2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 23-27, 2022, Kyoto, Japan

Bounded Rational Game-theoretical Modeling of Human Joint Actions with Incomplete Information

Yiwei Wang[1], Pallavi Shintre[2], Sunny Amatya[2], Wenlong Zhang[2†]

***Abstract*— As humans and robots start to collaborate in close proximity, robots are tasked to perceive, comprehend, and anticipate human partners' actions, which demands a predictive model to describe how humans collaborate with each other in joint actions. Previous studies either simplify the collaborative task as an optimal control problem between two agents or do not consider the learning process of humans during repeated interaction. This idyllic representation is thus not able to model human rationality and the learning process. In this paper, a bounded-rational and game-theoretical human cooperative model is developed to describe the cooperative behaviors of the human dyad. An experiment of a joint object pushing collaborative task was conducted with 30 human subjects using haptic interfaces in a virtual environment. The proposed model uses inverse optimal control (IOC) to model the reward parameters in the collaborative task. The collected data verified the accuracy of the predicted human trajectory generated from the bounded rational model excels the one with a fully rational model. We further provide insight from the conducted experiments about the effects of leadership on the performance of human collaboration.**

I. INTRODUCTION

Physical human-robot interaction (pHRI) has become ubiquitous in robot-assisted rehabilitation, robotic surgery, and collaborative manufacturing [1]. In these applications, an intelligent robot needs to build a human model (often referred to as *theory of mind*) to anticipate human actions for proactive planning of its own actions [2]. Modeling humans is a challenging task not only because of the inherent uncertainties of human decision-making [3] but also due to the human learning and adaptation of robots mediated by physical interactions [4]. In order to investigate human uncertainty and adaptation mechanisms during the physical collaboration between humans and robots, it is critical to study of how humans collaborate with each other.

Inverse optimal control (IOC) methods, which assume human motor behavior follows an optimal control law, are promising approaches to describing human physical actions [5]. The oversimplification of human collaboration behavior limited the variety of human motion this model can describe. Inverse reinforcement learning (IRL) emphasizes the need to address the *value alignment problem* in the human decision-making process, which is trying to identify how humans allocate preferences on the elements of a *multi-attribute* cost function [6]. Previous work implemented IRL algorithms to model general human-human interaction or collaboration by adding the features that were related to the other human agents in the system to the cost function [7]. Although these studies demonstrated competent results of modeling human collaboration with IRL methods, none of these approaches considered the human prediction and adaptation of their partners when executing collaborative tasks.

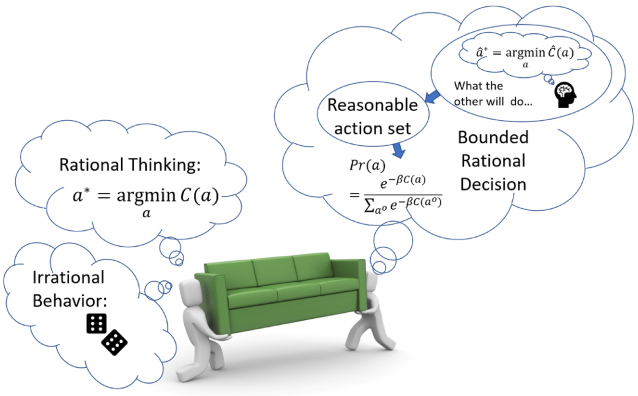


Fig. 1: Illustration of comparison between different rationality models. An irrational agent will choose an action at random. A rational thinker will choose an action with the best outcome. A bounded rational agent will build the theory of mind of the other agent and use bounded rationality to generate action.

Game-theoretic models have been widely adopted to describe interactive behaviors of humans [8]. Due to the nature of incomplete information of human-involved interactions, modeling how individuals predict their partners is important. In [9], the mutual adaption behavior was modeled with a multi-agent recursive least square learning model, and a *empathetic* intent inference framework was proposed to consider the mutual learning process of intelligent agents [10]. Li et al. proposed a differential game model for improving the robot controller by understanding the control strategy of human users in pHRI tasks [11].

Another key aspect in modeling human behavior is to understand its nature of randomness. Our previous research showed that human participants did not strictly follow the optimal strategies to complete the task [12]. Recent research adopted the idea of suboptimal behavior to describe human motor behavior [13]. The authors of [14] and [15] employed

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1944833.

[1]Y. Wang is with the State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan, China. He was with the School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, 85281, AZ, USA. Email: wang_yiwei@hust.edu.cn.

[2]P. Shintre, S. Amatya, and W. Zhang are with the School of Manufacturing Systems and Networks, Ira A. Fulton Schools of Engineering, Arizona State University, Mesa, 85212, AZ, USA. Email: {pshintre, samatya, wenlong.zhang}@asu.edu.

[†]Address all correspondence to this author.

2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-7927-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/IROS47612.2022.9982108

the assumption of the bounded rational behavior in the pHRI scenarios.

Prior studies on human-human cooperative motion conducted through haptic interfaces showed the capability of capturing the haptic behavior of human collaboration [16], [17]. The fact that the data sets from these experiments are rarely used for constructing a dynamic model of human collaborative motion generation inspired us to explore feasible methods to construct dynamic models describing cooperative pHRI scenarios.

A dynamic model that predicts human motion in a cooperative task would vastly help the robot to achieve safe and efficient collaboration with human partners in pHRI scenarios. The main contribution of this paper is threefold. First, we propose a novel human collaborative dynamic model based on game theory and bounded rationality, which is the first model to describe human dyadic behavior with the bounded rational theory. Second, we design a haptic interface with a virtual environment and conduct experiments with human subjects. The collected human data are used to construct the proposed models and verify their performance of predicting human cooperative behaviors. The third contribution is to observe and report evidence of human cooperative nature, such as the learning process between the human dyads. An interesting relationship between the leadership of the dyads and task performance is also reported in the paper, which asserts that clear role assignment during the collaboration could significantly improve the performance of dyads.

The rest of the paper is organized as follows. In Section II, the bounded rational game theoretical human cooperative model is introduced. The IOC method we adopt to learn the parameters of the reward function and probability distribution of trajectory is discussed. The design of the haptic virtual environment and the setup of the experiment are presented in Section III. The comparative results between our proposed game theoretic bounded rational human model and fully rational model with ground truth are presented in Section IV. Section V contains our conclusion and future plans.

II. GAME-THEORETIC AND BOUNDED RATIONAL HUMAN COLLABORATION MODEL

Figure 1 generally demonstrates the difference between bounded rationality, complete rationality, and irrational behavior. In this section, we start by defining the different strategies of a human agent before introducing our game-theoretic and bounded-rational human model. A human agent with *complete rationality* always chooses the optimal action, which is described as rational thinking in Fig. 1.

In this section, we describe the dynamic human cooperative model as a two-player non-zero-sum game. The optimal solution with complete information is discussed, based on the Nash Equilibrium of the game under the assumption of complete rationality. Then we propose the bounded rational game-theoretic human model, accommodating the rationality assumption of the opponents. The nature of human incomplete information in physical interaction is also discussed in this section. Lastly, we will introduce the IOC method to learn the reward and rationality parameters of our model based on observed data collected through the experiment.

A. Two-player joint object manipulation game

We set up a game to simulate a human-human cooperative task, where two human players are trying to move a heavy object, such as a couch, to the desired position as shown in Fig. 2. Each player is positioned at one end of the large long object which represents the couch in our setup. The grasping points on the object for both the participants are displayed as blue and yellow handles. The direction of the force is limited to be perpendicular to the object. This constraint demands collaboration of two players to complete the game.

As Fig. 2 shows, the position of the object's center of mass is presented as a vector $p = [p_x, p_y]^T$. Velocity vector is denoted as $v = [v_x, v_y]^T$. The orientation of the object is denoted as θ. Then we can define the state vector $x \in \mathbb{R}^6$ of the system as follows,

$$x = [p, v, \theta, \dot{\theta}]^T. \tag{1}$$

The system dynamics of the non-cooperative two-player non-zero-sum game in the discrete-time domain is presented as

$$x(k+1) = f(x(k), u_1(k), u_2(k)), \tag{2}$$

where $x(k+1) \in \mathbb{R}^6$ is the system state at time step $k+1$.The shared states are observable for both agents, which indicates the game has *complete information* setup.

B. Rational solution of human collaboration: A Nash equilibrium approach

1) Cost functions: A cost function for each player should capture these factors: the cumulative sum of the payoff of approaching the goal and the cost of actions. We denote the cost function of agent $i \in \{1, 2\}$ in a finite horizon quadratic form as follows,

$$\begin{aligned} C_1(x(k), \xi_1(k), \xi_2(k)) &= \sum_{n=0}^{H-1} \phi_1^T(k+n) W_1 \phi_1(k+n), \\ C_2(x(k), \xi_1(k), \xi_2(k)) &= \sum_{n=0}^{H-1} \phi_2^T(k+n) W_2 \phi_2(k+n), \end{aligned} \tag{3}$$

where we define the trajectory of u_i as $\xi_i(k) = \{u_i(k), u_i(k+1), ..., u_i(k+H-1)\}$. $H \in \mathbb{N}^+$ denotes the control horizon of the system. $W_i \in \mathbb{R}^{N \times N}$ is a constant weighting matrix, which represents the preference for each feature.

2) Nash equilibrium sets: According to its definition, a Nash equilibrium solution is defined as a pair of action trajectories from both agents $< \xi_1^*, \xi_2^* >$, which satisfies the following conditions [18],

$$\xi_1^*(k) = \underset{\xi_1(k) \in \Xi}{\arg\min}\, C_1(\xi_1(k), \xi_2^*(k), x(k)) \tag{4}$$

$$\xi_2^*(k) = \underset{\xi_2(k) \in \Xi}{\arg\min}\, C_2(\xi_1^*(k), \xi_2(k), x(k)). \tag{5}$$

where Ξ is the permissible action trajectory sets for the two agents. We define a set $\Xi^*(k)$ that contains all the Nash equilibrium pairs at time k. Then they pick a solution that

minimizes its own cost from $\Xi^*(k)$. This optimal solution $\xi_i^o(k)$ is denoted as follows,

$$\xi_1^o(k) = \underset{\xi_1(k), <\xi_1(k),\xi_2(k)> \in \Xi^*(k)}{\arg\min} C_1(\xi_1(k), \xi_2(k), x(k)) \tag{6}$$

$$\xi_2^o(k) = \underset{\xi_1(k), <\xi_1(k),\xi_2(k)> \in \Xi^*(k)}{\arg\min} C_2(\xi_1(k), \xi_2(k), x(k)) \tag{7}$$

By solving (6) and (7), we can obtain the optimal control trajectories. Then we choose the first element from each trajectory as their control input for current time step, hence $u_1^o(k) = \xi_1^o(k)[0]$ and $u_2^o(k) = \xi_2^o(k)[0]$. $\cdot[0]$ denotes the first element of the set.

C. Imperfect collaboration: rational solution under incomplete information

In practice, the human players are unaware of the other player's decision-making and motion generation process, which makes the Nash equilibrium solution provided in the prior subsection inaccessible. Thus, the human subjects need to estimate other players' decision-making processes. For this reason, from the perspective of agent 1, the optimal solution $\xi_i^o(k)$ in Eq. (6) and (7) becomes,

$$\hat{\xi}_1^o = \underset{\xi_1, <\hat{\xi}_1^{(1)}, \hat{\xi}_2^{(1)}> \in \hat{\Xi}^{*(1)}}{\arg\min} C_1(\xi_1, \hat{\xi}_2^{(1)}, x(k)) \tag{8}$$

$$\hat{\xi}_2^o = \underset{\xi_2, <\hat{\xi}_1^{(1)}, \hat{\xi}_2^{(1)}> \in \hat{\Xi}^{*(1)}}{\arg\min} \hat{C}_2(\hat{\xi}_1^{(1)}, \xi_2, x(k)) \tag{9}$$

where $\hat{\cdot}^{(1)}$ denotes the estimated values from agent 1's perspective. The estimated Nash equilibrium set is presented as $\hat{\Xi}^{*(1)} = \{< \xi_1^{*(1)}, \xi_2^{*(1)} >\}$ where $< \xi_1^{*(1)}, \xi_2^{*(1)} >$ satisfies,

$$\xi_1^{*(1)} = \underset{\xi_1 \in \Xi}{\arg\min} C_1(\xi_1, \xi_2^{*(1)}, x(k)) \tag{10}$$

$$\xi_2^{*(1)} = \underset{\xi_2 \in \Xi}{\arg\min} \hat{C}_2^{(1)}(\xi_1^{*(1)}, \xi_2, x(k)). \tag{11}$$

Essentially, agent 1 should estimate the parameter vector, which is denoted as $\hat{W}_2^{(1)}$, of agent 2's cost function $\hat{C}_2^{(1)}$. In the meanwhile, agent 2 would conduct similar process to estimate $\hat{W}_1^{(2)}$ to generate its estimated optimal control sequence $\hat{\xi}_2^o$.

D. Bounded rationality

The previous sections provided an overview of the game-theoretic model of the human motion generation process with *complete rationality* under incomplete information scenarios. Previous studies investigated human bounded rational behavior under circumstances where the time and information were limited [19], [13]. Under this assumption, the human decision-making process is no longer deterministic as Eq. (8) shows. For agent 1, the decision is made according to the following probability density function,

$$Pr(\xi_1) = \frac{exp(-\alpha_1 C_1(\xi_1, \hat{\xi}_2^{o(1)}, x(k)))}{\sum_{\xi_1^i \in \Xi} exp(-\alpha_1 C_1(\xi_1^i, \hat{\xi}_2^{o(1)}, x(k)))} \tag{12}$$

$\alpha_1 \in \mathbb{R}^+$ is the *rational coefficient*, which quantifies the level of agent's rationality. A higher α value makes the agent better at maximizing his reward making it highly rational, on the other hand when $\alpha = 0$, the agent is simply choosing actions uniformly at random. The prediction of other agent's behavior from agent 1's perspective is denoted as $\hat{\xi}_2^{o(1)}$.

Compare to [14], [15], where the human agents' cost functions are merely dependent on their own actions and states, the cost functions in our system rely on both agents' actions and the shared system states. This tangle of interests of both agents creates challenges for the human agent to evaluate its future actions, which also depend on its partner's actions. To tackle this difficulty, the human agent needs to develop a mechanism to make a reasonable prediction of its partner. Hence, we propose that the human agent assumes its partner as a *complete rational* agent, who behaves following the best action in the Nash equilibrium set of the game. Then, we combined the bounded rationality decision-making process that is shown above, with the game-theoretical model that provides a reasonable prediction of the partner's future actions. The novelty of our model is to reproduce human subjects' adaptation by considering the equilibrium solution of the game instead of each individual's optimality, which is akin to the human's decision-making process.

E. Inverse optimal control

The inverse optimal control (IOC) problem was formed to learn the parameters in the cost functions and probability distributions from the observation data we collected for each pair of participants. In the previous subsections, we have described how human behavior is governed by the bounded rational game-theoretic model. We can generate a series of states that describe the cooperative process of an experiment trial with a bounded rational action human model. We denote the series of states as $X = [x(0), x(1), ..., x(T)]$, where T presents the final time step of a trial.

According to the bounded rational game-theoretic model, if a set of parameters $\Psi = [W_1, \hat{W}_1, W_2, \hat{W}_2, \alpha_1, \alpha_2]$ and a proper initial condition $x(0)$ are given, a corresponding state trajectory of a trial, which can be denoted as $X_s(\Psi, x(0)) = [x_s(0), x_s(1), ..., x_s(T)]$, can be simulated by applying the proposed model.

Consider the observed system states of a trial in the experiment, to be $X^0 = [x^0(0), x^0(1), x^0(2), ..., x^0(T)]$, where T is the last time step of the collected data. The fitness function (the objective function) is defined as the square of the frobenius norm ($\|\cdot\|_F^2$) of difference between the observed and simulated trajectory, $g(\Psi) = \left\|X^0 - X_s(\Psi, x(0))\right\|_F^2$.

In our experiment, we constrained all the trails to have the same initial condition $x(0)$, which made X_s a function of Ψ. The IOC problem is formulated to be an optimization problem, which is to find the Ψ that makes the fitness function minimum, $\Psi^* = \arg\min_\Psi g(\Psi)$.

Prior work suggests using Genetic Algorithm (GA) to solve similar problems [5], [20]. By applying the GA algorithm to search for the optimal value of Ψ, eventually, we could learn a set of parameters that contains all the parameters of the bounded rational game-theoretical models for the human dyads.

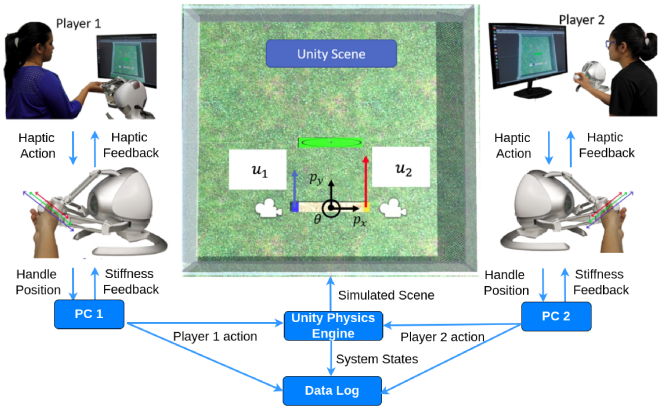


Fig. 2: Experiment setup during collaborative object pushing. Each player observes the current position of the object on the screen and inputs haptic action to move their end of the object using NOVIT Falcon input device. In real-time, the data is stored in a data log as well as updated through the unity physics engine to create the unity scene.

III. EXPERIMENT SETUP

In this section, we introduce an experimental study to analyze human behavior in a haptic virtual interface. The hardware setup for the aforementioned virtual interface is described. We further explain the participant recruitment detail and the instructions for the experiment.

A. Hardware configuration

The hardware setup is shown in Fig. 2. We used two NOVINT Falcon haptic controllers as the human input devices, which were connected to two PCs via USB connection. A C++ script was running on each PC to collect the raw data from NOVINT Falcon and convert them to the control input of each player in the system. The simulation of the dynamic system was implemented using the UNITY Physics Engine, which receives human inputs from haptic devices.

B. Participation recruitment and experiment process

The study was approved by the Institutional Review Board (IRB) of Arizona State University (STUDY00011502). A total of 30 participants (randomly assigned to 15 dyads) volunteered to participate in the experiment. None of the participants had experience with haptic devices before. The participants were brought to a quiet room and were asked to sit in front of two isolated PC, each of which connected to a haptic device. The two participants were facing the opposite direction with a makeshift curtain in between. Because they were not able to see or talk to each other during the entire study, they could only communicate with their actions. The participants were briefed about the hardware and joint object manipulation task and a consent form was signed. The participants were asked to move the object with the handle of the haptic device and try to reach the goal position of the object within 10 seconds. After each trial, the participants were given a 30-second break while the system was reset. Each pair were asked to repeat the same task 10 times, which counted as 10 trials of data. Since none of the participants had operated a haptic device before, we set the first two trials as the *learning phase*, and the data were not used to build human models.

IV. RESULTS AND DISCUSSION

A. General observation: human adaptation

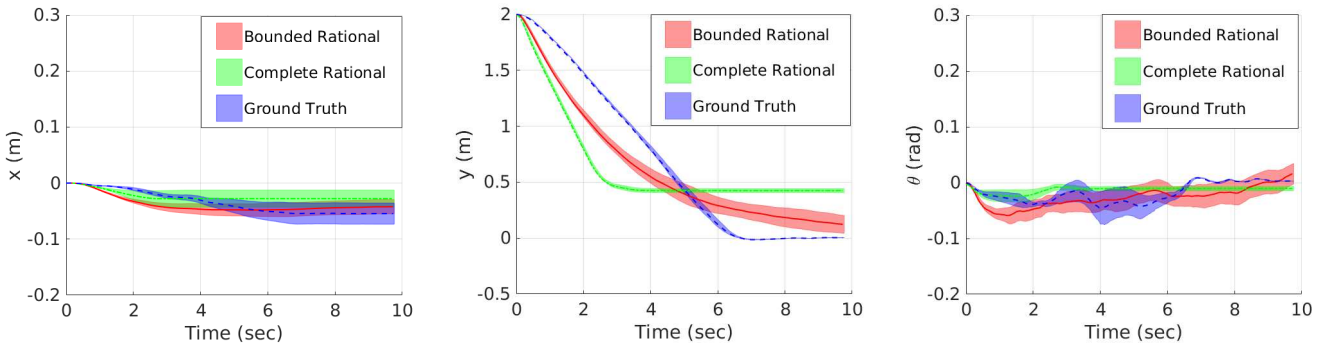
We start the discussion of results by highlighting two main observations from human data. The first one is the learning process of the human participants, which is indicated by the improvement of goal tracking performance in the first two trials. To check our hypothesis on the learning process, we compared the steady-state goal tracking errors, which contain translation and rotation errors, of the first two trials and the rest of all participants. The mean steady-state error of the first two trials is 0.3834, while the error of the rest trials is 0.1350. The t-test between the two means also suggests a significant difference ($p = 0.0049$). According to Fig. 4a, the participants were accustomed to the task after the second trial, which motivated us to mark the first two trials of the experiment as the *learning phase* as discussed in Sec. III. The human participants could not accomplish the task within the 10-second time limit in the first trial, while they could maintain acceptable performance to accomplish the task in trials 3 to 10. This observation agrees with the collaboration learning phenomenon between the dyads during haptic interactions with incomplete information [21].

Summary From Fig 4a, we can observe the task completion performance of the first two trials is longer than the rest of the trial and is termed as the *learning phase*.

B. The effect of bounded rationality

To justify the advantage of the bounded rational model, we applied the IOC method in Sec. II-D on the observed data to generate the bounded rational game-theoretical model for each of the participants. For each experiment, we combined their trial 3 to 10 data into a data set. Then we learned two bounded rational models, each of which is associated with one of the participants. We simulate the game with the two bounded rational human cooperative models to generate another data set, which contains the states' trajectories of 100 trials of the game. The difference between each generated trajectory to the corresponding participant's mean trajectory is defined as the fitting error.

In the meanwhile, we adopted a similar IOC process to learn two complete rational game-theoretic models, which we discussed in Sec. II-B. An example of the comparison of the model fitting results is presented in Fig. 3. In this figure, we choose pair 9 to demonstrate the advantage of the bounded rational model. The sub-figures of Fig. 3 represent the displacement of x and y directions and orientation of the object during an experimental trial, where the colored regions represent the mean and variance of the trajectories for all the trials. As Fig. 3 shows, the trajectory band for the bounded rational models is closer to the original data than the one of complete rational models, especially in the y-direction where the majority of the motion happened. We ran a t-test for the comparison of the mean steady-state errors between the complete rational model and the bounded rational model. As a result of the one-tailed student t-test, the fitting error of the bounded rational model was significantly lower than the complete rational model with $p < 0.003$ ($p = 4.6636e - 6$).



(a) Comparison of the x-direction displacement over time within trials of pair 9.

(b) Comparison of the y-direction displacement over time within trials of pair 9.

(c) Comparison of the object orientation over time within trials of pair 9.

Fig. 3: Comparison of the trajectories of the data sets generated with bounded rationality and complete rationality for pair 9. The bounded Rational model is able to generate a trajectory similar to ground truth data in the y direction.

The main reason why complete rational models had larger steady-state errors than the bounded rational ones is the assumption of human optimal behavior. The complete rationality assumption of human, which expected that human always chooses the optimal actions, would attribute the suboptimal actions of human participants to their lack of motivation, which diminishes the weighting factor on goal tracking in the learned human models. In the last subsection, we discussed the human suboptimal behavior, which is ubiquitous despite the human participants' strong desires. From the comparison of the simulated game playing under different assumptions of human mode, we could draw a conclusion that the bounded rational model can better describe human suboptimal behaviors when they interact with someone they are not familiar with.

Summary Figure 4b suggests that the bounded rationality models are closer to the ground truth cooperative behavior of the human dyads.

C. Leadership and performance of collaboration

In this subsection, we will report an important observation between the difference in the effort each individual invested in playing the game to the goal tracking errors, which evaluate the performance of the game. Before demonstrating the results, it is critical to define the roles in the game. In this experiment, we observed that the participants never started the game simultaneously; one participant, which we define as the *leader*, initiated the game by pushing the object first. Then the other participant, who was defined as the *follower*, would start his or her motion to collaborate. For a specific pair of participants, the roles assigned to each individual during every trial could be different. Due to the limited time, the roles were fixed within each trial, which indicated that the leader's actions would dominate the collaboration between the participants. To evaluate the effort each participant invested in each trial, we introduce the effort metric E_i for agent $i \in \{\text{“}Leader\text{”}, \text{“}Follower\text{”}\}$. We define E_i as follows,

$$E_i = \int_{t_o}^{t_f} ||u_i(t)||dt \tag{13}$$

where t_0 and t_g are the starting time and final time of the trial. $u_i(t)$ is the control input of agent i at time t. The goal tracking error of each trial was measured by calculating the distance between the actual final states of the object and the goal, which evaluated the quality of the collaboration. A lower goal tracking error indicated better performance during the specific trial. The difference in the effort metric between the leader and follower, hence $E_{diff} = E_{Leader} - E_{Follower}$, could reveal the style of leadership.

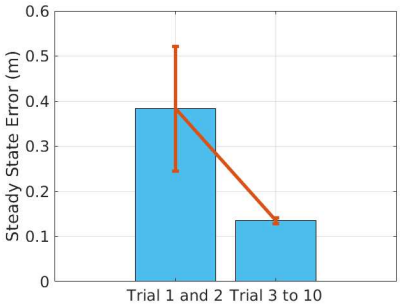
By examining the correlation between the effort differences and the goal tracking errors for all the participants, we found a significant negative correlation between them with $p \leq 0.003$ ($p = 0.0029$), which is shown in Fig. 4c. This observation strongly indicates that the performance of the dyad highly depends on the effort of the leader. If the leader took more responsibility for completing the game, which was reflected by a larger effort difference between the leader and the follower, a better result could be achieved by the dyad. This observation also agrees with the results reported in Messeri et al. [22].

Summary Figure 4c shows an inverse relationship between the tracking error and the difference in effort between the agents.

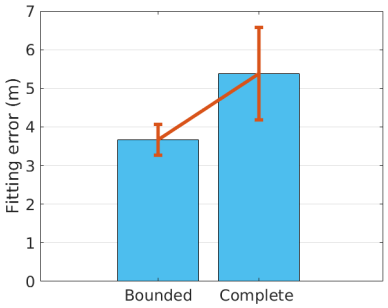
D. Discussion: from bounded rationality to pHRI

The results of this study reveal the stochastic nature of human cooperative behavior. As the ultimate objective is to help the robot comprehend the human partner's intent better, we believe the proposed model has the following contributions for the robot to befittingly collaborate with human partners.

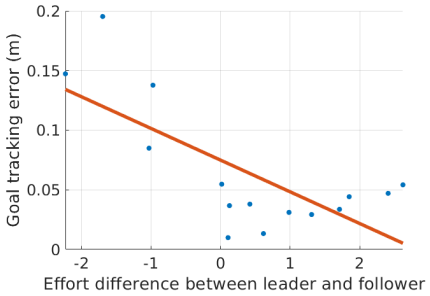
From the observation and verification, we have shown that the human participants' behavior follows bounded rationality under cooperative tasks. Unlike some prior research where the human manipulation was modeled as an optimization process [23], [11], the human partners show suboptimal behavior when they acknowledged that their partners were also human participants. When we design the cooperative algorithm for the robot in pHRI scenarios, this study suggests that it is critical for the robot to determine its role during the collaboration. As a leader, the robot should be more



(a) Comparison of steady-state errors of the first two trials and the remaining trails for all participants.



(b) Comparison of the fitting errors of complete rationality and bounded rationality for all pairs of participants.



(c) The goal tracking errors compared with the effort difference between leader and follower.

Fig. 4: Summary results for all the participants

active during the interaction, because stronger leadership could result in better performance.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we explored novelty methods to model human-human cooperative behaviors in joint actions. We designed a virtual joint object translation game, which required collaboration between human participants. A bounded rationality game-theoretical model was introduced to model human behavior during the game. By comparing results with bounded rational and complete rational assumptions of human actions, we found that the bounded rationality model fits the experimental data better, which suggested that the robots should be aware of suboptimal human behaviors while working with human partners. Another test demonstrated that if the leader paid less effort during the game, the performance of the pair could be worse. This observation highlighted the importance of leadership during collaboration.

We plan to extend our work to investigate how bounded rationality affects the pHRI scenarios with different goal positions. We plan to test how the bounded rationality game-theoretical human model could help the robot improve its performance while working with human partners.

REFERENCES

[1] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.

[2] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. A. Eslami, and M. Botvinick, "Machine theory of mind," in *International conference on machine learning*. PMLR, 2018, pp. 4218–4227.

[3] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, and C. F. Camerer, "Neural systems responding to degrees of uncertainty in human decision-making," *Science*, vol. 310, no. 5754, pp. 1680–1683, 2005.

[4] S. Ikemoto, H. B. Amor, T. Minato, B. Jung, and H. Ishiguro, "Physical human-robot interaction: Mutual learning and adaptation," *IEEE robotics & automation magazine*, vol. 19, no. 4, pp. 24–35, 2012.

[5] H. El-Hussieny and J.-H. Ryu, "Inverse discounted-based lqr algorithm for learning human movement behaviors," *Applied Intelligence*, vol. 49, no. 4, pp. 1489–1501, 2019.

[6] A. Y. Ng, S. J. Russell *et al.*, "Algorithms for inverse reinforcement learning." in *Icml*, vol. 1, 2000, p. 2.

[7] D. Hadfield-Menell, S. J. Russell, P. Abbeel, and A. Dragan, "Cooperative inverse reinforcement learning," *Advances in neural information processing systems*, vol. 29, pp. 3909–3917, 2016.

[8] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, 2017, pp. 323–331.

[9] C. Liu, W. Zhang, and M. Tomizuka, "Who to blame? learning and control strategies with information asymmetry," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 4859–4864.

[10] Y. Wang, Y. Ren, S. Elliott, and W. Zhang, "Enabling courteous vehicle interactions through game-based and dynamics-aware intent inference," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 217–228, 2019.

[11] Y. Li, G. Carboni, F. Gonzalez, D. Campolo, and E. Burdet, "Differential game theory for versatile physical human–robot interaction," *Nature Machine Intelligence*, vol. 1, no. 1, pp. 36–43, 2019.

[12] Y. Wang, G. J. Lematta, C.-P. Hsiung, K. A. Rahm, E. K. Chiou, and W. Zhang, "Quantitative modeling and analysis of reliance in physical human–machine coordination," *Journal of Mechanisms and Robotics*, vol. 11, no. 6, 2019.

[13] S. Schach, S. Gottwald, and D. A. Braun, "Quantifying motor task performance by bounded rational decision theory," *Frontiers in neuroscience*, vol. 12, p. 932, 2018.

[14] 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 avoidance1," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 250–265, 2020.

[15] M. Kwon, E. Biyik, A. Talati, K. Bhasin, D. P. Losey, and D. Sadigh, "When humans aren't optimal: Robots that collaborate with risk-aware humans," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 43–52.

[16] S. O. Oguz, A. Kucukyilmaz, T. M. Sezgin, and C. Basdogan, "Haptic negotiation and role exchange for collaboration in virtual environments," in *2010 IEEE haptics symposium*, 2010, pp. 371–378.

[17] C. E. Madan, A. Kucukyilmaz, T. M. Sezgin, and C. Basdogan, "Recognition of haptic interaction patterns in dyadic joint object manipulation," *IEEE transactions on haptics*, vol. 8, no. 1, pp. 54–66, 2014.

[18] T. Basar and G. J. Olsder, *Dynamic noncooperative game theory*. Siam, 1999, vol. 23.

[19] P. A. Ortega and A. A. Stocker, "Human decision-making under limited time," *arXiv preprint arXiv:1610.01698*, 2016.

[20] H. El-Hussieny, A. Abouelsoud, S. F. Assal, and S. M. Megahed, "Adaptive learning of human motor behaviors: An evolving inverse optimal control approach," *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 115–124, 2016.

[21] V. T. Chackochan and V. Sanguineti, "Incomplete information about the partner affects the development of collaborative strategies in joint action," *PLoS computational biology*, vol. 15, no. 12, p. e1006385, 2019.

[22] C. Messeri, A. M. Zanchettin, P. Rocco, E. Gianotti, A. Chirico, S. Magoni, and A. Gaggioli, "On the effects of leader-follower roles in dyadic human-robot synchronisation," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[23] M. Menner, P. Worsnop, and M. N. Zeilinger, "Constrained inverse optimal control with application to a human manipulation task," *IEEE Transactions on Control Systems Technology*, 2019.