Inference VS. Explicitness. Do We Really Need the Perfect Predictor? The Human-Robot Collaborative Object Transportation Case

J. E. Domínguez-Vidal and Alberto Sanfeliu

***Abstract*— When robots interact with humans, limitations in their internal models arise due to the uncertainty and even randomness of human behavior. This has led to attempts to predict human future actions and infer their intent. However, some authors argue for combining inference engines with communication systems that explicitly elicit human intention. This work builds on our Perception-Intention-Action (PIA) cycle, a framework that considers human intention at the same level as perception of the environment. The PIA cycle is used in a collaborative task to compare the effect on different human-robot interaction aspects of using a force predictor that infers human implicit intention versus a communication system that explicitly elicits human intention. A study with 18 volunteers shows that allowing humans to directly express themselves can achieve the same improvement as an intention predictor.**

***Index Terms*— Physical Human-Robot Interaction, Intent Detection, Human-in-the-Loop, User Study**

I. Introduction

Since the dawn of robotics, attempts have been made to enable robots to autonomously perform increasingly complex tasks. The Perception-Action (PA) cycle played a fundamental role in this evolution by enabling the functional decomposition of robot control [1]–[3]. This has led us to design and develop more and more elaborate control systems and architectures based to a greater or lesser extent on how the human brain works [4]. Thus, the correct perception, representation and understanding of the environment has become critical to enable the robot to make the correct decisions when navigating an urban environment [5] or choosing the appropriate tool [6].

However, when we made robots cease to be isolated machines and start interacting with humans, these systems started to encounter certain limitations, mainly motivated by the inherent uncertainty of human behavior [7]–[9]. Perception of the world ceases to be sufficient as the human's intention must also be known. This is when we start trying to predict the human's future actions with increasing success over the years and process these predictions to try to infer their intent [10], [11]. This process is also similar to how we humans tend to work. We learn to detect clues and subtleties in the behavior of our fellow humans by trying to "read" them, usually falling into countless misunderstandings. These same misunderstandings are reproduced in the inference models we have programmed causing errors in the robot's behavior that are usually blamed on limitations in the model or lack of sufficient data to train them when this errors can simply occur because the human has multiple ways of modeling the information they perceive [12], causing two agents to represent the same environment differently.

This has caused some authors [13]–[15] to start considering that it is not enough to achieve human-like robots with a correct perception of their environment but that it is necessary to combine inference engines with communication systems that allow explicit elicitation of the human's intention. This raises the following questions: Can we really achieve a perfect inference engine? And if so, is it really necessary? Would not it be more useful (and computationally less expensive) to foster a more fluid human-robot interaction that would allow obtaining the necessary information explicitly from the human?

This work arises to try to answer some of these questions. Specifically, it is the continuation of our previous work [16] in which we presented our Perception-Intention-Action (PIA) cycle as a framework that allows us to take into account the human intention obtained both implicitly and explicitly at the same level as the perception of the environment. Subsequently, it allows the combination of all sources of information by means of the Situation Awareness (SA) concept, thus keeping only the task-relevant information. This cycle was validated in a human-robot collaborative object transportation task. However, the implicit intention of the human was directly inferred from the actual force being exerted by the human. In this article, our cycle will be used in the same collaborative task to compare the effect on different human-robot interaction (HRI) aspects of using a force predictor that infers the human's implicit intention during the next second versus a communication system that explicitly elicits the human's intention.

In the remainder of the document, we present the related work in Section II. Section III includes our definition of implicit and explicit intention as well as the most relevant details about the two systems used in this work. Section IV presents the hypotheses we wish to test, the setup of the experiments and the results obtained. Finally, Sections V and VI present a brief discussion and the conclusions.

II. Related Work

It is relatively common to find different models in the literature that attempt to infer the human's intention [17]–[22]. In general, they use the human's gaze or their previous

Work supported under the European project CANOPIES grant number (H2020- ICT-2020-2-101016906) and by JST Moonshot R & D Grant Number JPMJMS2011. The first author acknowledges Spanish FPU grant with ref. FPU19/06582.

The authors are with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC). Llorens Artigas 4-6, 08028 Barcelona, Spain. {jdominguez, sanfeliu}@iri.upc.edu. The first one is the corresponding author.

movement, either of the whole body or just the hand in manipulation tasks, to generate a prediction of the object of interest or the trajectory they will follow and thereby infer the human's intention. These models use multiple architectures ranging from Gaussian Mixture Models (GMM) to more recent Artificial Neural Networks (ANN) in all their varieties. However, all of them have non-negligible error rates due to the reasons discussed above (limitations of the model, lack of data, etc.). It is also often pointed out that these errors can be reduced with more complex architectures [23] ignoring that allowing the other agent to explicitly indicate their intention when uncertainty is too high could have more beneficial effects.

This second point of view is the one used by Mullen et al. [24] in which they present a system in which a robot is performing a task autonomously but when it encounters a high uncertainty, it asks the human explicitly in order to reach its goal. Che et al [25] take this idea a step further in a navigation task and make a robot to indicate its intention both implicitly (through legible motions) and explicitly (through a vibration on a wristband on the wrist of each nearby human) when it is in the presence of humans whom it may disturb. While that work uses the opposite approach (it is the robot that reports its intention both implicitly and explicitly rather than asking the human), among their results is that the use of both communication systems increases trust in the robot. What they do not test is which of the two systems is preferred by the human or in what proportion they should be used. The same idea is used in [26] to improve object manipulation between two robots by communicating their plans implicitly through the force exchanged and explicitly by exchanging wireless messages. A final work that demonstrates the usefulness of explicitly relying on the human's intention is [27] in which they perform a collaborative search task in urban environments relying on a smartphone app [28] through which the human can indicate which areas they are going to explore or would like the robot to explore. Without this system, the occlusions of the environment would drastically reduce the performance of the task as each agent would be unaware of the areas explored by their partner.

In the case of tasks involving physical contact, and specifically in collaborative transportation, it is common to find models based on control techniques [29]–[31]. Some of these include some kind of prediction of the human's intention, understood as the human's desired trajectory [21] or the speed profile they would like to follow [32]. To the best of our knowledge, there is no study that analyzes the possibility of eliciting the human's intention explicitly in this task or the effect it has on human-robot interaction.

III. Implicit Inferable Intention VS. Explicit Intention

Starting from what in [13] is called explicit communication (direct communication in Che et al. [25]) and implicit communication (indirect in [25]), we introduce the concepts of implicit and explicit intention in [16]. We consider implicit intention to be that which can be inferred or deduced from the actions of the other agent. In the case of a collaborative transportation task, this intention, understood as the route that each agent wants to follow, can be inferred through the force they exerts on the transported object. However, the environment must also be taken into account, as the same force to the right may mean that the agent wants to turn to the right or simply that they want to avoid the obstacle on the left. Likewise, the previous experience collaborating with each agent is also decisive to know whether the same force to the back indicates the intention to brake completely or simply to slow down.

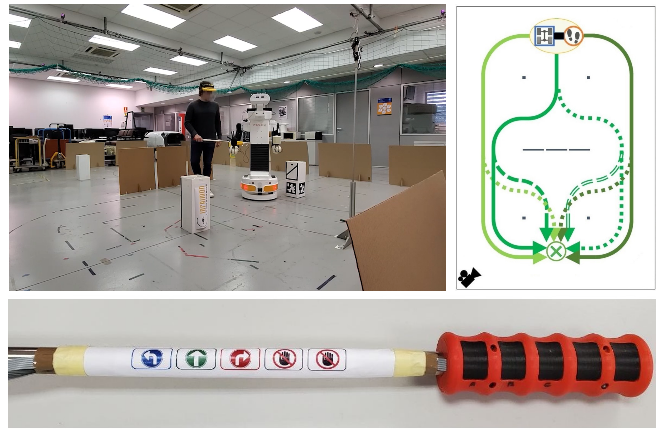

Fig. 1. **Human-robot pair with the transported object.** *Top Left* - Human-robot pair collaboratively transporting an aluminium bar. Goal marked with a chequered flag. *Top Right* - Designed setup for experiments. Different walls and columns that create at least eight routes to the goal. Camera icon represents picture's point of view on the left. *Bottom* - Handle of the transported object for better ergonomics for the human. Five buttons of which the first three are used to explicitly tell the robot which route the human wants. The last two buttons are not used. Meaning of each button annotated next to the handle.

On the other hand, we consider explicit intention to be that obtained using a direct communication channel between both agents using a code known to both. Thus, a system that took into account all of the above considerations when inferring the implicit intention of the human would still make errors simply because the casuistry is almost infinite. Whereas a system that obtained the explicit intention by directly asking the human which route to take when the uncertainty is too high, or simply allowing the human to indicate their intention when they deem it necessary, would reduce the uncertainty and thus the mistakes and the computational burden.

Due to all of the above, in this work we confront two systems. Both are based on the same navigation system used in [16] in which the force exerted by the human is combined with a representative force of the environment perceived by the robot (more technical details in [33]) to navigate through a complex scene with several routes to the goal (Fig. 1 - *Top*). However, each of this systems will obtain the human's intention in a different way.

A. Implicit Intention through Force Prediction

The first of the two systems makes use of our own force predictor to infer the force that the human will exert during the next 1 s and with this deduce their path. It uses a

Deep Learning model that combines Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) units. This model receives as input the evolution during the last 2 s of 4 factors: 1) the robot's LiDAR reading, 2) the representative force of the environment, 3) the robot's speed, 4) the human's exerted force. In this way, multiple aspects of the context are taken into account. For the sake of brevity, we will leave the details concerning the architecture of the predictor outside the scope of this article.

The experiments performed in [16] are used to train this predictor. Finally, we obtain an accuracy in the testset ranging from 94.4% for the force the human is about to exert in the next sampling period to 92.3% for the force to be exerted in 1 s. With this force prediction, the human's desired path is inferred and this inference is used to condition the robot's planner so that it adapts to the human's desires and replans as soon as it detects that these have changed.

B. *Explicit Intention trough Buttons*

The second system replaces the previous inference engine by allowing the human to express which route to follow when they consider it appropriate. For this purpose, five buttons are enabled on the handle created to provide greater ergonomics to the human carrying the object. These buttons allow establishing a direct communication channel through which the human can explicitly indicate when they want to go straight, turn right or left (Fig. 1 - *Bottom*).

IV. Evaluation

We conducted a round of experiments to test the extent to which using a predictor to achieve effective HRI can be beneficial and what aspects can be improved by allowing humans to express themselves explicitly.

A. *Hypothesis*

The first hypotheses we want to test or reject are in line with what can be found in the literature:

H1 - Adding a predictor to the robot's decision making system reduces the human's effort.

H2 - Adding a way to explicitly indicate the human's intention improves the perceived safety and trust in the robot.

H3 - The human prefers the robot to infer their intention rather than having to explicitly indicate it to the robot.

In addition, we posit a fourth hypothesis based on the thesis we advocate in the previous sections as well as in our previous work [16]:

H4 - A system that allows the human to directly express their intention improves multiple aspects of an effective HRI just as much as a system that attempts to infer it.

B. *Experiments Setup*

Three experiments are performed in which volunteers execute the same collaborative transportation task in the same scenario (Fig. 1 - *Top*). The first experiment serves as a baseline for statistical purposes and to allow the human to become familiar with the robot, including its control, its response and movement speed and its tendency to approach

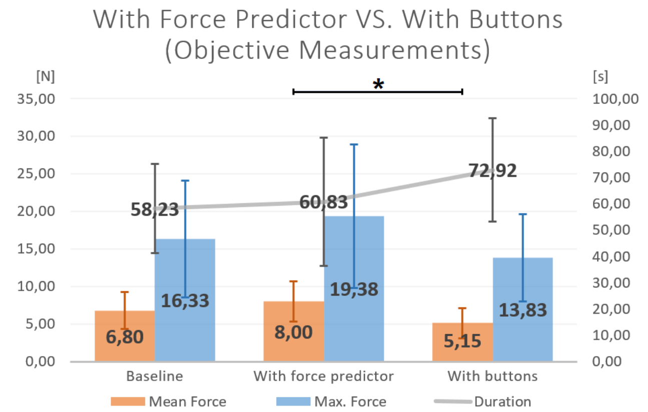

Fig. 2. **Three main objective measures considered.** Mean force exerted in orange, maximum force exerted in blue and duration in gray for the three experiments. Left axis in Newtons associated to both forces and right axis in seconds associated to duration. Statistical significance marked with *: $p < 0.05$

or move away from the obstacles. The second experiment uses the same navigation system (technical details in [33]) with the difference of adding the force predictor discussed in Section III-A. This predictor allows to obtain an estimation of the trajectory desired by the human. This allows to condition the robot's planner to try to adapt to the human's wishes.

The third experiment again makes use of the navigation system used in the first experiment. Instead of adding the previous predictor, it enables the buttons present on the handle discussed in Section III-B to allow humans to explicitly express themselves. This information is used to condition the planner so that at the next fork it is forced to follow the choice indicated by the human[1]. To avoid statistical distortions, the order of the second and third experiments is randomized so that approximately half of the volunteers perform first the experiment with the predictor and then with the buttons and the other half in reverse order. The robot used is the PAL Robotics' TIAGo++[2].

After each experiment, the volunteers are given a handmade questionnaire to valuate, both numerically and by choosing among the different experiments, different aspects of the interaction to evaluate them afterwards using ANOVA tests. All variables analyzed by variance tests are normally distributed according to the Shapiro-Wilk test unless otherwise indicated. Additionally, after finishing the questionnaire, a brief interview with open questions is performed with each volunteer allowing them to express their thoughts about the experiments.

C. *Participants*

Eighteen volunteers, aged between 21 and 55 ($\mu = 29.44$, $\sigma = 7.67$), were recruited from our research institute as well as from different schools of the partner university. Fourteen were male and their self-valuated subjective knowledge of robotics from 1 (none) to 7 (expert) was 3.63 ($\sigma = 1.36$). No volunteers were paid for participating in this study,

[1]Experiments example: `https://youtu.be/hriC-rz_fKY`
[2]`https://pal-robotics.com/robots/tiago/`

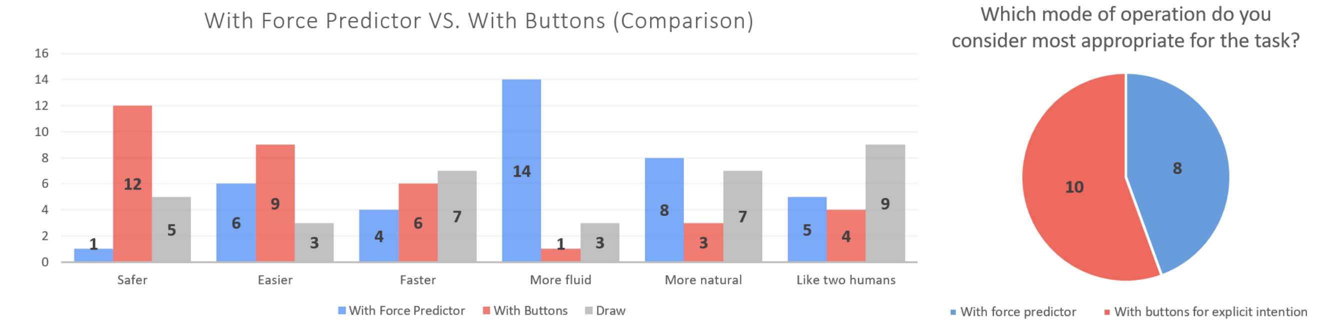

Fig. 3. **Comparison between using our force predictor and explicit intention communication.** *Left* - Election made by the volunteers instead of valuate aspects numerically *Right* - Election made by the volunteers with respect to which system they consider performs better at the task at hand. The maximum is 18 in both cases as it is the number of volunteers.

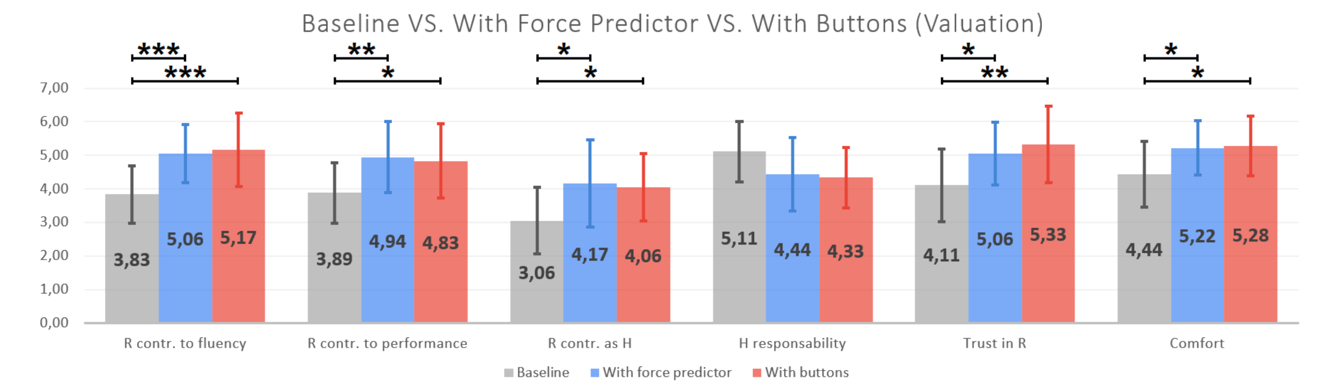

Fig. 4. **Assessment of the main aspects involved in the interaction.** Comparison among the baseline experiment (without predictor or explicit intention buttons) in gray, experiment with the force predictor in blue and experiment with the buttons in the handle in red. Valuation from 1 (very low) to 7 (very high). Statistical significance marked with *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

ensuring that there is no conflict of interest, and all of them have signed an informed consent form. They perform 54 experiments (exposing each volunteer to the 3 conditions). All the experiments reported in this document have been performed under the approval of the ethics committee of the Universitat Politècnica de Catalunya (UPC)[3] in accordance with all the relevant guidelines and regulations (ID: 2021.10).

D. Results

To demonstrate hypothesis H1 we performed several objective measurements taking advantage of the fact that the three experiments are performed in the same scenario. Specifically we looked at the mean and maximum force exerted as well as the duration of each experiment. Fig. 2 shows the values obtained.

The duration of each experiment has a high variability due to two main factors. First, the high number of possible routes means that the human does not always choose the shortest route. Second, each volunteer has a natural tendency to perform more or less force, which results in a higher or lower robot's movement speed. This implies that the variable "Duration" did not satisfy the Saphiro-Wilk test. A non-parametric Mann-Whitney U-test was performed without obtaining significant results. This same variability is also present in the maximum force so another Mann-Whitney U-test was necessary. No significant results were found either.

The mean force does meet the normality condition so that an ANOVA test can be run revealing a statistically significant variance: F=3.129, p-value=0.035. A *post hoc* Tukey's HSD (Honest Significant Difference) test was performed showing a statistically significant reduction according to the criterion of $p < 0.05$ (p=0.019) in the mean force between the experiment with the predictor (μ=8.00, σ=2.67) and with the buttons (μ=5.15, σ=2.01) but not between the baseline and the experiment with the predictor (p=0.052). Therefore, hypothesis H1 was rejected.

To test hypotheses H2 and H3, in addition to asking the volunteers to rate from 1 to 7 various parameters associated with an effective HRI, at the end of the three experiments they were asked to choose between the system with the predictor and the other with the buttons with respect to various aspects. They were also asked which mode of operation seems more appropriate for the task they were performing. Fig. 3 shows the results.

They considered safer and easier to execute the experiment in which they had the buttons over the option with the predictor, confirming part of hypothesis H2. This preference changed radically when they were asked which system they consider allows a more fluid or more natural interaction, opting for the predictor. Likewise, there were a technical

[3]Ethics committee URL: `https://comite-etica.upc.edu/en`

tie when choosing which system subjectively allows the task to be executed faster or which is more similar to how two humans interact. As for H3, Fig. 3 - *Right* shows that there were no preference for the system with the predictor when executing the task as a whole, so hypothesis H3 is rejected.

To test hypothesis H4, and to find out to what extent both systems can improve HRI, volunteers were asked to numerically rate different parameters after each experiment. The result is shown in Fig. 4. Applying ANOVA tests to the perceived contribution of the robot to the task, there is a statistically significant increase in both the contribution to fluency (F=16.57, $p < 0.001$) and performance (F=10.99, $p < 0.001$). Applying a Tukey's HSD test to both parameters, it was obtained that both systems present statistically significant increases with respect to the baseline for the contribution to fluency (with predictor: $p < 0.001$; with buttons: $p < 0.001$) and performance (with predictor: p=0.008; with buttons: p=0.022). This impacted in a statistically significant increase in the consideration that the robot contributes to the task in equal proportion to the human (F=8.904, $p < 0.001$) both for the system with predictor (p=0.014) and using buttons (p=0.033) if we compare them with the baseline.

Analyzing human's responsibility, that is, how responsible the human is for the task to be correctly executed, there was a reduction with both systems but it was not statistically significant (F=2.118, p=0.106). The same did not occur when we analyzed with the same previous procedure the trust that the human has in the robot. There was a statistically significant increase (F=12.94, $p < 0.001$) with both systems, being more pronounced using buttons to explicitly indicate intention (p=0.002) than with the force predictor (p=0.025). This ends up confirming hypothesis H2.

Finally, the comfort rating also showed a statistically significant increase (F=11.02, $p < 0.001$) both using our predictor (0.035) and with the buttons (p=0.021) with respect to the baseline. The coincidence in practically all the analyzed aspects corresponding to an effective HRI between the system with the predictor and the other one with the buttons makes hypothesis H4 be confirmed.

The post-experiment interview shed some light on the previous results. Volunteer 8 indicated "*I would like the robot to be able to do the task without me having to tell it anything*". This explains the perspective on the part of the volunteers that makes them choose the system with the predictor as the most appropriate for the task. However, volunteer 10 commented "*It's good that the robot can predict my intentions but it can be wrong*" in a clear allusion to the human's understanding that the robot is not perfect and therefore can make mistakes. This is what pushes the other half of the volunteers to choose the system with the buttons to express themselves explicitly when necessary. This was confirmed by volunteer 15: "*I prefer to have a way to take control when necessary, it makes me feel more relieved*". This justifies the choice of the system with buttons to indicate their intention as the safer of the two and reaffirms hypothesis H2. It also fulfills the point made in [16] that humans prefer to always have some way to take control.

V. General Discussion

We have implemented two systems to obtain the human's intention in a human-robot collaborative transportation task. The first system makes use of a force predictor to estimate the force to be exerted by the human and, with this estimate, to infer their intention, i.e., the route they wish to follow. The second system makes use of buttons that allow the human to explicitly express their intention when they deem it appropriate. Looking at Fig. 4, it can be stated that the use of a predictor improves the quality of the HRI. However, the interesting part is that allowing the human to express themselves directly achieves practically the same results.

First, it should be noted that the predictor used has an error rate that, although small, is not negligible. Furthermore, it is a force predictor that is used to estimate the human desired trajectory, not a trajectory predictor. This is because the entire robot control system is based on the correct combination of the force exerted by the human and the understanding of the environment through attractive and repulsive forces [33], making it much more natural to predict forces. The estimation of the human trajectory can therefore be improved, not only by improving the predictor used but also by adding other inputs that could be relevant such as the person's gaze. However, it is pertinent to ask whether we really want a perfect predictor.

That hypothesis H1 cannot be affirmed and hypothesis H3 is rejected goes against further improvement of our predictor. At the same time, the comments expressed by the volunteers advise against looking for a perfect predictor because they prefer to continue to have some way that makes them feel they are in control of the task, as we had already observed in [16]. This advises to have a good enough predictor (still necessary in situations where communication is not possible) and to go for better communication methods, which allow this exchange of explicit intentions in the most natural way possible, e.g., with natural language processing [34] or gesture communication [35].

In the end, this approach is closer to how we humans tend to work. It is true that we try to predict and infer as much as we can. However, when our task becomes complicated and the risk or cost of being wrong is high, we tend to prefer to ask our peers to reduce uncertainty or simply ask for help. If we aspire to create robots that are perceived as companions and not just machines, we should not forget this approach. In any case, this work should not be understood as being against the use of predictors, as they are still necessary when communication is not possible.

Finally, it is worth mentioning that one of the weaknesses of this work is that, although the population sample is varied in age and educational background, it may be small. In turn, the result in Fig. 3 regarding naturalness is largely dependent on the abstraction capacity of each volunteer to understand that the buttons used are equivalent to talking directly to the robot. Further experiments using a "more human-like" communication system (for example, with natural language processing) should be done to see if any difference occurs.

VI. CONCLUSIONS

In this article we have challenged the current trend of creating increasingly accurate predictors to try to infer more and more details of the human's intention or future actions. Through a collaborative transportation task we have found that, while using a force predictor can improve multiple aspects associated with effective HRI, giving the human the ability to express their intention explicitly obtains virtually the same results. Furthermore, the option with the predictor has not been shown to be preferred by volunteers.

These results support the idea that we should pivot towards methods that seek to improve human-robot communication rather than attempting to infer it in the best possible way. However, limitations observed in the task indicate that further experiments should be conducted on other tasks or using methods of eliciting the human's explicit intention that are more natural to the human to confirm this approach.

REFERENCES

[1] J. S. Albus, "A New Approach to Manipulator Control: The Cerebellar Model Articulation Controller (CMAC)," *Transactions ASME*, 1975.

[2] R. Brooks, "A robust layered control system for a mobile robot," *IEEE journal on robotics and automation*, vol. 2, no. 1, pp. 14–23, 1986.

[3] J. S. Albus *et al.*, "A Reference Model Architecture for Intelligent Systems Design," *An Introduction to Intelligent and Autonomous Control*, pp. 27–56, 1993.

[4] V. Cutsuridis and J. G. Taylor, "A cognitive control architecture for the perception-action cycle in robots and agents," *Cognitive Computation*, vol. 5, no. 3, pp. 383–395, 2013.

[5] A. Goldhoorn, A. Garrell, R. Alquézar, and A. Sanfeliu, "Searching and tracking people in urban environments with static and dynamic obstacles," *Robotics and Autonomous Systems*, vol. 98, pp. 147–157, 2017.

[6] N. Saito, T. Ogata, S. Funabashi, H. Mori, and S. Sugano, "How to select and use tools?: Active perception of target objects using multimodal deep learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2517–2524, 2021.

[7] A. D. Dragan, "Robot planning with mathematical models of human state and action," *arXiv preprint arXiv:1705.04226*, 2017.

[8] R. Choudhury, G. Swamy, D. Hadfield-Menell, and A. D. Dragan, "On the utility of model learning in hri," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 317–325.

[9] G. R. Ghosal, M. Zurek, D. S. Brown, and A. D. Dragan, "The effect of modeling human rationality level on learning rewards from multiple feedback types," *arXiv preprint arXiv:2208.10687*, 2022.

[10] F. J. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[11] P. Schydlo, M. Rakovic, L. Jamone, and J. Santos-Victor, "Anticipation in Human-Robot Cooperation: A Recurrent Neural Network Approach for Multiple Action Sequences Prediction," in *2018 IEEE International Conference on Robotics and Automation*. IEEE, 2018, pp. 5909–5914.

[12] P. N. Johnson-Laird, *Mental models*. The MIT Press, 1989.

[13] N. Gildert, A. G. Millard, A. Pomfret, and J. Timmis, "The need for combining implicit and explicit communication in cooperative robotic systems," *Frontiers in Robotics and AI*, vol. 5, p. 65, 2018.

[14] B.-J. Lee *et al.*, "Perception-Action-Learning System for Mobile Social-Service Robots using Deep Learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[15] S. Dar and U. Bernardet, "When agents become partners: A review of the role the implicit plays in the interaction with artificial social agents," *Multimodal Technologies and Interaction*, vol. 4, no. 4, p. 81, 2020.

[16] J. E. Domínguez-Vidal, N. Rodríguez, and A. Sanfeliu, "Perception-Intention-Action Cycle as a Human Acceptable Way for Improving Human-Robot Collaborative Tasks," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, p. 567–571.

[17] R. C. Luo and L. Mai, "Human Intention Inference and On-Line Human Hand Motion Prediction for Human-Robot Collaboration," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5958–5964.

[18] S. Jain and B. Argall, "Recursive bayesian human intent recognition in shared-control robotics," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3905–3912.

[19] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 83–90.

[20] I. Maroger, N. Ramuzat, O. Stasse, and B. Watier, "Human trajectory prediction model and its coupling with a walking pattern generator of a humanoid robot," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6361–6369, 2021.

[21] K. I. Alevizos, C. P. Bechlioulis, and K. J. Kyriakopoulos, "Physical human–robot cooperation based on robust motion intention estimation," *Robotica*, vol. 38, no. 10, pp. 1842–1866, 2020.

[22] A. Thobbi, Y. Gu, and W. Sheng, "Using human motion estimation for human-robot cooperative manipulation," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 2873–2878.

[23] D. Fridovich-Keil, A. Bajcsy, J. F. Fisac, S. L. Herbert, S. Wang, A. D. Dragan, and C. J. Tomlin, "Confidence-aware motion prediction for real-time collision avoidance," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 250–265, 2020.

[24] J. F. Mullen, J. Mosier, S. Chakrabarti, A. Chen, T. White, and D. P. Losey, "Communicating inferred goals with passive augmented reality and active haptic feedback," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 8522–8529, 2021.

[25] Y. Che, A. M. Okamura, and D. Sadigh, "Efficient and trustworthy social navigation via explicit and implicit robot–human communication," *IEEE Transactions on Robotics*, vol. 36, no. 3, pp. 692–707, 2020.

[26] N. Gildert, "Combining implicit and explicit communication in object manipulation tasks between two robots," Ph.D. dissertation, University of York, 2022.

[27] M. Dalmasso, A. Garrell, J. E. Domínguez-Vidal, P. Jiménez, and A. Sanfeliu, "Human-Robot Collaborative Multi-Agent Path Planning using Monte Carlo Tree Search and Social Reward Sources," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2021.

[28] J. E. Domínguez-Vidal, I. J. Torres-Rodríguez, A. Garrell, and A. Sanfeliu, "User-Friendly Smartphone Interface to Share Knowledge in Human-Robot Collaborative Search Tasks," in *30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2021, pp. 913–918.

[29] S. Tarbouriech, B. Navarro, P. Fraisse, A. Crosnier, A. Cherubini, and D. Sallé, "Admittance control for collaborative dual-arm manipulation," in *2019 19th International Conference on Advanced Robotics (ICAR)*. IEEE, 2019, pp. 198–204.

[30] A. Bussy, P. Gergondet, A. Kheddar, F. Keith, and A. Crosnier, "Proactive behavior of a humanoid robot in a haptic transportation task with a human partner," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 962–967.

[31] X. Yu, B. Li, W. He, Y. Feng, L. Cheng, and C. Silvestre, "Adaptive-constrained impedance control for human–robot co-transportation," *IEEE transactions on cybernetics*, vol. 52, no. 12, pp. 13 237–13 249, 2021.

[32] A. Al-Yacoub, Y. Zhao, W. Eaton, Y. M. Goh, and N. Lohse, "Improving human robot collaboration through force/torque based learning for object manipulation," *Robotics and Computer-Integrated Manufacturing*, vol. 69, p. 102111, 2021.

[33] J. E. Domínguez-Vidal, N. Rodríguez, R. Alquézar, and A. Sanfeliu, "Perception-Intention-Action Cycle in Human-Robot Collaborative Tasks," *arXiv preprint arXiv:2206.00304*, 2022.

[34] D. Mukherjee, K. Gupta, L. H. Chang, and H. Najjaran, "A survey of robot learning strategies for human-robot collaboration in industrial settings," *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102231, 2022.

[35] M. Peral, A. Sanfeliu, and A. Garrell, "Efficient hand gesture recognition for human-robot interaction," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 272–10 279, 2022.