You Are In My Way: Non-verbal Social Cues for Legible Robot Navigation Behaviors

Georgios Angelopoulos[1], Alessandra Rossi[2], Claudia Di Napoli[3], and Silvia Rossi[2]

Abstract— **People and robots may need to cross each other in narrow spaces when they are sharing environments. It is expected that autonomous robots will behave in these contexts safely but also show social behaviors. Thereby, developing an acceptable behavior for autonomous robots in the area mentioned above is a foreseeable problem for the Human-Robot Interaction (HRI) field. Our current work focuses on integrating legible non-verbal behaviors into the robot's social navigation to make nearby humans aware of its intended trajectory. Results from a within-subjects study involving 33 participants show that deictic gestures as navigational cues for humanoid robots result in fewer navigation conflicts than the use of a simulated gaze. Additionally, an increase in the perceived anthropomorphism is found when the robot uses the deictic gesture as a cue. These findings show the importance of social behaviors for people avoidance and suggest a paradigm of such behaviors in future humanoid robotic applications.**

I. INTRODUCTION

Humanoid robots are steadily moving into human-centered environments, including homes, offices, factories, or even disaster scenarios, and it has been suggested that robots will become more and more popular in non-industrial contexts [1]. In these environments, robots should be designed to interact with humans in situations where they have to cross each other in corridors, hallways, and sidewalks daily. If robots are not interactive and their behavior is not legible by humans, they will be viewed as merely obstacles rather than helpful assistants [2].

Most of the research on robots' navigation in shared space with humans focuses on human-aware motion planning capabilities, but these robots are not designed to communicate their intention to pedestrians. May et al. [3] observed that people are not comfortable when robots treat them as just dynamic objects and strive to avoid them. Instead, humans and robots must share a common understanding of the circumstance and each other's intentions. The simplest solution [4] for an acceptable behavior of robots is to adopt a strategy such as stopping whenever there is a pedestrian in view, with unfavorable results in terms of efficiency in reaching the goal to avoid collisions. Another example is the delivery and logistics mobile robots designed to move safely using natural language to inform the pedestrians [5]. However, people may want to interact with robots in a manner that is easily understandable and not constrained by the robot's own design [6]. Indeed, a robot must communicate its intended actions in order to reduce pedestrian's uncertainty. A way to solve this issue is to make the robots adherent to a set of social rules and show socially acceptable behaviors. Consequently, forms of natural non-verbal communication are the key to achieving this behavior. A first step, therefore, has been to study human behaviors, because while humans can naturally move in public spaces, they are also able to infer others' movement by observing pedestrians' subtle body language cues [7].

The presented work designs, tests, and compares naturalistic signaling mechanisms mimicking human communicative cues assuming that these behaviors do not require a demonstration to be understood. In particular, the study explores three navigational cues: two dissimilar simulated gaze patterns of the robot and a deictic gesture while crossing a human. We also explore the legibility of the cues and how these cues affect robot's social attribution. By increasing robots' behavior legibility, we can design more socially acceptable humanoids robots for public spaces.

II. RELATED WORK

Despite the increasing deployment of humanoid robots in our everyday life, the development of effective human-robot avoidance behavior has not, at least until very recently, received the attention it deserves. Consequently, humanoid robots navigating and maneuvering safely among humans while communicating their imminent actions are rarely found in the related research literature.

Fernandez et al. [8] present a study in which a non-humanoid robot navigates in a hallway and signals that it intends to pass a human participant using a strip of LEDs which acts as a turn signal within the field of view of the user. The authors found that people do not readily interpret LED turn signals when interacting with the non-humanoid robot but that a brief, passive signal demonstration (preliminary training) is sufficient to disambiguate its meaning.

Hart et al. [9] developed a mobile robot platform with a virtual agent head to measure the importance of gaze in coordinating people's navigation. They compared the performance of a robot with the virtual agent head and a robot that used a LED to turn signal. They showed that people are able to more easily interpret the gaze cue than the LED turn signal. Despite the encouraging results to

---

*This work has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955778.

[1] Georgios Angelopoulos is with the Interdepartmental Center for Advances in Robotic Surgery - ICAROS, University of Naples Federico II, Naples, Italy `georgios.angelopoulos@unina.it`

[2] Alessandra Rossi and Silvia Rossi are with the Department of Electrical Engineering and Information Technologies - DIETI, University of Naples Federico II, Naples, Italy `{alessandra.rossi, silvia.rossi}@unina.it`

[3] Claudia Di Napoli is with Istituto di Calcolo e Reti ad Alte Prestazioni, CNR, Naples, Italy `claudia.dinapoli@cnr.it`

impact people's navigational choices, the authors mention that interpreting gaze direction on a virtual agent's head turned out to be difficult and ineffective due to the so-called Mona Lisa Effect.

Reinhardt et al. [10] developed a back-off (BO) movement, that is a backward movement along the original robot trajectory, as a solution for a non-humanoid robot to yield priority to another party in spatial interaction. It allows the communication of robot's intention to yield priority to the interacting human when a collision has to be avoided. The BO was recognized as a yielding priority movement to the humans by almost twice as many people in the sample as a Stop and Wait movement. However, participants evaluated solely the legibility of the movement in a video study.

Hetherington et al. [11] designed two types of non-verbal robot motion legibility cues in a non-humanoid mobile robot: projected arrows and flashing lights. Their results showed that projected arrows were more socially acceptable and comprehensible than flashing lights.

Senft et al. [12] presented a humanoid robot passing humans in narrow corridors using the step-and-slide strategy. Despite their relevant results, the authors' implementation did not include the movement of extending the robot's arm on the step-and-slide strategy, as explained in [13]. Some participants felt pressured and observed and this could be due to the lack of this additional movement.

In the studies mentioned above, these proposed approaches generally need that people have some preliminary training to be clearly understood. In addition, the reported experimentations mainly concern non-humanoid robots. There is still a need for clear, natural, and efficient cues implemented into humanoid robots while they navigate in shared spaces with humans.

III. THE PROPOSED APPROACH

People communicate in complex ways using gestures such as pointing, showing, or drawing the attention of a social partner to a specific entity in the environment. Referential signals are used to share an interest in an object or to achieve a more defined aim, such as retrieving the object. Pointing is one of the most widely used human referential gestures [14]. Another type of signal is the communicative intent indicating behavior [15], in which the person looks at the observer's face and eyes, gesturing, or gazing, to show the observer's attentional status. Contrary to the referential signal, the person attempts to influence the other's behavior directly by drawing the attention. The communicative-intent indicators signal the sender's attempt to get the attention and to interact face-to-face.

The proposed study endows a robot with non-verbal mechanisms to convey navigational intention, inspired by human-human interaction, to increase the legibility of the robot and to diminish the need for preliminary training of the user. For this purpose, we designed three non-verbal cues to be used by a humanoid robot that can be understandable by people with different cultural and ethnic background without any demonstration. In particular, the navigational cues considered are:

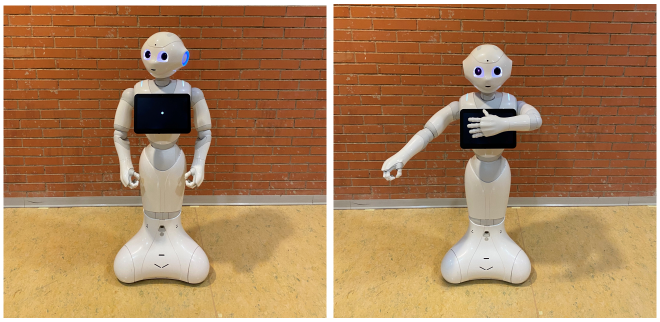
(a) NC1. (b) NC2.

Fig. 1: Non-Verbal Behavior for Directionality Intention.

- ***Navigation Cue 1*** **(NC1)**: Simulating Gaze by Head pose as shown in Figure 1a, with an intention to move toward the gazed direction.
- ***Navigation Cue 2*** **(NC2)**: Deictic Gesture where the humanoid extends its arm pointing to the intended direction and parallel to that it brings its hands toward to its chest to indicate itself, as shown in Figure 1b. The deictic gesture is a potent and pervasive tool used across cultures and contexts to express movement toward a target—someone, something, or somewhere.
- ***Navigation Cue 3*** **(NC3)**: Simulating Gaze by Head pose with attention. The gaze will indicate the intention to move toward the gazed direction but prior to that, the robot performs a direct gaze towards the human (communicative-intent indicator) as an attempt to influence the human's behavior directly by getting the attention.

In this work, we are interested in investigating the legibility of the described non-verbal signaling methods to convey the robot's intention to the human during a navigation task. In particular, our research question is "Which social cue (NC1, NC2, and NC3) is more legible (i.e., leading to fewer conflicts) to humans during a navigation task?". According to current literature, we expect that simulating gaze by head pose with attention (NC3) is a more legible navigational cue by resulting in fewer conflicts than the simulating gaze (NC1) (**Hypothesis 1**). Indeed, Senju et al. [16] showed that the gaze toward an object, preceded by direct gaze, leads to a more active, communicative role. Moreover, recent findings showed also that people prefer to use deictic gestures rather than gaze when responding to joint attention bids in human social interaction [17]. Therefore, we also expect that a deictic gesture (NC2) is a more legible navigational cue leading to fewer conflicts than the simulating gaze (NC1) (**Hypothesis 2**).

IV. METHODS

Detecting the environment is a requirement to effectively navigate and utilize the non-verbal signalling mechanisms properly.However, the environment detection requires heavy

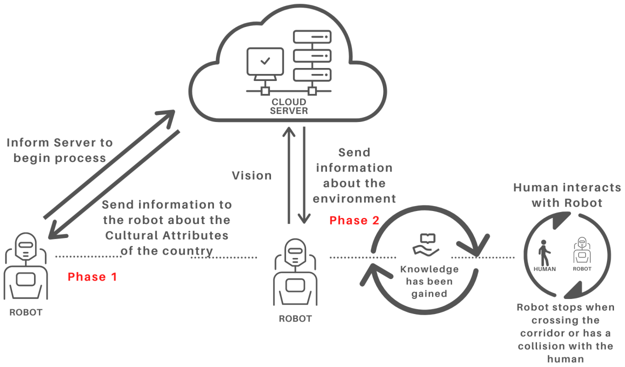

Fig. 2: The robot-server communication.

computation that is not always possible to be performed on the limited computational resources and memory of a robot, such as the one used for our study. Therefore, we rely on a cloud-based system consisting of a cloud server (Intel Core i9-10920X, 3.50GHz x 24) and the robot Pepper Y20 V18A that communicate synchronously. The only computation performed locally by the robot consists in identifying the distance between the walls and the human using Pepper's lasers.

A. Robot-Server Communication

The robot and server communicate through socket messages with a low overhead to avoid high-latency responses. The communication procedure consists of two phases, as shown in Figure 2.

The first phase allows the robot to acquire the knowledge of the Cultural Attributes of pedestrians in order to effectively interact [18]. In fact, the robot needs to comprehend the social norms of different countries for pedestrians when it navigates in a corridor or a hallway. For example, in most Western countries, such as Italy, people are used to walk on the right to avoid other pedestrians. In countries where people drive on the left, they prefer to walk on the left [19]. The server, therefore, transmits this country's cultural attributes to the robot at the beginning of each session. Depending on the country, the robot navigates on the left or on the right accordingly. In our pilot experiments, the robot's latency to get the above information was between 0.38 – 0.53 s.

The second phase allows the robot to obtain information about the environment. The robot needs to identify the presence, the location, and the direction of a person; we used a YOLOv3 algorithm for real-time detection.The weights of the neural network were initialized using a pre-trained model trained on the COCO dataset [20]. When the robot needs to update its knowledge of the environment, it transmits an image through sockets to the server to be processed. After that, the server recognizes the person's position in relation to the robot's field of view. In our pilot experiments, the robot's latency for this elaboration was in the range of 0.5 – 0.65 seconds.

Fig. 3: A snapshot of the experiment during execution. The participant already understood the behavior of the humanoid and changed the direction.

B. Navigation

The robot's navigation algorithm splits the environment into two traffic lanes to navigate. Firstly, the robot moves to a starting lane position by using its lasers located on the front and the side of its body. Afterward, Pepper moves into the corridor only if the system recognizes a person. When the robot moves within the social space of the human (set out at 2.5m accordingly to what was reported in [21]), the robot signals that it is about to change its lane using the non-verbal navigational cue according to the experimental condition. The distance at which the humanoid will execute its turn is two meters from the human. If the human understands such behavior, the robot continues until the end of the corridor; otherwise, the behavior is considered a failure, and the robot stops in order not to collide with the human. In a successful interaction, subsequently to the non-verbal behavior, we expect that the human will change its lane on the corridor. Conflicts are defined as the robot and the human coming to a complete stop because they approached each other too closely without making a decision, or the robot and the human participant are still in the same lane causing the human to make a swift correction. Pepper moves at a speed of 0.7 m/s, and it increases its velocity at 0.85 m/s when turning left or right, so that it is easier for the person to recognize the swift correction of the movement.

V. EXPERIMENTS

A user study was conducted to test which navigational cue was more legible for the participants and, therefore, produced fewer conflicts in order to investigate our hypotheses.

We designed a within-subjects study in a designated hallway environment at the University of Naples Federico II, as shown in Figure 3. The aisle's width and length were in pair with the ADA standards for hallways (i.e., 152.4cm minimum width) [22]. In particular, the width of this study's hallway was 190cm and the length was 800cm.

A. Procedure

Upon arrival, participants were asked to read and sign an informed consent form about the experiment's aims and procedure. Then, the robot and the experimental environment were introduced. Participants were told to walk along the hallway until the end of it, opposite to the robot's position, and that they were crossing each other path. Participants were left free to choose their starting position, walking speed, position in the hallway, and consequently the distance from the robot. To test and compare the effectiveness of the three robotic behaviors in coordinating navigation through the shared space, participants were tested with three experimental sessions, one for each social cue (NC1, NC2, and NC3), and a baseline session (NC4) in which the robot did not use any navigational cue. These behaviors were executed in random order.

For each condition, the robot was assigned a name to help participants to differentiate the navigational cues. The robot's names were chosen between unfamiliar Hawaiian gender-neutral names[1] to reduce possible gender biases. The robot displayed its name on the tablet located on its chest.

The experimental trial lasted 20-30 minutes. We video-recorded each session to further evaluate participants' behaviors during the trials.

B. Measurement

Participants were asked to complete a set of questionnaires about their level of English, demographic questions (i.e., age, gender, education), their previous experience with robots, and their perception of robots. We also wanted to evaluate possible negative bias of participants toward robots, and for this reason, we asked them to answer the following question "To what extent do you fear that machines will become out of control?".

To understand the legibility of the cues and how these cues affect the robot's social attributes, we adopted two well-established and validated questionnaires in the Human-Robot Interaction literature applying them in different periods during the experiment:

To measure people's judgments of the social attributes of the robot, participants were administered a brief post-interaction survey, at the end of each experimental session, comprising 7-point Likert and 18 cognitive-differences scale questions based on the RoSAS scale. The RoSAS is a psychometric instrument aimed towards measuring social perception and judgments of robots across multiple contexts and robotic platforms [23]. RoSAS considers the Warmth, Competence, Discomfort, and the Organic attributes of the robot.

To measure perceived robots' agency and experience, participants were asked to compare the four cues in one questionnaire, at the end of all the experimental sessions, comprising seven scale questions that are based on the well-established Mind Attribution Scale [24]. It allows to evaluate the personal judgments and the mental capacities of the robot.

C. Participants

We recruited 37 participants from the university's community; 3 participants were removed due to their insufficient knowledge of English and 1 participant for failing to participate in the study properly. The participant was zigzagging in front of the robot because, as he stated, he wanted to test the robot's capabilities. Consequently, we analyzed the responses of 33 participants, 22 males and 11 females, and aged between 18 and 35 (M=24.5, SD=3.211), granting us an effect size of d=0.25 with .90 power at an alpha level of .05.

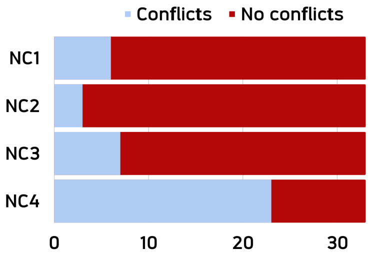

Fig. 4: Frequency of successful interactions and conflicts for each navigation cue.

VI. RESULTS

A. Reliability Analysis

Prior to conducting any analysis, we performed a Cronbach's α test to assess the internal reliability of the questionnaires. Cronbach's α for the Warmth dimension of the RoSAS questionnaires was $\alpha_{NC1} = 0.876$, $\alpha_{NC2} = 0.826$, $\alpha_{NC3} = 0.914$, $\alpha_{NC4} = 0.911$. The Cronbach's α for the Competence dimension are respectively $\alpha_{NC1} = 0.857$, $\alpha_{NC2} = 0.873$, $\alpha_{NC3} = 0.887$, $\alpha_{NC4} = 0.890$. For the Discomfort dimension, we observed $\alpha_{NC1} = 0.841$, $\alpha_{NC2} = 0.902$, $\alpha_{NC3} = 0.883$, $\alpha_{NC4} = 0.891$. Finally, the Cronbach's α related to the Mind Attribution Test was $\alpha = 0.750$. The majority of the participants (63.6%) stated that they had never interacted with robots before, while 36.4% of the participants already had previous experience with robots. We also observed that participants did not have a strong preference for walking on a particular side of the sidewalk. Indeed, 45.5% preferred to walk on the right side of the sidewalk/hallway, 21.2% preferred the left side, and the remaining did not express any preference.

B. Conflicts during interactions

A McNemar's test with continuity correction was run to compare the conflicts occurred in relation to the conditions. As shown in Figure 4, NC2 was the behavior that generated less conflicts (90.9%), followed by NC1 (81.8%), NC3 (78.8%), and NC4 (30.3%). We observed that the differences between NC1, NC2, and NC3 with the NC4 were statistically significant with $p < .0005$. Therefore, our hypothesis *H1* was not confirmed, as Simulating Gaze (NC1) had fewer conflicts than the Simulating Gaze by Head pose with attention (NC3).

C. Mind Attribution Scale Ratings

We also investigated participants' responses to the Mind Attribution Scale. Of the 33 participants recruited to the

---

[1] https://kidadl.com/baby-names/hawaiian/gender-neutral

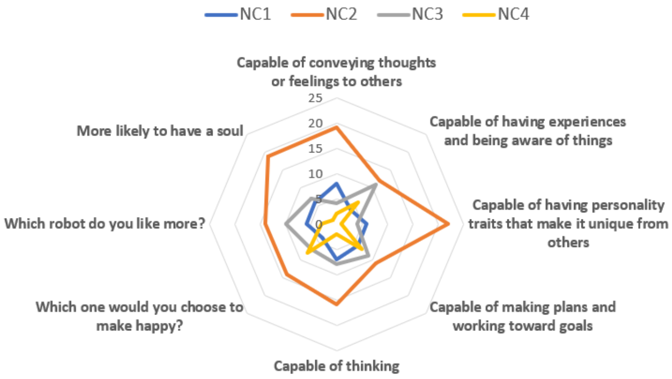

Fig. 5: Participants' outcomes demonstrated an increase in the robot's anthropomorphism when the NC2 cue was used.

study, 19 reported that NC2 could *convey thoughts or feelings to others*. A Chi-square goodness-of-fit test was conducted to determine the preferable cue used by the robot. The Chi-square test showed that the number of participants selecting NC2 was statistically significantly different ($\chi^2(3) = 20.339, p < .0005$), with just over half of the participants selecting NC2. The condition NC2 was considered more *capable of having personality traits that make it unique from others* for 22 participants ($\chi^2(3) = 32.091, p < .0005$). Furthermore, 16 participants reported that NC2 was more *capable of thinking* ($\chi^2(3) = 12.212, p = .007$). Participants (14 over 33) reported that they *liked* NC2 more ($\chi^2(3) = 12.212, p = .040$). Finally, 20 participants reported that NC2 was *more likely to have a soul* ($\chi^2(3) = 24.333, p < .0005$), with just over half of the participants selecting NC2.

Figure 5 illustrates the above results. Therefore, findings showed that the outcome of an interaction in which a humanoid using the Deictic Gesture (NC2) influences individuals' tendency to attribute minds to robots and therefore affects the psychological anthropomorphism of it.

D. Robotic Social Attributes Scale Ratings

A Wilcoxon signed-rank test was used to examine differences in the navigational cues on the RoSAS scale. We found statistically significant differences in the *Warmth* dimension between NC2 with NC4 and NC3 with NC4, and in the *Discomfort* scale between NC1 with NC2 as depicted in Figure 6. More particularly, in the *Warmth* scale, we observe a statistically significant difference between NC2 and NC4 ($z = -1.989, p = .047$), also between NC3 and NC4 ($z = -2.090, p = .037$). Furthermore, we distinguish in the *Discomfort* scale, a statistically significant difference between NC1 and NC2 ($z = -2.090, p = .037$). We did not notice any statistically significant difference in the *Competence* scale of the questionnaire with the most considerable to be between NC2 and NC4 ($z = -1.891, p = .059$). Interestingly, despite people perceiving the robot that uses a deictic gesture as a navigational cue (NC2) as warm, they rated NC2 as the most uncomfortable robot (increased ratings of discomfort) among the four.

In secondary exploratory analyses, as [25], [26], we compared ratings for each of the RoSAS subitems between the cue conditions as shown in Figure 7. The test showed a statistically significant difference in the *feelings* and *responsiveness* of the robot in NC2 compared to the NC4 ($z = -2.462, p = .014$ and $z = -2.118, p = .034$ respectively). We observed a statistically significant difference in the *emotional* aspect of the robot in NC2 compared to the NC1 ($z = -2.005, p = .045$). Pairwise comparisons were performed with a Bonferroni correction for multiple comparisons. In the scale of *interactivity*, we observe a statistically significant difference between NC2 and NC4 ($p = .022$). Comparisons on the other subitems between the navigational cues were not statistically significant.

Our results go beyond our hypotheses, especially for what concerns Hypothesis 2. Results indicate that a robot that uses deictic gesture is more favorable and more effective than a non-gesturing and a gazing robot, but this also increases the overall anthropomorphism of the humanoid robot.

The deictic gesture was the most effective navigational cue, albeit the participants considered it scarier and more awkward than the other navigational cues. In particular, a statistically significant difference in the *awkwardness* of the robot in NC2 compared to the NC1, with $z = -1.965, p = .049$ can be observed with the Wilcoxon signed-rank test. We also noticed a statistically significant difference in the *scariness* of the robot in NC2 compared to the NC3, with $z = -1.967, p = .049$. We assume, however, that these rating tendencies are due to the increasing human likeness and are inlined with the uncanny valley effects [27].

VII. CONCLUSIONS

This study aimed at integrating non-verbal behaviors in the robot's social navigation to cue nearby humans of its intended trajectory. Our results suggest that legible navigation cues are crucial for social robots to manage pass-by situations when crossing people in a corridor. Moreover, this paper evaluates whether a deictic gesture cue can outperform the so far proposed gaze cue, and a gaze cue with communicative intent. Results from our study showed that deictic gesture is more effective and increases the anthropomorphic perception of the robot.

Considering the satisfying results in terms of the number of resulting conflicts that this study produced, in future works we would like to refine our model to take into account other characteristics that may impact the interaction and the decision-making process. In particular, an adaptive selection of the navigational cues and a combination of the presented cues will be investigated. Future research should also address the generalizability of our findings regarding anthropomorphic inferences with other robotic platforms, such as non-humanoid robots or humanoid robots with eyes suitable for human-like gaze communication. Finally, our system should be deployed in real-world settings which are less controlled environments than the laboratory where the experiments were carried out.

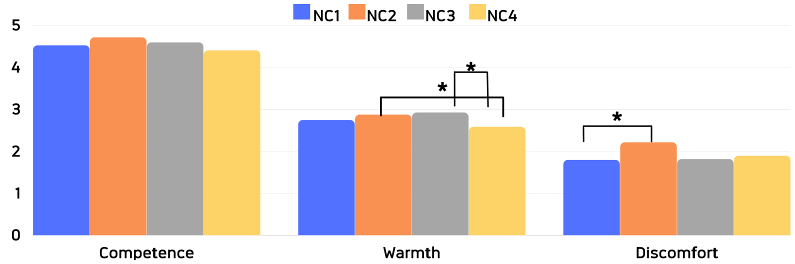

Fig. 6: RoSAS items, * indicates $p < 0.05$.

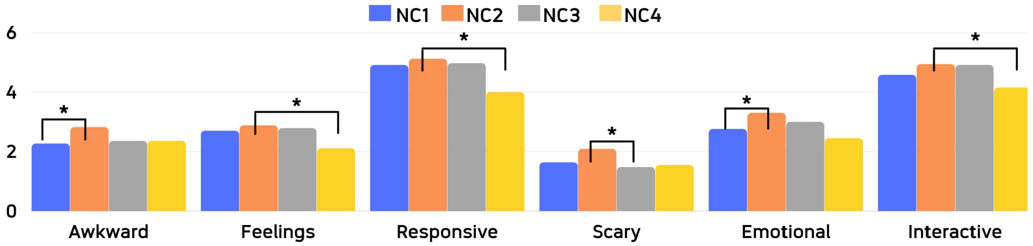

Fig. 7: RoSAS subitems which occur significant differences between navigational cues, * indicates $p < 0.05$.

REFERENCES

[1] M. Mara, J.-P. Stein, M. E. Latoschik, *et al.*, "User responses to a humanoid robot observed in real life, virtual reality, 3d and 2d," *Frontiers in Psychology*, vol. 12, p. 633 178, 2021.

[2] J. W. Hart, N. DePalma, M. W. Pryor, *et al.*, "Exploring applications for autonomous nonverbal human-robot interaction," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 728–729.

[3] A. D. May, C. Dondrup, and M. Hanheide, "Show me your moves! conveying navigation intention of a mobile robot to humans," in *2015 European Conference on Mobile Robots (ECMR)*, IEEE, 2015, pp. 1–6.

[4] A. Sciutti, M. Mara, V. Tagliasco, and G. Sandini, "Humanizing human-robot interaction: On the importance of mutual understanding," *IEEE Technology and Society Magazine*, vol. 37, no. 1, pp. 22–29, 2018.

[5] A. Rossi, F. Garcia, A. C. Maya, *et al.*, "Investigating the effects of social interactive behaviours of a robot on people's trust during a navigation task," in *Annual Conference Towards Autonomous Robotic Systems*, Springer, 2019, pp. 349–361.

[6] S. Rossi, A. Rossi, and K. Dautenhahn, "The secret life of robots: Perspectives and challenges for robot's behaviours during non-interactive tasks," *International Journal of Social Robotics*, vol. 12, no. 6, pp. 1265–1278, 2020.

[7] D. Helbing, P. Molnár, I. J. Farkas, and K. Bolay, "Self-organizing pedestrian movement," *Environment and planning B: planning and design*, vol. 28, no. 3, pp. 361–383, 2001.

[8] R. Fernandez, N. John, S. Kirmani, J. Hart, J. Sinapov, and P. Stone, "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)*, IEEE, 2018, pp. 234–239.

[9] J. Hart, R. Mirsky, X. Xiao, *et al.*, "Using human-inspired signals to disambiguate navigational intentions," in *International Conference on Social Robotics*, Springer, 2020, pp. 320–331.

[10] J. Reinhardt, L. Prasch, and K. Bengler, "Back-off: Evaluation of robot motion strategies to facilitate human-robot spatial interaction," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 10, no. 3, pp. 1–25, 2021.

[11] N. J. Hetherington, E. A. Croft, and H. M. Van der Loos, "Hey robot, which way are you going? nonverbal motion legibility cues for human-robot spatial interaction," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5010–5015, 2021.

[12] E. Senft, S. Satake, and T. Kanda, "Would you mind me if i pass by you?: Socially-appropriate behaviour for an omni-based social robot in narrow environment," in *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, 2020, pp. 539–547.

[13] P. Collett and P. Marsh, "Patterns of public behaviour: Collision avoidance on a pedestrian crossing," 1974.

[14] D. A. Leavens and W. D. Hopkins, "Intentional communication by chimpanzees: A cross-sectional study of the use of referential gestures.," *Developmental psychology*, vol. 34, no. 5, p. 813, 1998.

[15] I. Brinck, "The role of intersubjectivity in the development of intentional communication," *The shared mind: Perspectives on intersubjectivity*, pp. 115–140, 2008.

[16] A. Senju and G. Csibra, "Gaze following in human infants depends on communicative signals," *Current biology*, vol. 18, no. 9, pp. 668–671, 2008.

[17] N. Caruana, C. Inkley, P. Nalepka, D. M. Kaplan, and M. J. Richardson, "Gaze facilitates responsivity during hand coordinated joint attention," *Scientific reports*, vol. 11, no. 1, pp. 1–11, 2021.

[18] G. Trovato, M. Zecca, M. Do, *et al.*, "A novel greeting selection system for a culture-adaptive humanoid robot," *International Journal of Advanced Robotic Systems*, vol. 12, no. 4, p. 34, 2015.

[19] H. S. Jung and H.-S. Jung, "Survey of korean pedestrians' natural preference for walking directions," *Applied ergonomics*, vol. 44, no. 6, pp. 1015–1023, 2013.

[20] T.-Y. Lin, M. Maire, S. Belongie, *et al.*, "Microsoft coco: Common objects in context," in *European conference on computer vision*, Springer, 2014, pp. 740–755.

[21] L. Naik, O. Palinko, L. Bodenhagen, and N. Krüger, "Multi-modal proactive approaching of humans for human-robot cooperative tasks," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, IEEE, 2021, pp. 323–329.

[22] M. Hashemi and H. A. Karimi, "Indoor spatial model and accessibility index for emergency evacuation of people with disabilities," *Journal of Computing in Civil Engineering*, vol. 30, no. 4, p. 04 015 056, 2016.

[23] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (rosas) development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, 2017, pp. 254–262.

[24] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," *science*, vol. 315, no. 5812, pp. 619–619, 2007.

[25] A. Matsufuji and A. Lim, "Perceptual effects of ambient sound on an artificial agent's rate of speech," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 67–70.

[26] J. Jørgensen and M. B. Christiansen, "The sounds of softness. designing sound for human-soft robot interaction," *Frontiers in Robotics and AI*, vol. 8, 2021.

[27] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics & automation magazine*, vol. 19, no. 2, pp. 98–100, 2012.