2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) | 979-8-3315-8771-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/RO-MAN63969.2025.11217761

2025 34th IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 25-29, 2025. Eindhoven, The Netherlands

I Can't Help Myself! "Asking for Help" through an Elicitation Study in the Wild

Claire Liang*[1], Andy Elliot Ricci*[2], Malte F. Jung[3], Hadas Kress-Gazit[4]

***Abstract*— In this work, we examine robots "asking for help" in unpredictable human spaces. We focus on an open question particularly relevant for robots deployed in public– "how do people help robots?" We present an elicitation study that shows how asking for help in a real-world field study yields valuable and sometimes unexpected information. From our study, we examine strangers' responses toward a robot asking for spatial directions and extract valuable themes that can inform future asking-for-help systems. Our analysis provides a wide range of information, from geometric and topological information in natural language to details about rejection during an interaction. Further, we also provide anecdotes of valuable outlier behavior that can only be captured through a study in a real public space. Through our work, we highlight the importance of in-the-wild studies and discuss how the rich information they contribute will help robots effectively ask for help.**

I. Introduction

In real-world deployments, there can be gaps between expected performance and a robot's reality. In these scenarios, robots that can ask for help are better equipped to recover and succeed in their tasks. However, asking for help in the real world can be messy; the problems encountered in a lab study rarely capture the range of unexpected behaviors found in the real world. Often, these are most pronounced for robots deployed in public spaces because it is difficult to emulate organic interactions in the lab.

Some of these robots are currently deployed despite the difficulty of establishing ecological validity in a controlled study. For example, navigation robots in airports and other public spaces are already a billion-dollar industry [1]. Yet, these robots still have awkward interactions with the people they're supposed to assist [2]–[4], sometimes fail altogether, and struggle with the physical limitations of their systems [5]. If these navigating robots are to be deployed further, the success and adaptability of these robots may hinge on their ability to recover from failure in a human-friendly way.

In this paper, we show the value of investing in the problem of "asking for help" using an in-the-wild public study, especially for systems whose success may depend on brief interactions with strangers. Motivated by real-world failures of robots in airports, we use "direction giving" for robots navigating in human spaces as a case study. We show that by investigating the simple question of asking for directions, we can uncover a lot of unexpected, real, complex dynamics that can't be anticipated in theory or in the lab.

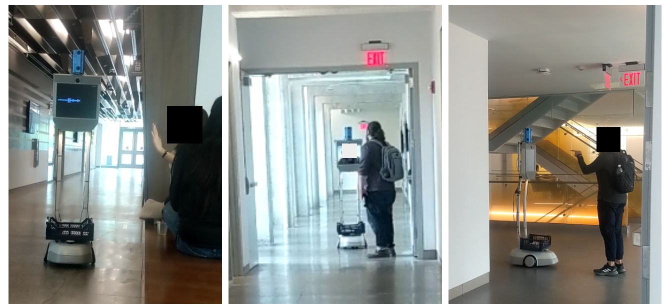


Fig. 1: A robot asking for directions during our in-the-wild study.

In this work, we analyze the natural language responses that robots' human counterparts provide as well as their reactions to different types of navigating robots via a post-interaction semi-structured interview. We present findings– such as how to express spatial information– that are important for relevant future navigating robot systems, but beyond that, we also include results that are informative for all systems that are deployed in these types of public spaces; some of the most valuable lessons from this study are the unexpected ones: communication outside of natural language utterances, angry outbursts, unanticipated group dynamics, and the complexity of rejection.

II. Related Work

Since there are a wide variety of scenarios in which a robot may ask for help, there is a large and diverse body of existing work about what strategies robots should take. Some algorithms enable robots to generate and respond to help requests [6], [7], systems that can execute the interaction for specific applications, and human-robot interaction studies exploring how people react to help requests [8], [9]. Each piece of existing research addresses important components of the robot asking for help process– such as how to formulate a help request or how to use a human response. This body of work provides many insights into human-robot interaction dynamics and algorithmic contributions for incorporating asking-for-help behavior in robot systems.

A. Aligning Psychology and Robotics: Understanding Asking for Help

To categorize existing work, we combine two perspectives: psychology and robotics theory. We find that the two angles

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA cyl48@mit.edu
[2]Digital & Computational Studies, Bates College, Lewiston, ME, USA aricci@bates.edu
[3]Department of Information Science, Cornell University, Ithaca, NY, USA malte.jung@cornell.edu
[4]The Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, USA hadaskg@cornell.edu
*Co-first authors. Supported by the National Science Foundation (IIS-1925100).

share key similarities, and by uniting them, we can understand the problem of asking for help from both the human and the robot's perspectives.

Theory from childhood cognitive development [10] provides a *general framework* for breaking down the process of asking for help. 1. "Initiation," 2. "Formulation" 3. "Expression," and 4. "Response evaluation and follow up." In [11]'s "*Checklist* for Needy Robots," the stages are: a. "Determining the robot needs help," b. "Who to ask, When to ask, and Where to ask," c. "What to ask and How to ask," and d. "What to do after receiving help".

We present a non-exhaustive collection of valuable existing asking-for-help literature through this "*checklist*" and "*framework*" lens.

1) Initiation: **Determining that the robot needs help**

For most "asking for help" systems, identifying when the robot is stuck or has encountered a failure [8] is equivalent to determining that the robot needs help [12]. All details of the process must be considered, such as anticipating an ask before failure [13] or optimizing timing [14].

2) Formulation: **Who, when, and where to ask**

Once the robot has determined that it needs to request help, it sets itself up to initiate an interaction. This means finding people to ask [15], determining when or if it is effective to ask a particular person for help [14], identifying the best location at which to ask [16], as well as when to and who to initiate an interaction [17].

3) Expression: **What to ask and how to ask**

Asking for help is more complex than a simple call of the word "help!" The way the robot calls attention to itself is a critical choice– [18] shows that while a robot might get more help when asking using natural language, requests communicated through beeping might evoke more empathy.

If using natural language, the execution of the request is important [19], as one must determine the correct level of specificity for the ask (lest the person deems the robot's request annoying [11]). Question generation is closely linked to language-generation systems as a whole [6], including template-based [20]and LLM approaches [21].

4) Response Evaluation and Follow Up: **What to do after receiving help**

Depending on the application, robots receive different kinds of help. In [12], the response to the robot's request for help is an action the human performs that allows the robot to achieve its goal. In other scenarios, the robot receives information, such as directions [22], rather than assistance.

When the help comes in the form of knowledge or information, the robot then has the opportunity to use it as it sees fit [23]–[26].

In other cases, the ask for help is not finished upon first exchange. The robot may need to explain its reason for failure [27] or follow up with clarification requests [28]. For example, [29] contributes a system that handles asking follow-up questions for a person's request to help disambiguate objects in the environment.

B. In-The-Wild Methodology

In human robot interaction communities, in-the-wild (sometimes referred to as field) studies are conducted in the real world, as opposed to a controlled laboratory setting [30] and are often used to *evaluate* real-world deployments of robot systems– for example, in airports [3]– and to *study how people interact* with robots– such as how people interact with delivery robots [31]. In-the-wild studies can be used to extract even more complex findings. For example, prior work has used emerging types of human-robot interactions to elicit responses, for example, to a robot trash barrel [32]. These studies provide valuable insights into human-robot interactions that inform robot design and deployment. For instance, [33] used prior in-the-wild HRI research findings to develop a placemaking framework for robot design. Overall, in-the-wild methods allow us to study robots in the real world, which can provide valuable insights into real-world human-robot interactions, providing a greater amount of ecological validity than controlled lab studies [30].

Because in-the-wild studies are less controlled than lab studies, the conclusions we can draw are different [30]. For example, in-the-wild studies allow us to learn about types of interactions that we would not have expected, including bullying behavior [34], [35]. The field of HRI has been reckoning with the question of how to balance the need for generalization and verifiability that comes from controlled studies with statistical significance and the ecological validity and findings from real-world studies [36], [37].

This paper is interested in how humans interact with robots in public spaces. Therefore an in-the-wild study is well-suited for this work. Our approach is similar to prior in-the-wild studies exploring how people interact with robots [38], [39].

III. Our Open Questions

Through our study we want to examine what humans bring to an asking-for-help interaction. Namely, what we should expect from them to best perform "Formulation" and "Expression." Having an understanding of these interactions would also inform "Response Evaluation and Follow Up'.'

As a result, we picked two directions of investigation that are relevant topics of interest shared across several of the existing methods and systems from the related work listed above.

The Language We Use to Help Robots We have systems that process help from humans and use that information in ways such as, failure recovery and map updates. However, the form of help expected in these systems can be narrow.

For example, most system contributions and deployed robots assume communication to primarily be natural language utterances, either written or spoken [40]. Those that go beyond natural language utterances may consider other inputs, such as gestures, to improve an interaction [23], [41], [42].

However, assumptions about natural language go beyond the primary use of utterances. It sometimes extends to the grammatical structure and content of the utterance. For example, a common implicit assumption for spatial tasks, such as navigation, is that robots should default to using landmarks in their natural language descriptions of paths. For example, while [43]'s compound action specification actions (travel, turn, face, verify, and find) do not require the use of landmarks in directions, when sentence planning, the iteration is "constrained to use only objects and properties of the environment visible to the follower" implicitly stating that the use of landmarks is an assumption for this model.

Although rooting directions in landmarks is common and caters well to semantic maps, using landmarks is not universal. [23] finds that people initially gave spatial directions with metric information and then shifted to landmark-based communication throughout a longer interaction.

In our study, we seek to pull on these threads and understand how conveying spatial information may impact how humans communicate with robots and what part natural language plays in their instructions.

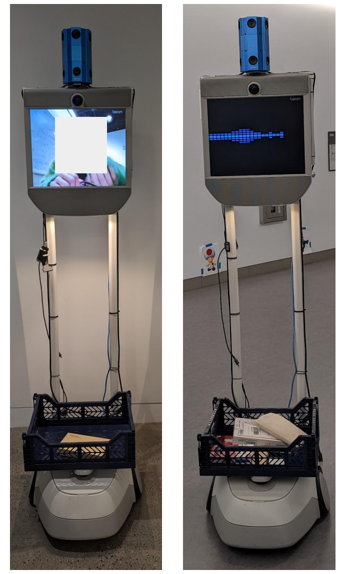
Robot Form's Impact on Human Behavior: We know that changing the robot's physicality can impact human-robot interaction. Prior work in asking for help has studied the impact factors such as the robot's visual cues [44], the robot's behavior [42], and the robot's level of autonomy [45]. [45] found that people helped a robot they perceived to be fully autonomous more quickly than one they believed to be teleoperated.

In HRI more broadly, we see some discussion of ideas about how robot factors influence interactions in theory on perceptions and attitudes towards social robots [46], [47], including anthropomorphism [48], and depiction theory [49]. In this work, we are interested in how two robot forms already deployed in real systems affect humans' responses, especially regarding the language they use and their attitudes toward the robot's request.

[45]'s finding about a difference in reactions towards telepresence robots and perceived autonomous robots is particularly relevant for robots asking for help in the wild, as telepresence robots are already widely deployed in public spaces today. To investigate the ecological validity of their findings further, we selected these two robot forms for further investigation.

IV. Our study

We conducted an in-the-wild study with two robot forms to understand how humans give directions to robots. We focus on a specific type of help request, spatial directions [22], [23], and explore both "the language of help" and "the impact of robot form". In the study, we remotely operated a Beam robot to approach people in public spaces and ask for directions to other locations in the building. To explore robot form, we ran the study under two different robot conditions already used in real-world systems: a telepresence robot and an autonomous robot. We used a Wizard of Oz approach for the autonomous robot condition [24]. An early version of this work without results and analysis is available in [50].



Telepresence Autonomous

Fig. 2: The two study conditions. Both robots are equipped with a basket of items for delivery. The telepresence robot is shown on the left with the operator's face on the screen. The perceived autonomous robot shows an example of the sound visualization mid-speech.

A. Telepresence Robot

We used a Beam Pro telepresence robot for both study conditions, as shown in Fig. 2. An experimenter, referred to as the robot operator or remote user, remotely operated the robot using the provided interface, shown in Fig. 3. The experimenter had video streams from two cameras on the robot and interacted with participants through the video-conferencing capabilities. In the telepresence condition, the experimenter's face was shown on the robot's screen, and their voice was not changed. In the autonomous condition, the experimenter's face was replaced with a sound wave visualization of their speech, and the experimenter's voice was augmented to sound robotic. The visualization was generated in real-time using OBS Studio [51], and the voice was modulated with VoiceMod [52].

B. Study Location

We conducted the study in two university buildings, Building A and Building B. We used only public spaces and conducted each study session on two floors of one of the buildings. We conducted three sessions in Building A and one session in Building B. Building A has an L-shaped hallway that connects multiple open spaces, and Building B has an H-shaped layout with one large open space and hallways connecting multiple closed rooms.

The experimenter chose the goal locations based on the robot's current location. Before the study, we selected a set of locations for each building, including elevators, water fountains, lounges, and lecture halls. The robot operator asked for directions to goal locations that were between 2 and 20 meters from the robot's location.

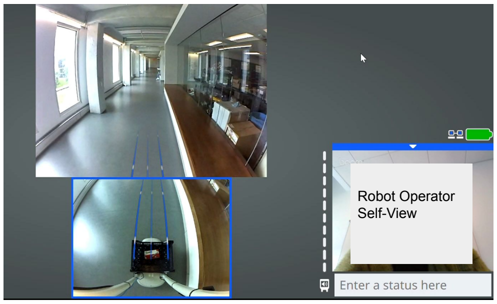


Fig. 3: The control interface for the Beam robot. The front camera view is shown on the top left, the downward-facing camera is shown on the bottom left, and the robot operator's view– what a participant sees on the screen in the telepresence condition– is on the bottom right.

C. Procedure

During the study, one researcher was the robot operator and remotely controlled the robot. The other researcher was with the robot to obtain consent and conduct interviews. In each study session, the robot operator drove the robot around the study location to look for people to ask for directions. When the robot operator saw someone, they approached them and asked for directions. The opening script was:

'Hi, can I ask for directions?"

If the person said yes or otherwise indicated that they were willing to engage with the robot, the operator proceeded to ask for directions to a goal location:

"Do you know how to get to [goal location]?".

After the participant gave directions, the operator drove away, following the provided directions as much as possible. Then, the local experimenter approached the participant, obtained consent, and asked if they could conduct a brief interview. Not all participants provided interviews.

In cases where a person was prompted for directions and either ignored the robot, said no, or otherwise indicated that they didn't want to give directions, the operator did not ask them for directions again. These people did not give consent and are not counted as study participants. Thus, we are unable to analyze the video collected during those interactions.

If a person asked what the robot was doing or questioned why it asked for directions, the operator explained that it was performing a delivery task. The local experimenter periodically changed the item in the robot's basket to provide visual evidence of the delivery task.

The robot continuously moved throughout the study space during each session and changed floors. This movement helped us sample participants who had not observed the robot interact with other people.

1) Semi-Structured Interview: The local experimenter conducted optional semi-structured interviews with participants who gave directions to the robot. The two guiding interview questions were the following:

- "How was it interacting with this robot and giving directions?"
- "Do you think you give directions differently to the robot than you would to me?"

The experimenter adjusted the interview questions based on the participant's responses to understand their experiences with the robot.

D. Data Collection & Analysis

We recorded video from the robot's cameras to capture all of the interactions with the robot and audio-recorded the semi-structured interviews. We used the footage from each consenting participant to create a transcript of the directions given, both nonverbally and through gestures (such as pointing). We used the audio recordings to transcribe the semi-structured interviews.

We performed a qualitative analysis of the transcribed interview data for the directions given and how participants perceived the robot's identity. The two experimenters analyzed the transcribed data using a thematic analysis process [53] and iteratively constructed codes and themes.

For the "directions given," we grounded our data analysis in the existing literature, including the three forms of spatial arrangement maps described in [9]. After breaking the direction transcripts into sentences, we assigned codes as follows (more specific definitions are in Section IV-D.1):

- A fixed set of phrases, "this/that way", "go straight",..., "turn left/right" are labeled as topological. Each interaction is labeled as "topological" or "metric" if all sentences are consistent in one category. Any with both are labeled as "hybrid."
- Landmarks are destinations or objects described with enough detail that they are uniquely distinguishable and used inextricably from the direction instructions.
- Gestures are labeled as "crucial" if the utterances are ambiguous without the gesture's inclusion.

We did not evaluate the directions for correctness since we were only interested in how participants gave directions and not how well they knew the study locations.

1) Definitions for data analysis:

Topological vs Metric directions We consider a sentence in a set of directions to be metric if it contains metric information, such as a distance or an angle of rotation– directing towards a single specific trajectory.

We consider topological directions to be directions that dictate a class of paths up to interpretation for the robot. A sentence is counted as topological if either 1. It uses landmarks 2. It assumes that the robot can determine how far it needs to go or how much it needs to turn (oftentimes, this is paired with gestures). For point 2, key phrases are *'"this way", "that way"*, *"down [that] way"*, turn *"left"* or *"right"*, or *your "left" or "right"*, *"turn around"*, *"go straight"*, *"over there"* with no further clarification.

We take each participant's interaction and analyze their utterances sentence by sentence. Each sentence is considered as "metric" or "topological." For the entire interaction, we assign one of three labels: "metric", "topological", or "hybrid". A set of directions is considered metric if it exclusively uses metric sentences, it is considered topological if it exclusively uses topological directions, and it is considered hybrid if it contains a mixture of both.

Landmarks We count a destination or object as a landmark if it is described with enough detail that it is distinguishable in the scene and is also used in an inextricable way from the direction instructions. For example, in the phrase: "Turn at the first doors on the left," the doors are a landmark since enough information is given that the doors are identifiable (they are the first on the left), and the action "turn" is tied to the doors. The phrase: "The door is over there" does not contain a landmark because although the doors are uniquely identifiable, there are no direction instructions associated with the doors.

Gesture We categorized gestures into two categories: "crucial" and "non-crucial." Gestures are considered "crucial" if the utterances are ambiguous without adding the information from the gesture. Gestures are "non-crucial" if the utterances can be interpreted independently and the gestures align with the directions.

E. Participants

To best capture natural interactions, we engaged with people in university building public spaces and then informed participants that the robot was part of a study only after their interactions. We did not collect any demographic data from the participants; however, since the study was conducted in a university building, it was likely that most participants were students, faculty, and staff.

A researcher approached participants after they gave directions to the robot to give them an information sheet about the study and to obtain verbal consent. The interactions of people who did not give verbal consent are excluded from the analysis. We recorded video from the robot during the study. We posted "Recording in Progress" signs in the study locations with the researcher's contact information, allowing people to opt out of being recorded. The study was approved by Cornell University IRB (IRB0145990).

We conducted two study sessions per condition across four different days. We operated for about 2.5 hours per condition for about 5 hours. We exclude a fifth session due to a video-recording error. A total of 111 people gave directions to the robot and consented to analysis.

Our dataset contains 48 transcribed interviews, 24 per condition; not all participants agreed to be interviewed. We excluded 18 participants from the analysis because they either knew the robot operator, knew about the research project, or had previously participated in the study. After the exclusions, we had 46 participants in the telepresence condition and 47 in the autonomous condition. Some participants gave directions in groups. Thus, we had 34 direction-giving interactions per condition and 68 total.

V. Results

A. Communication of spatial information

1) Metric vs Topological: We found that only 3 out of the 68 interactions contained hybrid instructions, 2 in the autonomous condition, and 1 in the telepresence condition. All other instructions were strictly topological.

Condition	# Crucial	# Non-Crucial Only	# No Gesture
Telepresence	27	7	0
Autonomous	20	12	2

TABLE I: Interactions that contained crucial gestures, non-crucial gestures only, or no gestures.

Condition	#Gest only	#Land only	#Gest + Land	#Neither
Telepresence	11	1	16	6
Autonomous	12	5	8	9

TABLE II: Interactions that included crucial gesture only, landmarks only, both crucial gesture and landmarks, or neither

Direction Example 1:

> "You're gonna want to turn, uhh, what– let's say 130 degrees from where to the right from where you now are. You're gonna want to go straight about 20 feet and make a left... go straight until you can't go anymore and then make another left. Just keep going and you'll see the water fountains." (P6, Telepresence Condition)

We present Direction Example 1 to show metric and topological language side by side; we refer to this as hybrid directions. Here, the metric indicators are the 130 degrees and the 20 feet. Topological indicators are "make another left" and "keep going" and the landmark indicator "until you can't go anymore."

The use of topological directions and landmarks does align the structure of directions described in [9] as well as the dialogue in studies such as [25], [54]. This common recurring assumption in the literature has ecological validity supported by the findings of this elicitation study.

2) Gesture: We found that participants gestured while giving directions to the robot in all but two interactions. Furthermore, we observed that *some gestures contained critical information that was not communicated verbally*. We refer to these as crucial gestures.

For example, the verbal directions "Yeah, that way" are incomplete without the accompanying pointing gesture (P 56, Telepresence Condition). The number of interactions containing gestures is shown in Table I.

Another notable finding of the field study was how integrated gestures and landmarks were to the participants' instructions. As seen in Table II, for the telepresence condition, 82% of the interactions used crucial gestures or landmarks, and 44% used both. In the autonomous condition, 74% of the interactions contained crucial gestures or landmarks, and 18% used both.

Direction Example 2

> "Um, I guess for you you'd have to go up the ramp. It would be best to [pointing] go over there all the way around past the tables go up the ramp and then you get to the vending machines outside the bathrooms, kind of more towards the

end [gesturing backwards] than the elevator end."
(P 53, Autonomous Condition)

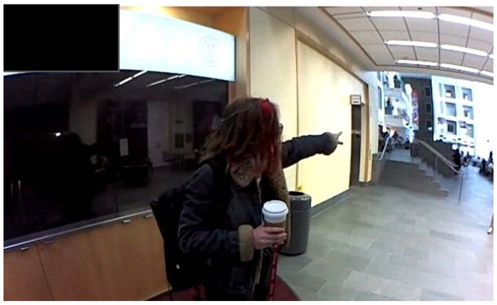


Fig. 4: A participant using a crucial pointing gesture when giving directions in a public space.

This is an example of the use of a landmark and two crucial gestures. The phrase: "the tables" represents a unique identifying object and, therefore, is a landmark. The pointing, shown in Fig. 4 before "over there" is part of the unique identification of the landmark because there are many groups of tables in the location. Therefore, the pointing associated with "over there" is necessary to identify the landmark and is counted as a crucial gesture. The "gesturing backwards" is also a crucial gesture because excluding the gesture from the directions makes the meaning of "towards the end... than the elevator end" unclear. The inextricability of the use of gesture and natural language utterances is particularly interesting and discussed more in Section VI

B. In-the-wild public deployment: Rejection

Condition	# Asked	# Gave Directions	% Gave Directions
Telepresence	51	34	67%
Autonomous	151	34	23%

TABLE III: Number of study participants. "% Gave Directions" is the percentage that gave directions after we asked.

Although some people went the extra mile when providing help, extending beyond what the robot requested, for instance, offering to walk the robot to its destination or open doors (15%: 3% AI, 26% telepresence). One of the most interesting results from this study was the "rejections". Much of the existing work investigating asking-for-help interactions assumes that the person is invested in assisting. However, we find that when deployed in public spaces, the reality is not as optimistic.

We analyzed when people turned down the robot when it asked for help. Telepresence impacted people's willingness to help the robot and how people talked about the robot during interviews. Only 23% of people gave directions when the robot was autonomous, versus the 67% in telepresence; breakdowns of the number of people asked compared to the number that provided directions are shown in Table III. Still, there was not a detectable difference in the directions themselves. This differed from our expectations. Our findings contrasted with our assumptions that were based on [45]'s findings from their teleoperated robot. This may be influenced by the difference in task or robot form.

We also found that interactions with the robot did not fall into a binary dichotomy of "ignored" or "provided help." People were taking pictures of the robot and staring at it, as seen in other relevant in-the-wild studies [31], but also other more complex social interactions.

For example, in one interaction, the robot approached two people to ask for help, and one person (A) rejected the robot's request and hid from the robot, while the other (B) engaged in an interaction with a third person (C) who was passing by. The first person (A) then joined the other conversation (with B and C) and B began to answer our request for help while the other two (A and C) continued conversing.

Additionally, there was a *range in what a "rejection" could look like*. A rejection could be a person entirely ignoring the robot's request or expressing they would not help the robot. Of the latter category, there was also a range in responses; most rejected the robot by saying "no" or "I'm sorry," or "I don't know," maintaining some level of politeness [55] despite not being prosocial.

There were also more violent rejections of the robot as well. For example, we had one person face the robot in the telepresence condition, look at the operator through the screen, and say "f*** off evil robot and die," implicitly *rejecting the robot's request for help, eschewing politeness, not being prosocial, and doing so directly to the face of the operator*.

VI. Discussion

In this work, we conducted an elicitation study to explore the nature of how people provide directions to robots in public spaces. Through this, we show a way to probe underlying assumptions that persist through existing asking-for-help work, tie in theory from psychology and linguistics, and address open questions about how people interact with these systems.

A. The Language for Spatial Information

The finding that everyone used topological language when giving directions provides strong evidence that planning topologically and creating interfaces that leverage topological information rather than geometric information [5] can lead to more fluid and natural interactions. In fact, many existing asking-for-help systems already use landmarks for their topological or semantic maps. These design choices can now be motivated by real-world evidence from our field study.

The results from this study also stress the importance of gesture interpretation tools [56], [57] for real-world interactions. The fact that crucial gestures were so pervasive in our data calls to linguistics literature. In linguistics indexicals [58] contain a set of demonstrative pronouns such as "this" and "that". The use of these indexicals integrated with spatial representation makes them spatial deictic terms. Spatial deixis are inherently grounded in space, centered around the location object they are referring to [59]. Therefore, the meaning of "that" as used in our study, is baked into the location of the "hallway" it refers to. Empirical research in children's speech development claims that the use of spatial deixis, “accompanied by a gesture from the eyes, head, or hands, towards the entity or event in question . . . this description of terms. . . [is] among the first to appear in children’s speech." [60] It seems that from an early age, humans entangle indexicals with gesture, and our study further emphasizes this fact [61].

The finding that so many participants used gestures and landmarks inextricably, and that certain landmarks only became uniquely identifiable with gesture, shows that the two may be even more integrated for robots asking for help than previously thought. This finding suggests that when creating robot systems for direction generation and understanding, it may be useful to *consider landmarks and gesture as an intentionally integrated unit* when defining abstractions.

B. The Complexity of Rejection

Through our study, we found interactions that demonstrate how multifaceted the act of “rejection” can be. For example, while people were less likely to reject a robot that was not perceived as autonomous, our most extreme rejection actually occurred with a real person’s face on the screen.

Rejection also becomes complicated when people are already in a group, such as when the robot approaches two people to ask for help Section V-B. There wasn’t a clear answer for who rejected the robot and when. Beyond that, it was sometimes unclear when each person’s interaction with the robot “began.” We often think about these interactions in dyads, but *asking a group versus an individual can increase the complexity of these interactions*; that is not explored extensively in the current literature.

Our experience with the one person who told the robot to “f*** off [...] and die" calls back to in-the-wild deployments of different robots have reported unexpected bullying behavior [34], which also reflects the extreme forms of rejection we saw in our work, for example, [35] reported that people punched, kicked, and slapped their robot. From the combination of our findings alongside the existing work in this space, perhaps *willingness to help is more complex than a binary dichotomy* that is assumed in other literature. Some rejections are harsher than others, and some forms of help are more generous. Future systems may need to anticipate these incidents on more continuous scales of rejection type and severity.

C. Limitations & Future Work

The design of the “autonomous” robot influenced participant interactions. The robot had a black screen when the operator was not speaking. This may have influenced participant interactions. Future work includes investigating the impact of a wider range of robot forms.

While our conclusions— that there exist behaviors in the real world not found in a lab study— extend to a general population, our work is still limited by being run exclusively at a university with a narrow sample size. Investigating a more general population has great potential in identifying even more valuable insights and ecological validity.

VII. Conclusion

Through this in-the-wild study of a robot asking for directions in a public place, we not only found meaningful answers to our posed questions about the use of landmarks and the impact of robot form, but we also discovered how inextricable gestures are from people’s utterances, and the complexity of rejection. The study data will be made available (https://cyl48.github.io/). This work helps us expand upon how we think about robots asking for help and identifies informative findings for designing future robotic systems that can expect the unexpected.

References

[1] G. M. Insights, “Airport robots market size, share and forecast analysis - 2032.” [Online]. Available: https://www.gminsights.com/industry-analysis/airport-robots-market

[2] M. Joosse, M. Lohse, and V. Evers, “How a guide robot should behave at an airport insights based on observing passengers,” *CTIT Technical Report Series*, no. TR-CTIT-15–01, 2015.

[3] M. Joosse and V. Evers, “A guide robot at the airport: First impressions,” in *ACM/IEEE International Conference on Human-Robot Interaction*, 2017.

[4] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, D. Cremers, V. Evers, M. Fiore *et al.*, “Spencer: A socially aware service robot for passenger guidance and help in busy airports,” in *Field and service robotics*. Springer, 2016, pp. 607–622.

[5] C. Liang, R. A. Knepper, and F. T. Pokorny, “No map, no problem: a local sensing approach for navigation in human-made spaces using signs,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 6148–6155.

[6] S. Tellex, N. Gopalan, H. Kress-Gazit, and C. Matuszek, “Robots that use language,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 25–55, 2020.

[7] M. Marge, C. Espy-Wilson, N. G. Ward, A. Alwan, Y. Artzi, M. Bansal, G. Blankenship, J. Chai, H. Daumé III, D. Dey *et al.*, “Spoken language interaction with robots: Recommendations for future research,” *Computer Speech & Language*, vol. 71, 2022.

[8] S. Honig and T. Oron-Gilad, “Understanding and resolving failures in human-robot interaction: Literature review and model development,” *Frontiers in psychology*, vol. 9, p. 861, 2018.

[9] C. Landsiedel, V. Rieser, M. Walter, and D. Wollherr, “A review of spatial reasoning and interaction for real-world robotics,” *Advanced Robotics*, vol. 31, no. 5, pp. 222–242, 2017.

[10] S. Ronfard, I. M. Zambrana, T. K. Hermansen, and D. Kelemen, “Question-asking in childhood: A review of the literature and a framework for understanding its development,” *Developmental Review*, vol. 49, pp. 101–120, 2018.

[11] N. Backhaus, P. H. Rosen, A. Scheidig, H.-M. Gross, and S. Wischniewski, “Somebody help me, please?!” Interaction design framework for needy mobile service robots,” in *IEEE workshop on advanced robotics and its social impacts (ARSO)*. IEEE, 2018.

[12] R. A. Knepper, S. Tellex, A. Li, N. Roy, and D. Rus, “Recovering from failure by asking for help,” *Autonomous Robots*, vol. 39, pp. 347–362, 2015.

[13] R. Papallas and M. R. Dogar, “To ask for help or not to ask: A predictive approach to human-in-the-loop motion planning for robot manipulation tasks,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 649–656.

[14] M. Kimoto, Y. Yasumatsu, and M. Imai, “Help-Estimator: Robot requests for help from humans by estimating a person’s subjective time,” *International Journal of Social Robotics*, pp. 1–14, 2022.

[15] S. Rosenthal and M. Veloso, “Mobile robot planning to seek help with spatially-situated tasks,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, no. 1, 2012, pp. 2067–2073.

[16] S. Rosenthal, M. Veloso, and A. K. Dey, “Is someone in this office available to help me? Proactively seeking help from spatially-situated humans,” *Journal of Intelligent & Robotic Systems*, 2012.

[17] J. Baraglia, M. Cakmak, Y. Nagai, R. P. Rao, and M. Asada, “Efficient human-robot collaboration: when should a robot take initiative?” *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 563–579, 2017.

[18] D. G. Holm, R. P. Junge, M. Østergaard, L. Bodenhagen, and O. Palinko, "What will it take to help a stuck robot? exploring signaling methods for a mobile robot," in *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 797–801.

[19] S. Tellex, R. Knepper, A. Li, D. Rus, and N. Roy, "Asking for help using inverse semantics," *Robotics: Science and Systems X*, 2014.

[20] T. Fong, C. Thorpe, and C. Baur, "Robot, asker of questions," *Robotics and Autonomous systems*, vol. 42, no. 3-4, pp. 235–243, 2003.

[21] A. Z. Ren, A. Dixit, A. Bodrova, S. Singh, S. Tu, N. Brown, P. Xu, L. Takayama, F. Xia, J. Varley *et al.*, "Robots that ask for help: Uncertainty alignment for large language model planners," *arXiv preprint arXiv:2307.01928*, 2023.

[22] A. Weiss, J. Igelsböck, M. Tscheligi, A. Bauer, K. Kühnlenz, D. Wollherr, and M. Buss, "Robots asking for directions—the willingness of passers-by to support robots," in *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 23–30.

[23] M. Marge, C. Bonial, A. Foots, C. Hayes, C. Henry, K. Pollard, R. Artstein, C. Voss, and D. Traum, "Exploring variation of natural human commands to a robot in a collaborative navigation task," in *Proceedings of the first workshop on language grounding for robotics*, 2017, pp. 58–66.

[24] L. D. Riek, "Wizard of Oz studies in HRI: a systematic review and new reporting guidelines," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 119–136, 2012.

[25] S. Hemachandra, F. Duvallet, T. M. Howard, N. Roy, A. Stentz, and M. R. Walter, "Learning models for following natural language directions in unknown environments," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5608–5615.

[26] S. Patki, A. F. Daniele, M. R. Walter, and T. M. Howard, "Inferring compact representations for efficient natural language understanding of robot instructions," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 6926–6933.

[27] D. Das, S. Banerjee, and S. Chernova, "Explainable ai for robot failures: Generating explanations that improve user assistance in fault recovery," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 351–360.

[28] T. Williams, F. Yazdani, P. Suresh, M. Scheutz, and M. Beetz, "Dempster-shafer theoretic resolution of referential ambiguity," *Autonomous Robots*, vol. 43, pp. 389–414, 2019.

[29] F. I. Doğan, I. Torre, and I. Leite, "Asking follow-up clarifications to resolve ambiguities in human-robot conversation," in *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 461–469.

[30] M. Jung and P. Hinds, "Robots in the wild: A time for more robust theories of human-robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 7, no. 1, pp. 1–5, 2018.

[31] A. Dobrosovestnova, I. Schwaninger, and A. Weiss, "With a little help of humans. an exploratory study of delivery robots stuck in snow," in *31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 1023–1029.

[32] B. Brown, F. Bu, I. Mandel, and W. Ju, "Trash in motion: Emergent interactions with a robotic trashcan," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, 2024, pp. 1–17.

[33] H. R. M. Pelikan, B. Mutlu, and S. Reeves, "Making sense of public space for robot design," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '25. IEEE Press, 2025, p. 152–162.

[34] D. Brščić, H. Kidokoro, Y. Suehiro, and T. Kanda, "Escaping from children's abuse of social robots," in *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*, 2015, pp. 59–66.

[35] P. Salvini, G. Ciaravella, W. Yu, G. Ferri, A. Manzi, B. Mazzolai, C. Laschi, S.-R. Oh, and P. Dario, "How safe are service robots in urban environments? Bullying a robot," in *IEEE International Symposium in Robot and Human Interactive Communication*, 2010.

[36] G. Hoffman and X. Zhao, "A primer for conducting experiments in human–robot interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, pp. 1–31, 2020.

[37] K. Fischer, "Effect confirmed, patient dead: A commentary on Hoffman & Zhao's primer for conducting experiments in HRI," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 1, 2021.

[38] A. Fallatah, B. Chun, S. Balali, and H. Knight, """ would you please buy me a coffee?" how microculturesimpact people's helpful actions toward robots," in *Proceedings of the ACM Designing Interactive Systems Conference*, 2020, pp. 939–950.

[39] P. Heikkilä, H. Lammi, M. Niemelä, K. Belhassein, G. Sarthou, A. Tammela, A. Clodic, and R. Alami, "Should a robot guide like a human? a qualitative four-phase study of a shopping mall robot," in *11th International Conference on Social Robotics*. Springer, 2019.

[40] V. Blukis, C. Paxton, D. Fox, A. Garg, and Y. Artzi, "A persistent spatial semantic representation for high-level natural language instruction execution," in *Conference on Robot Learning (CoRL)*, 2022.

[41] K. Chandan, V. Kudalkar, X. Li, and S. Zhang, "Arroch: Augmented reality for robots collaborating with a human," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3787–3793.

[42] K. Fischer, B. Soto, C. Pantofaru, and L. Takayama, "Initiating interactions in order to get help: Effects of social framing on people's responses to robots' requests for assistance," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 999–1005.

[43] A. F. Daniele, M. Bansal, and M. R. Walter, "Navigational instruction generation as inverse reinforcement learning with neural machine translation," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 109–118.

[44] M. Westhoven, T. van der Grinten, and S. Mueller, "Perceptions of a help-requesting robot-effects of eye-expressions, colored lights and politeness of speech," in *Proceedings of Mensch und Computer*, 2019, pp. 43–54.

[45] V. Srinivasan and L. Takayama, "Help me please: Robot politeness strategies for soliciting help from humans," in *Conference on human factors in computing systems (CHI)*, 2016.

[46] H. Mieczkowski, S. X. Liu, J. Hancock, and B. Reeves, "Helping not hurting: Applying the stereotype content model and bias map to social robotics," in *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 222–229.

[47] B. Reeves, J. Hancock *et al.*, "Social robots are like real people: First impressions, attributes, and stereotyping of social robots," 2020.

[48] P. A. Ruijten, A. Haans, J. Ham, and C. J. Midden, "Perceived human-likeness of social robots: testing the Rasch model as a method for measuring anthropomorphism," *International Journal of Social Robotics*, vol. 11, pp. 477–494, 2019.

[49] H. H. Clark and K. Fischer, "Social robots as depictions of social agents," *Behavioral and Brain Sciences*, vol. 46, p. e21, 2023.

[50] C. Liang, A. E. Ricci, H. Kress-Gazit, and M. F. Jung, "Lessons from a robot asking for directions in-the-wild," in *Companion of the International Conference on Human-Robot Interaction*, 2023.

[51] OBS Project, "OBS studio." [Online]. Available: obsproject.com

[52] Voicemod, "Voicemod." [Online]. Available: www.voicemod.net/

[53] V. Clarke, V. Braun, and N. Hayfield, "Thematic analysis," *Qualitative psychology: A practical guide to research methods*, vol. 3, pp. 222–248, 2015.

[54] A. F. Daniele, M. Bansal, and M. R. Walter, "Natural language generation in the context of providing indoor route instructions," in *Proceedings Robotics: Science and Systems Workshop on Model Learning for Human-Robot Communication*, 2016.

[55] F. H. Al-Hindawi and M. A. R. Alkhazaali, "A critique of politeness theories," *Theory and Practice in Language Studies*, vol. 6, no. 8, p. 1537, 2016.

[56] D. Shukla, O. Erkent, and J. Piater, "Probabilistic detection of pointing directions for human-robot interaction," in *International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 2015, pp. 1–8.

[57] M. Van den Bergh, D. Carton, R. De Nijs, N. Mitsou, C. Landsiedel, K. Kuehnlenz, D. Wollherr, L. Van Gool, and M. Buss, "Real-time 3D hand gesture interaction with a robot for understanding directions from humans," in *International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2011, pp. 357–362.

[58] J. Perry, "Indexicals and demonstratives," *A Companion to the Philosophy of Language*, pp. 586–612, 1997.

[59] S. C. Levinson, "Deixis," in *The handbook of pragmatics*. Blackwell, 2004, pp. 97–121.

[60] C. Tanz, "Studies in the acquisition of deictic terms," *Cambridge Studies in Linguistics London*, vol. 26, pp. 1–184, 1980.

[61] F. Esseili, "Spatial deixis and gesture in English: Adults vs. children," *The Purdue Linguistic Association Working Papers*, vol. 1, no. 1, 2008.