Understanding Robots: Making Robots More Legible in Multi-Party Interactions

Miguel Faria[1]   Francisco S. Melo[1]   Ana Paiva[1,2]

Abstract— In this work we explore implicit communication between humans and robots—through movement—in multi-party (or multi-user) interactions. In particular, we investigate how a robot can move to better convey its intentions using legible movements in multi-party interactions. Current research on the application of legible movements has focused on single-user interactions, causing a vacuum of knowledge regarding the impact of such movements in multi-party interactions. We propose a novel approach that extends the notion of legible motion to multi-party settings, by considering that legibility depends on all human users involved in the interaction, and should take into consideration how each of them perceives the robot's movements from their respective points-of-view. We show, through simulation and a user study, that our proposed model of multi-user legibility leads to movements that, on average, optimize the legibility of the motion as perceived by the group of users. Our model creates movements that allow each human to more quickly and confidently understand what are the robot's intentions, thus creating safer, clearer and more efficient interactions and collaborations.

I. INTRODUCTION

Robots are no longer restricted to factory settings, we can find robots in different contexts, such as assisting in tasks of health-care [14], education [18], entertainment [4], among others. However, to integrate robots in society, they need to be able to correctly interact and communicate with humans. Communication is paramount in activities that require collaboration between humans and robots to achieve a common goal [9]; namely activities like belonging to surgery teams during surgeries, where timely clear communication and precise actions are required for surgery success [10] or in furniture assembly, where incorrect communication can lead to different actors try to simultaneously use the same resources hindering the collaboration.

Body movements play an important role in the efficiency and fluidity of a task, because they occur naturally and allow humans to get cues regarding their partners motives and intentions [17]. *Legible movements* are a type of body movement designed to improve the information conveyed by a robot, focusing on making a robot's movements easier "to read" by human users. Leveraging on humans assuming a robot will move as efficiently and rationally as possible towards its intended objective, Dragan et. al. [6] use principles from animation to allow robots to convey more information in the earlier parts of the trajectory. Legible movements have

been shown to improve robot's expressiveness in human-robot interactions: Dragan et. al. [7] explore the impacts of legible movements in comparison with other motion types in a coffee shop scenario of a human-robot team fulfilling orders, evaluating how quick a human can deduce the robot's intentions and the trust placed by humans on robots. Knepper et. al. [11] and Mavrogiannis et. al. [13] explore applications of legibility in robot navigation scenarios, using the notion of legibility while moving to correctly communicate the direction of travel and help the navigation of other humans, reducing the need for re-planning.

Most works on legibility explore the uses and advantages of legible movements in single-user scenarios [1–3, 11, 12], however, when integrated in society, a robot must be expressive both in single and multi-party scenarios. Consider the example of a robot, in a surgery scenario, that needs to collaborate with other humans during an intervention on a patient, the robot's movements need to be precise and clear for the team members to execute their respective tasks, without requiring the robot to explicitly communicate its current objectives.

In this work we expand on our previous exploratory study [8], investigating how a robot's movement should be optimized, in multi-party interactions, to minimize confusion among users regarding the intentions of the robot. To that purpose we propose a model that extends the notion of legibility to multiple user scenarios, dubbed *multi-user legibility (MUL)*. MUL extends the notion of legibility to account for different users observing the robot's motions from different points-of-view and possibly interpreting the robot's intentions differently. Our proposal considers that the legibility of a movement should not be taken individually for each user but as the average of the legibilities for all users. We show, with a user study over M-Turk, that human users, when observing movements from different perspectives, find movements created with MUL clearer and are able to predict the robot's objective faster than when observing movements created with previous single-user definitions.

With this work we expand on the literature of legibility, by contributing with a novel definition of legible movements, applied to multi-party interaction scenarios. With this definition, a robot's movement can be shaped to be simultaneously legible to all human users simultaneously interacting with a robot. The resulting trajectories are clearer for users observing from different points-of-view (PoVs) than using previous definitions of legibility for single-user scenarios, thus improving the group's understanding of the robot's objectives and intentions.

[1] All authors are with INESC-ID and Instituto Superior Técnico, University of Lisbon, Portugal. E-mail: `miguel.faria@tecnico.ulisboa.pt` and `{fmelo,ana.paiva}@inesc-id.pt`.
[2] Ana Paiva is with Radcliffe Institute for Advanced Study, Harvard University.

II. LEGIBLE MOTION

Consider a trajectory ξ, defined in Cartesian space, as

$$\xi = [[x_1, y_1, z_1], [x_2, y_2, z_2], \ldots, [x_T, y_T, z_T]]$$

where T is the number of time points in the trajectory. A legible movement, when observed by a human, allows that human to quickly infer the objective G_R given the observed trajectory ξ, maximizing

$$Legibility(\xi) = \frac{\int P(G_R | \xi_{S \to \xi(t)}) f(t) dt}{\int f(t) dt}. \quad (1)$$

In (1) $P(G_R | \xi_{S \to \xi(t)})$ gives the likelihood of reaching objective G_R with the observed trajectory between the start point S and the current point $\xi(t)$, denoted as $\xi_{S \to \xi(t)}$. The function f is a weighting function, designed to give more weight to earlier parts of the trajectory. $P(G_R | \xi_{S \to \xi(t)})$ is modeled using a max-entropy distribution

$$P(G_R | \xi_{S \to \xi(t)}) = \frac{1}{Z} \exp \left\{ - C(\xi_{S \to \xi(t)}) \right\},$$

where Z is a normalizing constant and C is a cost function modeling how a human expects the robot to move. Following [5], we use this as the cost function the sum of squared velocities. Using the definition of legibility in (1), it is possible to generate legible trajectories using a gradient ascent approach that, in each iteration, improves the legibility score of the trajectory ξ:

$$\xi_{i+1} = \xi_i + \frac{1}{\eta} M^{-1} \nabla Legibility(\xi), \quad (2)$$

where M is used to measure the norm of a trajectory, $\|\xi\|_M^2 = \xi^T M \xi$, and η defines the learning rate.

Legible movement as defined in [6] assumes humans have an omniscient view of the workspace. This assumption can lead to movements that go outside the field-of-view (FoV) of a human, go through human blind spots or through obstructed parts of the workspace. To solve this problem, Nikoladis et al [16] proposed an extension to the original notion of legibility that uses a modified cost function $\tilde{C}$, defined as

$$\bar{C}(\xi) = C(_{2D}T^W(\xi)),$$

where $_{2D}T^W$ transforms ξ from world coordinates to the human referential and then projects the trajectory into a 2D representation in the human's viewport. With the extension, the legibility metric becomes dependent of the point of view of the human user. The transformation $_{2D}T^W$ to the human's viewpoint allows the optimization procedure to improve the legibility, creating trajectories that are always within the user's FoV. We henceforth refer to this approach as *single-user legibility (SUL)*.

III. MULTI-USER LEGIBLE MOTION

There are several scenarios where a robot must interact with multiple humans at the same time: in Correia et. al. [4] a robot plays a game of cards, simultaneously interacting with three human users; Kaplan et. al. [10] describes a scenario in which a robot is deployed as part of a surgical team to support the staff and Faria et. al. [8] describes a scenario where a robot sequentially serves cups of water to different human users. In these types of scenarios, for a robot to be legible, it must be able to generate movements that are simultaneously legible for all partners involved. Otherwise, it could optimize the legibility for one partner but reduce the legibility for the others, causing deception regarding its intentions. The presence of multiple human partners, causes multiple different perspectives over a movement and it to be perceived differently from each. Thus, in a multi-party scenario, the legibility metric should be influenced by how legible the movement is perceived from each perspective.

Thus, we propose that the standard SUL model of Nikoladis et. al. should be extended to multi-party scenarios, where the legibility metric is a combination of the perceived legibilities for each point-of-view. We dub this extension *multi-user legibility (MUL)*.

To correctly capture the group's legibility, MUL needs to incorporate information regarding the perceived legibility of each of the task's users. However, the integration of the different perceived legibilities must be such that no user is especially favored over the other users, otherwise we could fall back in the situation of single-user legibility where the robot's movement gives more information to part of the users. Thus, MUL averages the perceived legibilities of the users giving equal weight to all users.

Giving equal weight to all the users is motivated by different perspectives over the workspace contributing differently, giving better or worse information regarding the robot's intentions. Take the example in Figure 1: a person in User 1's position may have a better perspective of the robot's movement when compared to a person in User 2's perspective, because in User 1's perspective a side movement of the robot can be sufficient for the user to understand the robot's objective, while for User 2 the robot needs to execute a more complex movement combining movements along different axis. However not all user layouts allow to easily understand which perspectives give better information and, without previous knowledge of which perspectives offer better information, attributing different weights could give more importance to perspectives that offer worse information and create trajectories that could decrease legibility. So, by giving the same weight to all users we balance the perspectives with worse view of the movement with those with better view. Finally, giving the same importance to all perspectives guarantees that the movement is kept in the field-of-view of all users, keeping it always visible to all users, by preventing each user to excessively influence the shaping of the trajectory in one specific direction that could go outside the view of another user.

Another option we considered for combining the perspectives was to consider that the movement's legibility was given by the user with worse legibility. Following this approach, legible movements in multi-party interactions would maximize the legibility of the user that, at the time, would have more difficulty in understanding the robot's motion intent. However, after comparing the performance

of this approach with the user average we observed that, first, the legibility achieved by both methods was similar and, second, that trajectories optimized with the second approach would occasionally result in movements that would go outside the PoV of one user, an aspect not reflected in the movements obtained using the user average approach. Thus, we concluded that the user average approach was safer to interact with since it kept the movement always within the PoVs of the users, although in setups where users are not equally distributed throughout the workspace the movement can become slightly biased by taking the user average.

A. Definition of Multi-User Legibility

Under MUL, the legibility of trajectory ξ in a setting comprising N users is defined as

$$Legibility_{MUL}(\xi) = \frac{1}{N} \sum_{n=1}^{N} Legibility_{SULn}(\xi), \quad (3)$$

where $Legibility_{SULn}(\xi)$ is the single-user legibility as perceived by user n. Plugging the legibility expression of MUL in (2), the update step becomes

$$\xi_{i+1} = \xi_i + \frac{1}{\eta} M^{-1} \nabla Legibility_{MUL}(\xi), \quad (4)$$

with

$$\nabla Legibility_{MUL}(\xi) = \frac{1}{N} \sum_{n=1}^{N} R_n^{-1} \cdot \nabla Legibility_{SULn}(\xi).$$

R_n^{-1} is the inverse of the rotation from the coordinate space where ξ was defined to the n-th user's coordinate space, and $\nabla Legibility_{SULn}(\xi)$ is the gradient of the legibility for the nth user, as defined by Nikoladis et al. [16]. Applying R_n^{-1} to $\nabla Legibility_{SULn}(\xi)$ converts the legibility from the n-th user coordinate system to the world coordinate system, allowing to combine the gradients computed in the different users' perspectives.

B. Simulations

To guarantee that the trajectories generated by MUL perform adequately, we compared the performance of MUL with SUL by simulating various scenarios of collaborative interactions. In these simulations we compared the final group legibility of each trajectory. We chose a scenario with three different objects organized on top of a table and three human observers. Considering that in this scenario we had three observers, we optimized the movement once using the MUL model and three times using the SUL model — each time taking into account a different observer's PoV.

To test the resilience of the model, we optimized the movements for each of the three objects, accounting for changes caused from reaching for different objects. Besides optimizing for the different objects, we tested different configurations of the objects on the table, accounting for different relative placements of the objects and their impact on the resulting movements. We obtained 48 trajectories, 12 using MUL and 36 using SUL — 12 for each human PoV, for which we computed the average group legibility.
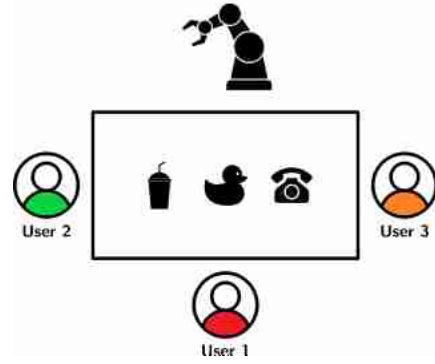

Fig. 1: Setup of the user study. In the center a table with three objects on it – a soda can, a rubber duck and a telephone – and around the table three users looking towards the objects and the robot. Each user has a different point-of-view allowing to evaluate the perception of each user regarding the movements of the robot.

We compared each MUL trajectory with the SUL trajectories for the same pair of configuration and target, resulting in 36 comparisons. The comparisons show that MUL achieved a better group legibility in 69% of the trajectories (25 out of 36). Also, several trajectories using SUL resulted in trajectories that would go outside the FoV of one or multiple of the human users for whom the model was not focusing. In the case of MUL, none of the trajectories passed outside the view of the human users, resulting in safer movements.

IV. EXPERIMENTAL EVALUATION

We conducted a study through Amazon's M-Turk to evaluate the impact in human observers of using MUL against the use of SUL in multi-party interactions. In this study, each participant watched 3 sets of videos of movements, where the participant would either watch movements optimized using multi-user legibility or using single-user legibility.

A. Setup

To evaluate the impact of MUL in the perceived legibility of a robot's movement in a multi-party interaction, the task setup had to be designed to have multiple different perspectives over the same movement while minimizing distractions from those movements.

The scenario, presented in Figure 1, features a robot moving to grab one of three objects placed on top of a table, while human users observe the movement and predict which object the robot is going to grab. The three objects — a soda can, a rubber duck and a telephone — were placed in a single line, evenly spaced between them and with the humans equally distant from the table. The scene also features 3 users: user 1 placed across from the robot, looking in its direction and Users 2 and 3 are placed on each side of the robot, facing each other. Having the users in different sides of the table, induces differences on how each movement is perceived, allowing to better study the impact the different optimizations have on the communication of intent.

B. Design and Hypotheses

In this study we explore the question:

"In interactions with multiple users simultaneously, does combining their perceived views of a robot's movement improve the movement's legibility?"

To support our exploration of the problem and answering the research question, we postulate the following working hypotheses:

H1 *Participants will consider a movement generated using multi-user legibility clearer than when generated using single-user legibility.*

H2 *Participants will understand quicker and more confidently the robot's target, when faced with a movement generated using multi-user legibility than with one generated with single-user legibility.*

The study followed a between-subjects design, with each group being a different optimization approach — either optimization with MUL or with SUL. Each participant watched 3 sets of 3 videos, each set was composed of videos with increasing length — 6, 12 and 18 seconds. All 3 sets used either only MUL or SUL generated movements and between each the robot's objective was randomized between all possible objects to prevent any bias to be created by the participant. Between each set of videos we also randomized the participant's perspective to evaluate the impact of the different perspectives on the movement's legibility.

After each video the participant was asked to predict what object the robot was going to grab and rate how confident they were in their prediction. At the end of each set, the participant would watch a 20 seconds video with the full movement and was asked if the movement matched its prediction. If not, the participant was asked to explain why he predicted another object. At the end of the 3 sets of videos the participant would rate how clear the movements were.

All videos were recorded using the WeBots simulator, an open source simulator developed by Cyberbotics [15]. Figure 2 shows a comparison between movements of the simulated robot towards the soda can on the table, as viewed from the different perspectives, recorded in the WeBots software.

C. Sample

We recruited 315 participants using Amazon's Mechanical Turk (M-Turk), with approximately 98% from the USA and the remaining distributed across Canada, Australia and the UK. We restricted the participation on the questionnaire to these four countries to reduce language barrier problems. The participants' age varied from 23 to 76 years old and an average of 42 years old. An analysis of the education level of the participants shows that 60% have a higher education degree and 90% have finished at least high school. Regarding the occupation of the participants, 94% are employed and 2% are students.

D. Results and Analysis

For each participant we measured: the perceived movement clearness, the average time to predict the robot's target, the average confidence in the prediction and the number of correct predictions. Both the prediction time and the prediction confidence were measured for each set of videos, while the perceived clearness and the number of correct predictions were measured only at the end of all 3 sets.

To measure the prediction times, first we measured the set prediction time: for correct target predictions, we considered the time as the earliest a participant answered correctly without wrongly predicting in subsequent videos; while for wrong predictions, we considered that the prediction time was 20 seconds, the full length of the movement. We then averaged the set prediction times to obtain the average prediction time for each participant. Both the perceived clearness of the movement and prediction confidence were measured using 10 point Lickert scales. We analyzed the confidence in the prediction by combining the scores for each 6, 12 and 18 second videos as in [16]. For the clearness of the movements, as each participant only rated the movements at the end of the study, we did not pre-process the data.

To compare the performance of both optimizations, we grouped the results for each measure as being either from a MUL or a SUL optimization, grouping all the SUL results together. The grouping is needed because SUL optimization focus on improving the understanding of one specific user and depends on that user's PoV over the task. Thus by analyzing the results for each SUL optimization individually, we would compare the performance of MUL with specific instances of SUL, instead of the performance of MUL with the performance of SUL.

Finally, we conducted a normality test that showed us that all three measures — perceived clearness, time taken and confidence in prediction — did not follow a normal distribution, so all the analyses used non-parametric tests.

The analysis of the perceived clearness allows to answer hypothesis **H1**. We conducted a Mann-Whitney test that showed MUL was considered significantly clearer than SUL, $U = 7277.5, p = 0.006$, with MUL achieving an average clearness of 7.338 and SUL achieving 6.4580, thus supporting hypothesis **H1**. Figure 3 shows a boxplot comparison between both models.

To answer hypothesis **H2**, we analysed the time taken to predict the robot's target and the confidence associated with said prediction. For the time taken, a Mann-Whitney test showed people took significantly less time to correctly predict the robot's target with MUL than with SUL, $U = 75722, p = 0.037$. Figure 4 shows that, on average, participants paired with MUL took 10.234 seconds, while participants paired with SUL took 11.331 seconds.

A follow-up analysis of the prediction time showed that 79.6% of the participants paired with MUL needed only 6 seconds to correctly predict the object against 68.5% when paired with SUL, a 10% increase in prediction speed for participants paired with MUL. A Chi-Square test, $\chi^2(2) = 11.012, p = 0.004$, supports the significance of the difference. In this analysis we only considered correct predictions, since our interest was in understanding how early a participant could correctly predict the object with each model.
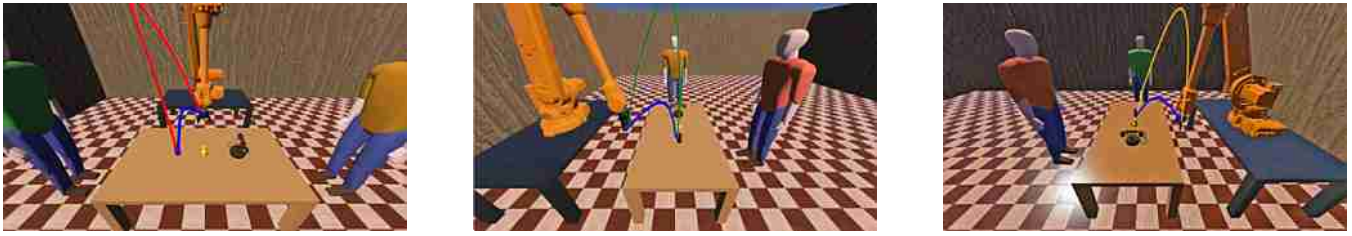
Fig. 2: Example of movements towards the soda can, observed from different PoVs. In blue we have the movement using MUL, while in red, green and orange we have the movement optimized using SUL focused on User 3. On the left we have the movements seen by User 1, in the middle User 2 and on the right User 3. By observing the images, we can see that the SUL movement, while clearly seen by User 3, when seen from the PoVs of Users 1 and 2, goes outside their field of view and sometimes is not as clear as the MUL movement.
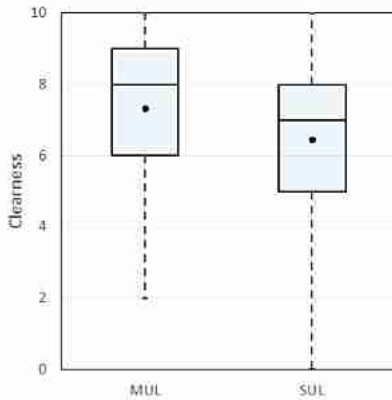

Fig. 3: Boxplot comparing the results for the perceived clearness of the MUL model and the aggregated SUL models. The average for each model is marked with a dot.

Fig. 4: Average time to correctly predict the target according to the model used, with standard deviation.

Regarding the confidence in prediction, a Mann-Whitney test showed participants are significantly more confident in their predictions with MUL than with SUL, $U = 43971.5, p < 0.001$. An analysis of Figure 5 shows that when paired with MUL, on average, participants rated their confidence in the prediction as 6.046 out of 10, opposed to 5.163 when paired with SUL. An analysis of Figure 5 also shows that the ratings for SUL have a bigger dispersion than the ratings for MUL. The results for the confidence and time to correctly predict support hypothesis **H2**.

Finally, regarding the number of correct predictions a Chi-Square test, $\chi^2(1) = 0.018, p = 0.893$, showed no statistical difference between the number of correct predictions between MUL and SUL. Participants paired with MUL correctly predicted the robot's target in 80.5% of the movements, while participants paired with SUL predicted correctly 80.1% of the movements.

E. Discussion

The results in Section IV-D show the positive impact of using MUL to model legibility in multi-party interactions. By combining the different perspectives, the resulting movements are influenced by the perspectives of all the user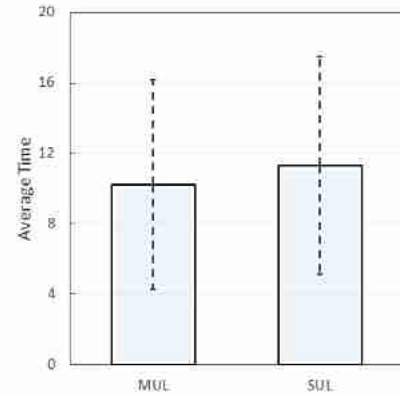s', creating movements that were shown to improve the information conveyed and the human understanding of the robot's intentions.

The results of the user study show that when faced with MUL generated movements, users on average take less time to understand the robot's intentions. The shorter time to understand a robot's intentions creates safer and efficient interactions, namely during collaborations, because it gives humans more time to adapt to the robot and also more time to gather information regarding what the other human partners are trying to achieve and act accordingly. The results also show a higher prediction confidence with lower dispersion of ratings. The consistency of high confidence ratings shows the positive impact MUL has in multi-party interactions because it decreases human need to control a robot's actions, leading to faster decision times and allowing to better focus on the task at hand. A high prediction confidence also plays the important role of guiding humans partners less certain in their understanding of the robot's intentions, who feel the need to look to the other human partners in search of more hints regarding a robot's intentions.

One particular finding that shows the usefulness of the MUL model over the SUL model is that participants showed a 10% increase in the time needed to correctly predicting the robot's target. This improvement is particularly interesting because one of the principal aspects of legibility is the
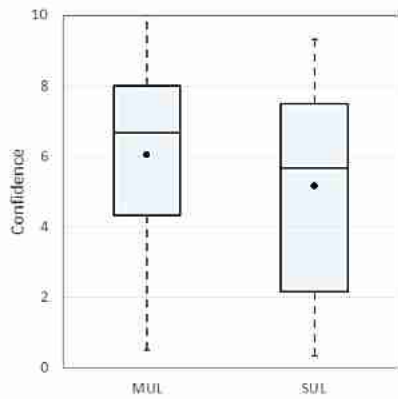
Fig. 5: Boxplot comparing the confidence in the predictions between the MUL model and the aggregated SUL models. The average for each model is marked with a dot.

ability to convey more information in the early portions of the movement. Thus, a 10% increase on the number of participants who needed the least amount of time to correctly predict the robot's objective shows how MUL increases the legibility of movements in multi-party settings.

V. Conclusion

In this work we proposed a model for legibility - MUL - focused on creating movements that improve the legibility of the group instead of each user's individually. MUL improves the group's legibility by taking each user's perceived legibility and combining them into a group legibility metric.

A user study, conducted through M-Turk, validated that jointly considering the different perspectives allows for users in multi-party interactions to infer the robot's intentions faster and more confidently than using the standard single-user approaches. Although MUL may generate less individually legible motions, by improving the group's average legibility MUL improves the group's general understanding of a robot's intention, thus improving team efficiency and safety in the interaction.

MUL is a promising model for applications in entertainment areas such as theatre where movement is an important communication tool and where people observe the movements from different perspectives or healthcare areas like surgery teams where a robot must communicate correctly and clearly for the rest of the medical staff to perform adequately.

Acknowledgments

References

[1] B. Busch et al. "Learning legible motion from human–robot interactions". In: *Int. J. of Social Robotics*. Vol. 9. 5. Springer, 2017, pp. 765–779.

[2] D. Carton et al. "Measuring the effectiveness of readability for mobile robot locomotion". In: *Int. J. of Social Robotics* 8.5 (2016), pp. 721–741.

[3] Y. Che et al. "Efficient and Trustworthy Social Navigation via Explicit and Implicit Robot–Human Communication". In: *IEEE Transactions on Robotics* 36.3 (2020), pp. 692–707.

[4] F. Correia et al. "A social robot as a card game player". In: *Proc. Artificial Intelligence and Interactive Digital Entertainment Conf.* Vol. 13. 1. 2017.

[5] A. Dragan et al. "Generating legible motion". In: *Robotics: Science and Systems*. 2013.

[6] A. Dragan et al. "Legibility and predictability of robot motion". In: *ACM/IEEE Int. Conf. Human-Robot Interaction*. IEEE. 2013, pp. 301–308.

[7] A. Dragan et al. "Effects of robot motion on human-robot collaboration". In: *ACM/IEEE Int. Conf. Human-Robot Interaction*. IEEE. 2015, pp. 51–58.

[8] M. Faria et al. ""Me and You Together" Movement Impact in Multi-User Collaboration Tasks". In: *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*. 2017, pp. 2793–2798.

[9] N. Gildert et al. "The need for combining implicit and explicit communication in cooperative robotic systems". In: Frontiers, 2018, p. 65.

[10] K. Kaplan et al. "Toward human-robot collaboration in surgery: performance assessment of human and robotic agents in an inclusion segmentation task". In: *Proc. IEEE Int. Conf. Robotics and Automation*. IEEE. 2016, pp. 723–729.

[11] R. Knepper et al. "Implicit communication in a joint action". In: *ACM/IEE Int. Conf. Human-Robot Interaction*. 2017, pp. 283–292.

[12] M. Kwon et al. "Expressing robot incapability". In: *ACM/IEEE Int. Conf. Human-Robot Interaction*. 2018, pp. 87–95.

[13] C. Mavrogiannis et al. "Social momentum: A framework for legible navigation in dynamic multi-agent environments". In: *ACM/IEEE Int. Conf. Human-Robot Interaction*. 2018.

[14] F.S. Melo et al. "Project INSIDE: towards autonomous semi-unstructured human–robot social interaction in autism therapy". In: *Artificial intelligence in medicine* 96 (2019), pp. 198–216.

[15] O. Michel. "Cyberbotics Ltd. Webots™: professional mobile robot simulation". In: *Int. J. Advanced Robotic Systems* (2004).

[16] S. Nikolaidis et al. "Viewpoint-Based Legibility Optimization". In: *ACM/IEEE Int. Conf. Human-Robot Interaction*. 2016, pp. 271–278.

[17] S. Saunderson et al. "How robots influence humans: A survey of nonverbal communication in social human–robot interaction". In: *Int. J. of Social Robotics*. Vol. 11. 4. Springer, 2019, pp. 575–608.

[18] E. Yadollahi et al. "When deictic gestures in a robot can harm child-robot collaboration". In: *ACM Conf. Interaction Design and Children*. 2018, pp. 195–206.