Robotic autonomy**.

ACM Reference Format:
Elliott Hauser, Yao-Cheng Chan, Geethika Hemkumar, Parth Chonkar, Daksh Dua, Efren Mendoza Enriquez, Tiffany Kao, Shikhar Gupta, Huihai

Wang, Junfeng Jiao, Reuth Mirsky, Justin Hart, Joydeep Biswas, and Peter Stone. 2023. "What's That Robot Doing Here?": Perceptions Of Incidental Encounters With Autonomous Quadruped Robots. In *First International Symposium on Trustworthy Autonomous Systems (TAS '23), July 11–12, 2023, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3597512.3599707

1 INTRODUCTION

The emerging study of human-robot encounters (HRE) investigates the experience of people who do not initially plan to use, interact, or collaborate with robots but appear to be coincidentally present with robots. Memorably nicknamed 'the Forgotten' by Rosenthal-von der Pütten et al., *incidentally co-present persons* (InCoPs) [70] is an under-studied and poorly understood constituency in the field of human-robot interaction (HRI), despite their increasing numbers. We contend that incidental encounters with autonomous service robots represent not just a pressing ethical concern but also a distinct and exciting scientific challenge for the development of trustworthy autonomous robotic systems. This paper presents methodologies and results from a pilot study and revised study designed to identify how canine metaphors of control impact participants' experiences of quadruped robot encounters. These exploratory studies help refine and characterize the emerging and related fields of HRE and trustworthy autonomous systems (TAS).

Autonomous quadruped service robots are a suitable platform to use in this research for three reasons. Firstly, they are mobile, making them suitable for service applications. Mobile service robots operating in public spaces generate more HREs than HRIs with their users. Although wheeled robots are currently the most prevalent, quadruped platforms' agility and ability to traverse varied kinds of terrain suggest that they will be increasingly common in the future. Secondly, the canine morphology of quadruped service robots makes them inherently evocative [see, e.g., 83]. Finally, this canine resemblance can be reasonably expected to have complex effects on the perception of human-quadruped dyads by InCoPs, evoking the idiom of pet owners, service dog users, etc. Our experimental design takes advantage of all three of these factors.

This study is designed to identify how the presence or absence of an accompanying human and their implied relationship with the quadruped influence InCoPs perceptions of incidental encounters. We hypothesized that participants who encountered a quadruped service robot with a human handler and a leash would report a more positive experience than those who encountered the robot alone. In addition to testing this hypothesis, we present other related factors identified via analysis of participant interviews.

In the pilot study, we utilized an intra-participant design with five conditions, with primary data collection via a questionnaire. In the second experiment (subsequently, revised study), we adopted an inter-participant design with two conditions, a longer version of the questionnaire, and added an interview protocol. The survey data from both studies yielded largely statistically insignificant results between the conditions. However, the interview data appeared to suggest that the presence of humans elicited positive reactions. Many of the participants' descriptions of their experiences can be explained by their ability or inability to answer implicit, contextually dependent questions like 'what is the robot doing here?'.

This paper presents several tangible contributions toward a better understanding of pedestrian experiences during encounters with autonomous quadruped robots. These include a foundation for designing in-the-wild HRE studies, interview-based insights into implicit questions InCoPs had during their robot encounters, and articulation of unresolved challenges that HRE researchers collectively face. The paper concludes with recommendations for future research.

This study also continues a trend in social navigation and HRI more broadly of expanding beyond dyadic interactions into more complex configurations of humans and robots [11, 48, 63, 76]. As robot deployments become more common, HRI research must expand its scope to include those whose daily lives may be impacted by the robot's presence [88, 97].

2 RELATED WORK

Rosenthal-von der Pütten et al.'s workshop at HRI 2020 sought to bring HRI attention to incidental robot encounters, introducing the term InCoP to name the HRI constituency that had until that point gone unnamed. Notwithstanding the intervening pandemic, their efforts have helped link phenomena reported in seemingly unconnected research areas.

In autonomous vehicle (AV) research, scientists have urged the development of 'implicit' interfaces, designed to support incidental encounters rather than planned or expected interactions to promote pedestrian safety [e.g. 58]. Simultaneously, researchers have documented cases of violence [83], vandalism [82] and bullying *against* robots, suggesting fear or a kind of robo-xenophobia, even of helpless or innocuous devices. These extremes suggest that mutual misunderstanding persists within the space of human-robot encounters, justifying calls for more robust theories in HRI [36]. This brief literature review surfaces existing works studying perceptions of quadruped robots, explorations of how canine appearance and associations impact acceptance, and studies of pedestrian encounters with robots of varied platforms. The present studies are motivated by and intended to extend these lines of research.

Perceptions of quadruped service robots. The relative dearth of HRI studies of quadruped platforms in service applications means that public perceptions are poorly understood, even as they are entering wider use and visibility. Moreover, media portrayals of quadruped robots, which have been shown to influence perceptions of robots [5], have unknown impacts on these perceptions. Media coverage likely to influence public perceptions includes general interest pieces on "robot dogs" (often with an alarmist tone; e.g., [29, 91]), news coverage, and research reports on deployments by public health and safety organizations [8, 97], and these platforms' growing use in marketing campaigns, such as the 2022 Samuel Adams Superbowl ad featuring Boston Dynamics Spot [50]. The longer this gap in research persists, the more difficult it will be to identify and characterize changes to and influences upon perceptions of service quadrupeds, and to articulate how they vary across cultures and within subcultures [64].

HRI of Quadruped Service Robots and Canine Metaphors. Quadruped platforms intuitively suggest canine interaction metaphors and usage. Indeed, one of the most developed areas of research has been the study of their use as guide dogs to assist visually impaired users with navigation [9, 16, 21, 61, 73, 94]. Related research specifically targeting assistive use cases, but employing canine metaphors; includes cynomorphic and general zoomorphic expressions [20, 74, 81], leash-based interfaces [55, 94, 96], and dog-inspired interaction design [49, 71]. Recent works have also started to investigate humans' general perceptions of quadruped robots. For instance, Küster et al. [47] conducted a study and asked the participants to rate robots after watching videos of them. The results suggested that the robot's ability and appearance can elicit different and complex reactions. This study touched upon the perception of quadruped robots, however, the video-based interaction is not representative of HRE.

Pedestrian Encounters with Service-Robots. There is a growing literature in HRI and adjacent fields — such as smart cities — regarding pedestrian-robot interaction; much of it very recent [23, 51, 54, 70, 86, 98]. In light of this recency, many studies have adopted an exploratory approach [35]. Researchers outside HRI have also begun to examine these issues; often from critical, ethical, or justice-focused perspectives [2, 15, 19, 39, 60, 88]. This work builds upon and is motivated by work on long-term autonomy [26, 46, 90], which has increased the number of human-robot encounters.

While the literature around encounters — as opposed to the direct interaction that is the traditional object of HRI studies — continues to expand, encounters with quadruped service robots remain poorly understood. This represents an important gap in current knowledge, which is particularly urgent to address as quadruped deployments become more common, and in light of the inherent differences in perception of robots by platform and style of locomotion.

3 A STANDARDIZED HUMAN-ROBOT ENCOUNTER SCENARIO

To enable our ongoing research program, we sought to create a standardized HRE scenario in the laboratory. While encounters outside the laboratory will undoubtedly have higher ecological validity, we reason that the higher efficiency of laboratory experiments is appropriate to initial, exploratory work. We plan to extend the most promising experimental designs to real-world settings in future work (see Section 6.3).

We sought a reusable encounter scenario that would let the robot or human-robot dyad be the primary study variable, while preserving meaningful participant agency. To accomplish this, we prepared a simulated hallway in a laboratory setting, using room dividers. To standardize encounters within this setting, a custom social navigation stack was developed that allows the robot to traverse the hallway while avoiding oncoming pedestrians at a standardized distance. Combined with participant instructions to walk to the other end of the hallway and pass the robot however feels most natural, this design creates a standardized HRE context that preserves a meaningful degree of potential variability in participant walking speed, reaction distance, and behavior towards the robot and/or dyad.

3.1 Robot Platform

For the experiment, the Boston Dynamics Spot robot is equipped with a laptop (Ryzen 9 5900HS, RTX 3060) running Ubuntu 21.10 and ROS Noetic in a Docker container, a Velodyne VLP-16 LiDAR, and a Microsoft Azure Kinect point cloud camera. The robot as equipped for the study can be seen in Figure 1.

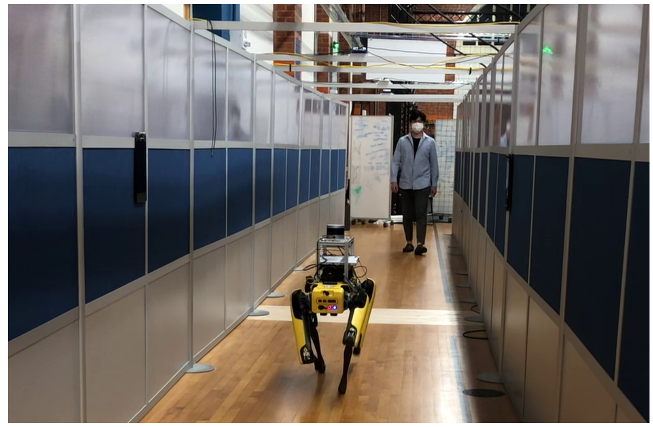


Figure 1: Photograph of study setting showing an impending human-robot encounter in the simulated hallway. The robot is shown in the Autonomous condition (see Figure 3)

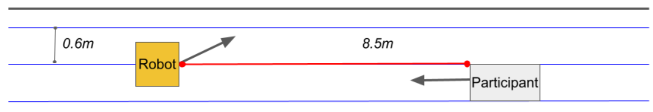


Figure 2: Diagram of the operation of the social navigation stack. The robot operates in a hallway that is 1.25m wide. Three lines representing "lanes" (following the lines, not between them) are laid out in the hallway, 0.6m apart. The robot moves to an opposing lane when it comes within a configurable distance of the participant.

3.2 Navigation

The system divides the hallway into three traffic lanes (similar to [17] and [24]), and the robot starts in the middle lane. A representation of the lanes is presented in Figure 2. When the robot reaches a threshold distance of the participant, configured as 8.5m in pilot study, it shifts to the opposite lane from the participant, continuing forward motion to pass the participant when able. The robot will stop to avoid collisions if the participant attempts to step in front of it, and continues to attempt to pass the participant when it determines there is space to do so, including via lateral motion when necessary.

The hallway is modeled as three traffic lanes, as illustrated in Figure 2. The hallway is 1.8m wide, and the lanes are modeled as lines, each 0.6m apart, with one lane in the middle of the hallway. The robot navigates by choosing waypoints that are 2.75m in front of it on the lane that it wishes to navigate on. This distance is hand-tuned to produce a smooth-looking lane-shifting behavior.

The robot is localized using an implementation of Episodic non-Markov localization (ENML) [4] using a map of the hallway area. Navigation goals are given to the robot as ROS twist messages, translated to the Spot's protocol using the Clearpath Spot ROS

Driver.[1] Visualization of the robot's state information is provided to robot operators (when initially setting up the robot) through Robofleet WebViz [80], an open source, browser-based visualizer that connects to ENML.

3.3 Pedestrian Detection and Control Algorithm Behavior

To detect people in the hallway, the navigation stack uses the Azure Kinect Body Tracking SDK[2]. The SDK provides the pose of the person relative to the camera as a track consisting of landmarks on the body. The chest is transformed into the global frame using the ROS TF2 service. The distance of this landmark from the left or right wall is computed in order to determine which side of the hallway the pedestrian is on. When the robot comes within a pre-programmed distance of the study participant, it shifts lanes to the lane opposite that which the participant is measured as being in.

4 PILOT STUDY: INTRAPARTICIPANT DESIGN

The emerging availability of quadruped robots opens a variety of interactions in which the robot is presented not solo but as part of a human-robot dyad. People are used to seeing dogs with *handlers*; pet owners, service dog users, and others. Our starting hypothesis was that the *invocation of visual signs of canine control, such as a leash or service harness, would positively influence participants' experience during incidental encounters with a robot*. The pilot study utilized a five-condition intra-participant design that we hypothesized would yield a series of increasing effects as the encounter conditions increasingly evoked the robot as a service dog. Limitations of this exploratory methodology are addressed in Section 6.6.

4.1 Pilot Study Methodology

The study design was approved as exempt from human subjects oversight by the University of Texas at Austin Institutional Review Board. We did not identify any ethical or participant safety concerns before or during the study. A total of 26 participants from the University of Texas at Austin campus were recruited via email distribution of an online recruitment form. One participant is excluded from data analysis for failure to complete the entire questionnaire.

4.1.1 Study Protocol and Conditions. After informed consent and an optional media release are obtained, participants are directed to one end of the test hallway, with the robot placed in the middle lane at the other end. The participants are then instructed to start walking to the other end of the hallway when they see the robot start walking, and to pass the robot whenever and however feels most natural. After the hallway interaction, the participants fill out the study questionnaire. This interaction is repeated 5 times, exposing each participant to each condition in randomized order to mitigate sequencing effects.

For each condition, the robot is presented differently. The robot, as outfitted for each condition, can be seen in Figure 3. The conditions are:

Autonomous The robot traverses the hallway by itself with no additional costuming.

Joystick The robot traverses the hallway with a researcher following behind, holding a game controller, pretending to control the robot.

Companion The robot traverses the hallway with a researcher walking next to it.

Service Vest The robot traverses the hallway with a researcher walking slightly behind it, holding onto a service dog harness. The robot wears a service dog vest.

Leashed The robot traverses the hallway with a researcher walking slightly behind it, holding onto a leash.

After completing the five conditions, participants complete one final survey, comparing the conditions to each other. After concluding the study, the participants are debriefed, revealing the purpose of the study and that the robot is always operating fully autonomously, even in the Joystick condition.

4.1.2 Questionnaire. The questionnaire administered to study participants between trials aimed to discern how study conditions impacted participants' perceptions of the robot, with all other variables held constant. In order to avoid participant fatigue from lengthy repeated questionnaires, we chose a small subset of Godspeed questions [1], as shown in Table 2. To these, we added exploratory questions including 5 Godspeed-like semantic difference questions, 4 emotions (Curious, Cautious, Calm, and Excited), and 3 questions about comfort encountering the robot in different contexts (Office, Campus, Home). These exploratory additions violate the ideal practice of utilizing standardized, validated questionnaires. In the absence of such instruments, we saw this choice as required to accomplish our study goals; limitations due to this choice are discussed in Section 6.6.

At the end of all five conditions, a final questionnaire was administered, asking participants to compare conditions on a subset of the questions administered between rounds.

4.2 Pilot Study Survey Results

The results for questionnaires administered between trials yield few statistically-significant results between conditions. Details are provided in Appendix A.

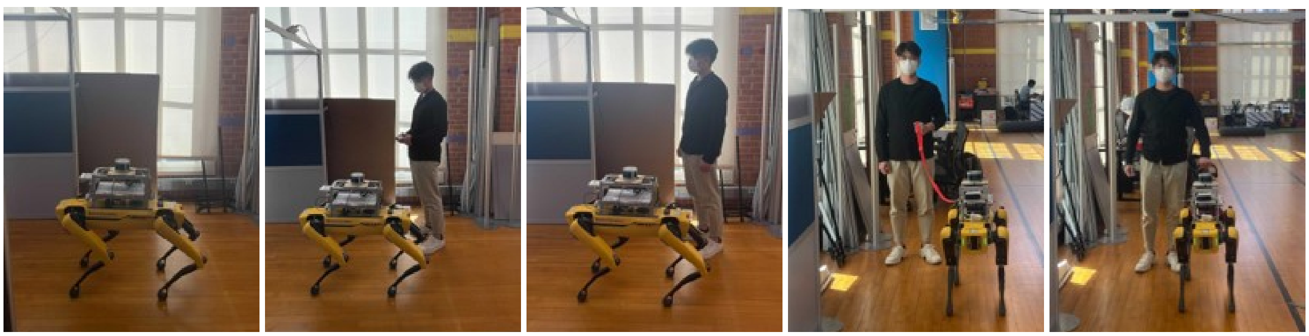
The results of the questionnaire administered at the end of the study are mostly significant. However, for the end questionnaire, participants are asked to rank the conditions against each other. Due to a configuration error, it was possible to (and some participants did) give two conditions the same ranking.

Statistical significance is computed using the Friedman Test. From these results, it can be seen that at least one of the interventions has significant effects on several metrics. Specifically, the Leashed and Service Vest conditions are viewed as the most Doglike; the Autonomous condition is seen as the most autonomous; and the Joystick, Service Vest, and Leashed conditions are seen as the safest. Joystick is seen as the most under control, but participants are most comfortable getting close to robots in the Leashed and Service Vest conditions. In addition to the initial testing of the hypothesis, the major goal of the pilot study was to identify two contrasting encounter conditions for the inter-participant revised study. Thus, we did not perform further statistical analysis of these results, such as ad-hoc analysis. Additional details and more complete results are reported in Appendix A.1.

[1] https://github.com/clearpathrobotics/spot_ros
[2] https://microsoft.github.io/Azure-Kinect-Body-Tracking/release/1.1.x/index.html

Table 1: Instructions given to research assistants for non-Autonomous conditions.

Instruction	Joystick	Companion	Leashed	Service Vest
Starting Position	*All Conditions: After attaining correct position, remain in place. Avoid looking at researcher who initiates autonomy sequence. Begin moving only after robot does, without surprise or startle.*			
	Keep joystick ready to suggest robot is autonomous	–	Maintain moderate leash slack	Hold harness consistently; do not drop
Position	One pace behind robot	One pace behind robot	One pace behind robot	Abreast of robot, harness in left hand
Gaze	Down hallway, glance at robot/joystick	Down hallway, glance at robot	Down hallway, glance at robot	Down hallway, glance at robot
Movement	*All Conditions: Follow lateral robot movements smoothly. Take care when passing participant. Pause or adjust motion as needed to avoid collisions.*			
	Mime joystick control of robot	–	Adjust following distance to maintain moderate slack	Move behind robot as needed to avoid participant



(a) Autonomous **(b) Joystick** **(c) Companion** **(d) Leashed** **(e) Service Vest**

Figure 3: The robot as outfitted for each of the five conditions in this study.

4.3 Pilot Study Discussion

Given the contrast between the significant overall ranking data and largely insignificant results of the sequentially administered questionnaires, we searched for methodological deficiencies to remedy in the revised study. We expected to be able to demonstrate differences in participant evaluation of the robots via the survey after incorporating these changes.

5 REVISED STUDY: INTERPARTICIPANT DESIGN WITH INTERVIEWS

The pilot study informed a refined methodology for studying human-robot encounters, which was implemented in the revised study. The revised study tests a more fully specified hypothesis that invocation of visual signs of canine control, operationalized as the Leashed condition, would positively influence participants' experience during incidental encounters with a robot, compared to the Autonomous condition as control.

5.1 Revised Study Protocol Revisions

This section provides a summary of the methodological adaptations we implemented. The revised study adopted an inter-participant design with only the Leashed and Autonomous condition, and shifted the Godspeed questionnaire to the six-point scale for semantic difference question types. The items adopted are shown in Table 2. We also included 15-30 minute semi-structured interviews, which initially consisted of 11 interview questions, sequenced to cover participant experience in the study, speculation about encountering the robot outside of the laboratory, speculation about 'other people's' reactions to the robot, and the participant's prior experience with robots and the robotics laboratory building, if any. This sequencing was designed to postpone any potentially suggestive content until after the participant had described their experience in the study. Each question contained brief follow-up cues to help the interviewer ask clarifying questions of interest. For this study, the interviews were not analyzed in their entirety. Instead, the authors consulted the interview transcripts only to confirm findings or help interpret surprising or unexplained findings from the quantitative study results. A subsequent full analysis of the interview data is detailed in a separate publication currently under review. Rationales behind these revisions not described here are provided in Appendix A.3 and limitations are discussed in Section 6.6.

The flow of the revised study was similar to the pilot study. Each participant was randomly assigned to only one condition, they were given a chance to interact freely with the robot after taking the Godspeed questionnaire, and they participated in a semi-structured interview afterward. As before, participants were recruited via the distribution of a web form via UT Austin email lists.

Table 2: Godspeed questions [1] used in each Study. *Doglike* Was substituted for Godspeed's *Humanlike* in user-facing questions and the *Anthropomorphism* category is labeled *Cynomorphism* (i.e., Doglike) here (participants were not shown category names).

Godspeed Question	Pilot	Revised
I. Cynomorphism		
Machinelike vs *Doglike*	Y	Y
Fake vs Natural	-	Y
Unconscious vs Conscious	Y	Y
Artificial vs Lifelike	-	Y
Moving Rigidly vs Elegantly	-	Y
II. Animacy		
Dead vs Alive	-	-
Stagnant vs Lively	-	-
Mechanical vs Organic	-	-
Artificial vs Lifelike	-	-
Inert vs Interactive	-	Y
Apathetic vs Responsive	-	Y
III. Likability		
Dislike vs Like	-	Y
Unfriendly vs Friendly	Y	Y
Unkind vs Kind	-	-
Unpleasant vs Pleasant	-	Y
Awful vs Nice	-	Y
IV. Perceived Intelligence		
Incompetent vs Competent	-	Y
Ignorant vs Knowledgeable	-	Y
Irresponsible vs Responsible	-	-
Unintelligent vs Intelligent	-	-
Foolish vs Sensible	-	-

5.2 Survey Results

Using one-way ANOVA, the Godspeed questionnaire administered in the revised study also yielded statistically insignificant results between conditions. In addition to the Godspeed questionnaire items, several exploratory items were also added to the survey, including proxemics, emotions, speed, public quadruped robot encounters, and services. All items and detail are provided in Appendix A. These exploratory items were all statistically insignificant between conditions. To the research team's surprise, some items even yielded the same mean value. These results indicate that the participants' perceptions of the robot, which were captured by Godspeed and other exploratory items, were the same between conditions. We thus turned to an analysis of participant interviews for help interpreting these unexpected null results.

5.3 Interview Results

Even though the null results from the surveys of both studies indicate that there was no difference between the conditions, the interview data provide signals that major factors influencing InCoPs' experience during HREs were not present as variables in the questionnaires and/or observable in the study design.

5.3.1 Effects of personal and intuited familiarity with robots. Several participants expressed that the presence of a human impacted their perceptions of the robot. Even more interestingly, participants who experienced the study's Autonomous condition speculated that the presence of a human with the robot would have helped them feel more at ease during the encounter.

The visual similarity of human-quadruped dyads to a familiar human-dog dyad seems to mitigate humans' cautiousness towards robots. P-20-A[3] illustrated this by saying, *"I think there's a sense of familiarity with, you know, having a human by a robot's side."* This participant recalled prior experiences of seeing a human walking next to a quadruped form, which mimicked the familiar sight of a human walking with a dog.

For some participants, the term "familiarity" indicated not only their personal familiarity, but also the intuited familiarity of the researcher walking next to the robot. *"There's this familiarity factor when you see a person and a leash, a person with the robot. [...] I can recognize a person and if a person's okay with this robot, it must not be scary or anything like that."* [P13-L] For other participants, the presence of a human with the robot also elicits higher perceived safety, regardless of personal familiarity. For instance, a participant reported:

> *I think I would definitely feel more comfortable if I see a human assisting the robot or assuming that it is assisting and walking with it. And I think the feeling of it being handled by someone somehow gives me more trust than it being on its own.* [P14-A]

As indicated by the qualifier "somehow", P14 was not confident that they could describe the precise dynamics of this effect. This was true of other participants as well, suggesting that the effect is initially subconscious and difficult to describe.

5.3.2 Summary of Interview Findings. The interview data indicate that there are characteristic differences in participant experiences of the Leashed and Autonomous condition. This is discrepant with the null results reported in Section 4.2 and Section 5.2.

5.4 Revised Study Discussion

In all of our conditions, the appearance and behavior of the robot remained the same. While important for the reproducibility of encounters in a laboratory setting, this consistency limits our survey's efficacy in differentiating the participants' perceptions of the robot and, especially, in informing causal explanations. The addition of participant interviews to the study protocol substantially increased our ability to triangulate our surprising survey results. The differences we observed in the interview data are primarily derived from the presence of humans and the social interactions inherent in encounters with a human-canine dyad. These differences cannot easily be captured between the scales, indicating the complex dynamics between humans and robots in HREs.

6 DISCUSSION

In HREs, the human-robot relationship is under-determined compared to traditional operator-robot, user-robot, or collaborator-

[3]Participant IDs are annotated with -L and -A to represent whether they encountered the Leashed or Autonomous conditions, respectively.

robot interactions. Our initial study design was intended to enable consistent replication of the ambiguous relationship between a mobile robot and the humans it encounters incidentally while navigating through shared space. We adapted our methodology between the pilot study and revised study to better account for the complexities introduced by this under-determined relationship in several ways, with varying degrees of success. We share here recommendations that we are following in our ongoing work and that we believe other human-robot interaction teams will benefit from as well.

6.1 Implicit Questions in Robot Encounters

HRE study participants must determine whether and how to react to the presence of the robot from the limited data available to them during the encounter. The limited data they're presented with are, roughly, the robot's appearance and behavior, any co-present humans' appearance and behavior, and the context of the encounter. To interpret this data, users must draw heavily from their prior experience.

In Section 5.3, we reported characteristic ways in which participants drew from prior experiences to interpret their encounter with the robot. By comparing participants' answers in interviews, we were able to discern several implicit questions that arose during the encounters:

- How will the robot react to me?
- Is the robot under control?
- Will the robot harm me?

Participants who reported negative emotions at any point in the encounter were likely to express some form of implicit doubt about the answers to these questions. No participants had extremely negative encounters with the robot, but those that were most ambivalent about it remained uncertain about these questions at the end of the interview. A participant observed, *"We're sort of allowing robots to come into our environment more and more I would definitely want to know what is the purpose behind it."* [P14-A] For this participant, knowing what a robot is doing is important, but they did not speculate on what those purposes might be.

6.1.1 Encounter Inflection Points. Some participants identified specific moments where their implicit uncertainties were resolved, which we term *inflection points* of the encounter. Characteristics of these moments were the resolution of the indeterminacy of the encounter and a sense of relief or resolution.

A common moment cited in this context was when the robot visibly reacted to the participant's growing proximity. A participant who encountered the Autonomous condition described the inflection point of the encounter as a realization that this was a "good robot": *"as soon as it sensed my movement, and it changed its direction. After that, I became extremely ..., like normal. I was like, Okay, it's a good robot."* [P7-A]

Another moment cited by multiple participants as an inflection point was when, after completing the questionnaire, the participants were offered an opportunity to approach the robot and ask any questions they had of the study staff. Participants who identified it as an inflection point described a resolution of ambivalence or caution into a more positive curiosity or calm.

For other participants, these questions did not arise. In describing why he wouldn't be worried to see the robot operating on the university's campus, a participant replied, *"maybe they're just using it for testing purposes, since we're at like a big research university."* [P2-L] When asked versions of these implicit questions later in the interview, participants who had not mentioned these topics responded with sophisticated and nuanced accounts of how they and others might feel when encountering the robot.

The presence of this kind of retrospectively elicited reasoning is a potential explanation of why implicit questions did not arise for participants that could provide them. Since they already had implicit answers to these basic questions, their attention turned naturally to other aspects of the encounter or beyond.

6.1.2 A Prototypical Implicit Question. In comparing the factors associated with having answers, or not, to these implicit questions, we observed that they can all be related to an overarching, prototypical question:

- What is that robot doing here?

This simple question juxtaposes the robot, its behavior, and a first-person expression of spatiotemporal context. Its phrasing is derived from a participant's response to a question about others encountering the robot; they speculated that others encountering the robot would first wonder *"what is that thing doing here?"*

Users who have an intuitive answer to this question gain corollary answers to the list of questions presented above. Intuitive or assumed answers to the prototypical question could come from prior experience with robots, or from visual cues. For instance, several participants envisioned an increased comfort level for themselves and others if the robot had been clearly marked as a delivery robot. This and other elicited examples of increased comfort with robots entailed implicit answers to one or more of the questions above. Participants (and their imagined others) for whom these implicit questions never arose implicitly were able to successfully move on to other reactions, such as curiosity, speculation, or adaptation of their behavior.

6.2 Augmenting Godspeed

Godspeed was developed as a functional-technical scale predominantly used to examine the usability and design of robots. However, in a research context such as HRE, humans' perception of robots is not merely a result of their usability but is also heavily influenced by the social factors of robots. The social interactions between humans and robots warrant measurements that are not simply imported from functional-technical scales.

Following the trend of Social Robotics studies, researchers have also started to develop scales that can be adopted in the field. These scales tend to differentiate from the technical-functional oriented scales by acknowledging and integrating humans' beliefs of robots' mental, social, and moral capabilities, meanwhile focusing on examining and characterizing the humans in HRI. Scales developed from Social Robotics can be adopted to augment Godspeed. For instance, the AMPH survey [12] can be used to categorize humans by identifying their tendency to anthropomorphize artifacts and natural objects. The authors argued that the way in which humans interact with technology is influenced by their own tendency to anthropomorphize technology or non-human objects. Echoing the concept of Sociomorphing [77], AMPH can also be used to challenge the traditional view of anthropomorphism. The synergy of both kinds of scales can measure the social influence of robots meanwhile making sure the functionality of robots is not neglected.

6.3 Recommendations for Human-Robot Encounters Research

Reflecting upon the challenges we have faced in evaluating our hypothesis, interpreting null and discrepant results, and revising our own methodology, we can offer some initial suggestions that we plan to follow in ongoing work in this area. We offer them a transparently contingent rationale in the hope that they might assist others seeking, as we did, guidance in the published literature. As the literature on human-robot encounters expands, we expect that many of the contingencies we note will be resolved empirically.

6.3.1 Triangulate Surveys with Interviews. We suggest that mixed research methodologies [12, 22, 27, 32, 66, 67, 69, 93] are essential for guiding exploratory HRE research. Mixed methods have been widely used in HRI studies on complex interactions such as human-robot teams [66, 95] and with specific populations such as older adults [27, 69] and children [59]. Combinations of methods can produce diverse data to inform the interpretation of complex phenomena [78].

In particular, we argue that quantitative studies informed by qualitative interpretation have advantages for HRE research, even compared to larger-scale studies employing quantitative metrics alone. This is because a key task for emerging fields of research is to identify and orient around large-magnitude effects. A mixed-method analysis of smaller-scale studies is more likely to accomplish this than large-scale studies, since the higher statistical power will typically only increase statistical confidence in low-magnitude effects (a power analysis further supporting this argument is presented in Section 6.6).

We see the social scientific technique of triangulation as particularly promising for HRE studies, with a simple modification. In social science, triangulation typically involves the confirmation of qualitative results using surveys or other quantitative data, and has been deployed in social studies of robotics [10]. In HRI studies of HREs, an inversion of this typical arrangement would accommodate the existing prevalence of metric- and questionnaire-based research in HRI. In this configuration, qualitative data from interviews and observation can confirm, contextualize, and better interpret quantitative results including survey instruments, proxemics, and temporal metrics of human and/or robot behavior, rather than primarily informing the design of confirmatory questionnaires.

6.3.2 Study Physical Human-Robot Encounters. We recommend that researchers primarily utilize in-person human-robot encounter study designs until prior work demonstrating parity of video- and image-based evaluation of robots in traditional HRI is replicated for human-robot encounters research. While it is likely uncontroversial that such study designs have higher ecological validity, the many experiential factors cited by our participants that are not reproducible in simulation leads us to view the validity of simulation-based research on human experiences of robot encounters as suspect until proven otherwise.[4]

The methodological challenges of HREs bear directly upon the selection of appropriate study designs. Although the present study cannot provide direct evidence to support this claim, we reason that prior evidence suggesting the parity of certain kinds of HRI study designs [13, 44] should not be presumed to apply to human-robot encounters until this evidence is replicated in the study of encounters. Until that time, we recommend that researchers use live encounters and, where feasible, in situ encounters over laboratory-based studies.

6.3.3 Start From Standardization. We applaud recent efforts to standardize metrics and methods in HRI [e.g., 72]. Our calls for the evaluation of existing methods as a prerequisite to the development of new methods are aligned with these efforts. This is why we chose the widely-used Godspeed questionnaire as the survey instrument in this study, despite its potential mismatch with our research goals.

We recommend that human-robot encounter researchers start with existing HRI methods and instruments. The process of new question and questionnaire development is arduous; necessarily so if the rigor of HRI research is to be maintained. Such efforts must be seen as the culmination of a program of research, not its beginning.

Our calls for the modification or development of new standardized instruments draw upon the evidence presented here that the Godspeed questionnaire's focus is not well-aligned with the full range of factors participants cited as influencing their encounter with the robot. For now, we encourage HRE researchers to join us in lightly adapting and evaluating existing methods and instruments where possible, and to seek opportunities for collaboratively modifying, developing, and evaluating standardized metrics and survey instruments specifically suited to the encounter context.

6.4 Applications to Autonomous Social Navigation for Robots

As HRE knowledge continues to expand, we expect that it will increasingly converge with and inform related areas of robotic autonomy research. Implications for social navigation research, which aims to enable robots to perceive, react to, and conform to social norms of movement [52, 56], are particularly promising. Prior work in social navigation has examined collision avoidance [14, 65, 87], comfort [33, 53, 89], smoothness of interaction [28, 40, 62], effort invested [18, 38] and other objective and subjective measures of social acceptability when pedestrians and a mobile robot move in a shared space. Social Navigation research often implements autonomy based upon identifications of social cues or socially-informed predictions of human movement [30, 43], and has even progressed to studies of robot ability to (socially) signal navigational intentions [17, 25, 31, 79, 87]. Social Navigation as a whole is deeply informed by empirical findings about [34, 68, 75, 84], models of [3, 6, 37], or training data containing [7, 41, 42, 85] human social navigation norms. This study provides empirical findings of human

[4]This strong view is both contingent and narrowly scoped to studies of *human experience* of incidental robot encounters. Simulations will likely remain an important tool in fields like social navigation that inherently involve incidental encounters. Ideally, though, close collaboration between HRI researchers and those studying robotic autonomy, control, and human factors for incidental HREs will encourage study designs involving physical encounters.

perception and reaction to quadruped service robot encounters that could inform future research in this key area of robotic autonomy.

6.5 Responsible Research and Innovation

The fast development of autonomous robots substantially increases HREs, which will heavily impact human society. As mentioned in previous sections, research has found cases of violence against robots [83]. Other recent works on HRE have found that people have concerns, such as collisions, inconvenience, and a lack of communication capabilities when they encounter an autonomous cleaning robot [23]. These concerns may not be issues to healthy and able-bodied individuals, but may potentially harm people with disability. On the other hand, law enforcement agencies have also started to deploy quadruped robots for surveillance duties, and pedestrians who encounter these robots may be subject to privacy invasion or feel stressed about the presence of the robots as advanced quadruped robots may appear as menacing [45]. In essence, robot deployments have merits and bring the potential for the greater good. However, it is also important to understand how robot deployments may affect pedestrians and what we can do to address the underlying ethical issues. This work took an exploratory approach to examine if adding visual cues to canine control can improve pedestrians' experience. The results can help roboticists and the authorities understand what the better ways are to design and deploy autonomous quadruped robots.

6.6 Limitations

Given the limited literature and knowledge of incidental encounters with autonomous robots in the field, this work is preliminary and took an exploratory approach. This section highlights the studies' most important limitations.

Firstly, our study adopted a lab-based design, which afforded us logistical efficiency and a way to isolate the study conditions from an encounter context. This undoubtedly reduced the study's ecological validity. The lab setting can have a framing effect and the complex dynamics inherent in natural social settings could not be reflected in study design. Secondly, the intra-participant design in the pilot study could have caused strong carryover effects as the participants experienced all conditions back-to-back. Conducting an intra-participant study design over a course of five days, for instance, would mitigate the carryover effects. Thirdly, an anonymous reviewer noted that the revised study was likely underpowered, and we concur. If we had increased the sample size, we might have found some significant differences between the Leashed and Autonomous conditions not visible in our survey data. Lastly, while required for the exploratory nature of our study, we acknowledge that the novel questions added to the questionnaire did not stem from a strong theoretical background and were not validated statistically.

These limitations bound our contributions to some extent. Regardless, future research can adopt our findings and consider our recommendations to develop HRE hypotheses and continue to work toward methodological standardization in this emerging field.

7 CONCLUSION

To test the hypothesis that the presence of canine metaphors of control would positively influence participants' encounters with quadruped service robots, we conducted two laboratory experiments ($n = 26, 22$). The survey data in both studies yielded largely null results. We confidently endorse the null hypothesis that no such influence exists when the participant's experience of the encounter is operationalized via Godspeed questions and Godspeed-like questions aiming to measure the participant's perception of the robot's qualities.

However, we identified rich and clear findings within the interview data collected in the revised study indicating that the participants have a better experience with the Leashed condition, specifically. The mechanism of this effect appears to be that the leash and the presence of another human help participants answer implicit questions of the general form 'what is that robot doing here?' Participants that had a positive experience, regardless of condition, did not seem preoccupied with this or related questions, and were able to provide sophisticated and contextualized answers to similar questions when prompted. We are undertaking studies determining the effects of apparent robot purpose on participant perceptions of incidental encounters with robots. We hope that this and future work will support the continued coalescence of the HRE research community [57, 70] and highlight the importance of robot encounters to HRI and the realization of trustworthy autonomous systems.

ACKNOWLEDGMENTS

This research was supported in part by NSF Award #2219236 and Living and Working with Robots, a core research project of Good Systems, a UT Grand Challenge.

REFERENCES

[1] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[2] Cynthia Bennett, Emily Ackerman, Bonnie Fan, Jeffrey Bigham, Patrick Carrington, and Sarah Fox. 2021. Accessibility and The Crowded Sidewalk: Micromobility's Impact on Public Space. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 365–380. https://doi.org/10.1145/3461778.3462065

[3] Niklas Bergstrom, Takayuki Kanda, Takahiro Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2008. Modeling of natural human-robot encounters. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2623–2629. https://doi.org/10.1109/IROS.2008.4650896

[4] Joydeep Biswas and Manuela M. Veloso. 2017. Episodic non-Markov localization. *Robotics and Autonomous Systems* 87 (2017), 162 – 176. https://doi.org/10.1016/j.robot.2016.09.005

[5] Ulrike Bruckenberger, Astrid Weiss, Nicole Mirnig, Ewald Strasser, Susanne Stadler, and Manfred Tscheligi. 2013. The Good, The Bad, The Weird: Audience Evaluation of a "Real" Robot in Relation to Science Fiction and Mass Media. In *Social Robotics*. Springer International Publishing, 301–310. https://doi.org/10.1007/978-3-319-02675-6_30

[6] Sachit Butail. 2015. Simulating the effect of a social robot on moving pedestrian crowds. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2413–2418. https://doi.org/10.1109/IROS.2015.7353704

[7] Yuying Chen, Congcong Liu, Bertram E Shi, and Ming Liu. 2020. Robot Navigation in Crowds by Graph Convolutional Networks With Attention Learned From Human Gaze. *IEEE Robotics and Automation Letters* 5, 2 (April 2020), 2754–2761. https://doi.org/10.1109/LRA.2020.2972868

[8] Zhiming Chen, Tingxiang Fan, Xuan Zhao, Jing Liang, Cong Shen, Hua Chen, Dinesh Manocha, Jia Pan, and Wei Zhang. 2021. Autonomous Social Distancing in Urban Environments Using a Quadruped Robot. *IEEE Access* 9 (2021), 8392–8403. https://doi.org/10.1109/ACCESS.2021.3049426

[9] Tzu-Kuan Chuang, Ni-Ching Lin, Jih-Shi Chen, Chen-Hao Hung, Yi-Wei Huang, Chunchih Teng, Haikun Huang, Lap-Fai Yu, Laura Giarré, and Hsueh-Cheng Wang. 2018. Deep Trail-Following Robotic Guide Dog in Pedestrian Environments for People who are Blind and Visually Impaired - Learning from Virtual and Real Worlds. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 5849–5855. https://doi.org/10.1109/ICRA.2018.8460994

[10] Bohkyung Chun and Heather Knight. 2020. The Robot Makers: An Ethnography of Anthropomorphism at a Robotics Company. *J. Hum.-Robot Interact.* 9, 3 (June 2020), 1–36. https://doi.org/10.1145/3377343

[11] Joe Connolly, Viola Mocz, Nicole Salomons, Joseph Valdez, Nathan Tsoi, Brian Scassellati, and Marynel Vázquez. 2020. Prompting Prosocial Human Interventions in Response to Robot Mistreatment. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction.* Association for Computing Machinery, New York, NY, USA, 211–220. https://doi.org/10.1145/3319502.3374781

[12] M F Damholdt, C Vestergaard, and J Seibt. 2020. Testing for 'Anthropomorphization': A Case for Mixed Methods in Human-Robot Interaction. In *Human-Robot Interaction: Evaluation Methods and Their Standardization*, Céline Jost, Brigitte Le Pévédic, Tony Belpaeme, Cindy Bethel, Dimitrios Chrysostomou, Nigel Crook, Marine Grandgeorge, and Nicole Mirnig (Eds.). Springer International Publishing, Cham, 203–227. https://doi.org/10.1007/978-3-030-42307-0_8

[13] Antonella De Angeli and Sheryl Brahnam. 2008. I hate you! Disinhibition with virtual partners. *Interacting with computers* 20, 3 (May 2008), 302–310. https://doi.org/10.1016/j.intcom.2008.02.004

[14] Gian Diego and Tipaldi Kai O Arras. 2011. Please do not disturb! Minimum interference coverage for social robots. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems.* 1968–1973. https://doi.org/10.1109/IROS.2011.6094867

[15] Anna Dobrosovestnova, Glenda Hannibal, and Tim Reinboth. 2021. Service robots for affective labor: a sociology of labor perspective. *AI & society* (April 2021), 1–13. https://doi.org/10.1007/s00146-021-01208-x

[16] Brian L Due. 2021. Interspecies intercorporeality and mediated haptic sociality: distributing perception with a guide dog. *Visual Studies* (Aug. 2021), 1–14. https://doi.org/10.1080/1472586X.2021.1951620

[17] Rolando Fernandez, Nathan John, Sean Kirmani, Justin Hart, Jivko Sinapov, and Peter Stone. 2018. Passive demonstrations of light-based robot signals for improved human interpretability. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Nanjing). IEEE, 234–239. https://doi.org/10.1109/roman.2018.8525728

[18] Gonzalo Ferrer, Anaís Garrell, and Alberto Sanfeliu. 2013. Social-aware robot navigation in urban environments. In *2013 European Conference on Mobile Robots.* 331–336. https://doi.org/10.1109/ECMR.2013.6698863

[19] Petra Gemeinboeck. 2022. Difference-in-relation: Diffracting human-robot encounters. *Matter: Journal of New Materialist Research* 3, 1 (Feb. 2022), 29–55. https://doi.org/10.1344/jnmr.v3i1.38958

[20] Moojan Ghafurian, Gabriella Lakatos, Zhuofu Tao, and Kerstin Dautenhahn. 2020. Design and Evaluation of Affective Expressions of a Zoomorphic Robot. In *Social Robotics.* Springer International Publishing, 1–12. https://doi.org/10.1007/978-3-030-62056-1_1

[21] Márta Gácsi, Sára Szakadát, and Ádám Miklósi. 2014. What Could Assistance Robots Learn from Assistance Dogs?. In *Bio-Inspired Models of Network, Information, and Computing Systems.* Springer International Publishing, 105–119. https://doi.org/10.1007/978-3-319-06944-9_8

[22] Andreas Kornmaaler Hansen, Juliane Nilsson, Elizabeth Ann Jochum, and Damith Herath. 2020. On the Importance of Posture and the Interaction Environment: Exploring Agency, Animacy and Presence in the Lab vs Wild using Mixed-Methods. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20).* Association for Computing Machinery, New York, NY, USA, 227–229. https://doi.org/10.1145/3371382.3378288

[23] K Hardeman. 2021. *Encounters with a seemingly autonomous sidewalk delivery vehicle: interviews with incidentally copresent pedestrians.* Ph.D. Dissertation. Utrecht University. http://mwlc.global/wp-content/uploads/2021/07/Thesis-Kevin-Hardeman-4150015.pdf

[24] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. 2020. Using Human-Inspired Signals to Disambiguate Navigational Intentions. In *International Conference on Social Robotics.* Springer, 320–331.

[25] Justin Hart, Reuth Mirsky, Xuesu Xiao, Stone Tejeda, Bonny Mahajan, Jamin Goo, Kathryn Baldauf, Sydney Owen, and Peter Stone. 2020. Using Human-Inspired Signals to Disambiguate Navigational Intentions. In *Proceedings of the 12th International Conference on Social Robotics.* Springer, 320–331. https://doi.org/10.1007/978-3-030-62056-1_27

[26] Nick Hawes, Christopher Burbridge, Ferdian Jovan, Lars Kunze, Bruno Lacerda, Lenka Mudrova, Jay Young, Jeremy Wyatt, Denise Hebesberger, Tobias Kortner, Rares Ambrus, Nils Bore, John Folkesson, Patric Jensfelt, Lucas Beyer, Alexander Hermans, Bastian Leibe, Aitor Aldoma, Thomas Faulhammer, Michael Zillich, Markus Vincze, Eris Chinellato, Muhannad Al-Omari, Paul Duckworth, Yiannis Gatsoulis, David C Hogg, Anthony G Cohn, Christian Dondrup, Jaime Pulido Fentanes, Tomas Krajnik, Joao M Santos, Tom Duckett, and Marc Hanheide. 2017. The STRANDS Project: Long-Term Autonomy in Everyday Environments. *IEEE robotics & automation magazine / IEEE Robotics & Automation Society* 24, 3 (Sept. 2017), 146–156. https://doi.org/10.1109/MRA.2016.2636359

[27] Denise Hebesberger, Tobias Koertner, Christoph Gisinger, Juergen Pripfl, and Christian Dondrup. 2016. Lessons learned from the deployment of a long-term autonomous robot as companion in physical therapy for older adults with dementia a mixed methods study. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* ieeexplore.ieee.org, 27–34. https://doi.org/10.1109/HRI.2016.7451730

[28] D Helbing and P Molnár. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 51, 5 (May 1995), 4282–4286. https://doi.org/10.1103/physreve.51.4282

[29] Karen Heller. 2021. Spot is the $74,500 robot dog of our dystopian dreams. *The Washington Post* (Aug. 2021). https://www.washingtonpost.com/lifestyle/style/spot-dog-robot-boston-dynamics/2021/08/06/81b2b780-f475-11eb-9068-bf463c8c74de_story.html

[30] Blake Holman, Abrar Anwar, Akash Singh, Mauricio Tec, Justin Hart, and Peter Stone. 2021. Watch Where You're Going! Gaze and Head Orientation as Predictors for Social Robot Navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (Xi'an, China). 6183–6190. https://doi.org/10.1109/icra48506.2021.9561286

[31] Mohammed Moshiul Hoque, Tomomi Onuki, Yoshinori Kobayashi, and Yoshinori Kuno. 2011. Controlling human attention through robot's gaze behaviors. In *2011 4th International Conference on Human System Interactions, HSI 2011.* 195–202. https://doi.org/10.1109/HSI.2011.5937366

[32] Ashatu Hussein. 2009. The use of Triangulation in Social Sciences Research. *Comparative social work* 4, 1 (April 2009), 106–117. https://doi.org/10.31265/jcsw.v4i1.48

[33] Phillip Jeffrey and Gloria Mark. 1998. Constructing social spaces in virtual environments: A study of navigation and interaction. In *Workshop on personalised and social navigation in information space.* researchgate.net, 24–38. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.551.6582&rep=rep1&type=pdf

[34] Michiel Joosse, Manja Lohse, and Vanessa Evers. 2014. Sound over matter: the effects of functional noise, robot size and approach velocity in human-robot encounters. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (Bielefeld, Germany) *(HRI '14).* Association for Computing Machinery, New York, NY, USA, 184–185. https://doi.org/10.1145/2559636.2559822

[35] Swapna Joshi and Selma Šabanović. 2019. A tactical urbanist approach to facilitate exploratory HRI research in public spaces. In *Proceedings of the 17th European Conference on Computer-Supported Cooperative Work: The International Venue on Practice-centered Computing and the Design of Cooperation Technologies.* European Society for Socially Embedded Technologies (EUSSET). https://doi.org/10.18420/ECSCW2019_P05

[36] Malte Jung and Pamela Hinds. 2018. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *J. Hum.-Robot Interact.* 7, 1 (May 2018), 1–5. https://doi.org/10.1145/3208975

[37] Mitsuhiro Kamezaki, Yusuke Tsuburaya, Taichi Kanada, Michiaki Hirayama, and Shigeki Sugano. 2022. Reactive, Proactive, and Inducible Proximal Crowd Robot Navigation Method Based on Inducible Social Force Model. *IEEE Robotics and Automation Letters* 7, 2 (April 2022), 3922–3929. https://doi.org/10.1109/LRA.2022.3148451

[38] Akira Kanazawa, Jun Kinugawa, and Kazuhiro Kosuge. 2019. Adaptive Motion Planning for a Collaborative Robot Based on Prediction Uncertainty to Enhance Human Safety and Work Efficiency. *IEEE Trans. Rob.* 35, 4 (Aug. 2019), 817–832. https://doi.org/10.1109/TRO.2019.2911800

[39] Takayuki Kanda. 2017. Enabling Harmonized Human-Robot Interaction in a Public Space. In *Human-Harmonized Information Technology, Volume 2: Horizontal Expansion*, Toyoaki Nishida (Ed.). Springer Japan, Tokyo, 115–137. https://doi.org/10.1007/978-4-431-56535-2_4

[40] Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. 2014. Universal power law governing pedestrian interactions. *Phys. Rev. Lett.* 113, 23 (Dec. 2014), 238701. https://doi.org/10.1103/PhysRevLett.113.238701

[41] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Soeren Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. 2022. Socially Compliant Navigation Dataset (SCAND). https://doi.org/10.18738/T8/0PRYRH

[42] Haresh Karnan, Anirudh Nair, Xuesu Xiao, Garrett Warnell, Sören Pirk, Alexander Toshev, Justin Hart, Joydeep Biswas, and Peter Stone. 2022. Socially CompliAnt Navigation Dataset (SCAND): A Large-Scale Dataset of Demonstrations for Social Navigation. *IEEE Robotics and Automation Letters* 7, 4 (Oct. 2022), 11807–11814. https://doi.org/10.1109/LRA.2022.3184025

[43] Kapil D Katyal, Gregory D Hager, and Chien-Ming Huang. 2020. Intent-Aware Pedestrian Prediction for Adaptive Crowd Navigation. In *2020 IEEE International Conference on Robotics and Automation (ICRA).* 3277–3283. https://doi.org/10.1109/ICRA40945.2020.9197434

[44] Merel Keijsers and Christoph Bartneck. 2018. Mindless Robots get Bullied. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* ieeexplore.ieee.org, 205–214. https://ieeexplore.ieee.org/abstract/document/9473529/

[45] Carolin Kemper and Michael Kolain. 2022. K9 Police Robots - Strolling Drones, RoboDogs, or Lethal Weapons? (Aug. 2022). https://doi.org/10.2139/ssrn.4201692

[46] Piyush Khandelwal, Shiqi Zhang, Jivko Sinapov, Matteo Leonetti, Jesse Thomason, Fangkai Yang, Ilaria Gori, Maxwell Svetlik, Priyanka Khante, Vladimir Lifschitz,

J K Aggarwal, Raymond Mooney, and Peter Stone. 2017. BWIBots: A platform for bridging the gap between AI and human–robot interaction research. *The International journal of robotics research* 36, 5-7 (2017), 635–659. https://doi.org/10.1177/0278364916688949 arXiv:https://doi.org/10.1177/0278364916688949
[47] Dennis Küster, Aleksandra Swiderska, and David Gunkel. 2021. I saw it on YouTube! How online videos shape perceptions of mind, morality, and fears about robots. *new media & society* 23, 11 (2021), 3312–3331.
[48] Hee Rin Lee, Eunjeong Cheon, Chaeyun Lim, and Kerstin Fischer. 2022. Configuring Humans: What Roles Humans Play in HRI Research. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan) *(HRI '22)*. IEEE Press, 478–492. https://dl.acm.org/doi/abs/10.5555/3523760.3523824
[49] Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Maya Cakmak, and Siddhartha Srinivasa. 2011. Predictability or adaptivity? Designing robot handoffs modeled from trained dogs and people. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 179–180. https://ieeexplore.ieee.org/document/6281284
[50] Sasha Lekach. 2022. Boston Dynamics robots let loose with afterwork beers in Super Bowl ad. https://mashable.com/video/boston-dynamics-sam-adams-super-bowl-ad. https://mashable.com/video/boston-dynamics-sam-adams-super-bowl-ad Accessed: 2022-4-13.
[51] Alexis Linard, Ilaria Torre, Anders Steen, Iolanda Leite, and Jana Tumova. 2021. Formalizing Trajectories in Human-Robot Encounters via Probabilistic STL Inference. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 9857–9862. https://doi.org/10.1109/IROS51168.2021.9635951
[52] Christoforos Mavrogiannis, Francesca Baldini, Allan Wang, Dapeng Zhao, Pete Trautman, Aaron Steinfeld, and Jean Oh. 2021. Core Challenges of Social Robot Navigation: A Survey. (March 2021). arXiv:2103.05668 [cs.RO] http://arxiv.org/abs/2103.05668
[53] Alyxander David May, Christian Dondrup, and Marc Hanheide. 2015. Show me your moves! Conveying navigation intention of a mobile robot to humans. In *2015 European Conference on Mobile Robots (ECMR)*. 1–6. https://doi.org/10.1109/ECMR.2015.7324049
[54] David C May, Kristie J Holler, Cindy L Bethel, Lesley Strawderman, Daniel W Carruth, and John M Usher. 2017. Survey of Factors for the Prediction of Human Comfort with a Non-anthropomorphic Robot in Public Spaces. *International Journal of Social Robotics* 9, 2 (April 2017), 165–180. https://doi.org/10.1007/s12369-016-0390-7
[55] Reuth Mirsky and Peter Stone. 2021. The seeing-eye robot grand challenge: rethinking automated care. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021)*.
[56] Reuth Mirsky, Xuesu Xiao, Justin Hart, and Peter Stone. 2021. Prevention and Resolution of Conflicts in Social Navigation – a Survey. (June 2021). arXiv:2106.12113 [cs.RO] http://arxiv.org/abs/2106.12113
[57] Frederik Moesgaard, Lasse Hulgaard, and Mads Bødker. 2022. Incidental Encounters with Robots. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 377–384. https://doi.org/10.1109/RO-MAN53752.2022.9900591
[58] Dylan Moore, Rebecca Currano, G Ella Strack, and David Sirkin. 2019. The Case for Implicit External Human-Machine Interfaces for Autonomous Vehicles. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 295–307. https://doi.org/10.1145/3342197.3345320
[59] Terran Mott and Tom Williams. 2022. Community-Situated Mixed-Methods Robotics Research for Children and Childhood Spaces. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan) *(HRI '22)*. IEEE Press, 1167–1169. https://dl.acm.org/doi/abs/10.5555/3523760.3523960
[60] Lu Vinh Nhat, Jochen Wirtz, Werner H Kunz, Stefanie Paluch, Thorsten Gruber, Antje Martins, and Paul G Patterson. 2020. Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? *Journal of Service Theory and Practice* 30, 3 (Jan. 2020), 361–391. https://doi.org/10.1108/JSTP-04-2019-0088
[61] Nahal Norouzi, Kangsoo Kim, Myungho Lee, Ryan Schubert, Austin Erickson, Jeremy Bailenson, Gerd Bruder, and Greg Welch. 2019. Walking Your Virtual Dog: Analysis of Awareness and Proxemics with Simulated Support Animals in Augmented Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 157–168. https://doi.org/10.1109/ISMAR.2019.000-8
[62] Billy Okal and Kai O Arras. 2016. Learning socially normative robot navigation behaviors with Bayesian inverse reinforcement learning. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 2889–2895. https://doi.org/10.1109/ICRA.2016.7487452
[63] Raquel Oliveira, Patrícia Arriaga, and Ana Paiva. 2021. Human-Robot Interaction in Groups: Methodological and Research Practices. *Multimodal Technologies and Interaction* 5, 10 (Sept. 2021), 59. https://doi.org/10.3390/mti5100059
[64] Irena Papadopoulos and Christina Koulouglioti. 2018. The Influence of Culture on Attitudes Towards Humanoid and Animal-like Robots: An Integrative Review. *Journal of nursing scholarship: an official publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau* 50, 6 (Nov. 2018), 653–665. https://doi.org/10.1111/jnu.12422
[65] Jin Hyoung Park, Francisco Arturo Rojas, and Hyun Seung Yang. 2013. A collision avoidance behavior model for crowd simulation based on psychological findings. *Comput. Animat. Virtual Worlds* 24, 3-4 (May 2013), 173–183. https://doi.org/10.1002/cav.1504
[66] Jonas E Pedersen, Kristoffer W Christensen, Damith Herath, and Elizabeth Jochum. 2020. I Like the Way You Move: A Mixed-Methods Approach for Studying the Effects of Robot Motion on Collaborative Human Robot Interaction. In *Social Robotics*. Springer International Publishing, 73–84. https://doi.org/10.1007/978-3-030-62056-1_7
[67] Cristina Pereira, Vìtor Pinheira, Maria João Guardado Moreira, Paulo Gonçcalves, and Simão Silva. 2018. A methodological approach to evaluate elderly-robot interactions. *Methodology: European journal of research methods for the behavioral & social sciences* 23, 3 (Aug. 2018), 205–213. https://doi.org/10.15405/ejsbs.241
[68] Matthias Rehm and Anders Krogsager. 2013. Negative affect in human robot interaction — Impoliteness in unexpected encounters with robots. In *2013 IEEE RO-MAN*. 45–50. https://doi.org/10.1109/ROMAN.2013.6628529
[69] Wendy A Rogers, Travis Kadylak, and Megan A Bayles. 2022. Maximizing the Benefits of Participatory Design for Human–Robot Interaction Research With Older Adults. *Human factors* 64, 3 (May 2022), 441–450. https://doi.org/10.1177/00187208211037465
[70] Astrid Rosenthal-von der Pütten, David Sirkin, Anna Abrams, and Laura Platte. 2020. The Forgotten in HRI: Incidental Encounters with Robots in Public Spaces. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 656–657. https://doi.org/10.1145/3371382.3374852
[71] Yea-Kyung Row, Se-Young Kim, and Tek-Jin Nam. 2020. Using pet-dog behavior traits to enhance the emotional experience of in-car interaction. *International Journal of Design* 14, 1 (2020), 19–34. https://koasas.kaist.ac.kr/handle/10203/274327
[72] Matthew Rueben, Shirley A Elprama, Dimitrios Chrysostomou, and An Jacobs. 2020. Introduction to (Re)Using Questionnaires in Human-Robot Interaction Research. In *Human-Robot Interaction: Evaluation Methods and Their Standardization*, Céline Jost, Brigitte Le Pévédic, Tony Belpaeme, Cindy Bethel, Dimitrios Chrysostomou, Nigel Crook, Marine Grandgeorge, and Nicole Mirnig (Eds.). Springer International Publishing, Cham, 125–144. https://doi.org/10.1007/978-3-030-42307-0_5
[73] Shozo Saegusa, Yuya Yasuda, Yoshitaka Uratani, Eiichirou Tanaka, Toshiaki Makino, and Jen-Yuan (james) Chang. 2010. Development of a Guide-Dog Robot: Leading and Recognizing a Visually-Handicapped Person using a LRF. *Journal of Advanced Mechanical Design, Systems, and Manufacturing* 4, 1 (2010), 194–205. https://doi.org/10.1299/jamdsm.4.194
[74] Vanessa Sauer, Axel Sauer, and Alexander Mertens. 2021. Zoomorphic Gestures for Communicating Cobot States. *IEEE Robotics and Automation Letters* 6, 2 (April 2021), 2179–2185. https://doi.org/10.1109/LRA.2021.3060416
[75] Trenton Schulz, Rebekka Soma, and Patrick Holthaus. 2021. Movement acts in breakdown situations: How a robot's recovery procedure affects participants' opinions. *Paladyn, Journal of Behavioral Robotics* 12, 1 (Jan. 2021), 336–355. https://doi.org/10.1515/pjbr-2021-0027
[76] Sarah Sebo, Brett Stoll, Brian Scassellati, and Malte F Jung. 2020. Robots in Groups and Teams: A Literature Review. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 1–36. https://doi.org/10.1145/3415247
[77] Johanna Seibt, Christina Vestergaard, and Malene F Damholdt. 2020. Sociomorphing, Not Anthropomorphizing: Towards a Typology of Experienced Sociality. In *Frontiers in Artificial Intelligence and Applications*, Marco Nørskov, Johanna Seibt, and Oliver Santiago Quick (Eds.). Frontiers in artificial intelligence and applications, Vol. 335. IOS Press, Amsterdam, 51–67. https://doi.org/10.3233/faia200900
[78] Johanna Seibt, Christina Vestergaard, and Malene F Damholdt. 2021. The Complexity of Human Social Interactions Calls for Mixed Methods in HRI: Comment on "A Primer for Conducting Experiments in Human-robot Interaction," by G. Hoffman and X. Zhao. *J. Hum.-Robot Interact.* 10, 1 (Feb. 2021), 1–4. https://doi.org/10.1145/3439715
[79] Moondeep C Shrestha, Tomoya Onishi, Ayano Kobayashi, Mitsuhiro Kamezaki, and Shigeki Sugano. 2018. Communicating Directional Intent in Robot Navigation using Projection Indicators. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 746–751. https://doi.org/10.1109/ROMAN.2018.8525528
[80] Kavan Singh Sikand, Logan Zartman, Sadegh Rabiee, and Joydeep Biswas. 2021. Robofleet: Open Source Communication and Management for Fleets of Autonomous Robots. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 406–412. https://doi.org/10.1109/IROS51168.2021.9635830
[81] Ashish Singh and James E Young. 2013. A Dog Tail for Utility Robots: Exploring Affective Properties of Tail Movement. In *Human-Computer Interaction – INTERACT 2013*. Springer Berlin Heidelberg, 403–419. https://doi.org/10.1007/978-3-642-40480-1_27
[82] Smith and Zeller. 2017. hitchbot: The risks and rewards of a hitchhiking robot.

Suomen Antropologi: Journal of the Finnish Anthropological Society 43, 3 (2017), 63–65. https://journal.fi/suomenantropologi/article/download/69710/30848

[83] Robert Sparrow. 2016. Kicking a robot dog. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 229–229. https://doi.org/10.1109/HRI.2016.7451756

[84] Ruth Maria Stock and Moritz Merkle. 2017. A service Robot Acceptance Model: User acceptance of humanoid robots during service encounters. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 339–344. https://doi.org/10.1109/PERCOMW.2017.7917585

[85] Li Sun, Zhi Yan, Sergi Molina Mellado, Marc Hanheide, and Tom Duckett. 2018. 3DOF Pedestrian Trajectory Prediction Learned from Long-Term Autonomous Mobile Robot Deployment Data. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 5942–5948. https://doi.org/10.1109/ICRA.2018.8461228

[86] Sofia Thunberg and Tom Ziemke. 2020. Are People Ready for Unexpected Encounters With Social Robots?. In *Workshop at the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. diva-portal.org. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1422130

[87] Vaibhav V Unhelkar, Claudia Pérez-D'Arpino, Leia Stirling, and Julie A Shah. 2015. Human-robot co-navigation using anticipatory indicators of human walking motion. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. 6183–6190. https://doi.org/10.1109/ICRA.2015.7140067

[88] A van Wynsberghe. 2016. Service robots, care ethics, and design. *Ethics and information technology* 18, 4 (Dec. 2016), 311–321. https://doi.org/10.1007/s10676-016-9409-x

[89] Dizan Vasquez, Billy Okal, and Kai O Arras. 2014. Inverse Reinforcement Learning algorithms and features for robot navigation in crowds: An experimental comparison. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1341–1346. https://doi.org/10.1109/IROS.2014.6942731

[90] Manuela Veloso, Joydeep Biswas, Brian Coltin, and Stephanie Rosenthal. 2015. CoBots: Robust Symbiotic Autonomous Mobile Service Robots. In *Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15)*. AAAI Press, Buenos Aires, Argentina, 4423–4429.

[91] Toby Walsh. 2017. *It's alive!: artificial intelligence from the logic piano to killer robots*. La Trobe University Press.

[92] Astrid Weiss and Christoph Bartneck. 2015. Meta analysis of the usage of the Godspeed Questionnaire Series. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. ieeexplore.ieee.org, 381–388. https://doi.org/10.1109/ROMAN.2015.7333568

[93] Ya-Huei Wu, Jérémy Wrobel, Mélanie Cornuet, Hélène Kerhervé, Souad Damnée, and Anne-Sophie Rigaud. 2014. Acceptance of an assistive robot in older adults: a mixed-method study of human-robot interaction over a 1-month period in the Living Lab setting. *Clinical interventions in aging* 9 (May 2014), 801–811. https://doi.org/10.2147/CIA.S56435

[94] Anxing Xiao, Wenzhe Tong, Lizhi Yang, Jun Zeng, Zhongyu Li, and Koushil Sreenath. 2021. Robotic Guide Dog: Leading a Human with Leash-Guided Hybrid Physical Interaction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 11470–11476. https://doi.org/10.1109/ICRA48506.2021.9561786

[95] Sangseok You and Lionel P Robert. 2022. Subgroup formation in human-robot teams: A multi-study, mixed-method approach with implications for theory and practice. *Journal of the Association for Information Science and Technology* 74, 3 (Feb. 2022), 323– 338. https://doi.org/10.1002/asi.24626

[96] James E Young, Youichi Kamiyama, Juliane Reichenbach, Takeo Igarashi, and Ehud Sharlin. 2011. How to walk a robot: A dog-leash human-robot interface. In *2011 RO-MAN*. 376–382. https://doi.org/10.1109/ROMAN.2011.6005225

[97] Azalea Yunus and Stacy A Doore. 2021. Responsible use of agile robots in public spaces. In *2021 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS)*. 1–5. https://doi.org/10.1109/ETHICS53270.2021.9632682

[98] Selma Šabanović. 2020. We're in This Together: Social Robots in Group, Organizational, and Community Interactions. In *Proceedings of the 8th International Conference on Human-Agent Interaction* (Virtual Event, USA) *(HAI '20)*. Association for Computing Machinery, New York, NY, USA, 3–4. https://doi.org/10.1145/3406499.3422314

A METHODOLOGICAL DETAIL

A.1 Pilot Study Details

Subsequent semantic difference questions not adapted from Godspeed were included to help characterize the Godspeed results. They included "The robot moved": "Too Close - Too Far." "The robot's motion was": "Erratic - Under Control."

Finally, a section of the questionnaire employed Likert scales measuring degrees of agreement with affective states and comfort with hypothetical encounters with the robot outside the laboratory setting. These were "When I encountered the robot I felt" ("Strongly disagree" to "Strongly agree"): "Curious," "Cautious," "Calm," "Excited." "I would be comfortable seeing this robot" ("Extremely uncomfortable" to "Extremely comfortable"): "Walking in an office," "Providing delivery services on campus," and "Providing delivery services to my home."

At the end of all five conditions, a final questionnaire was administered, asking participants to compare conditions on a subset of the questions administered between rounds. The scales are: "Doglike," "Friendly," "Autonomous," "Safe," "Under control," "Comfortable to get close to," "Comfortable seeing walk around an office," and "Comfortable seeing providing delivery services."

A.1.1 Pilot Study Additional Results. The scale "Remote-Controlled - Autonomous" is statistically significant by a one-way Analysis of Variances (ANOVA) ($F_{4,125} = 14.073$, $p < 0.001$). Using the Tukey post-hoc criterion, only contrasts against the "Remote-Controlled" (higher is more autonomous) condition are significant (all at $p < 0.001$; mean difference: "Fully-Autonomous" - 1.769, "Companion" - 1.846, "Leading" - 1.423, "Guided" - 1.615). The scale "Unconscious - Conscious" is statistically significant ($F_{4,125} = 4.428$, $p = 0.011$). Using the Tukey post-hoc criterion, most of the contrasts against the "Remote-Controlled" (higher is more autonomous) condition are again significant (mean difference: "Fully-Autonomous" - 1.038, $p = 0.33$; "Companion" - 1.192, $p = 0.009$; "Leading" - 0.808, $p = 0.161$; "Guided" - 0.923, $p = 0.077$).

Significant differences in average rankings on specific questions are shown in Figure 4.

A.2 Revised Study

The Godspeed questionnaire items used are reported in Table 2 and used a 6-point forced choice scale. The exploratory items include "The robot moved: Too Close - Too Far", "The robot moved: Too slow - Too fast", "The robot was Controlled - Autonomous", these three items also used 6 point forced choice scale. "When I encountered the robot, I felt Curious, Cautious, Calm, Excited, Nervous, Scared". These items used 5 point Likert scale ("Does not describe my feelings" to "Clearly describes my feelings"). "I would be comfortable seeing this robot walking in an office, classroom, library", "I would be comfortable seeing this robot providing delivery services on campus, to my home, serving food in a restaurant". These items used 5 point Likert Scale ("Extremely Uncomfortable" to "Extremely comfortable"). The ANOVA analysis results are reported in Table 3.

A.3 Protocol Changes for Revised Study

In addition to the most important changes described in Section 5.1, this section completes our account of changes made when revising the revised study protocol.

A.3.1 Encounter Protocol and Conditions. A possible explanation of the surprising null results from the pilot study advanced by several members of the research team was that the five-condition intra-participant design was subject to confounding factors such as a novelty effect, questionnaire fatigue, and/or satisficing, and herding. An inter-participant design would require selecting just two conditions to maintain study feasibility.

We selected the Autonomous and Leashed conditions on the basis of their significance ratings on the ranking survey. First, we excluded the Joystick condition as a candidate, since it was designed as a conceptual anchor for the other conditions. Of the remaining four, the Leashed condition was significantly ranked the most Dog-like, least autonomous, most under control, and most comfortable being close to. The Autonomous condition was at the opposite end of each of these measures. These results suggested to us that these two conditions had the highest likelihood of producing observable effects in the revised study.

A.3.2 Questionnaire Revision. Given our desire for enhanced methodological rigor and the fact that an inter-participant design alleviated participant fatigue considerations, we decided to substantially expand our questionnaire. We also shifted to the six-point scale for semantic difference question types.

Include more Godspeed questions. The intra-participant design required the repeated administration of the pilot study questionnaire, which made brevity an important concern. The inter-participant design alleviated these concerns, making the use of more Godspeed questions feasible.

Adopting a forced choice scale for Godspeed questions. While odd-numbered Likert scales are widely assumed to be preferable in all cases, semantic difference question types employed outside HRI typically utilize even-numbered scales to force participants to locate their perceptions on one side of the difference. Weiss and Bartneck's 2015 meta-analysis of the use of Godspeed questionnaires in HRI research [92] revealed a prevalent misconception of Godspeed's five-point scale as a Likert scale, clarified that Likert-type questions measure magnitude instead of difference, but still recommended the use of a five-point scale to preserve comparability with prior research. Despite these recommendations, we expected that an even-numbered, forced-choice design might reveal a signal obscured by the middle point of five or seven-point semantic difference scales.

Dedicated participant-facing research team and interview protocol. Interviews were considered as a possible component of the pilot study protocol but were not implemented due to the participant time investment required for 5 conditions and the training requirements. Having realized that interview data would be invaluable for interpreting surprising survey results, we decided to expand the study personnel to include experienced interviewers. Although the protocol was developed by an interdisciplinary team, the pilot study was conducted entirely by robotics researchers. Especially given the addition of an interview protocol, we established participant-facing and technical research teams, ensuring that the staffing of each was sufficient for each participant. This enabled authors with social science and interview experience to better contribute their expertise

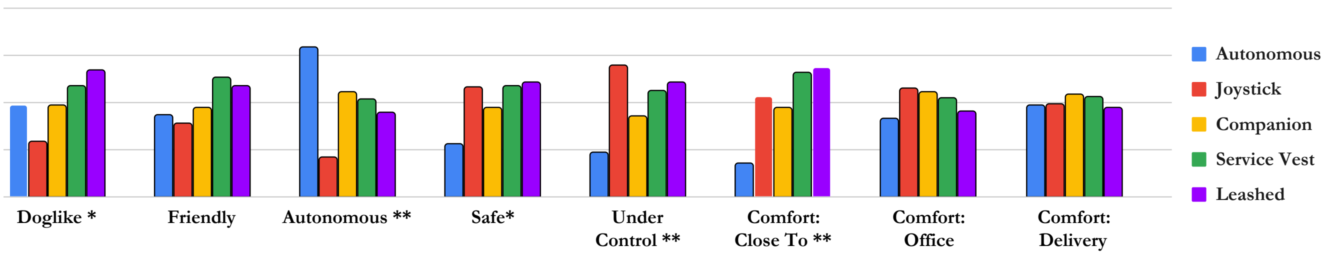


Figure 4: Rankings across each metric in the final questionnaire administered to study participants. Scales are reversed to simplify presentation: in this chart, 5 represents the highest raking, while 1 represents the lowest. * **- Significant at** $\alpha = 0.05$, ** **- Significant at** $\alpha = 0.01$

to qualitative data collection. It also helped standardize participant-researcher interactions, reducing the likelihood that these would be confounding variables. The utilization of a semi-structured interview after the collection of quantitative data allowed us greater freedom in targeting our questions and follow-up questions to better understand emerging trends and anomalies from our data, without biasing participants' responses. The protocol was modified slightly as the study went on, notably asking the participants to voluntarily share demographic information and asking about their tendency to anthropomorphize inanimate objects (inspired by the AMPH survey's success in supplementing Godspeed questions with participant baselines [12]). These changes were made towards the end of the guide when possible, to preserve parity of sequencing effects.

Simplification of robotic hardware and software setup. The robot platform in this study was simultaneously used in other studies, some of which had different configurations. What's more, the initiation of the social navigation stack was initially very complicated. Together, these factors led to delays and the need to debug hardware and software configurations at several points during the pilot study. In addition to decreasing the pace of research, the team reasoned that participant exposure to delays or technical difficulties could be another confounding factor. Thus, the robot's configuration was simplified to prevent difficult-to-resolve configuration problems and decrease the number and complexity of steps to initiate the robotic autonomy for the study.

Decrease avoidance initiation threshold. Upon reviewing the recordings of participant encounters, we determined that the $8.5m$ threshold used in the first study was too large to allow participant reactions to influence the robot's behavior.

Table 3: Full revised study survey results. Categories marked (Overall) sum relevant results above that point; all other categories sum following results. The only statistically significant result is Awful vs. Nice, but given the number of questions we believe this is not a significant result.

	Leashed		Autonomous			
	Mean	**σ**	**Mean**	**σ**	***F* value**	***p* value**
Cynomorphism	**3.455**	**0.89**	**3.546**	**0.88**	**0.06**	**0.81**
Fake v.s. Natural	3.818	1.08	3.727	1.35	0.03	0.86
Machine like v.s. Doglike	3.546	1.29	2.909	1.38	1.25	0.28
Unconscious v.s. Conscious	4.000	1.00	4.455	1.37	0.79	0.38
Artificial v.s. Lifelike	2.727	1.49	3.091	1.38	0.35	0.56
Rigid v.s. Elegant	3.182	1.47	3.546	1.29	0.38	0.55
Animacy	**4.000**	**1.07**	**4.227**	**1.33**	**0.19**	**0.66**
Inert v.s. Interactive	3.636	1.36	4.273	1.19	1.36	0.26
Apathetic v.s. Responsive	4.364	1.29	4.182	1.60	0.09	0.77
Likability	**4.477**	**0.94**	**3.932**	**0.75**	**2.27**	**0.15**
Dislike v.s. Like	4.818	0.98	4.000	1.10	3.40	0.08
Unfriendly v.s. Friendly	3.636	1.43	3.636	0.92	<0.01	1.00
Unpleasant v.s. Pleasant	4.546	1.13	4.091	1.14	0.89	0.36
Awful v.s. Nice	4.909	1.04	4.000	0.63	6.10	0.02
Perceived Intelligence	**4.636**	**0.45**	**4.409**	**0.86**	**0.60**	**0.45**
Incompetent v.s. Competent	4.636	0.50	4.455	0.82	0.39	0.54
Ignorant v.s. Knowledgeable	4.636	0.67	4.364	1.21	0.43	0.52
Godspeed (Overall)	**4.065**	**0.63**	**3.968**	**0.71**	**0.12**	**0.74**
Exploratory Questions	**4.000**	**0.61**	**3.879**	**0.50**	**0.26**	**0.62**
Unsafe v.s. Safe	4.455	1.21	4.818	0.98	0.60	0.45
Controlled v.s. Autonomous	5.000	1.00	4.182	1.33	2.66	0.12
Too close v.s. Too far	3.182	1.25	3.727	0.65	1.65	0.21
Too slow v.s. Too quick	3.818	0.60	3.727	0.47	0.16	0.70
Feelings (Neutral)	**4.000**	**0.77**	**4.000**	**0.92**	**<0.01**	**1.00**
Curious	4.727	0.65	4.636	0.92	0.07	0.79
Cautious	3.273	1.10	3.364	1.21	0.03	0.86
Feelings (Positive)	**3.500**	**0.97**	**3.364**	**0.90**	**0.12**	**0.74**
Calm	3.182	1.08	2.727	1.27	0.82	0.38
Excited	3.818	1.47	4.000	1.00	0.11	0.74
Feelings (Negative)	**2.046**	**0.99**	**2.546**	**1.11**	**1.25**	**0.28**
Nervous	2.455	1.21	3.000	1.26	1.07	0.31
Scared	1.636	1.03	2.091	1.22	0.89	0.36
Feelings (Overall)	**3.182**	**0.74**	**3.303**	**0.52**	**0.20**	**0.66**
Encounter	**2.818**	**1.06**	**3.061**	**0.66**	**0.41**	**0.53**
Office	3.182	1.25	3.000	1.00	0.14	0.71
Classroom	2.636	1.29	3.000	0.89	0.59	0.45
Library	2.636	1.29	3.182	1.17	1.08	0.31
Service	**3.970**	**1.04**	**4.030**	**0.94**	**0.02**	**0.89**
Campus	4.364	1.21	4.455	0.69	0.05	0.83
Home	4.091	1.38	4.000	0.89	0.03	0.86
Restaurant	3.455	1.13	3.636	1.50	0.10	0.75