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

Yi Chen¹, Lei Zhang¹, Tanner Merry¹, Sunny Amatya², Wenlong Zhang², and Yi Ren¹

Abstract—Human-robot interactions (HRI) can be modeled as dynamic or 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 consider approximated solutions composed of parameter estimation and motion planning steps, in order to decouple the belief and physical dynamics. In parameter estimation, current approaches often assume that the reward parameters of the robot are known by the humans. We argue that by falsely conditioning on this assumption, the robot performs non-empathetic estimation of the humans’ parameters, leading to undesirable values even in the simplest interactions. We test this argument by studying a two-vehicle uncontrolled intersection case with short reaction time. Results show that when both agents are unknowingly aggressive (or non-aggressive), empathy leads to more effective parameter estimation and higher reward values, suggesting that empathy is necessary when the true parameters of agents mismatch with their common belief. The proposed estimation and planning algorithms are therefore more robust than the existing approaches, by fully acknowledging the nature of information asymmetry in HRI. Lastly, we introduce value approximation techniques for real-time execution of the proposed algorithms.

I. INTRODUCTION

Human-robot interactions (HRI) have become ubiquitous in the past two decades, with applications in daily assistance, healthcare, manufacturing, and defense. Since humans and robots may not understand intents and strategies of each other when completing collaborative tasks, we consider HRI as dynamic general-sum games with incomplete information, where agents hold private parameters that determine their rewards. 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 dimensionality of the state, action, and reward parameter spaces [2], [3]. As a result, most existing HRI studies either resort to simplified complete-information games [4]–[7], or adopt variants of an empirical solution composed of two steps, namely, parameter estimation and motion planning [8]–[11]. This paper focuses on the latter approach. In the parameter estimation step, each agent takes an estimate of all agents’ parameters, and each derives an estimate (and its uncertainty) of the parameters of their fellow agents. This estimate is then used by the agent to plan its actions.

A common approach to parameter estimation based on noisy (bounded) rationality [11] assumes a Boltzmann distribution on the action space, where the probability of taking any action is determined by the action-value (Q-value), i.e., the equilibrial cumulative reward if the action is taken. Since action-values are parameterized by agents’ reward parameters, the belief about these parameters can be computed through Bayes updates using the observed actions (Sec. III). Specific to HRI, existing implementations of parameter estimation fall into two categories: In the first, agents assume that their ego parameters are known to their fellows, and estimate fellows’ parameters conditioned on this assumption [8]–[10]. This approach set agents to be non-empathetic, since they do not consider the potential mismatch between their true parameters and the estimates by their fellows. The second approach removes the conditioning [12], [13]. Since the common belief is not necessarily consistent with the true parameters, agents become empathetic, i.e., allowing their
fellow to have wrong estimates about their parameters. It should be noted that empathy incurs higher computational loads, since it requires full knowledge of the probabilities over the joint parameter space of all agents. In comparison, non-empathetic estimation only requires partial knowledge due to the conditioning. The necessity of empathy was only recently demonstrated for hand-picked interaction settings (e.g., vehicle interactions with specific initial states [12], [14]), which motivates the central question of this paper:

**In what interactions does empathy of agents matter?**

This paper makes the following contributions towards answering this question (summary in Fig. 1). (1) **Methodology:** 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 (or disadvantages) of empathy. (2) **Knowledge:** 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) **Computation:** 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 Pontryagin’s Maximum Principle (PMP) [15] 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 parameter estimation, motion planning, and data-driven action-value approximation. Sec. [IV] introduces the case, experimental settings, hypotheses, and analyses. 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 [16]. 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 where the solutions are probability measures (rather than point values) conditioned on the agents’ private information. 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 parameter 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 been studied.

**Decision modeling:** Human decision models in HRI [17]–[19] follow studies in choice modeling [20] and behavioral economics [21], [22]. Risk models are introduced to explain seemingly non-optimal human actions [6]. Social value orientation is introduced in [7] to explain agents’ courtesy towards others in general-sum dynamic games. Similar courtesy models have been discussed in [14], [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], under the term of inverse reinforcement learning (IRL). For dynamic games, IRL falsely assumes that each agent uses optimal control while all of its fellow agents follow fixed trajectories. To address this, MIRL performs estimation under solution concepts of the game rather than assuming optimality of individual actions [27], [28]. Along the same vein, 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 equilibrium as a solution concept, this paper considers agents to take simultaneous actions and are thus Nash equilibrial.

**Value approximation:** Solutions to Hamilton-Jacobi Bellman (HJB) equations have recently been shown to be effective at approximating solutions to Hamilton-Jacobi-Bellman equations underlying optimal control problems [31] and Hamilton-Jacobi-Isaac (HJI) equations for two-player zero-sum games with complete information [32], thanks to the universal approximation capability of DNNs [33]. 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’ action-values during parameter estimation and motion planning. In comparison, [7] requires equilibria to be computed by iteratively solving the KKT problems during the estimation of agent parameters, while the proposed value approximation technique allows agents to memorize values offline, thus accelerating the online estimation.

## III. METHODS

This section introduces the parameter estimation and motion planning algorithms to be used in the case study. We also elaborate on the approximation techniques of action-values.

### A. Parameter estimation 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 \( \Theta \), and rationality set \( \Lambda \). 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 \( \beta_i := \langle \lambda_i, \theta_i \rangle > 0 \) be the parameters of agent \( i \), where \( \lambda_i \in \Lambda \) and \( \theta_i \in \Theta \). We denote the total parameter...
set by $B := \Lambda^N \times \Theta^N$. $\Theta$, $\Lambda$, $B$, and $\mathcal{U}$ are considered discrete in this study. To reduce notational burden, we use a single variable $a$ for the set $(a_1, \ldots, a_N)$ and define $a_{-i} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N)$. E.g., $\beta \in B$ contains parameters for all agents, $\beta_{-i}$ 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 common belief $p_0(\beta)$, which will be updated as $p_k(\beta) := p(\beta; D(k))$ at time step $k$ with observations $D(k) = \{(x(t), u(t))\}_{t=1}^k$.

**Nash equilibria for complete-information game:** If $\theta$ is known to all and unique Nash equilibria exist, agents can derive the equilibrial action-values $Q(\cdot; \cdot; \theta): \mathcal{X}^N \times \mathcal{U}^N \rightarrow \mathbb{R}^N$. For agent $i$, $Q_i(x; u; \theta)$ is the value of action $u_i$ in state $x_i$ when ego and fellow parameters are $\theta_i$ and $\theta_{-i}$, respectively. For the discrete set of joint parameters, $\Theta^N$, we can derive $Q^N := \{Q(\cdot; \cdot; \theta)\}_{\theta \in \Theta^N}$, which maps $\Theta^N$ to the coupled equilibrial action-values. E.g., for a two-agent game where $|\Theta| = 2$, we have $|Q^2| = 4$ pairs of action-values. We discuss the approximation of action-values in Sec. III-B.

**Belief update:** Given observations $D(k)$ at time step $k$, $p_k(\beta)$ follows Bayes update:

$$p_k(\beta) = \frac{p(u(k)|x(k); \beta)p_{k-1}(\beta)}{\sum_{\beta' \in B} p(u(k)|x(k); \beta') p_{k-1}(\beta')}$$

(1)

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

$$p(u_i|x; \beta) = \frac{\exp(\lambda_i Q_i(x_i, (u_i, u_{-i}^1); \theta))}{\sum_{\beta_i} \exp(\lambda_i Q_i(x_i, (u_i', u_{-i}^1); \theta))}$$

(2)

$\forall i = 1, \ldots, N$. Here $u_{-i}^1$ are the observed actions taken by all agents except $i$ at the previous time step. Notice that this update is shared among all agents, provided that they receive the same observations. Therefore $p_k(\beta)$ remains a common belief. It should also be noted that if $\beta \in B$ 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 $p_k(\beta)$ as

$$p_k(\beta) = (1 - \epsilon)p_k(\beta) + \epsilon p_0(\beta)$$

(3)

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

**Parameter estimation:** The point estimate

$$\hat{\beta}(k) = \arg \max_{\beta \in B} p_k(\beta)$$

(4)

is empathetic for agent $i$ in the sense that it allows others’ estimation of $\beta_i$, which is the $i$th element of $\hat{\beta}(k)$, to be different from the true parameters $\beta^*_i$. Notice that $\hat{\beta}(k)$ is shared among all empathetic agents. For a non-empathetic agent $i$, its own estimation of fellow agents follows

$$\hat{\beta}_{-i}(k) = \arg \max_{\beta_{-i} \in B_{-i}} p_k(\beta_i; \beta_{-i}^*)$$

(5)

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 agents take actions strictly following the common belief, the interactions will be solely determined by the prior $p_0(\beta)$ independent of the private parameters of agents. This is inconsistent with real-world interactions where agents express their own intents. Therefore, we model each agent $i$ to follow control policies parameterized by their own parameters and the estimates of others, i.e., $\hat{\theta} = (\theta^*_i, \theta_{-i})$ for the empathetic case and $\hat{\theta} = (\theta^*_i, \theta_{-i})$ for the non-empathetic case. Specifically, agent $i$ takes actions deterministically following

$$u_i = \arg \max_{u_i \in \mathcal{U}} Q_i(x_i, (u_i, u_{-i}^1); \hat{\theta})$$

(6)

**Simulated interactions:** Alg. 1 summarizes the simulation of an interaction, which is parameterized by the initial states $x_0$, the set of agent parameters $\beta^*$, and the prior belief $p_0(\beta)$. The simulation outputs the trajectories of states $x(k)$, actions $u(k)$, beliefs $p_k(\beta)$, and values $v(k)$ of all agents.

**Algorithm 1:** Multi-agent interaction

**input:** $x_0, \beta^*, p_0(\beta)$

**output:** $\{(x(k), u(k), p_k(\beta), v(k))\}_{k=1}^T$

1. set $k = 0$ and $x(0) = x_0$;
2. while $k \leq T$ do
   3. compute $\hat{\beta}_{-i}$ if $i$ is empathetic (or $\hat{\beta}_{-i}$ if non-empathetic) using Eq. (4) or Eq. (5);
   4. compute $u_i(k)$ from Eq. (6);
   5. compute $x(k + 1) = h(x(k), u(k))$;
   6. update $p_{k+1}(\beta)$ using Eq. (3) and Eq. (1);
   7. $k = k + 1$;
end

**B. Action-value approximation**

Theoretically, the action-values of incomplete-information dynamic games should be defined on the joint space of physical and belief states. As a simplification, we approximate the action-values using those of the complete-information version of the games, parameterized by the estimated agent parameters. Further, while we model the interactions as (discrete-time) dynamic games, we propose to derive the equilibria from the (continuous-time) differential games to fully exploit the known physical dynamics. In the following, we describe the approaches to solving the resultant BVPs and to learning a value approximation based on the BVP solutions. We will highlight case-specific challenges and our solutions after we introduce the uncontrolled intersection case in Sec. IV.

**BVP:** Following PMP and for fixed initial states, the equilibrial states $x^*(t)$, actions $u^*(t)$, co-states $\lambda^*_i(t) := \nabla_x V^*_i(x^*, t; \theta)$, and values $V^*(x^*, t; \theta)$ for $t \in [0, T]$ are solutions to the following BVP [15]:

$$x^*(t) = h(x^*(t), u^*(t))$$

$$x^*(0) = x_0$$

$$\lambda^*_i = -\nabla_x H_i(x^*, u^*, \lambda^*_i(t); \theta)$$

$$\lambda^*_i(T) = -\nabla_c c_i(x^*(T); \theta)$$

$$u^*_i(t) = \arg \max_{u_i \in \mathcal{U}} H_i(x^*, u_i, \lambda^*_i(t); \theta),$$

(7)

$$V^*(x^*, t; \theta) = f(x^*, u^*; \theta),$$

$$v^*(x^*, T; \theta) = c(x^*(T); \theta) \\forall i = 1, ..., N,$$
where \( H_i(x, u, \lambda_i; t; \theta) = \lambda_i^T h_i(x, u) - f_i(x, u; \theta) \) is the Hamiltonian for agent \( i \). \( x_0 \) is the initial state. Note that \( V^*(x, t; \theta) \) is parameterized by all agent parameters due to its implicit dependence on the equilibrual actions \( u^* \). We solve Eq. (7) using a standard BVP solver [34] with case-specific modifications to be introduced in Sec. IV

**Value approximation**: Solving the BVPs for given \( \theta \) and \( x_0 \) gives us \( V^* \) and \( \nabla_x V^* \) for all agents along an equilibrual trajectory starting from \( x_0 \) and \( t = 0 \). Let this set of values and co-states be \( D_\theta(x_0, \theta) \). We then collect the data \( D_\theta := \{ D_\theta(x, \theta) \mid x \in S_x, \theta \in \Theta^N \} \) where \( S_x \) is a finite mesh of \( \mathcal{X}^N \). This data allows us to build surrogate models for the equilibrual values: \( \hat{V}(\cdot; \cdot, \theta, w) : \mathcal{X}^N \times [0, T] \to \mathbb{R}^N \) by solving the following training problem with respect to the surrogate model parameters \( w \):

\[
\min_w \sum_{(x, t, V^*, \nabla V^*) \in D_\theta} \left( ||\hat{V}(x, t; \theta, w) - V^*||^2 + C \|
abla_x \hat{V}(x, t; \theta, w) - \nabla_x V^*\|^2 \right).
\]

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 \( \hat{V} \) 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 \( \theta \in \Theta^N \). The result is a set of value functions \( \tilde{V} := \{ \tilde{V}(\cdot; \cdot, \theta, w) \}_{\theta \in \Theta^N} \). Alg. 2 summarizes the value approximation procedure.

**Algorithm 2: Value approximation**

\[
\text{input} : S_x, \Theta^N, T \\
\text{output} : \{ \tilde{V}(\cdot; \cdot, \theta, w) \}_{\theta \in \Theta^N} \\
1 \text{ set } D_\theta = \emptyset, \mathcal{V} = \emptyset; \\
2 \text{ for each } (x_0, \theta) \in S_x \times \Theta^N \text{ do} \\
3 \quad \text{ solve Eq. (7) for } D_\theta(x_0, \theta) = \{x^*(t), \lambda^*(t), V^*(x^*, t; \theta)\}_{t \in [0, T]}; \\
4 \quad D_\theta \leftarrow D_\theta(x_0, \theta); \\
5 \text{ end} \\
6 \text{ for each } \theta \in \Theta^N \text{ do} \\
7 \quad \text{ solve Eq. (8) for } \tilde{V}(\cdot; \cdot, \theta, w); \\
8 \quad \mathcal{V} \leftarrow \tilde{V}(\cdot; \cdot, \theta, w); \\
9 \text{ end}
\]

**Action-value approximation**: We approximate the action-value at time \( t \) using the Hamiltonian. 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.

![Fig. 2. (a) Collision loss in \((d_1,d_2)\) (b) Equilibrual value of one agent along time when both agents are non-aggressive](image)

**A. Uncontrolled intersection**

This case models the interaction between two cars at an uncontrolled intersection specified in Fig. 1. 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

\[
f_i(x, u; \theta) = f^{(c)}(u_i) + f^{(\alpha)}(x; \theta_i),
\]

where \( f^{(c)}(u_i) = -u_i^2 \) is a negative effort loss and \( f^{(\alpha)}(x; \theta_i) = -b \prod_{i,j=(1,2),(2,1)} \sigma_1(d_i, \theta_i)\sigma_2(d_j) \) models a negative penalty for collision. Here

\[
\sigma_1(d, \theta) = (1 + \exp(-\gamma(d - R/2 + \theta W/2)))^{-1},
\]

\[
\sigma_2(d) = (1 + \exp(d - R/2 - W/2 - L))^{-1},
\]

\( b = 10^4 \) sets a high loss for collision; \( \gamma = 10 \) is a shape parameter; \( R, L, \) and \( W \) are the road length, car length, and car width, respectively (Fig. 1b). \( \theta_j \) denotes the aggressiveness (sensitivity to collision) of the agent. Fig. 2 visualizes \( f^{(\alpha)} \) along \( d_1 \) and \( d_2 \). The terminal loss is defined as \( c_i(x) = \alpha d_i(T) - (v_i(T) - v_0)^2 \), where \( \alpha = 10^{-4} \), i.e., the agent is rewarded for moving forward and restoring its initial speed at \( T \). We adopt a simple dynamical model:

\[
\dot{x}_i(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} d_i(t) \\ v_i(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_i(t).
\]

We set \( \Theta = \{1, 5\} \) and \( \Lambda = \{0.1, 0.5\} \) as common knowledge. Note that \( \theta_i = 1, \lambda_i = 5 \) represents an aggressive (non-aggressive) agent; \( \lambda_i = 0.1, \lambda_i = 0.5 \) represents a noisy (less-noisy) decision model. We solve BVPs on \( S_x \), 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 [35]. 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 \(\{t_1 ± 1.25 \times 10^{-6} k\}_{k=0}^{800}\). These time stamps along with the approximated agent actions provide an initial guess for the system states and co-states. Fig. 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. 2a-b 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\% (12.17\%) for non-aggressive (aggressive) cases), 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 agents can yield or move first.

**B. Driving scenarios**

An incomplete-information driving scenario is a tuple \(s = (x_0, p_0(\beta), \theta^*, I)\) specifying the initial state, prior belief, true parameters, and estimation types. We pick initial states from \(S_x\), and parameters (aggressiveness) from \(\Theta\). We use \(a\) for aggressive and \(na\) for non-aggressive agents, and \(n\) for noisy and \(ln\) for less-noisy agents during belief updates. The prior common belief set \(P_0\): With the above parameter settings, each element of the prior belief set \(P_0\) is a 4-by-4 matrix containing joint probabilities for all 16 agent parameter combinations. Each dimension of the matrix follows the order \((a, n), (a, ln), (na, n), (na, ln)\), e.g., the 1st row and 2nd column of the matrix represents \(\Pr(\beta_1 = (a, n), \beta_2 = (a, ln))\). To constrain the scope of the studies, we set to prior belief that agents are mostly rational \(\Pr(\lambda_{1,2} = ln) = 0.8\), and explore two cases on \(\theta\): (1) Everyone believes that everyone is most likely non-aggressive \(\Pr(\theta(\lambda_{1,2} = na) = 0.8)\) or most likely aggressive \(\Pr(\theta(\lambda_{1,2} = a) = 0.8)\). This reduces \(P_0\) to \(p_0^{na}(p_0^{na})\), where \(p_0^{na}(p_0^{na})\) is the common prior where everyone is believed to be non-aggressive (aggressive).

**Parameter estimation type**: We set \(L = \{(e, e), (ne, ne)\}\) where \(e\) stands for “empathetic” and \(ne\) for “non-empathetic”. Using Alg. 1 and by setting a time interval of 0.05\(s\), 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 social value, i.e., the sum of the individual values at \(t = 0\): \(\bar{v}(s) = \sum_{i=1}^{N} v_i(s)\). 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 social 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 \((na, na)\)), \(p_0^{(1)} = p_0^{(2)} = p_0^{na}(p_0^{na})\). There exists \(X_0^{e} \subset X_0^{ne}\), such that for all \(x_0 \in X_0^{e}\), \(\bar{v}(s^{(1)}) > \bar{v}(s^{(2)})\).
2. **Empathy leads to higher social 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 higher social values (i.e., less collisions in the intersection case) when common belief is wrong. Hypothesis 2 passes the test (Fig. 5), although results suggest that when the 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. 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.

V. Conclusions

Using an uncontrolled intersection case, we studied the utility of empathetic parameter estimation in a two-agent incomplete-information dynamic game. We showed that empathy helped agents choose policies that led to higher social values when agents had common beliefs inconsistent with their true parameters. When belief and parameters were consistent, empathy did not hurt social values. 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 dynamic games.
