Enabling Courteous Vehicle Interactions through Game-based and Dynamics-aware Intent Inference

Yiwei Wang, *Student Member, IEEE,* Yi Ren, *Member, IEEE,* Steven Elliott, and Wenlong Zhang, *Member, IEEE*

***Abstract*—This paper concerns the open question of creating control policies of autonomous vehicles (AVs) that lead to courteous motion. The study is built on a two-agent interaction between two agents (M and H), where each agent plans its motion by optimizing a trade-off of goal fulfillment, safety, and courtesy losses. The paper has three contributions: First, the "double-blindness" issue in intent inference, i.e., inferring H's intent requires knowledge about H's inference of M's intent, is addressed. An empathetic intent inference algorithm is proposed, where H's intent, along with its inference of M's intent, are jointly inferred. Second, vehicle dynamics is explicitly incorporated into the intent inference to acknowledge its influence on decision making in driving through the drivers' knowledge about dynamical properties of surrounding vehicles. Lastly, a courtesy loss that leverages intent inference is introduced. This loss measures the expected additional loss to H caused by M's motion from a baseline where M behaves rationally and in favor of H. Simulation studies are conducted to demonstrate that (1) joint inference and knowledge about vehicle dynamics are important for the performance of intent decoding and motion planning, and (2) the proposed courtesy definition leads to more rational motions than those from an existing study.**

***Index Terms*—autonomous vehicles, Bayesian games, intent inference, human-machine interactions, courtesy.**

I. Introduction

The past decade has witnessed several major milestones in autonomous driving. It is expected that autonomous vehicles (AVs) and human-driven vehicles (HVs) will co-exist on road in the near future, and that AV-AV and AV-HV interactions will become ubiquitous. Scenarios such as lane changing, lane merging, and traffic intersections, where AV-involved interactions occur, are widely investigated [9], [26], [6]. Such interactions are not fully collaborative, thus requiring AVs to continuously decode the intents and predict future actions of surrounding vehicles in order to plan its own actions accordingly. However, intents of human drivers and other AVs can be unanticipated (e.g., due to cultural differences) or dynamically shifting (e.g., due to environment changes), and false inference of driver intents could lead to incorrect predictions of how interactions evolve, which, in turn, may cause uncourteous or even dangerous actions of AVs.

To connect dots from intent inference to courteous driving, this paper starts by investigating two key issues in driver intent inference that are only partially addressed in existing literature, namely, (1) the double-blindness of intents, and (2) the influence of vehicle dynamics on intent inference. The proposed inference algorithm then enables a refined definition of courteous motion that enforces rationality of the courtesy. Below we provide an overview of the key challenges, related work, and our approaches.

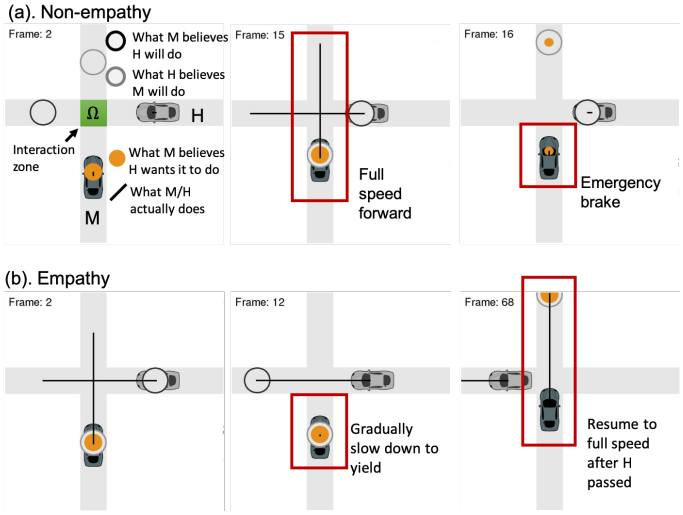


Fig. 1. A comparison between interactions at an intersection, with empathetic and non-empathetic M: By allowing H to have a potentially incorrect understanding of M's intent, M manages to recognize H's aggressiveness earlier, thus avoiding the situation where it has to perform an emergency brake.

(1) Double-blindness of intents: To infer the intent of a surrounding vehicle (denoted by "H" hereafter) based on the interaction history, one needs to know how H infers the intent of the host vehicle (denoted by "M" hereafter). In other words, M needs to be *empathetic* about the fact that H is unaware of M's actual intent, and therefore H only behaves according to its own expectation of M's future actions. The importance of acknowledging the double-blindness of intents is illustrated in Fig. 1, where M would have applied an emergency brake as it approaches an aggressive H (i.e., who cares little about collision), has M held the belief that H understands its true intent; in contrast an empathetic M would otherwise recognize H's aggressiveness earlier and choose to yield (see Sec. IV for detailed analysis).

The example here can be formalized as a Bayesian game without common knowledge priors [17]. Existing work in Bayesian games and autonomous driving have mostly circumvented this issue by assuming certain degree of common knowledge and model simplification: Some (e.g., [12], [14])

Manuscript received ...

Y. Wang, Y. Ren, and S. Elliott are with the School for Engineering of Matter, Transport and Energy, Ira A. Fulton Schools of Engineering, Arizona State University, Tempe, AZ, 85281 USA. E-mail:Yiwei.Wang.3@asu.edu, yiren@asu.edu, stevenelliott10@gmail.com.

W. Zhang is with The Polytechnic School, Ira A. Fulton Schools of Engineering, Arizona State Univeristy, Mesa, AZ, 85212 USA. Email: wenlong.zhang@asu.edu.

merely considered M as a bystander of the interaction, neglecting the influence of M's past actions on H's. Nikolaidis et al. [11] modeled H's intent as inferred based on past actions of both M and H, yet assumed that H's motion planning is independent from M's future actions, and thus H's understanding of M's intent was not considered in M's inference[1]. Again in a two-agent setting, Sun et al. [21] proposed to incorporate M's future actions into the inference of H's future actions. However, their approach assumed that the actual actions to be taken by M was known by H. Peng et al. introduced Bayesian persuasion to autonomous driving, where M manipulated H's belief of M's intent through actions [13]. The inference, however, is heuristic rather than game-theoretic[2]. Liu et al. [8] explicitly investigated the double-blindness of intents in a collaborative game with quadratic control objectives. The study demonstrated that by "blaming all", i.e., considering H's potential misunderstanding of M's intent while inferring H's intent, parameter estimation (intent inference) showed improved convergence. However, it is yet to know how double-blind intent inference should be done in non-cooperative games.

In this paper, we develop an **empathetic intent inference algorithm** where the intents of both agents are jointly inferred, e.g., when performed by agent M, the inference outcome includes the probability distributions of H's intent and H's expectation of M's intent, conditioned on past state trajectories of both agents. The inference will rely on a likelihood function that measures the difference between the observed and the inferred action of H. The latter is derived based on the assumption that both agents model each other using a baseline driving strategy where an agent chooses its motion uniformly from the Nash equilibrium set of the game (see Sec. II-A).

(2) Knowledge about vehicle dynamics: Both intent inference and motion planning of human drivers rely on the drivers' knowledge about physical properties of the driving environment. Fig. 2 compares two cases: In the first case, M yields to H by recognizing that H, a heavy-duty vehicle, has less acceleration or deceleration ability; in the second one, where M does not have the correct knowledge about the vehicle dynamics of H, M reaches an incorrect inference of H's intent and causes a collision. To acknowledge the importance of physics knowledge in learning and inference, there has recently been a surge of studies on integrating such knowledge into predictive models [23], [16]. For example, [24] proposed to regularize the prediction of intrinsic physical properties of objects through a physical simulator. Similarly, [20] regularized the prediction of object trajectories and human movements through physical constraints. Specific to autonomous driving, however, we have seen few studies that investigate the influence of physics knowledge, or the lack of it, on intent inference.

[1]We note that the assumption (human beings are reactive and do not consider the preference of a machine) is arguably sound given the specific human-robot interaction setting in [11], yet there is no evidence that it generalizes to driving interactions.

[2]E.g., following Peng et al., the aggressiveness of H is negatively related to its distance from M. This prior could be misleading, e.g., when an aggressive H is far from M yet approaching fast.

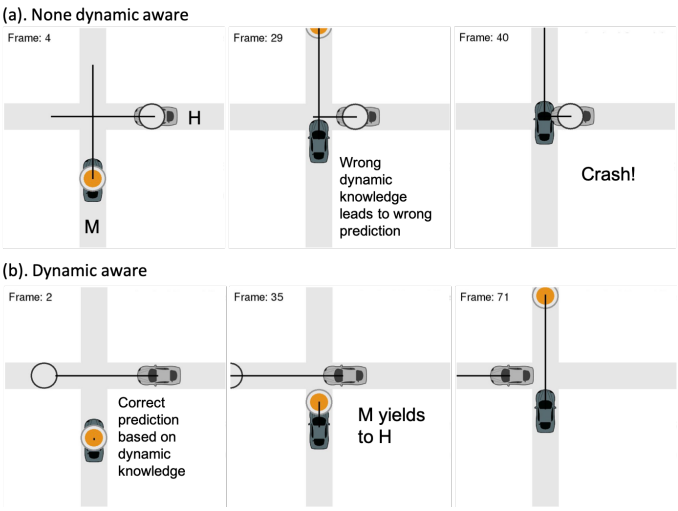


Fig. 2. A comparison between interactions with and without M's correct knowledge about H's physical properties: When M fails to incorporate H's slow acceleration and deceleration into its intent inference, it believes that H is able to stop as it approaches the intersection, thus reaching a wrong inference of H's intent and future actions.

In this paper, we model the action set as acceleration (or deceleration) rates, and allow agents to have different action sets. We consider the awareness of others' action set as the knowledge of vehicle dynamics. This knowledge is necessary for intent inference, since the computation of game equilibria relies on the payoff values of each agent, which are defined by the control objectives parameterized by the actions to be taken. We demonstrate through simulation studies that knowing the acceleration range of the other agent improves the inference of its intent.

(3) Courtesy in driving: The lack of social gracefulness of driving has recently been brought up as a potential issue of AVs in interactions with human drivers [4], [5]. Within a more focused scope of courteous driving, Sun et al. [22] defined courtesy as to minimize the inconvenience caused by M for H (see details in Sec. IV). While plausible, the authors assumed H's intent to be known by M. However, we found that this courtesy definition may create *irrationally* courteous behavior. An example is illustrated in Fig. 3: While M is much closer to the intersection, it chooses to yield to H if the courtesy loss is sufficiently weighted. This behavior of M is not expected by H, since it is sub-optimal for M (without considering the courtesy loss) under any future motion of H, i.e., even if H rushes through, moving forward is still a better choice for M than braking.

This paper investigates a new courtesy definition that better bounds the rationality of courteous behavior. We follow Sun et al. [22], where the courtesy loss is defined as the difference between H's actual payoff and its hypothetical payoff when M is fully collaborative. Differently, however, we require M to only collaborate using motions from its inferred equilibrium motion set, i.e., M collaborates rationally. Similar to [19], [22], [3], we allow M to proactively predict H's future motion as a reaction to M's, and combine this proactive control objective with the courtesy loss. Through an intersection case study, we show that a proactive M pretends to be aggressive and

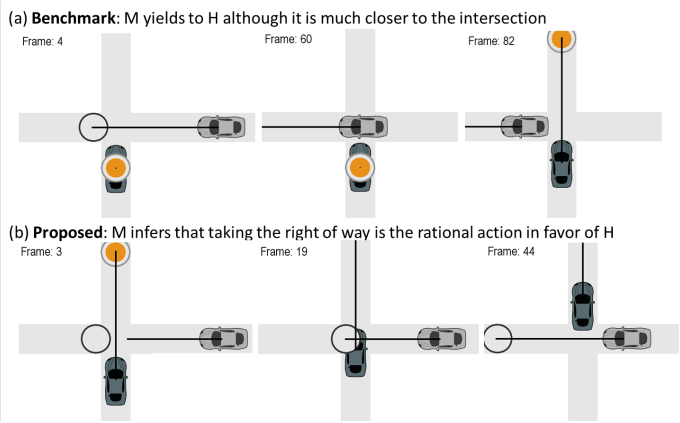


Fig. 3. A comparison between interactions (a) with an existing courtesy loss [22] and (b) the proposed loss

forces H to yield, whereas a courteous M only claims the right of way when doing so is inferred to be rational and in favor of H. We also demonstrate through a simulation that the developed courteous motion planning algorithm can be extended to interactions with three or more vehicles.

Below we summarize new contributions of this paper over the existing studies[3]: **Inference contents**: [14], [11], [8], [21] and this paper perform *inference* of both H's actions and intent, while [19], [22] assume H's intent to be known by M. Other studies (e.g., [7]) model M's controller as a black box that does not explicitly depend on H's intent. **Inference mechanism**: [11], [8], [19], [22], [21] and this paper make inference based on mechanistic models of driver decision making, e.g., model predictive control or bounded memory adaptation [11]. In parallel, researchers have studied data-driven approaches to intent and motion prediction, e.g., based on deep neural networks [27], [2], [14]. Although these work present promising solutions, purely data-driven approaches often require a large collection of labeled camera and Lidar data, and ignore the underlying Bayesian game-theoretic nature of interactions. These drawbacks of data-driven methods could lead to less verifiable generalization or explainability in predictions. Besides these supervised approaches, there have been notable efforts in learning to predict vehicle motion through imitation learning [10], [25], [1]. To the authors' knowledge, however, there have been few studies along this line that explicitly address the double-blindness challenge in inference or incorporate the driver's knowledge of the dynamics of surrounding vehicles. In addition, the control objectives or policies learned through imitation do not guarantee any social-adept driving performance. The present work, along with [8], is among the only few studies that explicitly address the issue of double-blind intents during vehicle interactions, and the presented work is the first to investigate this issue for non-cooperative games in intelligent vehicle settings. The presented work is also the first to highlight the importance of knowledge about vehicle dynamics in intent inference, and to investigate the rationality of courteous driving enabled by empathetic intent inference.

[3]It should be noted that the short survey here is by no means complete. Only representative and most relevant work are included for comparison.

A preliminary version of this paper has appeared in [18], and this paper significantly improves the width and depth of the study. In addition to [18], this paper (1) examines the effect of empathetic intent inference with rigorous statistical tests; (2) examines the effect of knowledge of vehicle dynamics; (3) proposes a more reasonable formulation of the courtesy loss (see Sec. III-C); (4) extends the proposed framework to a three-agent case study; and (5) uses a more realistic action space defined on vehicle acceleration rather than displacement. This paper also provides additional analyses of case studies that are not found in [18].

The rest of the paper is structured as follows: In Sec. II, we formulate the intent inference problem and introduce an empathetic inference algorithm. Courteous motion planning will be introduced in Sec. III, and case studies in Sec. IV. Sec. V discusses limitations of the proposed methods and future directions. Sec. VI concludes the paper.

II. Intent Inference

In this section, we will formulate the intent inference problem and propose an algorithm for a vehicle M to estimate the intent of the other vehicle H. The inference by M is based on the assumptions that H uses a *baseline* motion planning policy as introduced below, and that H also believes that M uses the same policy.

A. The interaction game and the baseline control policy

State and action: Denote the *state* and the *action* of agent $i \in \{H, M\}$ at time t as $\mathbf{s}_i(t)$ and $\mathbf{u}_i(t)$, respectively. In case studies (Sec. IV), the state of a vehicle is represented by its x- and y-coordinates ($\mathbf{s}_i(t) \in \mathbb{R}^2$), and the action by its acceleration rate ($\mathbf{u}_i(t) \in \mathbb{R}$). Vehicle *motion* is defined as a sequence of actions of a fixed length L starting from the current time step t: $\xi_i(t) = [\mathbf{u}_i(t), \mathbf{u}_i(t+1), \cdots, \mathbf{u}_i(t+L-1)]$. We allow i to have a finite set of candidate motions to choose from, and denote this set as Ξ_i. The dependency of variables on time t will be omitted for brevity when necessary. **Control objective**: For two agents i and j, the *instantaneous* loss of agent i at time t is denoted as $c(\xi_i; \xi_j, \theta_i, \mathbf{s}_i, \mathbf{s}_j, t)$, which is a function of i's future motion ξ_i, and parameterized by the other agent j's future motion ξ_j, i's intent θ_i, and the current states of both agents ($\mathbf{s}_i$ and $\mathbf{s}_j$). Specifically, the instantaneous loss is modeled as the weighted sum of the safety loss $c_{\text{safety}} \in \mathbb{R}$ and the task loss $c_{\text{task}} \in \mathbb{R}$:

$$c(\xi_i; \xi_j, \theta_i, \mathbf{s}_i, \mathbf{s}_j, t) = c_{\text{safety}}(\xi_i; \xi_j, \mathbf{s}_i, \mathbf{s}_j, t) + \theta_i c_{\text{task}}(\xi_i, \mathbf{s}_i, t). \tag{1}$$

For the intersection case, c_{safety} measures the future distances between i and j, and c_{task} penalizes motions that fail to move the ego agent across the intersection within L time steps (see Sec. IV for detailed definitions). The intent parameter $\theta_i \in \mathbb{R}$ represents the aggressiveness of the vehicle i. While intent can be defined as all parameters that jointly shape the control objective of an agent, this paper focuses on aggressiveness due to multiple practical reasons. First, aggressiveness represents the balance between the two representative losses of self goal and safety; second, introducing other intent parameters could further complicate the case studies, without significantly

strengthening the contribution of the paper. Note that we use "aggressiveness" to attribute motion planning, and "empathy" for intent inference. Therefore, an agent can be both "empathetic" and "aggressive".

Note that c_{safety} involves the distance between i and j, and thus requires ξ_i and ξ_j as inputs, the latter of which is unknown by i. The dependency of variables on $\mathbf{s}_i$ and $\mathbf{s}_j$ will be omitted for brevity.

The interaction game: The payoff for i in the game is defined by the *accumulative loss* $C(\xi_i;\xi_j,\theta_i) = \sum_{\tau=t}^{t+L-1} c(\xi_i;\xi_j,\theta_i,\tau)$ for feasible ξ_i. Denote $\hat{\theta}_j$ as i's estimation of j's intent. The game at time t yields a Nash equilibrium set $\mathcal{Q}(\theta_i,\hat{\theta}_j,t) = \{(\xi_i^*,\hat{\xi}_j^*)\}$, where $\hat{\xi}_j^*$ is i's estimation of j's planned motion. Each element of $\mathcal{Q}$ satisfies

$$\begin{aligned} \xi_i^* &= \underset{\{\xi_i\in\Xi_i\}}{\arg\min}\ C(\xi_i;\hat{\xi}_j^*,\theta_i), \\ \hat{\xi}_j^* &= \underset{\{\xi_j\in\Xi_j\}}{\arg\min}\ C(\xi_j;\xi_i^*,\hat{\theta}_j). \end{aligned} \tag{2}$$

As an example, at an intersection without a stop sign, either M or H yielding to the other will satisfy Eq. (2).

It should be noted that for i to infer the intent of j, i needs to put itself in j's shoes, i.e., to see the equilibrium set from j's perspective. Since j does not know θ_i, a necessary correction is to introduce $\tilde{\theta}_i$ and $\tilde{\xi}_i^*$ as i's inference of j's inference of i's intent and planned motion, respectively, and define $\mathcal{Q}(\hat{\theta}_j,\tilde{\theta}_i,t) = \{(\hat{\xi}_j^*,\tilde{\xi}_i^*)\}$ as the version of the equilibrium set that j could see according to i [4].

The baseline control policy: We model i to believe that j plans its motion by choosing *uniformly* from $\mathcal{Q}(\hat{\theta}_j,\tilde{\theta}_i,t)$, i.e., from i's perspective, j's motion follows the probability mass function:

$$p(\xi_j;\tilde{\theta}_i,\hat{\theta}_j,t) \propto |\{\xi_j;\xi_j\in(\xi_j,\xi_i) \text{ and } (\xi_j,\xi_i)\in\mathcal{Q}(\hat{\theta}_j,\tilde{\theta}_i,t)\}|. \tag{3}$$

We call this control policy a *baseline* to the variations to be introduced in Sec. III.

B. Inference of intent and motion

Eq. (4) formulates the inference problem. The key idea is to find combinations of intents of i and j such that the resulting motion of j with the highest probability mass at time $t-1$, denoted by $\xi_j^\dagger(t-1)$, should have its first action (denoted by $\hat{\mathbf{u}}_j^\dagger(t-1)$) match the observed action of j at time $t-1$ (denoted by $\mathbf{u}_j(t-1)$). In other words, what j could have done should match what it actually did.

$$\begin{aligned} \min_{\tilde{\theta}_i,\hat{\theta}_j}\quad & ||\hat{\mathbf{u}}_j^\dagger(t-1)-\mathbf{u}_j(t-1)||_2^2 \\ \text{subject to}\quad & \xi_j^\dagger(t-1) = \underset{\xi_j\in\Xi_j}{\arg\max}\, p(\xi_j;\tilde{\theta}_i,\hat{\theta}_j,t-1) \end{aligned} \tag{4}$$

In this paper, we use a finite intent set Θ for both agents to represent different levels of aggressiveness in driving. Since both Θ and Ξ are finite, the inference is made through an enumeration over the joint intent space $\Theta\times\Theta$. The outcome of the enumeration is a set $\mathcal{S}(t) = \left\{(\tilde{\theta}_i^*,\hat{\theta}_j^*)_k\right\}_{k=1}^K$, where each element is a global solution to Eq. (4). To quantify the uncertainty in inference and later incorporate it into motion planning, we assign equal probability mass $(1/K)$ to each of the solutions, based on which we can compute the empirical joint distribution $p(\tilde{\theta}_i,\hat{\theta}_j;t)$ defined on $\Theta\times\Theta$, and the marginals $p(\tilde{\theta}_i;t)$ and $p(\hat{\theta}_j;t)$ defined on Θ, by counting the appearances of all elements of Θ in $\mathcal{S}(t)$. Formally, these distributes are defined as

$$p(\tilde{\theta}_i,\hat{\theta}_j;t) \propto \begin{cases} 1/K, \text{ if } (\tilde{\theta}_i,\hat{\theta}_j)\in\mathcal{S}(t) \\ 0, \text{ otherwise} \end{cases} \tag{5}$$

and

$$\begin{aligned} p(\tilde{\theta}_i;t) &= \sum_{\hat{\theta}_j\in\Theta} p(\tilde{\theta}_i,\hat{\theta}_j;t), \\ p(\hat{\theta}_j;t) &= \sum_{\tilde{\theta}_i\in\Theta} p(\tilde{\theta}_i,\hat{\theta}_j;t). \end{aligned} \tag{6}$$

From the joint distribution $p(\tilde{\theta}_i,\hat{\theta}_j;t)$, i can infer j's planned motion and j's expectation of i's planned motion, as explained below.

a) Distribution of j's planned motion: Recall that each $(\tilde{\theta}_i^*,\hat{\theta}_j^*)\in\mathcal{S}(t)$ deduces a conditional distribution of j's motion starting from $t-1$ ($p(\xi_j;\tilde{\theta}^*,\hat{\theta}_j^*,t-1)$) through $\mathcal{Q}(\hat{\theta}_j^*,\tilde{\theta}_i^*,t-1)$. We can calculate the marginal $p(\xi_j;t-1)$ based on $p(\xi_j;\tilde{\theta}^*,\hat{\theta}_j^*,t-1)$ and $p(\tilde{\theta}_i,\hat{\theta}_j;t)$. Note that this is the distribution of j's motion at $t-1$ rather than t since we formulated the game at $t-1$. Although one can formulate a new game at t using $\mathbf{s}_i(t)$ and $\mathbf{s}_j(t)$ instead of $\mathbf{s}_i(t-1)$ and $\mathbf{s}_j(t-1)$ to derive $p(\xi_j;t)$, in this paper we will approximate $p(\xi_j;t)$ using $p(\xi_j;t-1)$ to simplify the computation.

b) Distribution of j's expectation of i's planned motion: Similarly, we can also deduce from $\mathcal{S}(t)$ a conditional distribution of i's motion starting from t, which is denoted by $p(\tilde{\xi}_i;\tilde{\theta}_i^*,\hat{\theta}_j^*,t)$, and the marginal $p(\tilde{\xi}_i;t)$. We shall emphasize that $p(\tilde{\xi}_i;t)$ is not the distribution of motion that i will follow, but only represents i's understanding of what j expects i to do.

C. Leveraging past observations

Note that if we consider agents' intent to be time invariant during the interaction, all previous observations can be leveraged for the inference, leading to the following modified problem:

$$\begin{aligned} \min_{\{\tilde{\theta}_i(\tau)\}_{\tau=1}^{t-1},\hat{\theta}_j}\quad & \sum_{\tau=1}^{t-1} ||\hat{\mathbf{u}}_j^\dagger(\tau)-\mathbf{u}_j(\tau)||_2^2 \\ \text{subject to}\quad & \xi_j^\dagger(\tau) = \underset{\xi_j\in\Xi_j}{\arg\max}\, p(\xi_j;\tilde{\theta}_i(\tau),\hat{\theta}_j,\tau) \\ & \text{for } \tau\in\{1,\cdots,t-1\}. \end{aligned} \tag{7}$$

Here we allow $\tilde{\theta}_i$ to freely change along time since j may change its mind about i, while keeping $\hat{\theta}_j$ fixed for all time steps. It is possible to impose a structure on $\{\tilde{\theta}_i^\tau\}_{\tau=1}^{t-1}$ so that $\tilde{\theta}_i$ changes gradually, i.e., j does not change its mind about i abruptly. We will leave this exploration to future studies.

An important insight is that solutions to Eq. (7), denoted by $\bar{\mathcal{S}}(t)$, can be found in a recursive way based on solutions to Eq. (4). To explain, consider an intent candidate

[4]Note that we designate the first argument of $\mathcal{Q}(\cdot,\cdot,t)$ to the ego agent, thus i and j are flipped when we put i in j's shoes.

$\hat{\theta}_j$ that exists in both $\bar{\mathcal{S}}(t-1)$ and $\mathcal{S}(t)$. Let solutions in $\bar{\mathcal{S}}(t-1)$ that contain $\hat{\theta}_j$ be in the form of $(\mathbf{a}, \hat{\theta}_j)$, where $\mathbf{a} = [a_1, \cdots, a_{t-1}] \in \mathcal{A} \subset \Theta^{t-1}$ is *a time series* of intents of i. Similarly, let solutions in $\mathcal{S}(t)$ that contain $\hat{\theta}_j$ be in the form of $(b, \hat{\theta}_j)$, where $b \in \mathcal{B}$ is an intent of i. Also let the operation of appending b to the array $\mathbf{a}$ be $[\mathbf{a}, b]$. Then $([\mathbf{a}, b], \hat{\theta}_j)$ for all $\mathbf{a} \in \mathcal{A}$ and $b \in \mathcal{B}$ is a solution to Eq. (7) at t. Following this property, we have $\bar{p}(\hat{\theta}_j; t) \propto \bar{p}(\hat{\theta}_j; t-1) p(\hat{\theta}_j; t)$. Note that for any candidate $\hat{\theta}_j$ absent in $\bar{\mathcal{S}}(t-1)$ or $\mathcal{S}(t)$, it also does not exist in $\bar{\mathcal{S}}(t)$ and its corresponding probability mass in $\bar{p}(\hat{\theta}_j; t)$ is zero. The update of $\bar{p}(\hat{\theta}_j; t)$ will trigger that of the joint probability $p(\tilde{\theta}_i, \hat{\theta}_j; t)$ and the marginal $p(\hat{\xi}_j; t)$. Their updated counterparts are denoted by $\bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t)$ and $\bar{p}(\hat{\xi}_j; t)$, respectively. Specifically, these distributions are updated as

$$\begin{aligned} \bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t) &= p(\tilde{\theta}_i, \hat{\theta}_j; t) \frac{\bar{p}(\hat{\theta}_j; t)}{p(\hat{\theta}_j; t)} \\ \bar{p}(\hat{\xi}_j; t) &= \sum_{(\tilde{\theta}_i, \hat{\theta}_j) \in \Theta \times \Theta} p(\xi_j; \tilde{\theta}_i, \hat{\theta}_j, t) \bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t) \end{aligned} \tag{8}$$

In the following, we will assume that the true intent of both agents do not change during the interaction, and thus keep using the notation $\bar{p}$ to emphasize that the probability mass distributions are updated based on all past observations.

The intent inference algorithm is summarized in Alg. 1.

Algorithm 1: Algorithm for inferring j's intent at time t

input : states $\mathbf{s}_i(t-1)$ and $\mathbf{s}_j(t-1)$, observed action of j $\mathbf{u}_j(t-1)$, past inference results $\bar{p}(\hat{\theta}_j; t-1)$

output: solutions to Eq. (4) $\bar{\mathcal{S}}(t)$, joint intent pmf $\bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t)$, marginal intent pmf $\bar{p}(\hat{\theta}_j; t)$, action pmf $\bar{p}(\hat{\xi}_j; t)$

for $(\tilde{\theta}_i, \hat{\theta}_j) \in \Theta \times \Theta$ **do**
  Compute $\mathcal{Q}(\theta_i, \hat{\theta}_j, t-1)$ and $p(\xi_j; \tilde{\theta}_i, \hat{\theta}_j, t-1)$;
  Find $\xi_j^\dagger(t-1) = \arg\max_{\xi_j \in \Xi_j} p(\xi_j; \tilde{\theta}_i, \hat{\theta}_j, t-1)$;
  Compute $d(\tilde{\theta}_i, \hat{\theta}_j) = ||\hat{\mathbf{u}}_j^\dagger(t-1) - \mathbf{u}_j(t-1)||_2^2$;
end
Find $\mathcal{S}(t) := \{(\tilde{\theta}_i^*, \hat{\theta}_j^*)\}$ for all $(\tilde{\theta}_i^*, \hat{\theta}_j^*) = \arg\min_{(\tilde{\theta}_i, \hat{\theta}_j) \in \Theta \times \Theta} d(\tilde{\theta}_i, \hat{\theta}_j)$;
Compute $p(\tilde{\theta}_i, \hat{\theta}_j; t)$ using Eq. (5);
Compute $p(\hat{\theta}_j; t)$ using Eq. (6);
Compute $\bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t)$, $\bar{p}(\hat{\theta}_j; t)$, and $\bar{p}(\hat{\xi}_j; t)$ using Eq. (8);

III. Motion Planning

We now introduce three motion planning formulations (control policies) that incorporate the inference outcomes from Alg. 1, namely, reactive, proactive, and courteous planning. These policies are different from the baseline (Sec.II-A) in that the latter uses point estimates of intents (without uncertainty information), and is indifferent towards differences in ego payoffs across equilibria. The introduction of these policies thus creates an inevitable inconsistency between intent inference and motion planning, i.e., the policy used by H is different from what M believes what H uses. This leads to the question of whether an agent should infer control policies of other agents (and others' inference of their own policy) in addition to their intents. In this paper, however, we restrict the study by fixing agents' beliefs on others' policies. Extensions to a higher-level inference problem will be discussed in Sec. V.

A. Reactive motion

Given the distribution of j's future motions $\bar{p}(\hat{\xi}_j; t)$, a reactive agent plans its motion by minimizing the expected loss within a time window:

$$\min_{\xi_i \in \Xi_i} \quad C_i^{\text{reactive}}(\xi_i) := \mathbb{E}_{\hat{\xi}_j \sim \bar{p}(\hat{\xi}_j; t)} \left[C(\xi_i; \hat{\xi}_j, \theta_i) \right]. \tag{9}$$

Since Ξ_i is a finite set, we resort to an enumeration to solve Eq. (9). The same applies to proactive and courteous agents.

B. Proactive motion

A proactive agent i plans by taking into consideration the dependency of j's planning on i's next action. Specifically, i calculates the conditional distribution $\bar{p}(\hat{\xi}_j; \xi_i, t)$ based on ξ_i and $\bar{p}(\hat{\theta}_j; t)$, assuming that j will quickly respond to ξ_i. To do so, i first finds the set of optimal motions of j for every $\hat{\theta}_j \in \Theta$ given ξ_i. This set is denoted by $\mathcal{Q}_j(\xi_i) = \cup_{\hat{\theta}_j \in \Theta} \mathcal{Q}_j(\xi_i, \hat{\theta}_j)$ where $\mathcal{Q}_j(\xi_i, \hat{\theta}_j) = \{\hat{\xi}_j^*; \hat{\xi}_j^* = \arg\min_{\xi_j \in \Xi_j} C(\xi_j, \xi_i, \hat{\theta}_j)\}$. Then for each element $\hat{\xi}_j^* \in \mathcal{Q}_j(\xi_i)$, we can compute

$$\bar{p}(\hat{\xi}_j^*; \xi_i, t) = \sum_{\hat{\theta}_j \in \Theta} \frac{\bar{p}(\hat{\theta}_j; t) 1(\hat{\xi}_j^* \in \mathcal{Q}_j(\xi_i, \hat{\theta}_j))}{|\mathcal{Q}_j(\xi_i, \hat{\theta}_j)|}, \tag{10}$$

where $1(\cdot)$ is an indicator function. For $\hat{\xi}_j \in \Xi_j / \mathcal{Q}_j(\xi_i)$, we set $\bar{p}(\hat{\xi}_j; \xi_i, t) = 0$. We can now formulate the motion planning problem for a proactive agent:

$$\min_{\xi_i \in \Xi_i} \quad C_i^{\text{proactive}}(\xi_i) := \mathbb{E}_{\hat{\xi}_j \sim \bar{p}(\hat{\xi}_j; \xi_i, t)} \left[C(\xi_i, \hat{\xi}_j, \theta_i) \right]$$

C. Courteous motion

We start by defining a *rational courteous motion* of i, denoted by ξ_i^j, as one that belongs to the equilibrium set $\mathcal{Q}(\hat{\theta}_j, \tilde{\theta}_i, t)$ and is in favor of j. Formally, the set of all rational courteous motions, $\{\xi_i^j\}$, can be derived from the following set of motion pairs:

$$\mathcal{Q}^j(\hat{\theta}_j, \tilde{\theta}_i, t) = \{(\xi_i^j, \xi_j) | (\xi_i^j, \xi_j) = \underset{(\xi_i, \xi_j) \in \mathcal{Q}(\hat{\theta}_j, \tilde{\theta}_i, t)}{\arg\min} C_j\}. \tag{11}$$

Since $\hat{\theta}_j$ and $\tilde{\theta}_i$ are uncertain, we can calculate the conditional probability $p(\xi_i^j; \tilde{\theta}_i, \hat{\theta}_j, t)$ and the marginal $\bar{p}(\xi_i^j; t)$ using $\mathcal{Q}^j(\hat{\theta}_j, \tilde{\theta}_i, t)$ and $\bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t)$:

$$\begin{aligned} p(\xi_i^j; \tilde{\theta}_i, \hat{\theta}_j, t) &\propto |\{\xi_j; \xi_j \in (\xi_i^j, \xi_j) \text{ and } (\xi_i^j, \xi_j) \in \mathcal{Q}^j(\hat{\theta}_j, \tilde{\theta}_i, t)\}| \\ \bar{p}(\xi_i^j; t) &= \sum_{(\tilde{\theta}_i, \hat{\theta}_j) \in \Theta \times \Theta} p(\xi_i^j; \tilde{\theta}_i, \hat{\theta}_j, t) \bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t). \end{aligned} \tag{12}$$

Essentially, $\bar{p}(\xi_i^j; t)$ defines the distribution of motions that i believes would be in favor of j.

Following [22], the courtesy loss of i is defined as the non-negative expected difference between the actual control objective value of j (C_j) and a best-case objective (C_j^{best}):

$$L_i^{\text{courtesy}}(\xi_i) := \mathbb{E}_{\tilde{\theta}_i,\hat{\theta}_j\sim p(\tilde{\theta}_i,\hat{\theta}_j;t)}\left[\max\{0, C_j(\xi_j;\xi_i,\hat{\theta}_j) - C_j^{\text{best}}(\tilde{\theta}_i,\hat{\theta}_j)\}\right], \tag{13}$$

where the best-case value is defined as j's minimal expected control objective value when i picks a *rational* courteous motion from which j benefits the most.

$$C_j^{\text{best}}(\tilde{\theta}_i,\hat{\theta}_j) = \min_{\xi_i\in\mathcal{Q}^j(\hat{\theta}_j,\tilde{\theta}_i,t)} \mathbb{E}_{\xi_j\sim p(\xi_j;\tilde{\theta}_i,\hat{\theta}_j,t)}\left[C_j(\xi_j;\xi_i,\hat{\theta}_j)\right]. \tag{14}$$

As mentioned in Sec. II-B, $p(\xi_j;\tilde{\theta}_i,\hat{\theta}_j,t)$ in Eq. (14) is approximated by $p(\xi_j;\tilde{\theta}_i,\hat{\theta}_j,t-1)$.

A courteous agent solves the following problem where the courtesy loss is added to the proactive control objective:

$$\min_{\xi_i\in\Xi_i} \quad C_i^{\text{courtesy}}(\xi_i) := C_i^{\text{proactive}}(\xi_i) + \beta L_i^{\text{courtesy}}(\xi_i), \tag{15}$$

The weight $\beta \geq 0$ tunes the importance of courtesy in the objective. Compared to the courtesy definitions in [22], Eq. 15 incorporates the uncertainty of j's intent and future actions, and constrains the calculation of j's best-case value using rational motions of i. Therefore, i does not go out of its way to help j.

IV. Case studies

We present three case studies to show the importance of empathy and knowledge about vehicle dynamics in intent inference, and the merit of the proposed courtesy definition in motion planning. All studies are based on an interaction between M and H at a four-way intersection.

A. Simulation setup

The setup of the intersection case is illustrated in Fig. 1, where M moves up (along y-axis) and H moves left (along x-axis). The intent candidate set is set to $\Theta = \{1.0, 10^3\}$. Only for 10^3, the agent can tolerate contacts with the other agent.

Motion representation: To avoid high-dimensional planning problems within intent inference, we introduce a surrogate action a_i as a scalar that determines the initial acceleration of i for a finite horizon $L = 100$. The surrogate action set is $\mathcal{A}_i = \{-2.0, -1.0, 0.0, 1.0, 2.0, 3.0\} \times \alpha_i$, where the scalar α_i represents the acceleration ability of i. The unit of the action is m/δ^2 where a time step $\delta = 1/20$ seconds. We assume that ξ_i follows a linear function parameterized by a_i and satisfies the following constraints: (1) the initial acceleration at time t is $u(t) = a_i\alpha_i$; (2) the final acceleration at time $t + L - 1$ is zero: $u(t + L - 1) = 0$. These assumptions lead to the following computation of action $u(k)$ within ξ_i for $k = \{t, \cdots, t + L - 1\}$: $u(k) = a_i\alpha_i\left(1 - (t + L - 1)^{-1}k\right)$. Note that since agents move in straight lines, we use u to represent the *magnitude* of the acceleration. Lastly, for the first time step, we assume that both agents have observed $u = 0$ (constant speed) from each other.

Losses: c_{task} penalizes the agent if it fails to move across the intersection within L steps. Taking M as an example, the loss is defined as $c_{\text{M,task}} = \exp(-s_M^{(x)}(t+L-1)+0.4)$, where $s_M^{(x)}$ is the state of M in the x direction. The safety loss is defined as $c_{\text{safety}} = \exp\left(\gamma(-D+\phi)\right)$, where $D = ||\mathbf{s}_M - \mathbf{s}_H||_2^2$ when both cars are in the interaction area Ω (see Fig. 1), $D = \infty$. The interaction area is introduced so that the instantaneous safety loss will be zero at any time step (in the time window) as long as one vehicle does not enter Ω. This treatment of the safety loss is necessary: Without considering an interaction area, M will choose not to move across the intersection even if H stops constantly outside of the intersection, because there would be a high loss due to close distance to H had M passed by. γ is empirically set to 5.0 so that the safety penalty increases significantly as the two cars approach each other; ϕ is empirically set to $0.13w^2$ where $w = 4.5$ (unit: meters) is the length of the car. This setting creates a safe zone for an agent since the penalty quickly approaches zero when $D > b$. For the courteous agent, β is set to 0.1.

Implementation issues: It is worth mentioning that since the agents do not follow the baseline control policy, their motions do not necessarily belong to an equilibrium set. This leads to occasions where the probability mass of the true intent of j becomes zero as inferred by i, or where the updated probability distribution, e.g., $\bar{p}(\hat{\theta}_j;t)$, has all zero entries as every candidate intent value has been eliminated during the inference process. In the implementation, we set $\bar{p}(\hat{\theta}_j;t)$ back to the uniform distribution when the latter happens.

B. Intent inference with baseline motion

The intent inference we introduced depends on the assumption that both agents follow the baseline control policy. Therefore it is necessary to verify the performance of the intent inference method when applied to interactions between baseline agents.

Experiment Setup: We conduct a set of experiments where both M and H are modeled to follow the baseline policy, an assumption used by the inference algorithm (see Sec. II-A). The results are summarized in (Tab. I), with four combinations of aggressiveness of M and H being the rows, and empathetic v.s. non-empathetic the columns. For each element in the table, we track the probability $p(\hat{\theta}_H = \theta_H)$ and average it over an interaction period of 100 time steps. Due to the probabilistic nature of the baseline policy, we report the mean and standard deviations of the averaged probabilities over 50 independent interactions. A paired sample t-test is performed to evaluate the difference between the accuracy of empathetic and non-empathetic inferences, and the p-value is reported in the last column.

Summary: A few remarks on the results are as follows: (1) The averaged probabilities do not approach 1 since the inference at the beginning of the interaction is often wrong. (2) Statistical tests on the experimental results show that the effect of empathy is most significant when both agents are non-aggressive, and diminishes when either agent becomes aggressive. These observations are consistent with intuition: First, an aggressive H often takes actions (e.g., rushing through the intersection) that can only be explained by its aggressiveness, irrespective to its inference of M's intent ($\tilde{\theta}_M$); second, an aggressive M acts in a legible (yet non-courteous) way

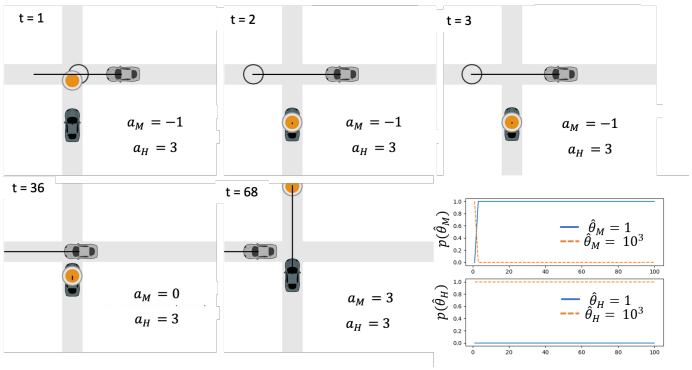


Fig. 4. Snapshots of the interaction between empathetic AV and the human drive vehicle.

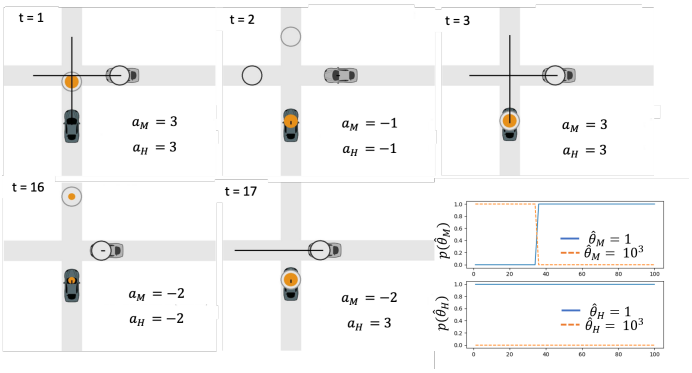


Fig. 5. Snapshots of the interaction between non-empathetic AV and the human drive vehicle.

that forces a non-aggressive H to yield, which reveals its non-aggressiveness. (3) Lastly, we should clarify that the presented algorithm allows $\tilde{\theta}_M$ to freely change over time, and therefore its inference only depends on the latest observation rather than the state-action history. This makes the inference of $\tilde{\theta}_M$ often incorrect. However, the contribution of empathy is in allowing $\tilde{\theta}_M$ to be different from the actual value θ_M, which opens up the hypothesis space of $\hat{\theta}_H$, i.e., certain values of $\hat{\theta}_H$ are only probable when we allow H to have wrong understanding of M.

TABLE I
MEAN (STANDARD DEVIATION) OF INTENT INFERENCE ACCURACY $p(\hat{\theta}_H = \theta_H)$ WITH AND WITHOUT EMPATHY

θ_M	θ_H	with empathy	without empathy	p-value
1	1	73.04% (15.84%)	53.38% (20.15%)	0.0343
1	10^3	83.74% (8.49%)	73.62% (7.60%)	0.1111
10^3	1	81.86% (12.34%)	85.16% (10.26%)	0.6203
10^3	10^3	81.00% (8.87%)	74.60% (12.24%)	0.3854

C. The importance of empathy in intent inference

We now investigate the importance of empathy in intent inference. We show that if M is modeled to believe that H already knows its true intent, then M will incorrectly infer H's intent.

Experiment setup: Both agents are set to be reactive, with $\theta_M = 1$ and $\theta_H = 10^3$. We set H to be more aggressive than M, so that when M incorrectly infers H's intent (as non-aggressive), it will fail to realize that H will not yield. In the non-empathetic case, we fix $\tilde{\theta}_M = 1$ for M's inference processes, and keep H as an empathetic agent.

Summary: Fig. 4 and Fig. 5 show interactions for the empathetic and the non-empathetic cases, respectively. We observe that M in the non-empathetic case moves forward until it is compelled to yield at $t = 17$. In the empathetic case, this abrupt change of motion does not happen. Instead, M realizes the aggressiveness of H earlier, and chooses to stop before it enters the interaction zone.

Analysis: Details of the interactions are explained below. We start with the non-empathetic case. At the first time step, as both vehicles observe $a = 0$, the equilibrium sets can be derived as in Tab. II. From H's perspective, $a_M = 0$ leads to $\hat{\theta}_M = 10^3$ following Alg. 1 (essentially, between $\hat{\theta}_M = 1$ and $\hat{\theta}_M = 10^3$, the latter has equiliria, $a_M = 3$ or $a_M = 0$, with a smaller average discrepancy to $a_M = 0$). Similarly, from M's perspective, fixing $\tilde{\theta}_M$ to 1 leads to $\hat{\theta}_H = 1$. Based on this initial inference and their states, both vehicles take $a = 3$ at $t = 1$. These actions are observed by each other at $t = 2$. Both agents keep their inference since $a = 3$ can be explained by their current beliefs of each others' intents, i.e., there exists an equilibrium for which $a_H = 3$ and $\hat{\theta}_H = 1$, and another for which $a_M = 3$ and $\hat{\theta}_M = 10^3$. Therefore the predicted actions are $\hat{a}_M = \hat{a}_H = 3$, causing both agents to brake with $a(t = 2) = -1$. This iteration between acceleration and braking continues until $t = 16$ when both agents choose to brake with $a = -2$. During this period, M consistently has the wrong inference of H due to its fixation on $\tilde{\theta}_M = 1$, and causes it to approach the interaction zone with a high speed.

In the empathetic case, the difference is that M starts by considering H as aggressive without the fixation. This causes M to immediately slow down, which also allows H to quickly correct its inference of M.

TABLE II
EQUILIBRIUM SETS AT $t = 1$, REACTIVE M VS. REACTIVE H, $\theta_M = 1$, $\theta_H = 10^3$

θ_M	θ_H	$\{a_M(t = 1), a_H(t = 1)\}$
1	1	{3, -1} or {-1, 3}
1	10^3	{-1, 3}
10^3	1	{3, -1}
10^3	10^3	{3, 0} or{0, 3}

D. The influence of knowledge about vehicle dynamics on intent inference

Experiment setup: To show the importance of knowledge about vehicle dynamics, we now model M and H to have different acceleration ability, specifically, $\alpha_M = 0.002$ and $\alpha_H = 0.0002$. Both M and H are reactive, with $\theta_H = \theta_M = 1$ to avoid confounding effects. For the case where M has incorrect knowledge about H's dynamics, we set $\hat{\alpha}_H = 0.002$ during M's intent inference session.

Observation: We use Fig. 6 and Fig. 7 to compare the outcomes of correct and incorrect $\hat{\alpha}_H$. Fig. 6 demonstrates the interaction when α_M and α_H are known to both agents: M realizes that H will not be able to stop due to its low deceleration rate, and thus chooses to yield. When M has the incorrect knowledge about $\hat{\alpha}_H$, which is shown in Fig. 7, M fails to predict H's future motion, which causes a collision. This simple example shows that vehicle dynamics needs to be carefully considered in intent inference. It is true, however, that

different aspects of dynamics may become relevant to other interaction settings. For example, in a winding road where M needs to pass H, M may need to understand stability properties of H, such as its height of the center of gravity, to predict H's driving style.

Analysis: Tab. III summarizes the equilibrium set at $t = 1$ when α_M and α_H are known to both agents. At the first time step, as M observes $a_H = 0$, it thinks that $\hat{\theta}_H = 10^3$ follows the same reasoning process we presented in the previous subsection. On the other hand, $a_M = 0$ in this setting does not offer H a clue about M's intent. Since it is difficult for H to stop in time due to its low deceleration, its best action through the whole interaction is to move forward as fast as possible ($a_H = 3$). At the same time, M will hold on to the believe that H is aggressive since it explains the observations of $a_H = 3$. Therefore M always chooses to yield to H until H passes the intersection zone. It is worth noting that even if M believes H to be non-aggressive, it will still yield since it understands that braking is not an option for H.

TABLE III
EQUILIBRIUM SETS AT $t = 1$, REACTIVE M VS. REACTIVE H, $\theta_M = 1$, $\theta_H = 1$, $\alpha_M = 0.002$, $\alpha_H = 0.0002$

θ_M	θ_H	$\{a_M(t=1), a_H(t=1)\}$
1	1	{-1, 3} or {3, -2}
1	10^3	{-1, 3}
10^3	1	{-1, 3} or {3, -2}
10^3	10^3	{3, -1} or{-1, 3}

Now we turn to the case where M does not acquire the correct knowledge about α_H. Specifically, M observes H's action as $a_H^M = u_H/\alpha_H$. With $\hat{\alpha}_H = 10\alpha_H$, the observed action of H is 10 times smaller than their actual values. This structured bias causes a series of incorrect inference during the interaction, as we explain below. At $t = 1$, the equilibrium set from M's perspective follows Tab. II instead of Tab. III. With the initial observation of $a_H = 0$, M infers that H is aggressive ($\hat{\theta}_H = 10^3$) and that it will keep its speed ($\hat{a}_H = 0$). As a result, M chooses to brake: $a_M(t = 1) = -1$. At $t = 4$, the slow acceleration of H makes M believe that H will brake, since between the equilibrial choices of $a_H = -1$ and $a_H = 3$, its observed action, $a_H^M = 0.3$, is closer to the former than the latter. Based on this misunderstanding, M starts to accelerate, which in turn, causes H to slow down, which further makes M to believe that H is non-aggressive at $t = 9$. Tab. IV shows the Nash equilibrium sets at $t = 9$ from M's perspective. M continues to move with $a_M = 3$, while believing that H will be able to stop before the intersection zone with $a_H = -2$ as the black circle indicates. However, due to its weak braking ability, H is not able to avoid the collision at $t = 37$.

TABLE IV
EQUILIBRIUM SETS FROM M'S PERSPECTIVE AT $t = 9$, REACTIVE M VS. REACTIVE H, $\theta_M = 1$, $\theta_H = 1$, $\alpha_M = 0.002$, $\alpha_H = 0.002$

θ_M	θ_H	$\{a_M(t=9), a_H(t=9)\}$
1	1	{-2, 3} or {3, -1}
1	10^3	{-2, 3}
10^3	1	{-2, 3} or {3, -1}
10^3	10^3	{-2, 3}

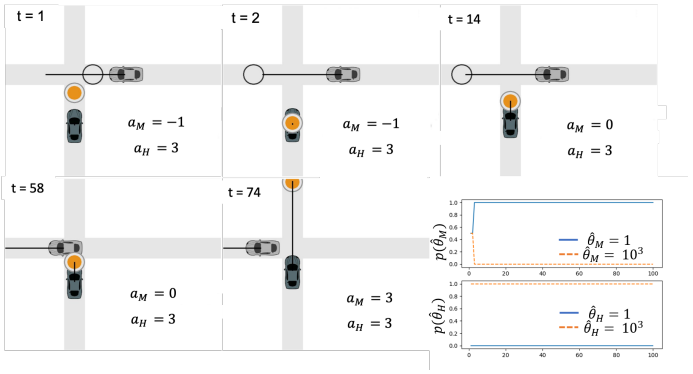


Fig. 6. Snapshots of the interaction between reactive M vs. reactive H with different ability ($\alpha_H = 0.0002$, $\alpha_M = 0.002$), and the change in $\tilde{\theta}_M$ by H and $\tilde{\theta}_H$ by M.

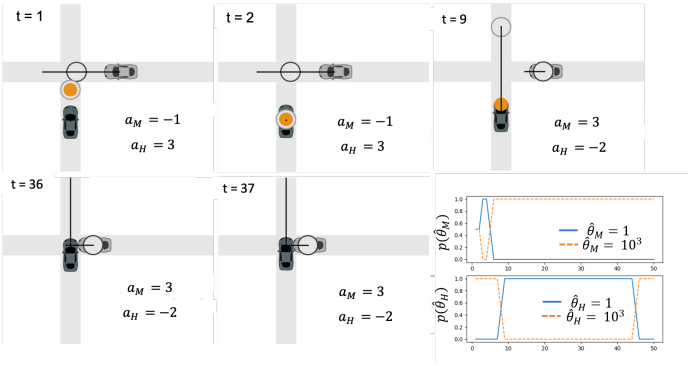


Fig. 7. Snapshots of the interaction between reactive M vs. reactive H with different ability ($\alpha_H = 0.0002$, $\alpha_M = 0.002$), while M thinks H's ability is 0.002.

E. The merit of the proposed courtesy definition

Here we reuse the intersection case to illustrate the merit of the proposed courtesy definition, in comparison with [22]. We show that our definition leads to more rational behavior of M.

Experiment setup: To implement [22] as a benchmark, we choose the best-case objective value of H to follow "alternative II" in the paper where M's motion planning were only to help H[5]. Specifically, the benchmark implementation uses:

$$C_j^{\text{best,bench}}(\tilde{\theta}_i, \hat{\theta}_j) = \min_{\xi_i \in \Xi_i} \mathbb{E}_{\xi_j \sim p(\xi_j; \tilde{\theta}_i, \hat{\theta}_j, t)} \left[C_j(\xi_j; \xi_i, \hat{\theta}_j) \right], \tag{16}$$

and follows Eq. (13) and Eq. (15) to form its control objective. Note that Eq. (16) is only different from Eq. (14) in the scope of ξ_i (full set of feasible motions as opposed to rational motions).

We use two cases to show the difference between the two courtesy definitions. The first setting is the same as before, where M and H starts at the same distance from the interaction zone. In the second setting, we set the initial states of M and H to $\mathbf{s}_M(0) = (0.0, -1.2)$ and $\mathbf{s}_H(0) = (3.0, 0.0)$, respectively.

[5]Note that in [22], the alternative losses are called "baseline". Here we use the term "best-case" to avoid confusion with the baseline control policy introduced in Sec. II-A

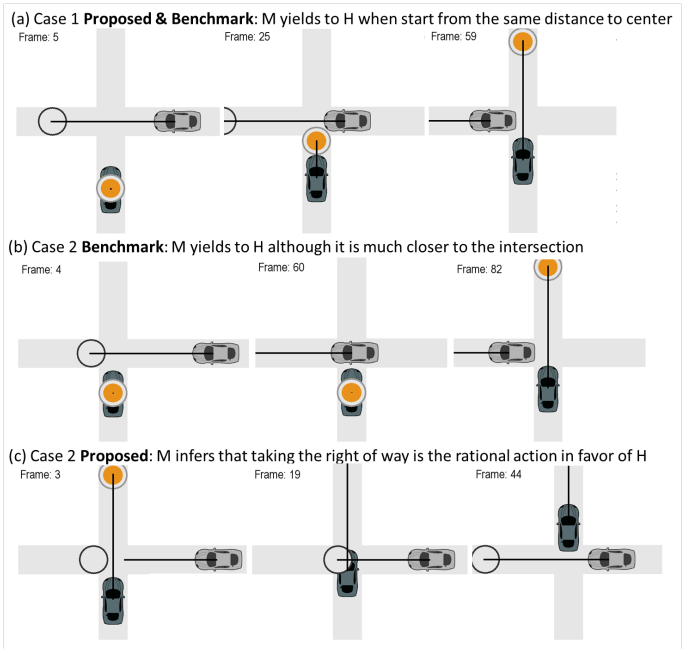


Fig. 8. Comparison between the proposed courtesy motion and the benchmark: (a) when M and H start from the same distance to the center, both produce the same courteous interactions; (b) when M starts much closer to the center than H, but still has enough space to yield, it takes a full brake using the benchmark courtesy loss; (c) under the same setting as (b) and using the proposed courtesy loss, M chooses to rush through, since this is the most courteous choice within its perceived equilibrium motion set.

This makes M closer to the interaction zone than H, yet the distance (1.2) is enough for M to take a full brake and stop outside of the intersection zone, if it plans to yield to H. We set $\theta_M = \theta_H = 1$, $\alpha_M = \alpha_H = 0.002$, $\beta = 10$ to encourage courteous behavior, and H as reactive.

Analysis: Fig. 8 compares the interactions from the two courtesy definitions and from the two cases. In the first case, M chooses to yield with both definitions. This verifies the correctness of the implementation. In the second case, and with the benchmark courtesy, M takes a full brake, since doing so would allow H to pass with minimal loss (thus minimizing L^{courtesy}). We argue that this behavior of M, however, is irrational. More concretely, full brake does not belong to the equilibrium set of motion pairs that M perceives based on its current inference of H. In fact, even if H chooses to rush through (by taking $a_H = 3$), M would still have a better payoff should it also choose to accelerate ($a_M = 3$) due to its shorter distance to the center. *Importantly, according to M, this knowledge is known by H*, and therefore M should believe that H will not expect M to take a full brake, but rather to rush through.

F. Study on the courtesy factor

In this subsection, we investigate the effect on AV's behavior by the courtesy factor β.

Experiment Setup: We set up three cases where β is set to 0, 1, or 10. M follows the courteous motion introduced in Sec. III-C. H is reactive. Both agents are non-aggressive: $\theta_M = \theta_M = 1$. The acceleration abilities are equal: $\alpha_M = \alpha_H = 0.002$.

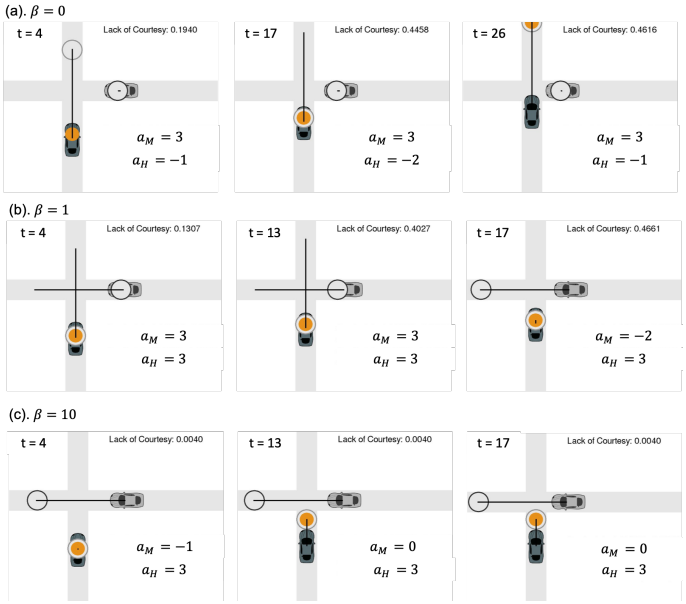


Fig. 9. A comparison between interactions (a) the courtesy loss weight $\beta = 0$, (b) the courtesy loss weight $\beta = 1$, (c) the courtesy loss weight $\beta = 10$

Observation: When $\beta = 0$ (Fig. 9(a)), M becomes proactive, i.e., it pretends to be aggressive, persuading H to believe that $p(\hat{\theta}_M = 10^3) = 1$ and to yield to M. When $\beta = 1$ (Fig. 9(b)), M tries to persuade H to yield until it has to yield at $t = 17$. Although M shows some courtesy by letting H cross first, it does not follow what H would like it to do. When $\beta = 10$ (Fig. 9(c)), M strictly follows the motion M believes to be favored by H.

Summary: By increasing the value of β, M behaves more courteously by taking actions closer to what is in favor of H; on the other hand, when β decreases, M behaves closer to a proactive agent.

G. Extension of knowledge and courtesy

Lastly, we demonstrate a case where we extend the interaction to three agents, as shown in Fig. 10. This case is adjusted from an accident involving an autonomous vehicle in Tempe, Arizona [15], and is set at an intersection where the vertical lanes have green lights but left-turn lanes have no privilege. H1, as a human-driven vehicle, attempts to turn left; while H2, the other human-driven vehicle, attempts to rush through the intersection. Neither of the cars realizes the existence of each other, since the view is blocked by M, an AV that is waiting on the left turn lane with no cars behind it. Ideally, M should realize the potential danger as it can observe both H1 and H2, and unblock the sight by slightly moving backwards. We show that this can be achieved by the proposed intent inference and motion planning framework, provided that M identifies the correct inference task and courtesy objective.

Specifically, we set the motion planning objective of M to be the sum of those of H1 and H2, and extend the inference to the visibility of H1 and H2 from each other. For simplicity, both agents are set as non-aggressive, and this knowledge is assumed to be known by all. Under this setting, M's inference enumerates over two possibilities: When H1 and H2 can see

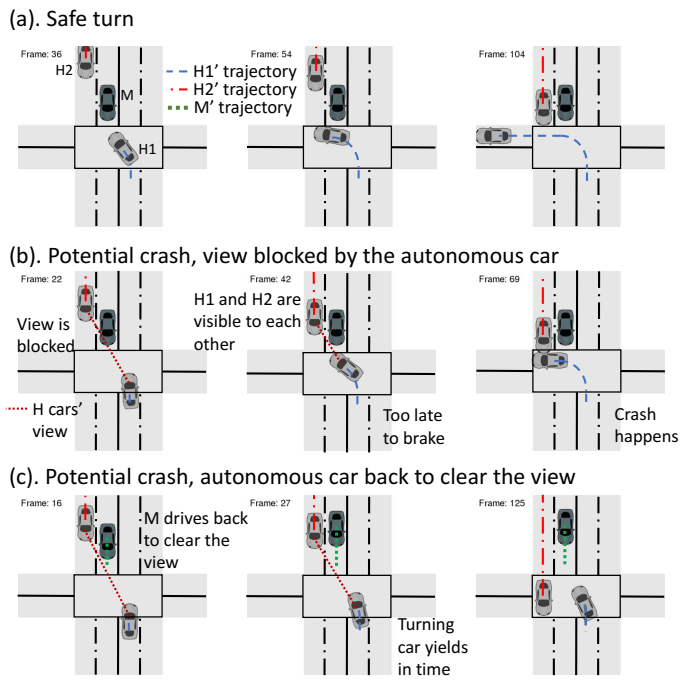


Fig. 10. Demonstration of an extension of knowledge and courtesy in multi-vehicle interactions: (a) When H1 and H2 have no potential collision, M does not move; (b) When H1 and H2 have potential collision while M is non-courteous, the collision happens; (c) M identifies the potential collision by realizing the blocking of sight between H1 and H2 through inference, and prevents the collision by moving backward to clear the view.

each other, the two should draw equilibrial actions from a game; otherwise, each should perform motion planning by assuming a fixed environment. M infers by matching the observed actions of H1 and H2 with those from these two hypothetical cases.

Observation: We show in Fig. 10 that the implementation achieves what we expected: When H1 and H2 have no future collision (Fig. 10a), M has uninformed inference of their visibility, and flat objective with respect to its own action. When there is a potential collision, M realizes the block of sight early on (Fig. 10c), since H1 should otherwise start to brake as it enters the interaction zone.

V. Discussion

We now discuss several limitations of the present study and propose future directions.

a) The inconsistency between intent inference and motion planning: One may notice that during motion planning, i incorporates the uncertainty of its inference about j's future motion. However, in intent inference, i assumes that j uses a point estimation of i's intent in its planning. More concretely, the planning of i relies on the joint probability mass function $\bar{p}(\tilde{\theta}_i, \hat{\theta}_j; t)$. However, when i puts itself in j's shoes during the inference, it believes that j chooses from equilibrium motions derived from games where i has a fixed intent. To address this inconsistency, i would need to model j to have considered a distribution of intents of i in j's planning, leading to an inference of the distribution (of $\tilde{\theta}_i$) by i. It is not yet clear whether considering this modeling complexity is necessary.

b) Provable necessity and sufficiency of empathy: We demonstrate in Sec. IV that a non-empathetic agent creates false inference of others' intent, which leads to undesirable consequences. However, we have not yet investigated the conditions under which empathy will be necessary or sufficient for the inference algorithm to achieve correct convergence.

c) Inference of the control policy: Discussion so far suggested intent inference without considering the variants of control policies may not be effective when one has a wrong guess of the others' policy. The question is then how inference of policies can be incorporated, for example, to differentiate a proactive agent that pretends to be aggressive from an aggressive agent. In a discrete setting as presented in this paper, the inference can be done by enumerating over all candidate policies. This, however, will not accommodate the estimation of hyper-parameters such as β. Another potential approach is to introduce a meta-objective as a weighted sum of loss functions collected from all policies, and to infer the true policy by estimating these weights, in addition to the intent parameters. The same approach could be used for the inference of the degree of lawfulness of HVs with respect to traffic rules.

d) Computational challenges: Extending the proposed inference and planning algorithms to continuous domains will be necessary for their scalability. However, doing so will introduce computational challenges since both involve non-convex optimization problems. In addition, the incorporation of a high-dimensional distribution of inferred motions in motion planning can be intractable. One potential solution is to compute inference results offline through an enumeration over possible interactions in typical scenarios such as lane changing and intersection, and perform scenario-specific table lookup at runtime.

e) Scalability to multi-agent interactions: The proposed algorithm is algorithmically scalable to multi-agent interactions provided that all other agents only consider the ego agent as the only agent they interact with. Additional complexity is introduced when other agents interact with both the ego agent and a third agent. In this context, and strictly following the modeling of this paper, the ego agent would need to put on each of the other agents' shoes and infer their inference about others (including itself). This will lead to significantly higher computational cost than two-agent cases. One potential solution is to assume that other agents consider their surrounding agents, except the ego agent, as part of the environment, i.e., moving obstacles. Doing so reduces the computation of equilibrium for multi-agent games back to that of two-agent games. The validity of this approach, however, can only be tested by human studies.

VI. Conclusions

This paper had three novel contributions to the literature on intelligent vehicles. First, we explicitly addressed the double-blindness issue in intent inference for two-agent non-cooperative interactions. We showed empirically that allowing agents to have empathy in inference, i.e., to understand that others might have false understanding of their intents, can help to achieve correct intent inference. Second, we verified that the knowledge about vehicle dynamics was also important to

correctly infer intents. And lastly, we discussed the limitation of an existing courteous driving policy that avoided inconvenience caused by the AV, in that inconvenience in interactions could be expected by other drivers as part of the consequence of rational interactions. We proposed a new courteous policy that bounded the courteous motion of the AV using its inferred set of equilibrium motions.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation through NRI:FND:Scalable and Customizable Intent Inference and Motion Planning for Socially-Adept Autonomous Vehicles (CMMI-1925403) and the Center for Assured and Scalable Data Engineering at Arizona State University.

REFERENCES

[1] F. Codevilla, M. Miiller, A. López, V. Koltun, and A. Dosovitskiy, "End-to-end driving via conditional imitation learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4693–4700.

[2] N. Djuric, V. Radosavljevic, H. Cui, T. Nguyen, F.-C. Chou, T.-H. Lin, and J. Schneider, "Motion prediction of traffic actors for autonomous driving using deep convolutional networks," *arXiv preprint arXiv:1808.05819*, 2018.

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

[4] T. Holstein, G. Dodig-Crnkovic, and P. Pelliccione, "Ethical and social aspects of self-driving cars," *arXiv:1802.04103*, 2018.

[5] B. Kuipers, "How can we trust a robot?" *Communications of the ACM*, vol. 61, no. 3, pp. 86–95, 2018.

[6] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *Robomech Journal*, vol. 1, no. 1, p. 1, 2014.

[7] N. Li, D. W. Oyler, M. Zhang, Y. Yildiz, I. Kolmanovsky, and A. R. Girard, "Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems," *IEEE Transactions on control systems technology*, vol. 26, no. 5, pp. 1782–1797, 2018.

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

[9] C. Liu, C.-W. Lin, S. Shiraishi, and M. Tomizuka, "Distributed conflict resolution for connected autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 1, pp. 18–29, 2018.

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

[11] 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.

[12] J. Nilsson, J. Silvlin, M. Brannstrom, E. Coelingh, and J. Fredriksson, "If, when, and how to perform lane change maneuvers on highways," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 4, pp. 68–78, 2016.

[13] C. Peng and M. Tomizuka, "Bayesian persuasive driving," *arXiv:1809.09735*, 2018.

[14] N. C. Rabinowitz, F. Perbet, H. F. Song, C. Zhang, S. Eslami, and M. Botvinick, "Machine theory of mind," *arXiv:1802.07740*, 2018.

[15] R. Randazzo, "Who was at fault in self-driving uber crash? accounts in tempe police report disagree," *The Republic—azcentral.com*, 2017. [Online]. Available: https://www.azcentral.com/story/money/business/tech/2017/03/29/tempe-releases-police-report-uber-crash/99797486/

[16] Y. Rasekhipour, A. Khajepour, S.-K. Chen, and B. Litkouhi, "A potential field-based model predictive path-planning controller for autonomous road vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1255–1267, 2017.

[17] E. Rasmusen and B. Blackwell, "Games and information," *Cambridge, MA*, vol. 15, 1994.

[18] Y. Ren, S. Elliott, Y. Wang, Y. Yang, and W. Zhang, "How shall I drive? interaction modeling and motion planning towards empathetic and socially-graceful driving," in *2019 International Conference on Robotics and Automation*. IEEE, 2019, pp. 4325–4331.

[19] 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.

[20] R. Stewart and S. Ermon, "Label-free supervision of neural networks with physics and domain knowledge." in *AAAI*, vol. 1, no. 1, 2017, pp. 1–7.

[21] 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.

[22] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, "Courteous autonomous cars," *arXiv preprint arXiv:1808.02633*, 2018.

[23] V. Turri, Y. Kim, J. Guanetti, K. H. Johansson, and F. Borrelli, "A model predictive controller for non-cooperative eco-platooning," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 2309–2314.

[24] J. Wu, J. J. Lim, H. Zhang, J. B. Tenenbaum, and W. T. Freeman, "Physics 101: Learning physical object properties from unlabeled videos." in *BMVC*, vol. 2, no. 6, 2016, p. 7.

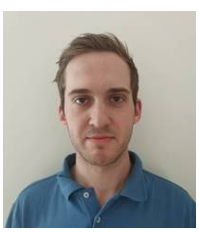
[25] J. Zhang and K. Cho, "Query-efficient imitation learning for end-to-end autonomous driving," *arXiv preprint arXiv:1605.06450*, 2016.

[26] M. Zhang, N. Li, A. Girard, and I. Kolmanovsky, "A finite state machine based automated driving controller and its stochastic optimization," in *ASME 2017 Dynamic Systems and Control Conference*. ASME, 2017, pp. V002T07A002–V002T07A002.

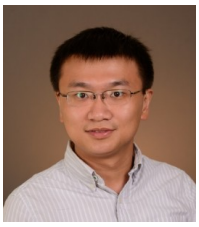
[27] A. Zyner, S. Worrall, and E. Nebot, "A recurrent neural network solution for predicting driver intention at unsignalized intersections," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1759–1764, 2018.

Yiwei Wang (S'15) received his B.Eng. degree in Huazhong University of Science and Technology, Wuhan, China, in 2013, and M.Sci. degree in Arizona State University, Tempe, USA, in 2014. He is a research assistant in Robotics and Intelligent Systems Laboratory, Arizona State University. His research interests include physical human–robot interaction, human-robot collaboration, dynamic and control in robotics.

Yi Ren is an Assistant Professor in Mechanical Engineering at Arizona State University. He received his Ph.D. in Mechanical Engineering from the University of Michigan in 2012 and his B.Eng. degree in Automotive Engineering from Tsinghua University in 2007. From 2012 to 2014 he was a postdoctoral researcher at the University of Michigan, leading projects on automated hybrid powertrain design, large-scale consumer preference learning, and government policy design for electric vehicle markets. Dr. Ren's current research focuses on the development of robust machine learning mechanisms for risk-sensitive tasks, with particular focus on accelerated computational engineering design and verifiable human-machine interactions. His work on optimal design through crowdsourcing received the Best Paper Award in Design Automation at the 2015 ASME International Design and Engineering Technical Conferences.



Steven Elliott received his B.Eng. and Master's degrees in Aerospace Engineering from Arizona State University in 2017 and 2018, respectively. He has since been working at Northrop Grumman as a Guidance, Navigation, and Control Engineer.



Wenlong Zhang received the B.Eng. degree (Hons.) in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2010, and the M.S. degree in mechanical engineering in 2012, the M.A. degree in statistics in 2013, and Ph.D. degree in mechanical engineering in 2015, all from the University of California, Berkeley, CA, USA. He is currently an assistant professor in the Polytechnic School at Arizona State University, where he directs the robotics and intelligent systems laboratory (RISE Lab). Dr. Zhang's research interests include dynamic system modeling and control, human-machine collaboration, and statistical learning. Dr. Zhang received the Best Paper Award at the IEEE Real-time System Symposium in 2013 and was a finalist of the Semi-plenary Paper Award at the ASME Dynamic Systems and Control Conference (DSCC) in 2012. He received a Bisgrove Early Tenure-Track Faculty Award from Science Foundation Arizona in 2016.