When Shall I Be Empathetic? The Utility of Empathetic Parameter Estimation in Multi-Agent Interactions

Yi Chen[1], Lei Zhang[1], Tanner Merry[1], Sunny Amatya[2], Wenlong Zhang[2], and Yi Ren[1]

***Abstract*— Human-robot interactions (HRI) can be modeled as differential games with incomplete information, where each agent holds private reward parameters. Due to the open challenge in finding perfect Bayesian equilibria of such games, existing studies often decouple the belief and physical dynamics by iterating between belief update and motion planning. Importantly, the robot's reward parameters are often assumed to be known to the humans, in order to simplify the computation. We show in this paper that under this simplification, the robot performs non-empathetic belief update about the humans' parameters, which causes high safety risks in uncontrolled intersection scenarios. In contrast, we propose a model for empathetic belief update, where the agent updates the joint probabilities of all agents' parameter combinations. The update uses a neural network that approximates the Nash equilibrial action-values of agents. We compare empathetic and non-empathetic belief update methods on a two-vehicle uncontrolled intersection case with short reaction time. Results show that when both agents are unknowingly aggressive (or non-aggressive), empathy is necessary for avoiding collisions when agents have false believes about each others' parameters. This paper demonstrates the importance of acknowledging the incomplete-information nature of HRI.**

I. Introduction

Human-robot interactions (HRI) have become ubiquitous in the past two decades, with applications in daily assistance, healthcare, manufacturing, transportation and defense. Since humans and robots may not understand the intents of each other during interactions, we consider HRI as differential general-sum games with incomplete information, where agents hold private reward parameters. Finding perfect Bayesian equilibria (PBE) of such games is an open challenge [1] due to the entanglement of physical and belief dynamics, and existing solutions (e.g., structured PBEs) do not scale well with the dimensionalities of the state, action, or belief spaces [2], [3]. As a result, most existing HRI studies resort to simplified optimal control formulations or complete-information games [4]–[7] for motion planning, and some use belief update to adapt the planned motion [8]–[11]. While this approach does not necessarily produce PBEs (due to the ignorance of belief dynamics during the planning), it is a tractable attempt at modeling the coupled dynamics of beliefs and physical states, and is therefore the focus of this paper.

This work was supported by the National Science Foundation under Grant CMMI-1925403.

[1]Y. Chen, L. Zhang, T. Merry, and Y. Ren are with Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ, 85287, USA. Email: ychen837@asu.edu; tmerry@asu.edu; lzhan300@asu.edu; yiren@asu.edu

[3]S. Amatya and W. Zhang are with The Polytechnic School, Ira A. Fulton Schools of Engineering, Arizona State University, Mesa, AZ, 85212, USA. Email: samatya@asu.edu; wenlong.zhang@asu.edu

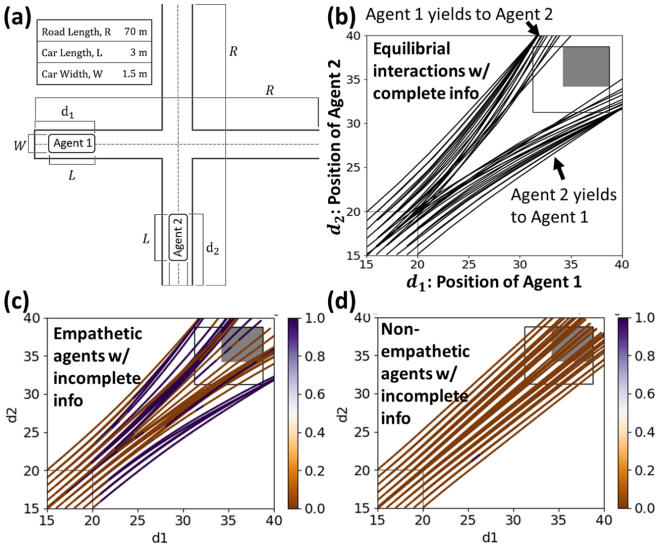

Fig. 1. (a) Schematics of a two-agent uncontrolled intersection case. d_1 and d_2 are positions of agents. (b) Nash equilibrial state trajectories when both agents know they are non-aggressive. Empathetic (c) and non-empathetic (d) non-aggressive agents when they have incorrect initial common beliefs that both are aggressive. Here aggressive means less sensitive to close distances. Empathetic agents are more likely to avoid collisions due to their better estimation of others' reward parameters and choice of the correct policies. Correct estimation is color coded as purple. Solid and hollow boxes at the top right corner represent "close-distance" (including collision) states from the perspectives of the aggressive and non-aggressive agents, respectively. Bottom left hollow box represents the initial states. Trajectories are mirrored along the diagonal line due to intrinsic symmetry.

We consider *empathy*, and the lack of it, as two different ways of performing belief update: Empathetic agents follow Harsanyi's assumption [12], i.e., they share the same common belief about everyone's parameters, and therefore acknowledge the fact that others have uncertainty about their own parameters. In contrast, a non-empathetic agent believes that everyone else knows its parameters, and therefore holds its own version of the "common belief". This difference in empathy affects the belief dynamics and therefore the state trajectories and values of all agents.

We note that there is a modeling convenience when considering agents as non-empathetic (or even non-game-theoretic), which justifies such simplifications adopted by existing HRI studies. To understand this, we first note that updating the belief about an agent relies on the Hamiltonian (action-value) of that agent. From the perspective of a non-empathetic ego agent, its fellow agents play complete-information games with it, and therefore each fellow's Hamiltonian is a function of the states and its own parameter. This is also the case when the fellow agent is non-game-theoretic. To enable belief update, the Hamiltonian is usually solved offline for

the corresponding complete-information game or the optimal control problem: For the latter, the closed-loop Hamiltonian can be found through reinforcement learning in general; for the former, where closed-loop solutions are hard to be solved, one could use open-loop Hamiltonian solved through Pontryagin's Maximum Principle (PMP) [13] as an approximation. Since solving such open-loop solutions for all states can be costly, approximation of the open-loop Hamiltonian using data-driven approaches is needed [14]. In comparison, the Hamiltonian of an empathetic agent, due to its incomplete information, is a function of all agents' beliefs, in addition to their states and the agent's own parameters. Due to the additional involvement of the belief space, the computational cost of creating data-driven approximations of open-loop Hamiltonian becomes even higher.

The convenience of modeling agents as non-empathetic or even non-game-theoretic leads to the central question of this paper:

In what interactions does empathy matter?

This paper makes the following contributions towards answering this question (summary in Fig. 1). (1) We define an interaction space spanned by the initial system states, agent parameters, the common belief about the parameters, and the empathy of agents to systematically evaluate the potential advantages of empathy. (2) Through a two-vehicle uncontrolled intersection case, we show that empathy in parameter estimation leads to significantly better values when reward parameters and common beliefs are completely wrong (e.g., everyone being aggressive while believing all to be nonaggressive). (3) To enable fast parameter estimation and motion planning, we develop a learning architecture for approximating Nash equilibrial action-values for given agent parameters. The approximation model is trained on equilibrial interactions solved from the boundary-value problems (BVPs) following PMP on a meshgrid of the state space.

The rest of the paper is structured as follows: Sec. II reviews related work. Sec. III elaborates on empathetic and non-empathetic belief update, motion planning, and data-driven Hamiltonian approximation. Sec. IV introduces the case study. We conclude the paper in Sec. V.

II. Related Work

Multi-agent perfect Bayesian equilibrium: A PBE consists of a policy and belief pair that simultaneously satisfies sequential rationality and belief consistency [15]. It is known that there does not exist a universal algorithm for computing PBE due to the interdependence of policies and beliefs [3]. This open challenge is partially addressed recently in [3], which shows that a subset of PBEs can be computed recursively by solving fixed-point equations for each agent . Since the fixed-point equations are interdependent on agents' policies, the algorithm is non-scalable with respect to the number of agents, time, or the dimensionalities of the action, state, and belief spaces. Only solutions for two-agent two-step games have been demonstrated so far [2], [3]. The inverse problem, i.e., estimation of agent parameters given PBE demonstrations, has not yet been studied.

Decision modeling: Human decision models in HRI [16]–[18] follow studies in behavioral economics [19]–[21]. Risk models are introduced to explain seemingly non-optimal human actions [6]. Social value orientation is introduced to explain agents' courtesy towards others in general-sum dynamic games [7]. Similar courtesy models have been discussed in [22], [23]. In this paper we model agents to take Nash equilibrial actions deterministically without considering courtesy. We only use noisy rationality to compensate for modeling errors during belief updates, similar to [11].

Multi-agent inverse reinforcement learning (MIRL): Parameter estimation (for reward and policy) has been investigated for single-agent problems [24]–[26]. For dynamic games, multi-agent inverse reinforcement learning performs estimation under solution concepts of the game rather than assuming optimality of individual actions [27], [28]. Along the same vien, the belief update algorithm introduced in this paper extends the single-agent framework in [11] to games, while allowing noisiness of rationality to be estimated along with the agents' reward parameters. Compared with [7], where agents' parameters are estimated using Stackelberg equilibrial as a solution concept, this paper considers agents to take simultaneous actions and are thus Nash equilibrial.

Value approximation: Solutions to Hamilton-Jacobi equations often have no analytical forms, can be discontinuous, and only exist in a viscosity form [29], [30]. Deep neural networks (DNN) have recently been shown to be effective at approximating solutions to Hamilton-Jacobi-Bellman (HJB) equations underlying optimal control problems [14] and Hamilton-Jacobi-Isaac (HJI) equations for two-player zero-sum games with complete information [31], thanks to the universal approximation capability of DNNs [32]. In this paper, we extend this approximation scheme to values of general-sum complete-information differential games, and then use the resultant value networks to approximate agents' Hamiltonian during belief update and motion planning. In comparison, [7] requires equilibria to be computed by iteratively solving the KKT problems during parameter estimation, while the proposed value approximation method allows agents to leverage memorized value gradients, thus accelerating the estimation.

III. Methods

This section introduces the belief update and motion planning algorithms to be used in the case study. We also elaborate on the approximation of Hamiltonian.

A. Belief update and motion planning

Preliminaries and notations: For generality, we consider a multi-agent game with N agents. All agents share the same individual action set $\mathcal{U}$, state space $\mathcal{X}$, reward parameter set Θ, and rationality set Λ. Together, they share an instantaneous reward function $f(\cdot,\cdot;\theta) : \mathcal{X}^N \times \mathcal{U}^N \to \mathbb{R}^N$, a terminal reward function $c(\cdot;\theta) : \mathcal{X}^N \to \mathbb{R}^N$, a dynamical model $h : \mathcal{X}^N \times \mathcal{U}^N \to \mathcal{X}^N$, and a finite time horizon $[0, T]$. Let $\beta_i :=< \lambda_i, \theta_i >$ be the parameters of agent i,

where $\lambda_i \in \Lambda$ and $\theta_i \in \Theta$. We denote the total parameter set by $\mathcal{B} := \Lambda^N \times \Theta^N$. Θ, Λ, $\mathcal{B}$, and $\mathcal{U}$ are considered discrete in this study. To reduce notational burden, we use a single variable $\mathbf{a}$ for the set $(a_1, ..., a_N)$ and define $a_{-i} = (a_1, ..., a_{i-1}, a_{i+1}, ...a_N)$. E.g., $\boldsymbol{\beta} \in \mathcal{B}$ contains parameters for all agents, $\boldsymbol{\beta}_{-i}$ those except for agent i. We denote by a^* the true value of variable a, and $\hat{a}$ its point estimate. Lastly, we assume the existence of a prior belief $p_i^0(\boldsymbol{\beta})$ of agent i, which will be updated as $p_i^k(\boldsymbol{\beta})$ at time step k with observations $\mathcal{D}(k) = \{(\mathbf{x}(t), \mathbf{u}(t))\}_{t=1}^k$. When agents are empathetic, they share the same common belief $p^k(\boldsymbol{\beta})$.

Nash equilibria for a complete-information game: If $\boldsymbol{\theta}$ is known to all and unique Nash equilibria exist, we can derive the equilibrial Hamiltonian $H_i(\cdot, \cdot; \boldsymbol{theta}) : \mathcal{X}^N \times \mathcal{U} \to \mathbb{R}$ for every agent i, which is the value of action u_i in state $\mathbf{x}$ when ego and fellow parameters are θ_i and θ_{-i}, respectively. For the discrete set of joint parameters, Θ^N, we can derive $\mathcal{H}^N := \{\mathbf{H}(\cdot, \cdot; \theta)\}_{\theta \in \Theta^N}$, which maps Θ^N to the equilibrial Hamiltonian. E.g., for a two-agent game where $|\Theta| = 2$, we have $|\mathcal{H}^2| = 4$ pairs of action-values.

Belief update: Given observations $\mathcal{D}(k)$ at time step k, $p_k^i(\boldsymbol{\beta})$ follows a Bayes update:

$$p_k^i(\boldsymbol{\beta}) = \frac{p(\mathbf{u}(k)|\mathbf{x}(k); \boldsymbol{\beta}) p_{k-1}^i(\boldsymbol{\beta})}{\sum_{\boldsymbol{\beta}' \in \mathcal{B}} p(\mathbf{u}(k)|\mathbf{x}(k); \boldsymbol{\beta}') p_{k-1}^i(\boldsymbol{\beta}')}, \tag{1}$$

where $p(\mathbf{u}|\mathbf{x}; \boldsymbol{\beta}) = \prod_{i=1,...,N} p(u_i|\mathbf{x}; \boldsymbol{\beta})$ since actions are modeled to be drawn independently by agents, and

$$p(u_i|\mathbf{x}; \boldsymbol{\beta}) = \frac{\exp(\lambda_i H_i(\mathbf{x}, u_i; \boldsymbol{\theta}))}{\sum_{\mathcal{U}} \exp(\lambda_i H_i(\mathbf{x}, u_i'; \boldsymbol{\theta}))}. \tag{2}$$

$\forall i = 1, ..., N$. It should be noted that if elements of $\boldsymbol{\beta} \in \mathcal{B}$ is mistakenly assigned zero probability, e.g., due to noisy observations, this mistake will not be fixed by future updates. To address this, we modify $p_k^i(\boldsymbol{\beta})$ as

$$p_k^i(\boldsymbol{\beta}) = (1 - \epsilon) p_k^i(\boldsymbol{\beta}) + \epsilon p_0^i(\boldsymbol{\beta}) \tag{3}$$

before its next Bayes update, where ϵ represents the learning rate. This allows all $\boldsymbol{\beta}$ combinations to have non-zero probabilities throughout the interaction, provided that the prior p_0^i is non-zero on $\mathcal{B}$.

Parameter estimation: Recall that the open-loop equilibrial value of incomplete-information games is defined on the joint space of the agent's belief and parameters, and all agents' states. To keep the approximation of Hamiltonian (which relies on the spatial gradient of values) manageable, we will use point estimates, rather than the distributions ($p_k^i(\boldsymbol{\beta})$), for computing the equilibrial values (see Sec. III-B). Specifically, we use

$$\hat{\boldsymbol{\beta}}(k) = \arg\max_{\boldsymbol{\beta} \in \mathcal{B}} p_k^i(\boldsymbol{\beta}) \tag{4}$$

for empathetic agent i, and

$$\tilde{\boldsymbol{\beta}}_{-i}(k) = \arg\max_{\boldsymbol{\beta}_{-i} \in \mathcal{B}_{-i}} p_k^i(\boldsymbol{\beta}_{-i}|\beta_i^*), \tag{5}$$

for a non-empathetic agent. Different from empathetic agents, non-empathetic agents may have estimates different from each other, due to the conditioning on their own parameters.

Motion planning: If empathetic agents take actions strictly following the common belief, the interactions will be solely determined by the prior $p_0(\boldsymbol{\beta})$ independent of the private parameters of agents. This is inconsistent with real-world interactions where agents express their own intents. Therefore, we model empathetic agent i to follow control policies parameterized by its own parameters and the estimates of others, i.e., $\hat{\boldsymbol{\theta}} = (\theta_i^*, \hat{\boldsymbol{\theta}}_{-i})$. Similarly, the non-empathetic agent uses $\hat{\boldsymbol{\theta}} = (\theta_i^*, \tilde{\boldsymbol{\theta}}_{-i})$. Each agent then takes actions deterministically following

$$u_i = \arg\max_{u_i \in \mathcal{U}} H_i(\mathbf{x}, u_i; \hat{\boldsymbol{\theta}}). \tag{6}$$

Simulated interactions: Alg. 1 summarizes the simulation of an interaction, which is parameterized by the initial states $\mathbf{x}_0$, the set of agent parameters $\boldsymbol{\beta}^*$, and the prior belief $\mathbf{p}_0(\boldsymbol{\beta})$. The simulation outputs the trajectories of states $\mathbf{x}(k)$, actions $\mathbf{u}(k)$, beliefs $\mathbf{p}_k(\boldsymbol{\beta})$, and values $\mathbf{v}(k)$ of all agents.

Algorithm 1: Multi-agent interaction

input : $\mathbf{x}_0$, $\boldsymbol{\beta}^*$, $\mathbf{p}_0(\boldsymbol{\beta})$
output: $\{(\mathbf{x}(k), \mathbf{u}(k), \mathbf{p}_k(\boldsymbol{\beta}), \mathbf{v}(k)\}_{k=1}^T$
set $k = 0$ and $\mathbf{x}(0) = \mathbf{x}_0$;
while $k \leq T$ **do**
  update $\mathbf{p}_k(\boldsymbol{\beta})$ using Eq. (3) and Eq. (1) ;
  compute $\hat{\boldsymbol{\beta}}_{-i}$ if i is empathetic (or $\tilde{\boldsymbol{\beta}}_{-i}$ if non-empathetic) using Eq. (4) or Eq. (5);
  compute $u_i(k)$ from Eq. (6);
  compute $\mathbf{x}(k+1) = h(\mathbf{x}(k), \mathbf{u}(k))$;
  $k = k + 1$;
end

B. Action-value approximation

In the following, we describe the approaches to solving boundary value problems (BVPs) resulted from PMP for differential games and to learning value approximations based on the BVP solutions.

BVP: Following PMP and for fixed initial states, the equilibrial states $\mathbf{x}^*(t)$, actions $\mathbf{u}^*(t)$, co-states $\boldsymbol{\nu}_i^*(t) := \nabla_{\mathbf{x}} V_i^*(\mathbf{x}^*, t; \boldsymbol{\theta})$, and values $\mathbf{V}^*(\mathbf{x}^*, t; \boldsymbol{\theta})$ for $t \in [0, T]$ are solutions to the following BVP [13]:

$$\begin{aligned}
&\dot{\mathbf{x}}^* = h(\mathbf{x}^*(t), \mathbf{u}^*(t)) \\
&\mathbf{x}^*(0) = \mathbf{x}_0 \\
&\dot{\boldsymbol{\nu}}_i^* = -\nabla_{\mathbf{x}} H_i(\mathbf{x}^*, u_i^*, \nu_i^*(t); \boldsymbol{\theta}) \\
&\nu_i^*(T) = -\nabla_{\mathbf{x}} c_i(\mathbf{x}^*(T); \boldsymbol{\theta}) \\
&u_i^*(t) = \arg\max_{u_i \in \mathcal{U}} H_i(\mathbf{x}^*, u_i, \nu_i^*(t); \boldsymbol{\theta}), \\
&\dot{\mathbf{V}}^*(\mathbf{x}^*, t; \boldsymbol{\theta}) = f(\mathbf{x}^*, \mathbf{u}^*; \boldsymbol{\theta}), \\
&\mathbf{V}^*(\mathbf{x}^*, T; \boldsymbol{\theta}) = c(\mathbf{x}^*(T); \boldsymbol{\theta}) \ \forall i = 1, ..., N,
\end{aligned} \tag{7}$$

where $H_i(\mathbf{x}, u_i, \nu_i, t; \boldsymbol{\theta}) = \nu_i^T h_i(\mathbf{x}, u_i) - f_i(\mathbf{x}, u_i; \boldsymbol{\theta})$ is the Hamiltonian for agent i. $\mathbf{x}_0$ is the initial states. Note that $\mathbf{V}^*(\mathbf{x}, t; \boldsymbol{\theta})$ is parameterized by all agent parameters due to its implicit dependence on the equilibrial actions $\mathbf{u}^*$. We solve Eq. (7) using a standard BVP solver [33] with case-specific modifications to be introduced in Sec. IV.

Value approximation: Solving the BVPs for given $\boldsymbol{\theta}$ and $\mathbf{x}_0$ gives us $\mathbf{V}^*$ and $\nabla_{\mathbf{x}} \mathbf{V}^*$ for all agents along an equilibrial trajectory starting from $\mathbf{x}_0$ and $t = 0$. Let this set of values

and co-states be $D_v(\mathbf{x}_0, \boldsymbol{\theta})$. We then collect the dataset $\mathcal{D}_v := \{D_v(\mathbf{x}, \boldsymbol{\theta}) \mid \mathbf{x} \in \mathcal{S}_\mathbf{x}, \boldsymbol{\theta} \in \Theta^N\}$ where $\mathcal{S}_\mathbf{x}$ is a finite mesh of $\mathcal{X}^N$. This data allows us to build surrogate models for the equilibrial values: $\hat{\mathbf{V}}(\cdot, \cdot; \boldsymbol{\theta}, w) : \mathcal{X}^N \times [0, T] \to \mathbb{R}^N$ by solving the following training problem with respect to the surrogate model parameters w:

$$\min_w \sum_{(\mathbf{x}, t, \mathbf{V}^*, \nabla\mathbf{V}^*) \in \mathcal{D}_v} \Big(||\hat{\mathbf{V}}(\mathbf{x}, t; \boldsymbol{\theta}, w) - \mathbf{V}^*||^2 + C||\nabla_\mathbf{x}\hat{\mathbf{V}}(\mathbf{x}, t; \boldsymbol{\theta}, w) - \nabla_\mathbf{x}\mathbf{V}^*||^2 \Big). \tag{8}$$

Here C balances the matching of values and co-states, and $||\cdot||$ is the l_2-norm. To accommodate potential discontinuity in the values, we model $\hat{\mathbf{V}}$ using a deep neural network, and derive its co-states through auto-differentiation. Eq. (8) can then be solved using a gradient-based solver for all combinations of preferences $\boldsymbol{\theta} \in \Theta^N$. The result is a set of value functions $\mathcal{V} := \{\hat{\mathbf{V}}(\cdot, \cdot; \boldsymbol{\theta}, w)\}_{\boldsymbol{\theta} \in \Theta^N}$. Alg. 2 summarizes the value approximation procedure.

Algorithm 2: Value approximation

input : $\mathcal{S}_\mathbf{x}$, Θ^N, T
output: $\{\hat{\mathbf{V}}(\cdot, \cdot; \boldsymbol{\theta}, w)\}_{\boldsymbol{\theta} \in \Theta^N}$

```
set D_v = ∅, V = ∅;
for each (x_0, θ) ∈ S_x × Θ^N do
    solve Eq. (7) for
      D_v(x_0, θ) = {x*(t), ν*(t), V*(x*, t; θ)}_{t∈[0,T]} ;
    D_v ← D_v(x_0, θ);
end
for each θ ∈ Θ^N do
    solve Eq. (8) for V̂(·, ·; θ, w);
    V ← V̂(·, ·; θ, w);
end
```

Hamiltonian approximation: We approximate the Hamiltonian at time t using $\nabla_\mathbf{x}\hat{\mathbf{V}}$ as the co-states. Note that we need to consider time as part of the state since the game has a finite time horizon.

IV. Case Study

The goal of the case study is to identify interaction settings where empathetic agents together perform "better" than non-empathetic ones. In order to perform a thorough study and due to space limitation, we focus on an uncontrolled intersection case and discuss experimental settings, hypotheses and analyses as follows.

A. Uncontrolled intersection

This case models the interaction between two cars at an uncontrolled intersection specified in Fig. 1a. The state of agent i is defined by the agent's position d_i and its velocity v_i: $x_i = (d_i, v_i)$. The individual state space is set as $\mathcal{X} = [15, 20]m \times \{18, 18\}m/s$, where the initial velocity is fixed for visualization purpose and can be extended in future work. The action of agent i is defined as its acceleration rate, and the action space as $\mathcal{U} = [-5, 10]m/s^2$. The instantaneous reward function is

$$f_i(\mathbf{x}, u_i; \boldsymbol{\theta}) = f^{(e)}(u_i) + f^{(c)}(\mathbf{x}; \theta_i), \tag{9}$$

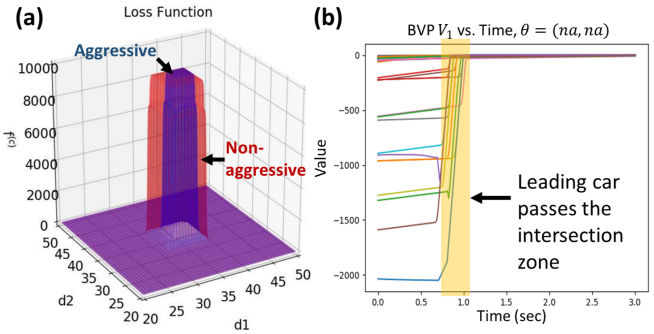

Fig. 2. (a) Collision loss in (d_1, d_2) (b) Equilibrial value of one agent along time when both agents are non-aggressive

where $f^{(e)}(u_i) = -u_i^2$ is a negative effort loss and

$$f^{(c)}(\mathbf{x}; \theta_i) = -b \prod_{i,j=\{(1,2),(2,1)\}} \sigma_1(d_i, \theta_i)\sigma_2(d_j) \tag{10}$$

models a negative penalty for collision. Here

$$\sigma_1(d, \theta) = (1 + \exp(-\gamma(d - R/2 + \theta W/2)))^{-1}; \tag{11}$$

$$\sigma_2(d) = (1 + \exp(\gamma(d - R/2 - W/2 - L)))^{-1}; \tag{12}$$

$b = 10^4$ sets a high loss for collision; $\gamma = 10$ is a shape parameter for σ, which is designed to recreate a rapid rise in loss when two cars are in contact; R, L, and W are the road length, car length, and car width, respectively (Fig. 1a). θ_j denotes the aggressiveness (sensitivity to collision) of the agent. Fig. 2a visualizes $f^{(c)}$ along d_1 and d_2. The terminal loss is defined as $c_i(\mathbf{x}) = \alpha d_i(T) - (v_i(T) - v_0)^2$, where $\alpha = 10^{-6}$ is the scaling factor of displacement reward at T, i.e., the agent is rewarded for moving forward and restoring its initial speed at T. We adopt a simple dynamical model:

$$\begin{bmatrix} \dot{d}_i(t) \\ \dot{v}_i(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} d_i(t) \\ v_i(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_i(t). \tag{13}$$

We set agent parameters $\Theta = \{1, 5\}$ and $\Lambda = \{0.1, 0.5\}$ as common knowledge. Note that $\theta_i = 1$ and $\theta_i = 5$ represents an aggressive and a non-aggressive agent; $\lambda_i = 0.1$ and $\lambda_i = 0.5$ represents a noisy and less-noisy decision model. We solve BVPs on $\mathcal{S}_\mathbf{x}$, which is a meshgrid of $\mathcal{X}$ with an interval of $0.5m$ for both d_1 and d_2.

Solving BVPs: BVP solutions for complete-information differential games are known to be dependent on the initial guess of state and co-state trajectories [34]. Specific to our case, it can be shown from Eq. (7) that collision avoiding behavior can only be derived when numerical integration over $\partial f^{(c)}/\partial d_i$ can be correctly performed. This integration, however, is challenging since $\partial f^{(c)}/\partial d_i$ resembles a mixture of delta functions, and therefore requires dense sampling in the space of (d_1, d_2) where the collision happens. To this end, we predict two time stamps, t_1 and t_2, respectively corresponding to (1) when the second car enters and (2) when the first car leaves the intersection zone. The prediction is done by assuming that the leading car moves at its initial velocity and the trailing car takes maximum deceleration. We then densely sample around t_1 using $\{t_1 \pm 1.25 \times 10^{-6} k\}_{k=0}^{800}$. These time stamps along with the approximated agent actions provide an initial guess for the system states and co-states.

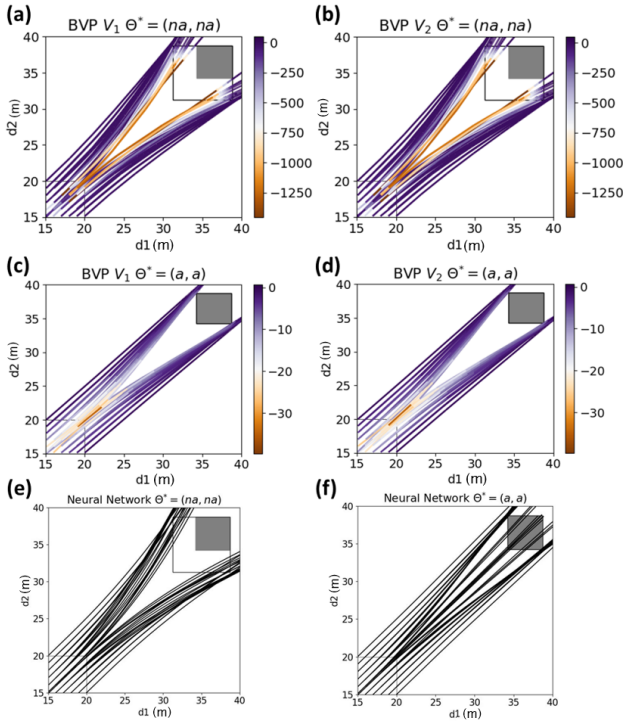

Fig. 3. Interactions b/w non-aggressive (a,b) and aggressive agents (marked as (na,na) and (a,a) respectively), colored by the equilibrial values of agent 1 (a,c) and 2 (b,d). Negative values due to collision penalty. Smaller values in (a,a) due to low sensitivity to close distances by aggressive agents. (e,f) Interactions reproduced through value approximation for (na,na) and (a,a). Gray box and white box represent the (a, a) collision zone and the (na, na) psuedo-collision zone, respectively.

Fig. 3a-d demonstrates equilibrial trajectories in the space of (d_1, d_2) when both agents are non-aggressive and aggressive.

Approximating values: We notice that there exist abrupt changes in the value along time and space in the BVP solutions, due to the high penalty of collision and close calls (Fig. 2b), i.e., after the two agents pass each other, which in some cases incurs high loss due to close distances between the two, the value increases significantly. We found that conventional fully-connected network architectures cannot effective learn this unique structure, and therefore propose the following value network architecture:

$$\hat{V}(\mathbf{x}, t; \boldsymbol{\theta}, w) = \eta f_1(\mathbf{x}, t; \boldsymbol{\theta}, w) + (1-\eta) f_2(\mathbf{x}, t; \boldsymbol{\theta}, w), \quad (14)$$

where f_1 and f_2 follow the same architecture: `fc5-(fc16-tanh)×3-(fc2-tanh)`, where `fc`n represents a fully connected layer with n nodes and `tanh` is the hyperbolic tangent activation. η is a sigmoid function that determines whether one of the agents have passed the intersection zone. Training data are collected from $|\mathcal{S}_\mathbf{x}| = 121$ BVP solutions, and test data from another 36 solutions sampled in $\mathcal{X}^2$ (on average 100 nodes per trajectory). We use ADAM [35] with learning rate 0.01 and the hyperparameter $C = 1$, which performed the best after testing a set of choices. Fig. 3e,f illustrate the approximated trajectories in (d_1, d_2) considering complete information, where actions are chosen by maximizing the approximated action-values using the resultant value networks, when both agents are respectively non-aggressive and aggressive. Some remarks: While value approximation is not perfect (test relative MAE of 15.64% and 12.17% for non-aggressive and aggressive cases, whereas the original neural network design has MAE around $10^{-1}\%$ to 1%), the approximated equilibria are mostly acceptable. We do face an intrinsic challenge in learning the values when both agents have the same initial states and are aggressive, potentially due to a combination of relatively high error in co-state approximation (83.73% relative MAE) and the nonuniqueness of equilibria in these scenarios, i.e., either of the agent can yield or move first.

B. Driving scenarios

An incomplete-information driving scenario is a tuple $s =< \mathbf{x}_0, \mathbf{p}_0(\boldsymbol{\beta}), \boldsymbol{\theta}^*, \mathbf{l} >$ specifying the initial state, prior belief, true parameters, and estimation types. We pick initial states from $\mathcal{S}_\mathbf{x}$, and parameters (aggressiveness) from Θ. We use a, na, n, ln for aggressive, non-aggressive, noisy, less-noisy, respectively. **The prior common belief set $\mathcal{P}_0$:** With the above parameter settings, each element of the prior belief set $\mathcal{P}_0$ is a 4-by-4 matrix containing joint probabilities for all 16 agent parameter combinations. Each dimension of the matrix follows the order $(na, n), (na, ln), (a, n), (a, ln)$, e.g., the 1st row and 2nd column of the matrix represents $\Pr(\beta_1 = (na, n), \beta_2 = (na, ln))$. To constrain the scope of the studies, we assume that agents are mostly rational ($\Pr(\lambda_{1,2} = ln) = 0.8$), and explore two cases of $\boldsymbol{\theta}$: Everyone believes that everyone is (1) most likely non-aggressive ($\Pr(\theta_{1,2} = na) = 0.8$) or (2) most likely aggressive ($\Pr(\theta_{1,2} = a) = 0.8$). This reduces $\mathcal{P}_0$ to $\{p_0^{na}, p_0^a\}$, where p_0^{na} and p_0^a are the common priors where everyone is believed to be non-aggressive and aggressive, respectively. **Parameter estimation type:** We set $\mathcal{L} = \{(e, e), (ne, ne)\}$ where e stands for "empathetic" and ne for "non-empathetic". Using Alg. 1 and by setting a time interval of $0.05s$, interaction trajectories can be computed for each driving scenario s. The resultant values at $t = 0$ are denoted by $v(s)$. **Evaluation metrics**: Lastly, we measure the goodness of empathetic and non-empathetic estimations using the sum of the individual values (total value) at $t = 0$: $\bar{v}(s) = \sum_{i=1}^{N} v_i(s)$. **Implementation**: Code is available here. See supplementary video for animated interactions.

C. Hypotheses

The following hypotheses concerning two driving scenarios $s^{(1)}$ and $s^{(2)}$ will be tested empirically:

1) *Empathy leads to higher total value when agents are unknowingly aggressive (or non-aggressive):* Let $l^{(1)} = (e, e)$, $l^{(2)} = (ne, ne)$, $\theta^{*(1)} = \theta^{*(2)} = (a, a)$ (or (na, na)), $p_0^{(1)} = p_0^{(2)} = p_0^{na}$ (or p_0^a). There exists $\mathcal{X}_0' \subset \mathcal{X}_0'$, such that for all $\mathbf{x}_0 \in \mathcal{X}_0'$, $\bar{v}(s^{(1)}) > \bar{v}(s^{(2)})$.
2) *Empathy leads to higher total value when agents are knowingly aggressive (or non-aggressive):* The same as Hypothesis 1, except that the common beliefs are set to be consistent with the truth parameters.

D. Results and analysis

Hypothesis 1 passes the test (Fig. 4), suggesting that being empathetic leads to less collisions in the intersection case when initial common belief is wrong. Hypothesis 2 passes the test (Fig. 5), although results suggest that when the initial

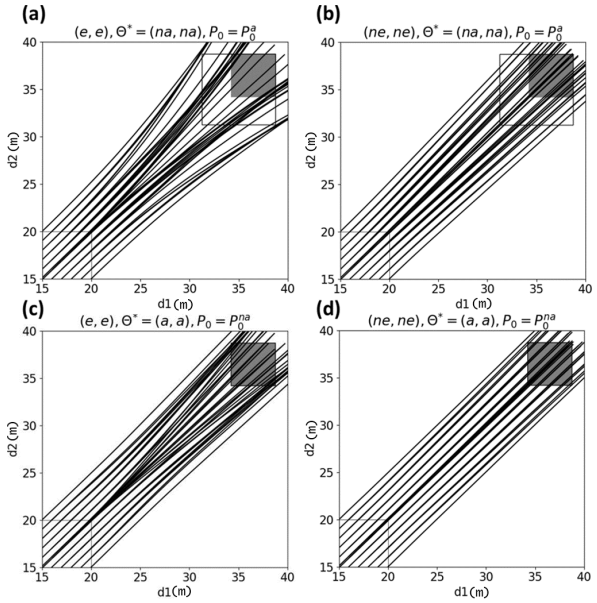

Fig. 4. Interactions when common belief mismatches with true reward parameters. (a,b) Unknowingly non-aggressive, (c,d) Unknowingly aggressive. (a,c) Empathetic, (b,d) Non-empathetic.

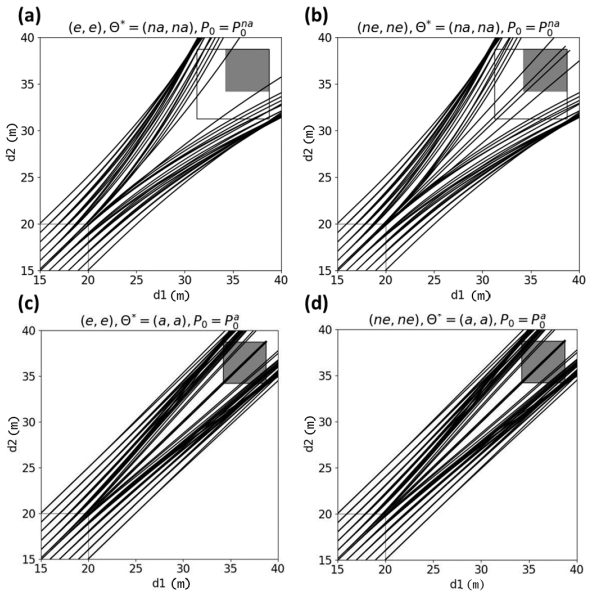

Fig. 5. Interactions when common belief matches with true reward parameters. (a,b) Knowingly non-aggressive, (c,d) Knowingly aggressive. (a,c) Empathetic, (b,d) Non-empathetic.

common belief is consistent with the true parameters, empathy does not play a significant role. Also notice that matching between belief and parameters help improve the interactions. To understand how empathy helps, we inspect whether agents choose the correct policies (among (na,na), (na,a), (a,na), and (a,a)) at each time step during the interaction following Alg. 1. Specifically, when agents are non-aggressive, the correct policy follows $\hat{\beta} = (na, na)$, vice versa. In Figs. 6 and 7, we color-code the correct (1) and incorrect (0) choices of policies for both agents. Results show that empathetic agents tend to choose the correct policies when they are trailing (Fig. 6). We conjecture that this is due to the fact that the actions of the leading agent are intrinsically more effective at signaling, i.e., its lower acceleration suggests that it does not care much about potential close distances and thus its high aggressiveness. On the other hand, non-empathetic agents never choose the correct policies (Fig. 7).

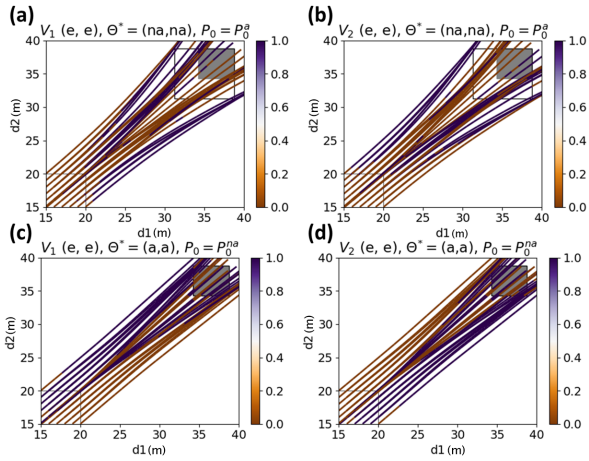

Fig. 6. Color-coding of the policy choices by empathetic agents, for non-aggressive (a,b) and aggressive (c,d) scenarios, where 1 (purple) represents when the policy is consistent with β^*.

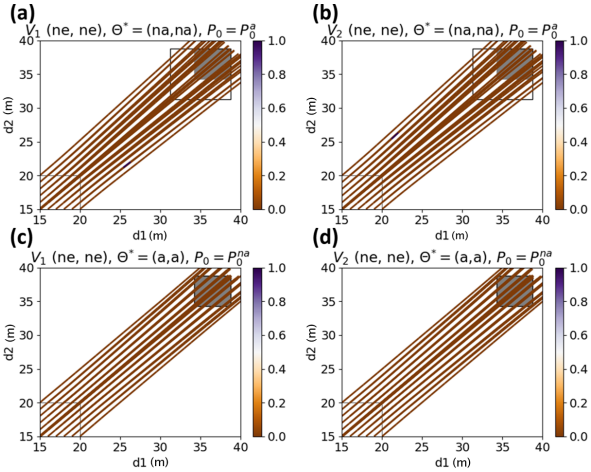

Fig. 7. Color-coding of the policy choices by non-empathetic agents, for non-aggressive (a,b) and aggressive (c,d) scenarios, where 1 (purple) represents when the policy is consistent with β^*.

V. Conclusions

Using an uncontrolled intersection case, we studied the utility of empathetic belief update in a two-agent incomplete-information differential game. We showed that empathy helped agents choose policies that led to higher total values when agents had common beliefs inconsistent with their true parameters. While its findings should be tested under a larger set of driving scenarios (e.g., roundabout and lane changing), this study provides a methodology for systematically evaluating the utility of empathy in incomplete-information differential games. The proposed interaction model can be improved in the following directions: (1) It is more reasonable to use beliefs (parameter distributions) rather than point estimates for motion planning, so that the planned actions take into account uncertainties of agents. (2) It is possible to improve Hamiltonian approximation by indirectly approximating the co-state trajectories, so that knowledge about the system dynamics and reward functions can be leveraged.

REFERENCES

[1] R. Buckdahn, P. Cardaliaguet, and M. Quincampoix, "Some recent aspects of differential game theory," *Dynamic Games and Applications*, vol. 1, no. 1, pp. 74–114, 2011.

[2] A. Sinha and A. Anastasopoulos, "Structured perfect bayesian equilibrium in infinite horizon dynamic games with asymmetric information," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2016, pp. 256–263.

[3] D. Vasal, A. Sinha, and A. Anastasopoulos, "A systematic process for evaluating structured perfect bayesian equilibria in dynamic games with asymmetric information," *IEEE Transactions on Automatic Control*, vol. 64, no. 1, pp. 81–96, 2018.

[4] J. N. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch, "Learning with Opponent-Learning Awareness," *arXiv:1709.04326 [cs]*, Sep. 2017, arXiv: 1709.04326.

[5] D. Sadigh, N. Landolfi, S. S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for cars that coordinate with people: leveraging effects on human actions for planning and active information gathering over human internal state," *Autonomous Robots*, vol. 42, no. 7, pp. 1405–1426, Oct. 2018.

[6] 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.

[7] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proceedings of the National Academy of Sciences*, vol. 116, no. 50, pp. 24 972–24 978, 2019.

[8] S. Nikolaidis, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in collaborative tasks: Models and experiments," *The International Journal of Robotics Research*, vol. 36, no. 5-7, pp. 618–634, 2017.

[9] L. Sun, W. Zhan, and M. Tomizuka, "Probabilistic prediction of interactive driving behavior via hierarchical inverse reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2111–2117.

[10] C. Peng and M. Tomizuka, "Bayesian persuasive driving," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 723–729.

[11] 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.

[12] J. C. Harsanyi, "Games with incomplete information played by "bayesian" players, i–iii part i. the basic model," *Management science*, vol. 14, no. 3, pp. 159–182, 1967.

[13] L. S. Pontryagin, "On the theory of differential games," *RuMaS*, vol. 21, no. 4, pp. 193–246, 1966.

[14] T. Nakamura-Zimmerer, Q. Gong, and W. Kang, "Adaptive deep learning for high dimensional hamilton-jacobi-bellman equations," *arXiv preprint arXiv:1907.05317*, 2019.

[15] D. Fudenberg and J. Tirole, "Perfect bayesian equilibrium and sequential equilibrium," *journal of Economic Theory*, vol. 53, no. 2, pp. 236–260, 1991.

[16] G. Antonini, M. Bierlaire, and M. Weber, "Discrete choice models of pedestrian walking behavior," *Transportation Research Part B: Methodological*, vol. 40, no. 8, pp. 667–687, 2006.

[17] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.

[18] A. Bobu, D. R. Scobee, J. F. Fisac, S. S. Sastry, and A. D. Dragan, "Less is more: Rethinking probabilistic models of human behavior," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 429–437.

[19] D. McFadden and K. Train, "Mixed mnl models for discrete response," *Journal of applied Econometrics*, vol. 15, no. 5, pp. 447–470, 2000.

[20] X. Su, "Bounded rationality in newsvendor models," *Manufacturing & Service Operations Management*, vol. 10, no. 4, pp. 566–589, 2008.

[21] G. I. Bischi and A. Naimzada, "Global analysis of a dynamic duopoly game with bounded rationality," in *Advances in dynamic games and applications*. Springer, 2000, pp. 361–385.

[22] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, "Courteous autonomous cars," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 663–670.

[23] 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, 2020.

[24] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 1.

[25] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[26] J. Fu, K. Luo, and S. Levine, "Learning robust rewards with adversarial inverse reinforcement learning," *arXiv preprint arXiv:1710.11248*, 2017.

[27] X. Lin, P. A. Beling, and R. Cogill, "Multi-agent inverse reinforcement learning for zero-sum games," *arXiv preprint arXiv:1403.6508*, 2014.

[28] X. Lin, S. C. Adams, and P. A. Beling, "Multi-agent inverse reinforcement learning for certain general-sum stochastic games," *Journal of Artificial Intelligence Research*, vol. 66, pp. 473–502, 2019.

[29] L. C. Evans and P. E. Souganidis, "Differential games and representation formulas for solutions of hamilton-jacobi-isaacs equations," *Indiana University mathematics journal*, vol. 33, no. 5, pp. 773–797, 1984.

[30] P.-L. Lions and P. E. Souganidis, "Differential games, optimal control and directional derivatives of viscosity solutions of bellman's and isaacs' equations," *SIAM journal on control and optimization*, vol. 23, no. 4, pp. 566–583, 1985.

[31] V. Rubies-Royo and C. Tomlin, "Recursive regression with neural networks: Approximating the hji pde solution," *arXiv preprint arXiv:1611.02739*, 2016.

[32] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, "The expressive power of neural networks: A view from the width," in *Advances in neural information processing systems*, 2017, pp. 6231–6239.

[33] J. Kierzenka and L. F. Shampine, "A bvp solver based on residual control and the maltab pse," *ACM Transactions on Mathematical Software (TOMS)*, vol. 27, no. 3, pp. 299–316, 2001.

[34] P. A. Johnson, "Numerical solution methods for differential game problems," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.