Lessons Learned from Interaction Studies with Delivery Robots

Heqiu Song
Chair for Individual and Technology
RWTH Aachen University
52062 Aachen, Germany
Email: heqiu.song@itec.rwth-aachen.de

Astrid M. Rosenthal-von der Pütten
Chair for Individual and Technology
RWTH Aachen University
52062 Aachen, Germany
Email: arvdp@itec.rwth-aachen.de

***Abstract*—The sharp rise in demand for last-mile delivery services has led to the development of delivery robots designed to navigate pedestrian pathways. Effective signaling cues between these robots and pedestrians, also called Incidentally Co-present Persons (InCoPs), are essential to prevent collisions in narrow passageways. This paper presents a range of investigations, including video-based laboratory experiments and field observations, to explore the impact of different signaling cues on human-robot interactions in real-world settings. This paper showcases some of the main results and highlights some key challenges when conducting studies with delivery robots.**

I. INTRODUCTION

Over the past decade, the rise of autonomous robots in urban environments has driven the development of autonomous delivery solutions designed to meet the increasing customer expectations for timely deliveries [10]. These robots, coexisting with pedestrians, face the challenge of safe interaction, especially with Incidentally Co-present Persons (InCoPs) [13], who are individuals not directly involved in the delivery process. Understanding how signaling cues affect collision avoidance is crucial. While research on autonomous vehicles (AVs) and pedestrian interactions provides useful insights [8, 15], the unique scenarios involving delivery robots on pedestrian pathways require further investigation.

Studies have investigated the impact of different signaling cues on pedestrians' perceptions and responses, for instance, Angelopoulos et al. [1] conducted a study using a humanoid robot (i.e., Pepper) and found that deictic gestures as navigational cues result in fewer navigation conflicts compared to a simulated gaze. Additionally, Kannan et al. [6] found that displays are preferred over lights for conveying intent, with the preferred type of content varying according to the scenario. These cues are essential for conveying a delivery robot's intentions, particularly in narrow passageways where clear communication is crucial to prevent collisions. However, pedestrian attentiveness, often compromised by digital distractions like mobile phone usage, can significantly affect the effectiveness of these signaling cues. Reduced awareness impairs pedestrians' ability to respond to cues effectively, emphasizing the need to explore the correlation between attentiveness levels and cue efficacy [12, 2]. Furthermore, environmental crowdedness significantly influences interaction dynamics. Research indicates that navigating dense crowds poses challenges for robots and potential safety hazards [18]. Therefore, understanding how signaling cues are perceived and responded to in crowded environments is vital for ensuring safe interactions.

Although researchers have tested different navigation cues [4], their studies have been limited to comparing only 2 to 3 cues. Additionally, there is limited research on the effects of digital distractions and environmental crowdedness on pedestrian paths. This paper presents a series of studies examining the dynamics of human-robot interaction by investigating the effects of signaling cues, digital distractions, and environmental crowdedness on pedestrians' reaction times, decision-making ease, and delivery robot safety. Through three structured within-subject video-based laboratory studies and one field study, valuable insights were obtained to inform the development of safer and more efficient delivery robot systems. Additionally, this research contributes to broader discussions on human-robot interaction, offering insights applicable to various autonomous robotic systems navigating shared spaces with humans safely.

II. METHOD

In this section, we present two video-based laboratory studies and one field study aimed at investigating the effects of signaling cues, digital distractions, and environmental crowdedness on pedestrians' reaction time, decision-making ease, and perceptions of delivery robot safety.

A. Video-based Laboratory Studies

1) Study Designs and Participants: First, we conducted a 5x2x2 within-subject video-based laboratory study using PsychoPy, incorporating three independent variables. The study involved viewing videos from a first-person perspective of an Incidentally Co-present Person (InCoP) turning a corner and walking toward a delivery robot. The independent variables were: 1) Type of visual signaling cues: gaze, arrow, blinking, text, no cue (selected based on earlier research [6, 5, 4] and equipment feasibility, Figure 2). 2) Phone distraction: attentive, distracted by text message (Figure 3). 3) Crowdedness: crowded, uncrowded (Figure 3). To mitigate the risks of order effects and participants predicting the direction of the robot's signal, each condition included the delivery robot signaling to

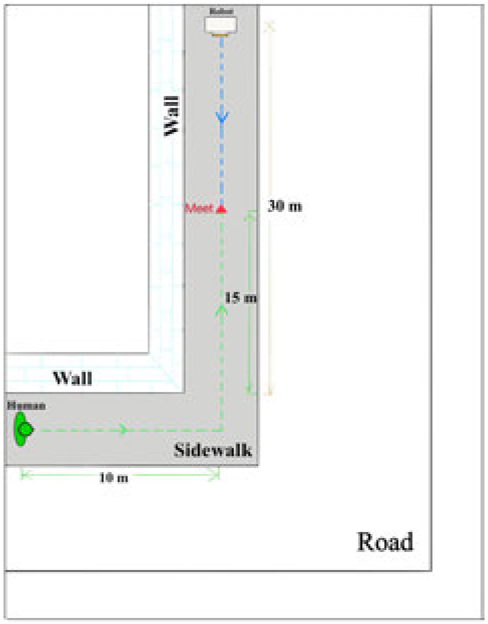

Fig. 1. Top View Layout of the Video Shooting Location.

Fig. 2. Different Visual Signaling Cues on the Delivery Robot (from Left to Right: Gaze, Blinking, Text, Arrow).

the right twice and to the left twice. In total, we produced 80 videos, presented in a randomized order. Before starting the experiment, participants went through a practice session where they watched a video of a human, instead of a robot, walking toward the camera. After each video, we measured participants' reaction time (recorded by PsychoPy), safety perception (”*How SAFE was the experience?*” on a 7-point Likert scale), ease of decision-making (”*How EASY was it to make the decision to turn?*” on a 7-point Likert scale), and a manipulation check for the distracted conditions (”*Which RESTAURANT was mentioned in the text message?*”, ”*Which TIME was mentioned in the text message?*”, ”*Which PERSON was mentioned in the text message?*”). After viewing all the videos, participants were asked about their interpretation of the signaling cues, followed by a short interview to discuss their preferred signaling cue system and the reasons for their preferences. A total of 53 volunteers participated in the first laboratory study. Two participants were excluded because they did not understand the instructions and did not make any choices during the experiment. The final sample consisted of 51 participants (29 females and 22 males; age: *M* = 26.25, *SD* = 6.03).

Next, Kannan et al. [6] found that certain intents are better shown with combinations. Therefore, we conducted the second study with a 3 (cue combinations: blinking with gaze, blinking with arrow, and blinking with text) * 2 perspectives (original/flipped angles of the video) * 2 (crowdedness: crowded, uncrowded) * 2 (digital distractions: attentive, distracted) within-subject study design. 31 participants participated in the study, three were excluded because they failed the manipulation check. The final sample consisted of 28 participants (13 female, 13 male, and 2 diverse; age: *M* = 32.86, *SD* = 13.44).

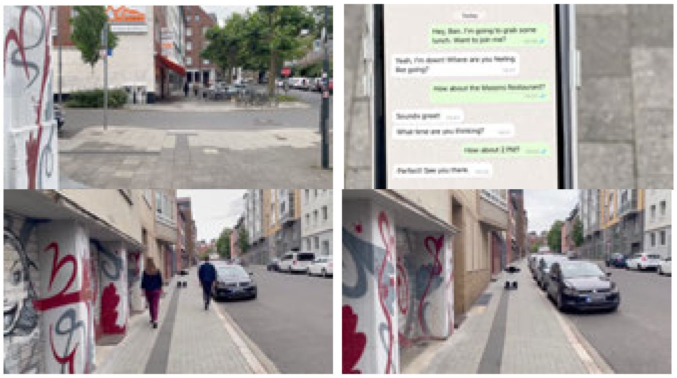

Fig. 3. Video Capture of the Attentive (up-left), Distracted (up-right), the Crowded (down-left), and Uncrowded (down-right) Conditions Conditions with No Cue Showing on the Delivery Robot.

B. Field Study

After testing the signaling cues in the video-based conditions, we conducted a within-subject field study to examine the effect of signaling cues and phone distraction on human perception in real-world settings. In the field study, each participant began by walking straight while either looking forward or looking at their phone. At a certain marked point on the ground at a turning corner, they turned and looked up, encountering the robot driving towards them and displaying a signaling cue (i.e., no cue, gaze, arrow, blinking, combination of gaze and blinking, and combination of arrow and blinking). The robot displayed signals pointing left and right randomly in each signaling condition. After each trial, participants were asked, ”*How SAFE was the experience?*” and ”*How EASY was it to make the decision to turn?*” on 7-point Likert scales. Their reaction time was also recorded by cameras. 23 participants participated in this field study, two of them were excluded because they did not understand the experiment requirements correctly, and another two participants were excluded because they were under the age of 18. Therefore, the final sample consisted of 19 participants (10 female, 9 male; age: *M* = 21,98, *SD* = 20.73).

III. RESULTS

A. Robot Navigation Cues and Pedestrian's Interpretation

The data collected was analyzed using SPSS (IBM, version 29.0.1.0). From the first video-based laboratory study, Friedman tests results revealed that participants found it easiest (χ^2 = 26.36, p <.001), safest (χ^2 = 17.72, p = .001), and fastest (χ^2 = 57.99, p <.001) to decide on a turn for avoidance

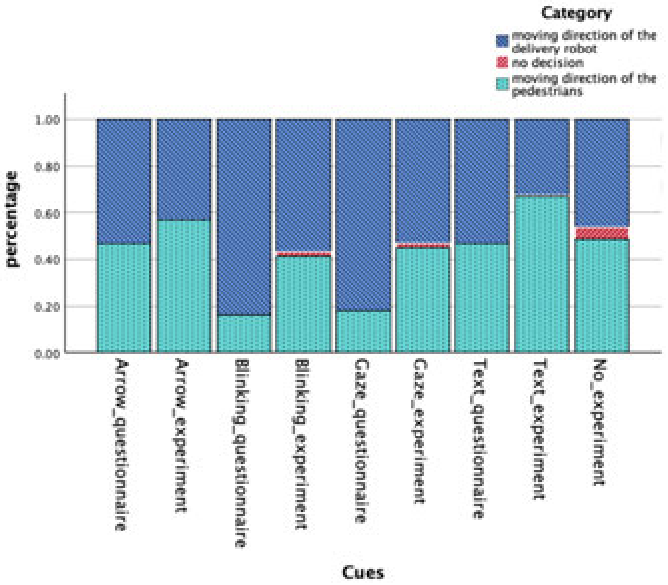

Fig. 4. A Stacked Bar Chart of Pedestrians' Interpretations from the Questionnaire and Experimental Data

when the delivery robot displayed arrow cues. Participants perceived the absence of any cue as the most challenging scenario for decision-making. Furthermore, when pedestrians were attentive, they were more likely to find the decision-making process easier (distractions: t (50) = 2.66, p = .01), safer (distractions: t (50) = 2.85, p = .006), and made quicker decisions (distractions: Z = -4.98, $p < .001$). Results also indicated that participants felt safer (t (50) = -3.36, p = .001) and found it easier (t (50) = -2.15, p = .04) to make the turning decision when it was not crowded. However, they reacted faster when they were in crowded conditions (Z = -2.45, p = 0.01).

In the second study, we investigated the effect of combinations of cues. Participants perceived the combination cues of the robot more as the direction of the robot. They found the combination of blinking and text cues the easiest (F (2,58) = 3.58, $p < .05$) and reacted fastest to the combination of blinking and arrow cues ($\chi^2(2)$ = 7.40, $p < 0.03$). Furthermore, they decided easier and faster and judged the interaction safer when they were attentive ($p < .05$). However, no significant results were found between the crowded and uncrowded conditions.

The results from the field study confirmed that participants perceive the arrow cues the fastest (F (4,72)=4.05, $p < .05$; M = 3.59, SD = .16). However, no significant differences were found between the attentive and distracted conditions.

Regarding pedestrians' interpretation, we can see in Fig. 4, the questionnaire data results from the first study revealed a clear preference among participants to associate the gaze cue (Z = -4.81, $p < .001$) and blinking cue (Z = -4.81, $p < .001$) with the movement direction of the delivery robot. However, there were no significant distinctions in the interpretations observed for the arrow cue (Z = -0.28, p = 0.78) and text cue (Z = - 0.28, p = 0.78).

B. Pedestrians' Preferred Visual Signal Cue System and Reasons

Different reasons were given for pedestrians' preference for different cue systems:

- **Gaze**: appearance, cute
- **Arrow**: familiarity, ease of use
- **Blinking**: familiarity, car-working features, easy to understand
- **Text**: easy to understand

C. Practical Lessons Learned Regarding Methodology

Transferring lab studies to field studies often presents significant challenges. Below are the challenges we experienced in conducting both lab studies and field studies:

- Environmental Restrictions and Observational Fears
- Weather Dependency and Equipment Protection
- Information Asymmetry
- Participant Distractions and Engagements

IV. DISCUSSION

A. Robot Navigation Cues and Pedestrian's Attentiveness

In our initial lab studies, we evaluated pedestrians' reactions to various visual signaling cues on a delivery robot, including gaze, arrow, blinking, text, no cues, and combinations of cues. Findings indicated that the arrow cue was perceived as the simplest and safest, yielding the quickest response times in avoiding the robot. This aligns with prior research; for example, Hetherington et al. [3] favored projected arrows over flashing lights for social acceptance, and other studies highlighted arrows for indicating robot turning intent [7]. Interviews revealed participants valued the arrow's *familiarity* and *ease of use*. Yet, interpretations suggested inconsistency in pedestrian perceptions of arrow cues, potentially posing challenges for real-world application despite its perceived effectiveness. Therefore, it is better to use gaze or blinking cues in real-life situations, even though they might be slightly slower, harder to perceive, and less safe for pedestrians than arrow cues. These cues are still preferable to text cues or no cues at all. Importantly, pedestrians perceive these two cues more consistently.

Additionally, we simulated scenarios mirroring real-life circumstances where pedestrians were distracted by text messages on their mobile phones. We found that distracted pedestrians, compared to attentive ones, diverted their cognitive resources significantly. This hindered their ability to promptly process the delivery robot's cues, aligning with cognitive load theory [16]. This observation parallels results from Lin and Huang [9], comparing reactions of distracted pedestrians with those focused on tasks unrelated to their mobile phones.

Furthermore, the environment's density affected pedestrians' perceptions. Salvini et al. [14] found that reduced space or potential contact raised safety concerns and decreased acceptance of the robot. However, unlike the findings of Zhang et al. [19], suggesting longer thinking times with a robot in a crowd, our study showed pedestrians reacted faster in crowded scenarios. This might be because of their tendency to align with others' movements in crowded environments [11].

This study expands on earlier research that examined only 2 to 3 navigation cues [6, 5, 4] by comparing 4 navigation signaling cues against a control condition (i.e., no cue). This broader comparison provides a more comprehensive understanding of how people perceive different signaling cues. Additionally, the study introduces factors from real-life situations, such as digital distractions and crowded environments, to investigate their impact on pedestrians' perception of the delivery robot's navigation signaling cues.

B. Difficulties on Transferring Lab Studies to Field Studies

Transferring lab studies to field studies often presents significant challenges, however, there is limited discussion on this topic in the literature. This discussion focused on the specific difficulties we encountered in this transition, including cultural perceptions, environmental factors, participant behavior, and the physical attributes of the robot used in the studies.

In Germany, there are restricted rules about conducting studies and recording in private areas including the university campus. A notable apprehension among employees at university campuses has been observed. This fear creates a substantial barrier to conducting field studies within these environments. Consequently, researchers are often restricted to public roads for their studies. This presents its own set of challenges, as the general public in Germany is not accustomed to the presence of cameras on the road. The unfamiliarity can lead to discomfort or altered behavior, skewing the data collected.

Another factor that we cannot ignore when conducting outdoor field studies is the weather conditions. Researchers must be flexible and prepared for varying weather conditions, which can impact both the study and the equipment. Protecting devices, especially prototypes that are not as robust as off-the-shelf products, becomes a critical concern. For instance, rain, snow, or extreme heat can damage electronic components or interfere with their operation, leading to inconsistent results or interruptions in the study. This dependency on weather conditions introduces a layer of unpredictability that is absent in the lab, where conditions can be controlled and replicated consistently.

Furthermore, in our lab studies, participants received a human practice session, giving them more information about what to expect and how to interact with the equipment or scenarios presented. This prior knowledge can lead to different behaviors compared to field studies, where participants typically have less information and no practice sessions. We found that the first trial in a field study is particularly crucial, as it can significantly influence subsequent sessions. Additionally, the novelty of encountering a robot in the field can lead participants to behave differently as they try to make sense of the situation. In an earlier study [17], participants frequently mentioned that InCoPs will need to understand how to behave or interact with delivery robots. They suggested that the solution is not only in making robot behavior transparent but also in educating society about this new technology (e.g., ”People should be informed about autonomous delivery vehicles, how they behave and how to behave towards them, when they start to be in widespread use, e.g. through local newspapers.”, ”Will need regulation. Society will need guidance on how to handle interactions.”). Moreover, participants were more likely to be focused on the task in lab studies, as they were in a controlled environment free from unexpected interruptions. However, in the field studies, participants often encountered real-world distractions, such as passing cars, other pedestrians, and ambient sounds. This distraction level made it challenging to ensure participants' full engagement with the study.

The physical size of the robot used in the studies also played a significant role in influencing participant behavior. The robot was larger than a typical starship robot, which made it more intimidating to participants. A larger robot may evoke a sense of unease or caution, leading to behaviors that would not be present with a smaller, less imposing robot, or a robot in the video. This factor highlights the importance of considering the physical design of robots in field studies.

In summary, transferring lab studies to field settings involves navigating a complex array of challenges. Each of the factors mentioned above can influence the outcomes of field studies in significant ways, underscoring the need for careful planning and consideration when designing and conducting such research.

C. Limitations and Future Work

In video-based lab studies, the results are limited because participants do not experience the interaction in real life. It is important to conduct studies that allow participants to interact with the robot directly. However, efforts should be made to find or create semi-controlled environments that blend the naturalistic elements of field settings with the controlled aspects of lab studies. This could involve using mixed-reality environments or dedicated outdoor research facilities where participant behaviors can be observed without causing discomfort. To address information asymmetry, future studies could include a brief orientation session for participants in field studies. This would help standardize the level of participant knowledge and reduce variability in behavior. Finally, experimenting with different robot sizes and designs will provide insights into how physical attributes influence participant interactions, allowing for the development of robots that are less intimidating and more beneficial to naturalistic behavior. By implementing these strategies, future research can achieve more reliable and generalizable results.

ACKNOWLEDGMENTS

We thank Xiying Li, Teibega Rajkumar, Simon Kroes, and Fredrik Konrad for data collection.

REFERENCES

[1] Georgios Angelopoulos, Alessandra Rossi, Claudia Di Napoli, and Silvia Rossi. You are in my way: Non-verbal social cues for legible robot navigation behaviors. In

2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 657–662, 2022. doi: 10.1109/IROS47612.2022.9981754.

[2] Christopher Chabris and Daniel Simons. *The invisible gorilla: How our intuitions deceive us*. Harmony, 2011.

[3] Nicholas J Hetherington, Elizabeth A Croft, and HF Machiel Van der Loos. Hey robot, which way are you going? nonverbal motion legibility cues for human-robot spatial interaction. *IEEE Robotics and Automation Letters*, 6(3):5010–5015, 2021.

[4] Nicholas J. Hetherington, Ryan Lee, Marlene Haase, Elizabeth A. Croft, and H. F. Machiel Van der Loos. Mobile robot yielding cues for human-robot spatial interaction. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3028–3033, 2021. doi: 10.1109/IROS51168.2021.9636367.

[5] Carla S Jakobowsky, Anna MH Abrams, and Astrid M Rosenthal-von der Pütten. Gaze-cues of humans and robots on pedestrian ways. *International Journal of Social Robotics*, 16(2):311–325, 2024. doi: 10.1007/s12369-023-01064-3.

[6] Shyam Sundar Kannan, Ahreum Lee, and Byung-Cheol Min. External human-machine interface on delivery robots: Expression of navigation intent of the robot. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, pages 1305–1312, 2021. doi: 10.1109/RO-MAN50785.2021.9515408.

[7] Shyam Sundar Kannan, Ahreum Lee, and Byung-Cheol Min. External human-machine interface on delivery robots: Expression of navigation intent of the robot. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, page 1305–1312. IEEE Press, 2021. doi: 10.1109/RO-MAN50785.2021.9515408. URL https://doi.org/10.1109/RO-MAN50785.2021.9515408.

[8] Yeti Li, Murat Dikmen, Thana G. Hussein, Yahui Wang, and Catherine Burns. To cross or not to cross: Urgency-based external warning displays on autonomous vehicles to improve pedestrian crossing safety. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '18, page 188–197, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359467. doi: 10.1145/3239060.3239082. URL https://doi.org/10.1145/3239060.3239082.

[9] Ming-I Brandon Lin and Yu-Ping Huang. The impact of walking while using a smartphone on pedestrians' awareness of roadside events. *Accident Analysis & Prevention*, 101:87–96, 2017. ISSN 0001-4575. doi: https://doi.org/10.1016/j.aap.2017.02.005. URL https://www.sciencedirect.com/science/article/pii/S0001457517300556.

[10] Nicola Marsden, Tobias Bernecker, Raoul Zöllner, Nico Sußmann, and Sebastian Kapser. Buga:log – a real-world laboratory approach to designing an automated transport system for goods in urban areas. In *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pages 1–9, June 2018. doi: 10.1109/ICE.2018.8436361.

[11] Stephen Reicher. *The Psychology of Crowd Dynamics*, chapter 8, pages 182–208. John Wiley & Sons, Ltd, 2001. ISBN 9780470998458. doi: https://doi.org/10.1002/9780470998458.ch8. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470998458.ch8.

[12] Ronald A Rensink. A function-centered taxonomy of visual attention. 2015.

[13] Astrid Rosenthal-von Der Pütten, David Sirkin, Anna Abrams, and Laura Platte. The Forgotten in HRI: Incidental Encounters with Robots in Public Spaces. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 656–657, Cambridge United Kingdom, March 2020. ACM. ISBN 978-1-4503-7057-8. doi: 10.1145/3371382.3374852. URL https://dl.acm.org/doi/10.1145/3371382.3374852.

[14] Pericle Salvini, Diego Paez Granados, and Aude Billard. Safety concerns emerging from robots navigating in crowded pedestrian areas. *International Journal of Social Robotics*, 14, 03 2022. doi: 10.1007/s12369-021-00796-4.

[15] Francisco Soares, Emanuel Silva, Frederico Pereira, Carlos Silva, Emanuel Sousa, and Elisabete Freitas. To cross or not to cross: Impact of visual and auditory cues on pedestrians ô crossing decision-making. *Transportation Research Part F: Traffic Psychology and Behaviour*, 82:202–220, 2021. ISSN 1369-8478. doi: https://doi.org/10.1016/j.trf.2021.08.014. URL https://www.sciencedirect.com/science/article/pii/S1369847821001911.

[16] John Sweller. Chapter two - cognitive load theory. volume 55 of *Psychology of Learning and Motivation*, pages 37–76. Academic Press, 2011. doi: https://doi.org/10.1016/B978-0-12-387691-1.00002-8. URL https://www.sciencedirect.com/science/article/pii/B9780123876911000028.

[17] Manuscript under review. 2024.

[18] David Weinberg, Healy Dwyer, Sarah E. Fox, and Nikolas Martelaro. Sharing the sidewalk: Observing delivery robot interactions with pedestrians during a pilot in pittsburgh, pa. *Multimodal Technologies and Interaction*, 7(5), 2023. ISSN 2414-4088. doi: 10.3390/mti7050053. URL https://www.mdpi.com/2414-4088/7/5/53.

[19] Bingqing Zhang, Javad Amirian, Harry Eberle, Julien Pettre, Catherine Holloway, and Tom Carlson. From hri to cri: Crowd robot interaction—understanding the effect of robots on crowd motion: Empirical study of pedestrian dynamics with a wheelchair and a pepper robot. *International Journal of Social Robotics*, 14, 04 2022. doi: 10.1007/s12369-021-00812-7.