A Theoretical Framework for Large-Scale Human-Robot Interaction with Groups of Learning Agents

Nicholas Teh
Dept of Computer Science,
National University of Singapore
nicholas.teh@u.nus.edu

Shuyue Hu
Dept of Computer Science,
National University of Singapore
husy@comp.nus.edu.sg

Harold Soh
Dept of Computer Science,
National University of Singapore
harold@comp.nus.edu.sg

ABSTRACT

Recent advances in robot capabilities have led to a growing consensus that robots will eventually be deployed at scale across numerous application domains. An important open question is how humans and robots will adapt to one another over time. In this paper, we introduce the model-based **T**heoretical **Hu**man-Robot **S**cenarios (**THuS**) framework, capable of elucidating the interactions between large groups of humans and learning robots. We formally establish THuS, and consider its application to a human-robot variant of the *n*-player coordination game, demonstrating the power of the theoretical framework as a tool to qualitatively understand and quantitatively compare HRI scenarios that involve different agent types. We also discuss the framework's limitations and potential. Our work provides the HRI community with a versatile tool that permits first-cut insights into large-scale HRI scenarios that are too costly or challenging to carry out in simulations or in the real-world.

CCS CONCEPTS

• **Theory of computation** → **Convergence and learning in games**; • **Computing methodologies** → **Cooperation and coordination**; • **Computer systems organization** → *Robotics.*

KEYWORDS

human-robot interaction, game theory, multi-agent learning

ACM Reference Format:
Nicholas Teh, Shuyue Hu, and Harold Soh. 2021. A Theoretical Framework for Large-Scale Human-Robot Interaction with Groups of Learning Agents. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’21 Companion), March 8–11, 2021, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3434074.3447220

1 INTRODUCTION

Recent years have witnessed a significant gain in robot capabilities, driven in-part by progress in artificial intelligence and machine learning. These advances have spurred usage in real-world scenarios, from autonomous vehicles for transportation [2] to robots that provide eldercare [1] and aid in the front-lines against the COVID pandemic [7]. These use-cases have led to a growing consensus that it is only a matter of time before robots are deployed at-scale on our roads, and in our homes and workplaces.

However, it remains an open question how large groups of humans and robots will *co-adapt* as they begin to interact with one another. What behaviors will *emerge*? What would be the *dynamics* of their interactions? What *outcomes* are likely to occur? These questions are critical as robots take on a greater role in our society: the answers have implications for a range of issues, from the design of robots — e.g., robot-learning methods we employ — to roboethical policies and human-robot trust [6, 19]. Unfortunately, these questions are also notoriously difficult to answer; large-scale experiments are challenging to carry out due to logistical and technological barriers. Agent-based simulation forms an possible alternative, but can be costly to develop and is computationally-expensive when considering large numbers of agents and scenarios.

In this paper, we adopt a theory-based approach, and take the first steps towards a model-based framework for elucidating the interactions between large groups of humans and robots. Our key contribution is the **T**heoretical **Hu**man-Robot **S**cenarios (**THuS**) framework, which extends recent results [3] in multi-agent systems and game theory to different *groups* of human and robots. THuS enables us qualitatively understand and quantitatively compare HRI scenarios that involve different *agent types*. For example, this extended abstract describes a case study where we consider how Q-learning robots and noisily rational humans can learn to coordinate. Compared to agent-based simulations, THuS can provide analytical forms that are amenable to theoretical analysis and has lower computational cost when dealing with very large populations, e.g., in the tens of thousands to millions of agents. As such, THuS can provide first-cut insights into large-scale HRI scenarios that are challenging to carry out in simulations or in the real-world.

2 BACKGROUND AND RELATED WORK

THuS builds upon game theory, which has been recently used to study HRI scenarios. For example, Li et al. [10] proposed a game-theoretic framework to study human-robot physical interaction, and Paeng et al. [13] explored trust in HRI through the coin entrustment game. These studies, together with other prior work that adopt a game-theoretic approach to HRI [9, 12], have largely focused on dyadic interactions. In contrast, we consider large numbers of humans and robot agents.

We leverage on the concept of population games [15], a compelling model for studying strategic decision-making in large populations. Prior work in robotics have used similar concepts (mean-field games [4, 8]), but mainly for controlling a large group of robots, e.g., a flock of UAVs in windy conditions [16], or mobile robots for evacuating humans [11]. These works make a common assumption

HRI ’21 Companion, March 8–11, 2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8290-8/21/03…$15.00
https://doi.org/10.1145/3434074.3447220

in the theory — that the population comprises of homogeneous optimal controllers. Here, we contribute a theoretical framework that can account for the trial-and-error learning process of humans and robots. Notably, THuS does not focus on any particular HRI setting, but rather, provides a means for us to examine systemic aspects of large-scale HRI scenarios.

3 THEORETICAL HUMAN-ROBOT SCENARIOS (THUS) FRAMEWORK

Modelling a large number of strategically interacting agents is typically nontrivial; the number of coupled equations required to be solved for individual decision-making poses a huge computational barrier [3] — it grows exponentially with the number of agents. This is particularly so if one seeks to study emergent population behaviour, the dynamics of agents' interaction, or the eventual equilibrium of the system (if any). Recently, Hu et al. [3] proposed a generalized theoretical framework for understanding the dynamics of independent learners in population games. Here, we extend their framework to the multiple-groups setting, and showcase how the theory can shed light on HRI scenarios. To the best of our knowledge, this is the first time a theoretical framework for large-scale HRI scenarios is presented.

Population Games with Groups. In this paper, we consider humans and robots engaging in a *population game with groups* [15]. Population games provide a unified way to model strategic interactions among a great number of "small" agents — each agent is under the aggregated effect of all the other agents, and is not affected by the behaviour of any individual agent. Within each group, agents are anonymous, in that exchanging their labels (or identities) does not lead to any difference in setting. More formally,

Definition 1. A population game with l groups is defined by a tuple $\langle \mathcal{G}, \mathcal{A}, (u_k) \rangle$. $\mathcal{G} = \{G_1, \ldots, G_l\}$ is a set of l disjoint sets of agents, and the total number of agents in the population is given by $n = \sum_{i=1}^{l} |G_i|$. $\mathcal{A} = \{a_1, \ldots, a_m\}$ is a set of m available actions for each agent. For each group $G_k \in \mathcal{G}$, let $\mathbf{o}_k = [o_k^1, \ldots, o_k^m]^\top$ denote its action profile, where o_k^j is the proportion of agents choosing action $a_j \in \mathcal{A}$ in the group. The payoff function $u_k(a, \mathbf{o})$ determines the payoff to any agent in the group G_k if the agent takes action $a \in \mathcal{A}$ given the population action profile $\mathbf{o} = [\mathbf{o}_1, \ldots, \mathbf{o}_l]$.

Independent Learners. In our setting, agents can *learn* through repeated participation in a population game over a time horizon T. Specifically, we consider *independent learners*. Every agent keeps track of a set of *critical parameters* that are used to inform its decision-making strategy. An agent can update its critical parameters based on the reward(s) it receives, or observations it makes.

Consider a group G_k of agents. For each agent $i \in G_k$, let $\mathbf{x}_{k,i}(t) = [x_{k,i}^1(t), \ldots, x_{k,i}^d(t)]^\top$ be its critical parameters, which changes over time t while the agent repeatedly applies some predefined update rule throughout their gameplay. At each time step t, each agent i takes an action denoted by $a_{k,i}(t)$, receives an immediate reward denoted by $r_{k,i}(t)$, and has access to an *information set*[1] denoted by $I_k(t)$ that is available to every agent in the group.

[1]Because agents are anonymous, $I_k(t)$ does not contain any information regarding the identities of agents. The information set is an empty set if agents have no additional information other than their own received reward.

Let $p_k(\mathbf{x}_k, t)$ denote the probability density function of the critical parameters in the group G_k. If each agent i makes use of $r_{k,i}(t)$ and $I_k(t)$ to update its critical parameters $\mathbf{x}_{k,i}(t)$, we can derive the following result based on Theorem 1 in [3]:

Theorem 2. *Consider a population game $\langle \mathcal{G}, \mathcal{A}, (u_k) \rangle$, where each group $G_k \in \mathcal{G}$ of agents adopt the same learning method. If there exists a function $\mathbf{f}_k = [f_k^1, \ldots, f_k^d]^\top$ and a function h_k for each group $G_k \in \mathcal{G}$, such that both equations*

$$\frac{d\mathbf{x}_{k,i}(t)}{dt} = \mathbf{f}_k(\mathbf{x}_{k,i}(t), r_{k,i}(t), I_{k,i}(t), t), \text{ and} \tag{1}$$

$$r_{k,i}(t) = h_k(\mathbf{x}_{k,i}(t), t), \tag{2}$$

hold for each agent $i \in G_k$, and f_k^j is differentiable at any $x_{k,i}^j(t)$ for all $j \in \{1, \ldots, d\}$, then the learning dynamics of the entire population is described by the following system of partial differential equations:

$$\frac{\partial p_k(\mathbf{x}_k, t)}{\partial t} = -\sum_{j=1}^{d} \frac{\partial}{\partial x_k^j} [p_k(\mathbf{x}_k, t) f_k^j(\mathbf{x}_k, h_k(\mathbf{x}_k(t), t), I_k(t), t)], \tag{3}$$

for $k = 1, \ldots, l$.

Intuitively, agents in each group use a common specified rule to update their critical parameters. If we are able to characterize each of these update rules with an ordinary differential equation (ODE) as in Eqn. (1), then, under the appropriate conditions, the probability of agents having each of the critical parameters can be described by the partial differential equation (PDE) in Eqn. (3).

In large populations, this framework allow us to characterise the dynamics of the population through a constant number of PDEs (depending on a fixed number of groups), rather than having an exponentially large number of coupled equations depending on the number of agents [3]. A constant number of PDEs is mathematically tractable. Thus, it becomes feasible to predict and to analyze the dynamics of large-scale HRI scenarios with the use of existing numerical tools for solving PDEs.

4 CASE STUDY: COORDINATION GAMES

In many real-world settings, agents have to organically coordinate to find and adhere to social conventions. For example, robots and humans have to coordinate when moving in a physical space (e.g., when driving on roads or navigating in a crowded airspace) to prevent collisions. Coordination games are a common class of games that model the aforementioned scenarios. For expositional simplicity, we consider a set $\mathcal{A} = \{A, B\}$ of two actions (A and B) available to each agent, and assume that agents of different groups have the same payoff function. Let n denote the total number of agents in the population. We define the common payoff function as

$$u(A, \mathbf{o}) = \frac{c_1}{n} \sum_{k=1}^{l} |G_k| o_k^A \quad \text{and} \quad u(B, \mathbf{o}) = c_2 + \frac{c_3}{n} \sum_{k=1}^{l} |G_k| o_k^B \tag{4}$$

where o_k^A and o_k^B are the proportions of agents taking the actions A and B, respectively, in the group k, and c_1, c_2 and c_3 are pre-defined constants. There are two pure strategy Nash equilibria in this game, i.e., all the agents take action A if $u(A, \mathbf{o}) > u(B, \mathbf{o})$ or all the agents take action B if $u(B, \mathbf{o}) > u(A, \mathbf{o})$.

We consider how a group of Bayes-rational human agents [5, 14, 17] co-learns with a group of Q-learning [18] robots during repeated plays of the above population coordination game. We selected these human and robot models for simplicity; more sophisticated models can be substituted in and studied in a similar manner. Due to space constraints, we focus on describing key high-level results and relegate details to a future full version of this paper.

4.1 Human Agent Learning & Decision-Making

Consider a group G_k of human agents. In our model, each human agent maintains a belief over the probability of other agents choosing action A or B, and this belief informs its own decision-making. For each human $i \in G_k$, at time t, this belief can be represented by the distribution over the average probability $\theta^j_{k,i}(t) \sim \text{Beta}(\alpha^j_{k,i}(t), \beta^j_{k,i}(t))$ that an agent in group G_j takes action A.

During gameplay, each human $i \in G_k$ iteratively updates their beliefs by updating the parameters $\alpha^j_{k,i}(t)$ and $\beta^j_{k,i}(t)$ for each group $G_j \in \mathcal{G}$ using Bayes rule. Since it may not be feasible for a human agent to observe the actions of *all* other agents in large populations, we assume they *sample* observations; at each time step t, a human agent observes the actions taken by z randomly sampled agents from every group $G_j \in \mathcal{G}$ (including its own). We define

$$\phi^j_{k,i}(t) = \frac{\alpha^j_{k,i}(t)}{\alpha^j_{k,i}(t)+\beta^j_{k,i}(t)}. \tag{5}$$

We assume the initial parameters $\alpha_{k,i}(0)$ and $\beta_{k,i}(0)$ sum to a constant c_4. The critical parameters are $\boldsymbol{\phi}_{k,i}(t) = [\phi^1_{k,i}(t), \dots \phi^l_{k,i}(t)]^\top$ for every human $i \in G_k$, and each $\phi^j_{k,i}(t)$ is given by

$$\phi^j_{k,i}(t) = \phi^j_{k,i}(t-1) \times \frac{c_4 + zl(t-1)}{c_4 + zlt} + \frac{\sum_{j=1}^{l} \lambda^j_{k,i}(t-1)}{c_4 + zlt}, \tag{6}$$

where $\lambda^j_{k,i}(t-1)$ is the number of agents sampled from group G_j at time $t-1$ that are observed to take action A. The constants z and l denote the sample size and the number of groups, respectively.

We assume that human agents are noisily rational, and know the relative group size and the payoff function of the game. Then, for each human $i \in G_k$, the estimated value of taking each action $a \in \mathcal{A}$ at time t is

$$v_{k,i}(a,t) = \begin{cases} \frac{c_1}{n} \sum_{j=1}^{l} |G_j| \phi^j_{k,i}(t), & \text{if } a = A \\ c_2 + \frac{c_3}{n} \sum_{j=1}^{l} |G_j| [1 - \phi^j_{k,i}(t)], & \text{if } a = B \end{cases} \tag{7}$$

Based on the value function, each human i selects each action $a \in \mathcal{A}$ with probability

$$\xi_{k,i}(a,t) = \frac{e^{\tau_k v_{k,i}(a,t)}}{\sum_{a' \in \mathcal{A}} e^{\tau_k v_{k,i}(a',t)}}, \tag{8}$$

where τ_k is the Boltzmann exploration constant representing the rationality of any agent in group G_k; a higher τ_k indicates that the agent will take the higher-valued action with higher probability (i.e. act more rationally).

4.2 Robot Learning & Decision-Making

The robots in our case-study are Q-learning agents. Consider a group G_g of Q-learners. Note that the state dependency of Q-values can be dropped as there is no explicit state transition in population games. For each Q-learner $i \in G_g$, we denote the vector of Q-values by $\mathbf{Q}_{g,i}(t)$. As in [3], we consider the vector of Q-values $\mathbf{Q}_{g,i}(t)$ to be the critical parameters. The Q-value of each action is iteratively revised during gameplay as follows:

$$Q_{g,i}(a,t) = \begin{cases} (1-\eta) Q_{g,i}(a, t-1) + \eta r_{g,i}(t-1), & \text{if } a = a_{g,i}(t-1) \\ 0, & \text{if } a \neq a_{g,i}(t-1) \end{cases} \tag{9}$$

where η is the learning rate, $a_{g,i}(t-1)$ denotes the action taken by the Q-learner at time $t-1$, and the received reward $r_{g,i}(t-1)$ is obtained through the payoff function of the game. Similar to the human agents, a Q-learner selects an action probabilistically using Eqn. (8).

4.3 Population Dynamics

Given the update rules in subsections 4.1 and 4.2, the corresponding ODEs representing individual agents' updating of critical parameters are as follows. For each human i in a group $G_k \in \mathcal{G}$,

$$\frac{d\phi^j_{k,i}(t)}{dt} = \frac{-zl\phi^j_{k,i}(t) + \sum_{j=1}^{l} \lambda^j_{k,i}(t)}{c_4 + zlt} \tag{10}$$

for $j = 1, \dots, l$. For each Q-learner i in a robot group $G_g \in \mathcal{G}$, we denote the Q-value of each action $a \in \mathcal{A}$ by $Q^a_{g,i}(t)$ (with a slight abuse of notation). Then, we have

$$\frac{dQ^a_{g,i}(t)}{dt} = \eta \frac{e^{\tau_k Q^a_{g,i}(t)}}{\sum_{a' \in \mathcal{A}} e^{\tau_k Q^{a'}_{g,i}(t)}} [r_{g,i}(t) - Q^a_{g,i}(t)] \tag{11}$$

for $a \in \{A, B\}$. By Theorem 2, we have the following PDEs describing the evolution of the probability densities of each group over their critical parameters

$$\frac{\partial p_k(\boldsymbol{\phi}_k, t)}{\partial t} = -\sum_{\ell=1}^{n} \frac{\partial}{\partial \phi^j_k(t)} \left(\frac{d\phi^j_k(t)}{dt} p_k(\boldsymbol{\phi}_k, t) \right), \tag{12}$$

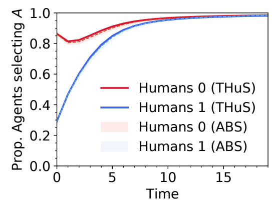
$$\frac{\partial p_g(\mathbf{Q}_g, t)}{\partial t} = -\sum_{a \in \{A,B\}} \frac{\partial}{\partial Q^a_g(t)} \left(\frac{dQ^a_g(t)}{dt} p_g(\mathbf{Q}_g, t) \right), \tag{13}$$

where (12) describes the human group, and (13) describes the robot group. These two equations are coupled through the observations humans make, as well as the immediate rewards received by robots. More specifically, at time t, the immediate rewards of robots are determined by the expected numbers of humans and robots taking action A. On the other hand, for each human i in the group G_k, at time t, the observation $\lambda^g_{k,i}(t)$ is approximately the expected number of robots in the group G_g taking action A.[2]

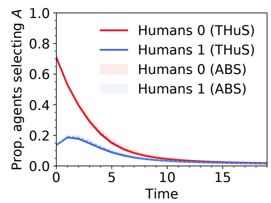
4.4 Results and Analysis

We instantiate $n = 2$ groups. The constants of the payoff function are $c_1 = 1$, $c_2 = 0.2$ and $c_3 = 0.6$, such that $u(A, \mathbf{o}) > u(B, \mathbf{o})$ if and only if the proportion of agents taking action A in the population is greater than 50%, i.e., the critical mass of action A is 50%. We use the finite difference method to numerically solve the PDEs presented in section 4.3. The results are presented along with corresponding agent-based simulations (20 runs, each with group size 5,000).

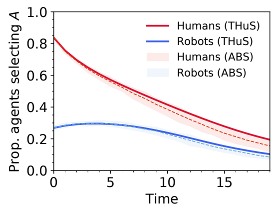
[2] Such an approximation will become more accurate as the sample size tends to the group size, i.e., the human observes all the robots in the group G_g.



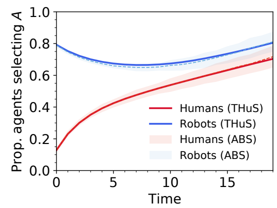
(a) Human-Human: critical mass of action A is reached initially



(b) Human-Human: critical mass of action A is *not* reached initially



(c) Human-Robot: critical mass of action A is reached initially



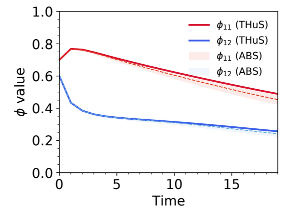
(d) Human-Robot: critical mass of action A is *not* reached initially

Figure 1: Population coordination game with two groups. THuS produces dynamics that closely track the agent-based simulations (ABS), and reveal that Human-Robot population dynamics differ from the Human-Human setting.

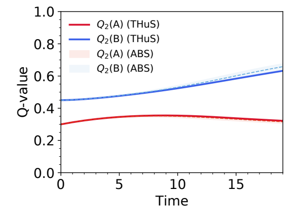
First, let us consider two human groups. Recall that in a population coordination game, there are two pure strategy Nash equilibria: all of the agents take action A, or all of them take action B. Fig. 1 shows that which action the human-human population eventually converges to take is largely influenced by initial conditions; all of the humans take action A if the initial proportion of agents taking action A exceeds 0.5 (Fig. 1(a)), but all of the humans take action B if that initial proportion is below 0.5 (Fig. 1(b)).

Now, let us change one of the groups to Q-learning robots. As shown in Figs. 1(c) and 1(d), it is surprising to note that given similar initial conditions, the human-robot population develops a trend that is *completely different* than the human-human population. Even though the initial proportion of agents taking action A in the population exceeds the critical mass 50%, the human-robot population tended towards action B (as shown in Fig. 1(c)). Similar phenomena can be observed in Fig. 1(d), indicating that the humans are adapting their behaviours to the robots. While such phenomena are observed in empirical HRI experiments, THuS sheds light on the underlying mechanism from a theoretical perspective. As shown in Fig. 2, the Bayesian humans quickly changed their beliefs about robot actions, which led them to follow the robot group's majority action of choosing action B. In contrast, robots were slower to adapt their Q-values. In this specific scenario, the learning rate of robot agents affected the dynamics of the *entire system*; our subsequent experiments showed increasing the robot learning rate (from 0.2 to 0.5) led to similar behavior as in the human-human setting.

Table 1 compares the computation time taken by THuS, compared to the agent-based simulation ($T = 20$ with timings averaged over 20 runs). In practice, repeating the simulation multiple times is necessary due to stochasticity in the decision-making process. The computational cost of THuS is unaffected by the population or observation sample sizes, and only needs to be run *once*.



(a) Beliefs of humans



(b) Q-values of robots

Figure 2: Evolution of critical parameter values.

Table 1: A comparison of ABS and THuS run times with varying population and sample size.

Population Size			Computational Cost (seconds)	
Humans	Robots	Sample Size	ABS (per run)	THuS
1,000	1,000	30	4.71	5.43
2,500	2,500	30	11.20	5.50
5,000	5,000	30	23.93	5.46
10,000	10,000	50	57.70	5.46

5 DISCUSSION AND FUTURE WORK

Our case study demonstrates that THuS is an effective tool for understanding the dynamics of agent populations with multiple groups. In particular, it is nontrivial to intuit our findings about the effect of the robot learning rate from the outset.

To use THuS appropriately, it is important to understand both its strengths and limitations. Notably, THuS is *not* a replacement for real-world studies or detailed simulation; it is theoretical tool designed to provide insights into a human-robot population. Like other theoretical methods, THuS abstracts away details to focus on specific core aspects of interactions; this abstraction enables us to gain insights into learning and policy dynamics, but comes at a cost. THuS treats humans and robots as anonymous agents in that they are indistinguishable apart from their critical parameters. This assumption is a simplification but enables tractability and is suitable when dealing with large population masses.

We see THuS as a *first* step towards filling a gap: there is a dearth of mathematical tools available for the analysis of large-scale HRI. THuS can be particularly beneficial when used before embarking upon more costly studies, and to examine scenarios that are impossible to conduct in the real-world (e.g., due to ethical concerns or logistical limitations). We look forward to more sophisticated tools and specialized variants that can be built using THuS as a foundation.

ACKNOWLEDGEMENTS

This research is supported by the National Research Foundation Singapore under its AI Singapore Programme (Award Number: AISG-RP-2019-011).

REFERENCES

[1] David Fischinger, Peter Einramhof, Konstantinos Papoutsakis, Walter Wohlkinger, Peter Mayer, Paul Panek, Stefan Hofmann, Tobias Koertner, Astrid Weiss, Antonis Argyros, and Markus Vincze. 2016. Hobbit, A Care Robot Supporting Independent Living At Home: First Prototype and Lessons Learned. *Robotics and Autonomous Systems* 75 (2016), 60–78.

[2] Nan Gu, Zhouhua Peng, Dan Wang, Yang Shi, and Tianlin Wang. 2019. Antidisturbance Coordinated Path Following Control of Robotic Autonomous Surface Vehicles: Theory and Experiment. *IEEE/ASME Transactions on Mechatronics* 24, 5 (2019), 2386–2396.

[3] Shuyue Hu, Chin-Wing Leung, Ho-fung Leung, and Harold Soh. 2020. The Evolutionary Dynamics of Independent Learning Agents in Population Games. *arXiv preprint arXiv:2006.16068* (2020).

[4] Minyi Huang, Roland P Malhamé, Peter E Caines, et al. 2006. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems* 6, 3 (2006), 221–252.

[5] Koosha Khalvati, Seongmin Park, Saghar Mirbagheri, Remi Philippe, Mariateresa Sestito, Jean-Claude Dreher, and Rajesh P. N. Rao. 2019. Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science Advances* 5, 11 (2019).

[6] Bing Cai Kok and Harold Soh. 2020. Trust in Robots: Challenges and Opportunities. *Current Robotics Reports* 1, 4 (2020), 297–309. https://doi.org/10.1007/s43154-020-00029-y

[7] Samuel Lalmuanawma, Jamal Hussain, and Lalrinfela Chhakchhuak. 2020. Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review. *Chaos, Solitons & Fractals* 139 (2020), 110059.

[8] Jean-Michel Lasry and Pierre-Louis Lions. 2007. Mean field games. *Japanese Journal of Mathematics* 2, 1 (2007), 229–260.

[9] Kang Woo Lee and Jeong-Hoon Hwang. 2008. Human–robot interaction as a cooperative game. In *Trends in Intelligent Systems and Computer Engineering*. Springer, 91–103.

[10] Yanan Li, Keng Peng Tee, Rui Yan, Wei Liang Chan, and Yan Wu. 2016. A Framework of Human–Robot Coordination Based on Game Theory and Policy Iteration. *IEEE Transactions on Robotics* 32, 6 (2016), 1408–1418.

[11] Zhiyu Liu, Bo Wu, and Hai Lin. 2018. Coordinated Robot-Assisted Human Crowd Evacuation. In *Proceedings of the 57th IEEE Conference on Decision and Control (CDC)*. 4481–4486.

[12] Stefanos Nikolaidis, Swaprava Nath, Ariel Procaccia, and Siddhartha Srinivasa. 2017. Game-Theoretic Modeling of Human Adaptation in Human-Robot Collaboration. In *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 323–331.

[13] Erin Paeng, Jane Wu, and James Boerkoel. 2016. Human-Robot Trust and Cooperation through a Game Theoretic Framework. In *Proceedings of the 13th AAAI Conference on Artificial Intelligence (AAAI)*. 4246–4247.

[14] Michael Rescorla. 2021. Bayesian modeling of the mind: From norms to neurons. *WIREs Cognitive Science* 12, 1 (2021), e1540.

[15] William Sandholm. 2015. Chapter 13 - Population Games and Deterministic Evolutionary Dynamics. Handbook of Game Theory with Economic Applications, Vol. 4. 703 – 778.

[16] Hamid Shiri, Jihong Park, and Mehdi Bennis. 2019. Massive Autonomous UAV Path Planning: A Neural Network Based Mean-Field Game Theoretic Approach. *arXiv preprint arXiv:1905.04152* (2019).

[17] Harold Soh, Yaqi Xie, Min Chen, and David Hsu. 2020. Multi-task trust transfer for human–robot interaction. *The International Journal of Robotics Research* 39, 2-3 (2020), 233–249.

[18] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction.*

[19] Yaqi Xie, Indu P Bodala, Desmond C Ong, David Hsu, and Harold Soh. 2019. Robot Capability and Intention in Trust-Based Decisions Across Tasks. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 39–47.