2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

When Shall I Estimate Your Intent? Costs and Benefits of Intent Inference in Multi-Agent Interactions

Sunny Amatya[1], Mukesh Ghimire[2], Yi Ren[2], Zhe Xu[2], and Wenlong Zhang[1*]

***Abstract*— This paper addresses incomplete-information dynamic games, where reward parameters of agents are private. Previous studies have shown that online belief update is necessary for deriving equilibrial policies of such games, especially for high-risk games such as vehicle interactions. However, updating beliefs in real time is computationally expensive as it requires continuous computation of Nash equilibria of the sub-games starting from the current states. In this paper, we consider the triggering mechanism of belief update as a policy defined on the agents' physical and belief states, and propose learning this policy through reinforcement learning (RL). Using a two-vehicle uncontrolled intersection case, we show that intermittent belief update via RL is sufficient for safe interactions, reducing the computation cost of updates by 59% when agents have full observations of physical states. Simulation results also show that the belief update frequency will increase as noise becomes more significant in measurements of the vehicle positions.**

I. Introduction

Humans and robots have been increasingly interacting with each other in sophisticated tasks such as manufacturing, personal care, and autonomous driving. For such interactions to be safe and efficient, understanding the intents of other agents is critical. For example, failure to understand and anticipate the intent of a human driver (H) can result in unexpected behaviors of the autonomous vehicle (AV), which contributes to distrust and disuse of the technology [1]. To this end, game-theoretic approaches become necessary to design policies for human-robot interactions (HRI) [2].

Game-theoretic modeling has primarily been applied to HRI tasks in two ways. The first idea is to simplify the motion planning problem as an optimal control problem or model the interaction as a complete-information game [3]. Taking this approach, researchers have used belief updates to adjust the planned motion [4]. However, this idea may not work for autonomous driving, as often times a human driver also needs to infer the AV's intent online. Here, intent is defined as a non-observable goal-directed parameter [5]. The second approach is to explicitly consider mutual intent inference, and enable the AV to build a Theory of Mind (ToM) of the human, so the AV will infer both *what your intent is* and *what you think my intent is*. ToM models

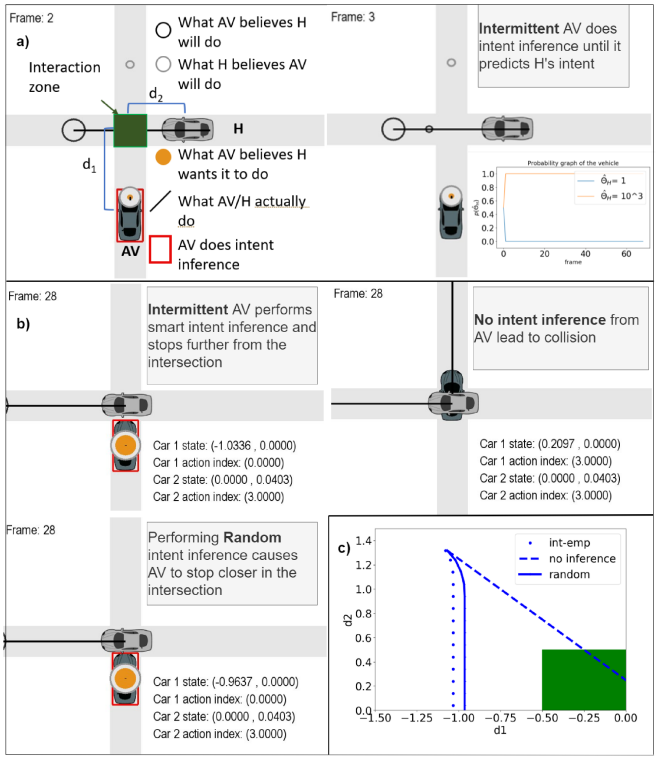


Fig. 1. (a) (left) Schematics of the two-agent intermittent intent inference, and collision occurs when both AV and H are in interaction zone at the same time. (right) The AV performs intent inference until it is able to identify H's intent. (b) A motivating example of AV performing intent inference at different intervals: (top left) Intermittent intent inference leads to safer interaction, (top right) no intent inference leads to a collision, (bottom left) performing intent inference randomly leads to the AV stopping closer to the intersection. (c) Paths of the two vehicles with different intent inference algorithms. The intermittent empathetic (int-emp) agent is further from the interaction zone compared to random and no intent inference.

describe how humans update their beliefs and make decisions when interacting with other agents allowing an agent to reason about herself and others as rational entities [6] [7]. In [8], this *empathetic intent inference* approach allows an AV to correctly infer the human's intent and gracefully negotiate with the human driver to resolve potential conflicts in uncontrolled intersections.

Despite progress in game-theoretic modeling of HRI, it remains computationally challenging to execute belief update algorithms in real-time HRI tasks. On one hand, due to the mutual intent inference, a robot has to run intent inference at a higher level than the human, i.e., *what your intent is* and *what you think my intent is*, in order to efficiently coordinate with the human. As a response, a human can infer intent at the same level, demanding the robot to go a higher level, and this process to go ad infinitum [9]. On the other hand,

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

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

[2] M. Ghimire, Y. Ren, and Z. Xu are with the School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ, 85287, USA. Email: {mghimire,yiren,xzhe1}@asu.edu

* Address all correspondence to this author.

the states and actions in many HRI tasks are continuous. The ToM framework of observing the ego and other agents' behaviors for belief updates becomes computationally expensive and even intractable in this case [10]. These challenges have motivated researchers to explore other representations of intents. For example, machine learning has been used to learn latent representation of the human's policy, which allows the robot to reason about and potentially influence the human's actions [11].

In this paper, we will address the computational challenges of belief update by taking inspiration from humans [12]. Studies have shown that humans employ intermittent control for many motor tasks, such as balancing [13], walking [14], and even driving [15]. They only adjust their motor control inputs as required for safety or stability, similar to the event-triggered control scheme for physical systems [16]. Existing work on intermittent and/or event-triggered control focuses on single-agent systems. In this paper, we extend this concept to intent inference in multi-agent systems (i.e., HRI). Our *hypothesis* is that a robot may not need to update its belief at each sampling time, especially when the system is in an equilibrium state where there is a consensus of intent estimate. In contrast, a robot only needs to conduct belief update when its current estimate cannot explain the behavior of other agent, or when it can improve the team performance by updating its policy, which depends on intent estimate. As a result, the key question to be answered in this paper is:

When should a robot infer other agents' intent?

This paper makes the following contributions towards answering this question (see summary in Fig. 1): 1) We consider the triggering of intent inference as a high-level controller, which is learned through reinforcement learning (RL); 2) For a two-vehicle uncontrolled intersection case, we show that a learned intent inference policy decreases the average computation time from 4.87 to 2.85 seconds per sampling period, while maintaining safety performance; 3) We test the performance of the proposed algorithm in a high-risk zone of collisions in the presence of measurement noises, and show that the proposed algorithm performs more frequent belief updates when the noise level is higher, which aids in reducing the chance of collision.

II. Related Work

Intent representation and inference: Some earlier works of intent recognition in collaborative robotics required inference of the goal as well as the latent parameters in pursuit of the goal [17]. As all agents in multi-agent interaction depend upon each other for task completion, intent inference enables the agents to determine the intent as well as the possible trajectories of the other agent which aids in faster mutual adaptation. The notion of driving style or attention level as an intent inference parameter has been used in two-agent games [18]. Social value orientation has been used as a marker of driver intent for reformulating the interdependent optimization problems as a local single-level optimization using Karush-Khun-Tucker conditions [3]. The notion of double blindness in intent inference was developed by [19], and an empathetic intent inference method was further explored for autonomous vehicles in [8], [20]. These further layers of computation of intent makes intent inference algorithms generally expensive to compute in real time.

Event-triggered and intermittent control: Intermittent control implies a discontinuous control with active and passive phases compared to a continuous active control [16]. Human motor control has been modeled with event-triggered intermittent control activated via noise or input threshold and is used for a variety of balancing task ranging from stick balancing to postural balancing [13], [14]. Intermittent control behaviors have also been observed in various driving task such as car following [15] and ground vehicle steering [21]. Assuming that the same form of control can be used in intent inference for decision making in a driving task, it is important to observe how event-triggered intent inference performs compared to continuous inference, and which state parameters should be used to trigger the event.

Game-based models for multi-agent interactions: One early application of game-theoretic approaches for multi-agent interaction is the pursuit-evasion game, in which two players have opposing goals [2]. Depending on the state of the system (discrete vs. continuous), the methodology for solving the game can vary from computing minimum cost in the search tree [22] to computing solution for Hamilton-Jacobi-Issac (HJI) equations [8]. For interactions where not all information about each agent is publicly available (incomplete- or imperfect-information games), they were modeled as partially observable stochastic games [18]. Hence, an agent has to use the other agent's actions as observations of the other agent's underlying utility function parameters. Modeling agents with incomplete information and reasoning over other agent's intent, function parameter, or policy is an iterative process which can become computationally intractable [11]. At the same time, such parameter estimation does not necessarily lead to continuously changing policies as seen in [22], where the agents do not often change their policy and maintain a motion planning strategy for significant amount of time throughout the interaction. This motivates us to perform computationally expensive belief update only when it is necessary.

III. Methods

In this section, we define the Markov Decision Process (MDP) for the RL formulation of our intermittent intent inference algorithm. With this formulation, the RL agent (AV) learns to perform intent inference (essentially update its beliefs) whenever deemed necessary. In this paper, we demonstrate a high-level strategy that can be applied to any intent inference algorithm of the developer's choice. Here, we adopt a belief update approach based on empathetic intent inference, whose benefits have been highlighted in our past work [20], [23]. We combine the empathetic belief update with the proposed high-level RL-based controller, and further show its performance compared to non-empathetic algorithm which is widely used in existing HRI literature.

TABLE I
NOMENCLATURE

H	human-driven vehicle
AV	autonomous vehicle
Θ	intent set, ($\Theta = [1, 1000]$)
S	state space
U	discrete input space ($U = [-2, -1, 0, 1, 2, 3]$)
A	action space for the RL Agent ($A = [0, 1]$)
s_{AV}, s_H	position and velocity state of the AV and H
u_{AV}, u_H	AV and human control actions
L	predefined finitie time horizon
ξ_{AV}, ξ_H	sequence of control actions ($\xi_i = [u_i(t), u_i(t+1), \cdots u_i(t+L-1)]$)
ξ^*	Nash equilibrium solution of the action
C_{AV}, C_H	objective function of interaction game
θ_{AV}, θ_H	intent parameter of AV and H
$\hat{\theta}_H, \hat{\theta}_{AV}$	AV's estimate of H's intent, H's estimate of AV's intent
$\tilde{\theta}_{AV}, \tilde{\theta}_H$	AV's inference of H's inference of AV's intent, H's inference of AV's inference of H's intent
c_{safety}, c_{task}	safety loss, task loss
$p(\tilde{\theta}_i, \hat{\theta}_{-i}; t)$	empirical joint probability distribution of intent
$p(\hat{\theta}_{-i}, t)$	marginal probability distribution of other's intent
$p(\tilde{\theta}_i, t)$	marginal probability of self's intent
$\bar{p}(\tilde{\theta}_i, \hat{\theta}_{-i}; t)$	updated joint probability distribution after bayes update
$p(\xi_{-i}; t)$	inferred motion of agent $-i$ based on baseline policy Nash equilibrium set
$\bar{p}(\xi_{-i}; t)$	updated inferred motion of agent $-i$ after bayes update

A. RL-based Intent Estimation

Following [23], our setup considers a two agent game $i \in \{H, AV\}$, where both agents share the same input set U, state space S and intent set Θ. A standard MDP is defined by a tuple $< S, A, T, R >$, which consists a set of states $S = S_p(t-1) \times U_p(t-1) \times B_\theta(t)$ where $S_p(t-1) \ni (s_{AV}(t-1), s_H(t-1))$ contains the physical states (position and velocity) of the agents at time $t-1$, $U_p(t-1) \ni (u_{AV}(t-1), u_H(t-1))$ is the acceleration used by the agents in time $t-1$, $B_\theta \ni (\bar{p}(\tilde{\theta}_{AV}, \hat{\theta}_H, t), \bar{p}(\tilde{\theta}_H, \hat{\theta}_{AV}, t))$ are the joint belief over intent parameter θ of the agents. The RL problem is, whether, at a given time step, to perform belief update or not. The RL agent in our simulation is the AV. The RL agent's action set is defined as $A = \{0, 1\}$, where 1 means that the AV will update its belief at the current step and 0 otherwise.

We assume that both the H and AV are optimal planners with the goal of minimizing cumulative loss over the horizon L. The loss functions of the AV and human are defined as:

$$C_i(\xi_{AV}, \xi_H, \theta_i) \quad = \quad \sum_{t}^{t+L-1} c(\xi_i; \xi_{-i}, \theta_i, s_i, s_{-i}, t), \quad (1)$$

where $i \in \{H, AV\}$. At time t, the instantaneous loss of an agent i, c_i, is modeled as the weighted sum of the safety loss c_{safety} and the task loss c_{task}:

$$c_i(\xi_i; \xi_{-i}, \theta_i, s_i, s_{-i}, t) = c_{safety}(\xi_i; \xi_{-i}, s_i, s_{-i}, t) + \theta_i c_{task}(\xi_i; s_i, t). \quad (2)$$

The reward, R, of the MDP is a function of the cumulative sum of instantaneous loss, C_i, of both agents as defined in (1) and effort of belief update, e_b, where $R = -\sum C_i - e_b$.

The state transition function for the MDP, T, consists of two parts: 1) Physical state transition: $T(S_p'|S_p, U_p)$ updates the physical states of the agents based on the applied action $U_p(t)$ according to the vehicle dynamics; 2) Belief update transition: $T(B_\theta'|B_\theta, A)$ updates the joint belief over the intent parameter. The belief update is described in Section III-B.1. Note that the motion planning is carried out at every timestep in this work.

B. Belief Update and Motion Planning

As mentioned above, the proposed RL formulation allows us to use any belief update algorithm and motion planner. Here, we highlight the non-empathetic and empathetic intent inference algorithms used for the belief update and the motion planner for physical state transition.

*1) **Belief update**:* In this section we discuss different aspects that are required for the agents to update their belief. The empathetic agents compute the joint belief distribution over both their and the other agent's intent while the non-empathetic agents compute only belief of the other agent.

Non-empathetic agent: A non-empathetic agent, i believes the other agent, $-i$, knows its true intent. Hence, it computes the Nash equilibrium solution of the game such that, $Q(\theta_i, \hat{\theta}_{-i}, t) = \{(\xi_i^*, \hat{\xi}_{-i}^*)\}$ where each element in Q satisfies

$$\begin{aligned} \xi_i^* &= \arg\min C(\xi_i; \hat{\xi}_{-i}^*, \theta_i), \\ \hat{\xi}_{-i}^* &= \arg\min C(\xi_{-i}; \xi_i^*, \hat{\theta}_{-i}) \end{aligned} \quad (3)$$

Empathetic agent: An empathetic agent, i, puts itself in the other agent's, $-i$, shoes and tries to see the other agent's perspective. As i estimates $-i$'s intent with $\hat{\theta}_{-i}$, it generates an estimate of $-i$'s estimate of itself with $\tilde{\theta}_i$ and finds the Nash equilibrium based on $Q(\tilde{\theta}_i, \hat{\theta}_{-i}, t) = \{(\tilde{\xi}_i^*, \hat{\xi}_{-i}^*)\}$

Baseline policy: Agent i believes that the agent $-i$ plans its motion choosing uniformly from the Nash equilibrium set. Hence for the agent i, the probability mass function of $-i$'s motion is given by:

$$p(\xi_{-i}) \propto \begin{cases} |\{\xi_{-i}; (\xi_{-i}, \xi_i) \in Q(\hat{\theta}_{-i}, \tilde{\theta}_i, t)\}|, \\ \text{if } i \text{ is empathetic} \\ |\{\xi_{-i}; (\xi_{-i}, \xi_i) \in Q(\hat{\theta}_{-i}, \theta_i, t)\}|, \\ \text{if } i \text{ is non-empathetic} \end{cases} \quad (4)$$

where $|\cdot|$ is the cardinality of a set.

Inference problem: The inference problem is solved by finding the resulting motion of the other agent $\xi_{-i}^\dagger(t-1)$ with the highest probability mass at time $t-1$.

$$\min_{\tilde{\theta}_i, \hat{\theta}_j} \left\| u_{-i}^\dagger(t-1) - u_{-i}(t-1)) \right\|_2^2 \quad (5)$$

$$\text{s.t.} \quad \xi_{-i}^\dagger(t-1) = \arg\max_{\xi_j \in \Xi_j} p(\xi_{-i})$$

Each solution of (5) is in set $S(t)$ and assigned equal probability mass $(1/K)$. The resulting joint probability distribution is provided by

$$p(\tilde{\theta}_i, \hat{\theta}_{-i}; t) \propto \begin{cases} 1/K, & \text{if } (\tilde{\theta}_i, \hat{\theta}_{-i}) \in S(t) \\ 0, & \text{otherwise} \end{cases}$$

and $p(\hat{\theta}_{-i}; t) = \sum_{\theta_i \in \Theta} p(\tilde{\theta}_i, \hat{\theta}_{-i}; t)$.

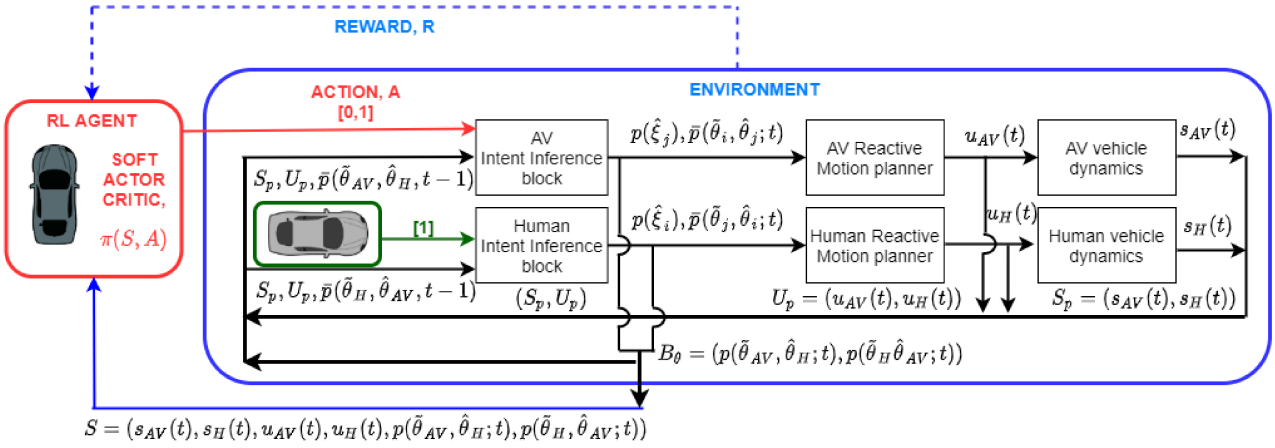


Fig. 2. Block diagram of the RL-based intermittent empathetic intent inference. The RL agent takes the states from the environment and provides **ACTION [0,1]**. Vehicle in green box represents H which gives a constant signal 1 to do belief update at every timestep. Current physical states, S_p, input states, U_p, and the current joint probability matrix, B_θ, are input for the intent inference block, which returns the distribution of the probability of the other agent and the marginal probability of the trajectory. The reactive motion provides action for the AV which incorporates the inferred motion and intent of the agents (see Section III 2-3 for details).

After calculating $p(\theta_{-i}; t)$, following Bayes rules, we update the marginal probability distribution as, $\bar{p}(\hat{\theta}_{-i}; t) \propto \bar{p}(\hat{\theta}_{-i}; t-1) p(\hat{\theta}_{-i}; t)$ [23]. In a similar manner, the posterior for the joint probability distribution and marginal probability of the estimated trajectory is updated as:

$$\bar{p}(\tilde{\theta}_i, \hat{\theta}_{-i}; t) = p(\tilde{\theta}_i, \hat{\theta}_{-i}; t) \frac{\bar{p}(\hat{\theta}_{-i}; t)}{p(\hat{\theta}_{-i}; t)} \tag{6}$$

$$\bar{p}(\hat{\xi}_{-i}; t) = \sum_{(\tilde{\theta}_i, \hat{\theta}_{-i}) \in \Theta \times \Theta} p(\xi_{-i}, \tilde{\theta}_i, \hat{\theta}_{-i}, t) \bar{p}(\tilde{\theta}_i, \hat{\theta}_{-i}; t) \tag{7}$$

*2) **Motion planning**:* We use the reactive planning strategy [23] that incorporates the inferred motion and intent of the agents. Given the distribution of $-i$'s future motions, $\bar{p}(\hat{\xi}_{-i}; t)$, a reactive agent plans its motion by minimizing the expected loss given as:

$$\min_{\xi_i \in \Xi_i} C_i^{reactive}(\xi_i) := E_{\hat{\xi}_{-i} \sim \bar{p}(\hat{\xi}_{-i}; t)}[C(\xi_i; \hat{\xi}_{-i}, \theta_i)] \tag{8}$$

In our reactive motion planner, the agents react to the inferred intent and motion to highlight the impacts of the proposed intermittent intent inference strategy.

IV. Case Study

The goals of the case study are to: 1) explore costs and benefits of running intent inference at a lower frequency compared to the sampling rate; 2) compare the performance of the proposed algorithm with the baseline intent inference algorithms; 3) test the algorithm with measurement noise.

A. Simulation Setup

*1) **Simulated RL interactions**:* We use the OpenAI Gym [24] library to configure the RL settings. We solve the RL problem using the Soft Actor Critic for Discrete action (SAC-Discrete) [25] setup. It was chosen due to the low number of hyperparameters to be tuned and its demonstrated robustness to different RL environments.

We use a fixed configuration for RL environment which generates a set of 1000 random initial states using uniform sampling for the physical and belief states of the agents. We use the first 750 data sets for training and the remaining 250 for testing. We parameterize each of the 750 data sets using the initial states $\mathcal{S}(0)$, which consisted of the physical states $S_p(0)$ and the prior belief $B_\theta(0)$. The simulation ran k steps until the end conditions were met (i.e. when one of the vehicles passed the intersection). Each k^{th} step of the simulation provided the trajectories of states $s(k)$, actions $u(k)$, updated belief $B_\theta(k)$, and reward $R(k)$. During training, the policy was updated based on the reward received by the agent as a result of the chosen action A for whether or not to perform belief update.

We perform an ablation study by varying e_b in the reward, R, primarily to determine the optimum penalty weight when the agent decides to update its belief. During the testing of the RL agent, we store the amount of memory used and time taken in each simulation.

*2) **Environment definition**:* We simulate the H-AV interactions at an uncontrolled intersection shown in Fig. 1a. The state of the agent i is defined by the agent's position x_i and its velocity v_i: $s_i = (x_i, v_i)$. The action of the agent i is defined as its representative acceleration rate.

Loss function: The instantaneous loss function is a weighted sum of the safety and task loss. The task loss is modeled to penalize when the agent fails to cross the intersection in the given horizon, and it is defined as $c_{task} = exp(-x_i(t+L-1)+0.6)$ where x_i is the position of the agent i. The safety loss is a function of the distance between the two agents and is defined as $c_{safety} = exp(\eta(-D+\phi))$, where $D = \|x_i - x_j\|_2^2$. Different from [23], η is empirically chosen to be 3.0 and $\phi = 0.13w^2$ where $w = 4.5$.

Constants and assumptions: We make an assumption that the true intent θ of both the AV and H do not change during the interaction. Note that $\theta_i = 1$ and $\theta_i = 1000$ indicate the agent to be non-aggressive and aggressive, respectively. In all our simulations we made our AV non-aggressive while our H can be either aggressive or non-aggressive.

We also assume both AV and H keep their inference

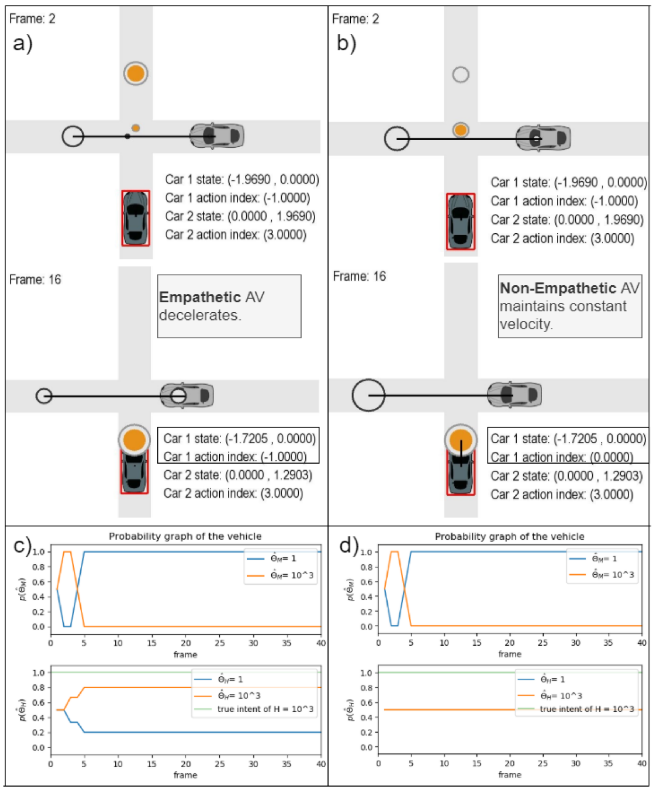


Fig. 3. (a) AV performs empathetic intent inference at every third interval and is able to predict the H as aggressive so it brakes. (b) AV performs non-empathetic intent inference at every third interval and is unable to predict the H and maintains a constant velocity. (c) Empathetic intent inference with a reduced update frequency shows AV taking significant time to predict the true intent of H. (d) Non-empathetic intent inference with a reduced update frequency shows AV being unable to predict the true intent of the H.

nature (empathetic or non-empathetic) constant during the simulations, and H does intent inference at every time step.

*3) **Evaluation metrics**:* We measure the computation time and memory used for the empathetic (E), non-empathetic (NE), and the proposed intermittent empathetic intent inference algorithm (I). We test the algorithm in 250 initial conditions generated from the 1000 uniformly distributed conditions mentioned in Sec. IV-A.1. We further examine the value and distance between the vehicles for each algorithm.

We also test the intermittent algorithm in collision-prone simulation by introducing Gaussian noise $n \sim \mathcal{N}(0, \sigma^2)$ in the physical state of the H as observed by the AV and making the initial position of the AV closer to the intersection. We test in these conditions, because a wrong estimate of the other agent's intent when they are in close proximity can lead to a collision. We evaluate the AV's behavior under different noise levels using collision ratio and inference ratio. Collision ratio is defined as the total number of collision cases divided by the total number of test cases. Inference ratio is defined as the average rate at which belief update occurs during each noise level.

B. Analysis of Different Belief Update Frequencies

We first study the impact of different intent inference frequencies by letting the AV perform intent inference once in every three time steps. Here, the AV was non-aggressive and the H was aggressive. We perform two simulations with both the AV and H as (1) empathetic, and (2) non-empathetic. In this test, we set the initial positions of both the vehicles 2 meters away from the intersection. As seen in Fig. 3a-b, the empathetic agent chooses to decelerate while the non-empathetic agent chooses to maintain its velocity. The frequency of intent inference affects the estimation of H's intent as seen in Fig. 3c. Furthermore, when the AV uses non-empathetic belief update, it is not able to correctly estimate the H's intent as seen in Fig. 3d. This provides preliminary evidence to use empathetic intent inference intermittently. We can see that by just choosing to increase the interval between the belief updates, one can get the correct inference of H's intent but it takes more instances of belief updates. This motivated us to model the intent inference algorithm as an RL problem, where, instead of reducing the interval between belief update, we let the RL agent decide whether or not to perform belief update at a given time step.

C. Performance Comparison of Intermittent Empathetic Intent Inference with Baseline Algorithms

Experimental setup: The baseline empathetic and non-empathetic intent inference algorithms are presented in Section III-B.1, and they ran at each time step in our study. Whereas, in the intermittent intent inference case, the AV decides whether it needs to conduct empathetic intent inference using the policy learned from the RL algorithm. Using the evaluation metrics discussed in Section IV-A.3, the baseline algorithms are compared with the intermittent empathetic (I) intent inference algorithm. The initial position x_H is sampled randomly from $x_H \in [2.5, 1.5]$ and x_{AV} is sampled randomly from $x_{AV} \in [-1.5, -2.5]$.

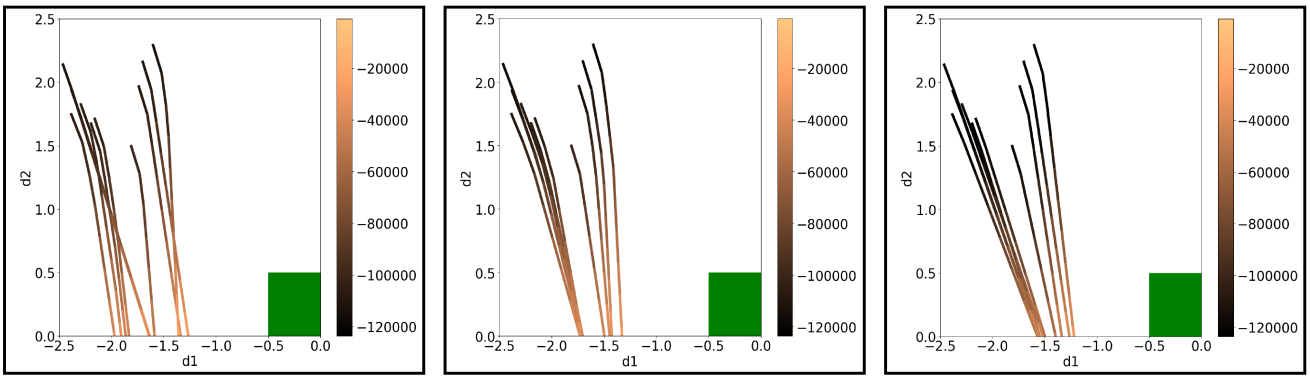
Observation: We find that intermittent intent inference has several benefits in terms of computational load, as shown in Table II. As expected, we see a decrease in the average memory used for computation when AV performed belief updates intermittently as opposed to the cases when it was updated at every time step. As a result, the average time to process one sample of measurement in the simulation also decreases. The results are promising as they substantiate our claim that updating beliefs intermittently, as necessary, is able to maintain safe vehicle interaction as characterized by the distance between the vehicles and the cumulative reward collected by them. The value, in fact, is higher when the AV updated beliefs intermittently.

The qualitative visualization of 10 random samples is shown in Fig. 4. We show the values for three cases - E, NE, and I, along with the intersection area marked as a green box. We want the trajectories in the figure to be far from the

TABLE II

COMPARISON BETWEEN THREE INTENT INFERENCE SCHEMES

Metrics	Emp (E)	N-Emp (NE)	Int (I)
Memory used (KB)	3.11±0.28	2.95 ± 0.95	1.27 ±0.34
Computation Time (s)	4.87±0.43	2.59 ±0.23	2.85 ±0.24
Value	-114088.72 ±6582.32	-132688.88 ± 6374.33	-109208.23 ± 4176.26
Distance	1.43±0.05	1.33 ± 0.07	1.42± 0.15



(a) Intermittent (b) Empathetic (c) Non-empathetic

Fig. 4. Color coding of the value for same 10 random simulations for intermittent, empathetic and non-empathetic intent inference. The trajectories for Intermittent cases (a) have lower values compared to empathetic (b) and non-empathetic cases (c). The trajectories in (a) also converge further from each other at the intersection and have higher variance compared to (b) and (c).

green region – the trajectory that passes through green box corresponds to a collision. While there are no collisions in the test case, we further observe that in NE cases the value of the RL agent throughout the trajectory is prominently low and converge around 1.5 meter at the intersection. Meanwhile, the trajectories for E cases have improved values and converge further away from each other at the intersection than in NE cases. Compared to the NE and E cases, we find that that I cases have significantly improved values denoted by the lighter trajectory colors. They also have a wider range of separation at the intersection than those from the NE and E cases. The variance is also higher because, for belief updates in I cases, the AV estimates the intent as well as the predicted trajectory of H, and uses the last predicted trajectory for their motion planner until a new prediction is needed.

These results indicate that performing intermittent intent inference is safe and highly cost-effective. It should be noted that the widely-adopted non-empathetic intent inference is unable to estimate the other agent quickly; hence there is a mean distance of 1.33 between the agents which is significantly lower compared to empathetic agents. With intermittent algorithm we are able to generate safety comparable to empathetic agents at similar computation time as non-empathetic agents.

D. Intermittent Intent Inference with Measurement Noise

Experimental setup: Noise in measurement is a prominent issue for autonomous driving especially in challenging weather conditions. In this part of the study, we carry out intent inference in presence of noise in the observation of physical states. In particular, we model the AV with measurements of the H's physical states. The initial positions of AV and H were set as $x_{AV} \in [-1.25, -1.0]$ and $x_H \in [1.35, 1.6]$, meaning both agents were closer to the intersection compared to previous case study. For all the test cases we choose both agents to be empathetic, based on the results of the previous studies. A Gaussian white noise $n \sim \mathcal{N}(0, \sigma^2)$ is introduced into the AV's observation of the H's position, with standard deviation of $\sigma = [0.0, 0.00625, 0.0125, 0.025, 0.05]$.

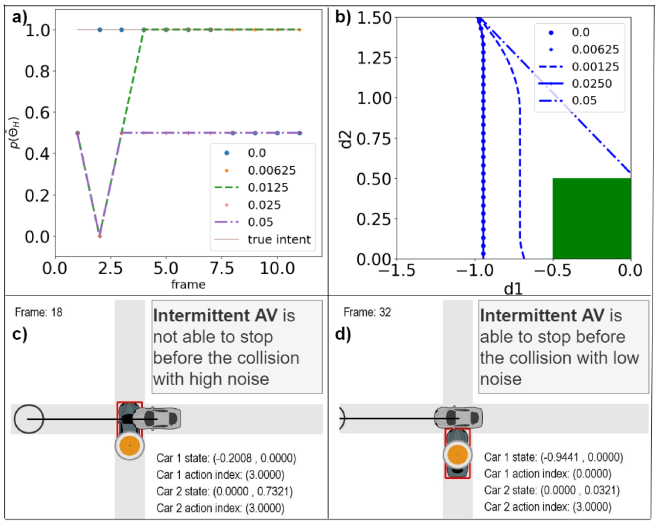


Fig. 5. (a) Different estimated intent of the H by the AV in presence of various levels of noise. The AV is able to accurately predict the intent of the H in presence of noise of $\sigma = 0.00625$ which leads to safer maneuver. (b) Based on the belief update and perceived states of H, AV collides with the H in presence of noise with higher variance. (c) In presence of noise with $\sigma = 0.05$, AV is unable to estimate the H's true intent which leads to collision. (d) When the noise variance is reduced to $\sigma = 0.00625$, the AV is able to estimate H's intent correctly and stop before collision.

Observation: In the case without noise, the AV infers the H at every timestep as seen in Table III. This denotes the best case policy for given scenario where the agents are closer to the intersection and have higher probability of collision. As we increase the noise, we see that the collision ratio increases as expected since the AV is updating its belief with increasing error. When there is zero noise in the observation ($\sigma = 0.0$), the AV performs intent inference at all steps to ensure safety since both vehicles start close to the intersection in this study. However, the inference ratio first decreases and then increases as the noise gradually increases from 0.0 to 0.05, with lowest inference ratio occurring at noise $\sigma = 0.0125$.

To further understand the "bowl-shape" inference ratio, we visualize the estimated intent in the first few samples for each noise level in Fig. 5a. It is seen that for $\sigma = 0.0125$ the AV is able to estimate the intent of the H accurately only after a few timesteps compared to lower σ values. While this does not

TABLE III

COMPARISON BETWEEN DIFFERENT LEVELS OF MEASUREMENT NOISE

Noise standard deviation (σ)	Collision ratio	Inference ratio
0.0	0.100	1.0±0.0
0.00625	0.129	0.847±0.035
0.0125	0.133	0.155±0.089
0.025	0.145	0.562±0.064
0.05	0.140	0.810±0.051

aid in avoiding the collision, the AV does realize that further belief update is not necessary, which causes the inference ratio to drop significantly. We also see that for higher σ values, the AV is not able to estimate the intent of H, leading to higher inference ratios. We further plot the paths of the two vehicles with different noise levels in Fig. 5b. Here we can see that the AV's inability to estimate H's intent results in a collision in the high noise case ($\sigma = 0.05$).

Figs. 5c-d show the same case with varying noise levels where there is a collision due to a higher σ (c) while no collision is observed with a lower $\sigma = 0.00625$ in (d). The results indicate that in collision-prone area when AV and H are closer to the intersection, performing belief update over the other agent frequently led to less collisions. These collisions often occur when the AV is not able to identify the true intent of H in a timely manner.

V. CONCLUSION

In this paper, we develop an RL-based intermittent empathetic intent inference algorithm. In an uncontrolled intersection case, we study the use of intermittent empathetic intent inference for belief updates. We first show that decreasing the frequency of intent inference degrades the accuracy of intent prediction, proving the necessity of an intelligent intent inference algorithm. We show that when the vehicles were farther away from the intersection, with intermittent empathetic intent inference, fewer belief updates were needed without compromising the safety. We find that as the noise increases, the collision ratio of the AV also increases. However, the inference ratio first decreases and then increases.

The proposed algorithm can be extended to include belief over H's trajectories. This can aid in generating trajectories from intermittent agents resembling empathetic agents. The algorithm can further be tested in more complex driving scenarios, such as roundabouts and lane changing. As this work only considers AV as the intermittent agent, we can further model both H and AV as intermittent agents. The generalizability of the proposed algorithm can also be tested using various belief update algorithms and further be compared with traditional style controller for performance.

REFERENCES

[1] N. L. Tenhundfeld, E. J. de Visser, A. J. Ries, V. S. Finomore, and C. C. Tossell, "Trust and distrust of automated parking in a Tesla model x," *Human factors*, vol. 62, no. 2, pp. 194–210, 2020.

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

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

[4] D. Fridovich-Keil, A. Bajcsy, J. F. Fisac, S. L. Herbert, S. Wang, A. D. Dragan, and C. J. Tomlin, "Confidence-aware motion prediction for real-time collision avoidance," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 250–265, 2020.

[5] R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, and M. Nicolescu, "Understanding activities and intentions for human-robot interaction," in *Human-Robot Interaction*. BoD–Books on Demand, 2010, pp. 288–305.

[6] N. Robalino and A. Robson, "The economic approach to 'theory of mind'," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1599, pp. 2224–2233, 2012.

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

[8] Y. Chen, L. Zhang, T. Merry, S. Amatya, W. Zhang, and Y. Ren, "When shall I be empathetic? the utility of empathetic parameter estimation in multi-agent interactions," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 2761–2767.

[9] W. Yoshida, R. J. Dolan, and K. J. Friston, "Game theory of mind," *PLoS computational biology*, vol. 4, no. 12, p. e1000254, 2008.

[10] C. Baker, R. Saxe, and J. Tenenbaum, "Bayesian theory of mind: Modeling joint belief-desire attribution," in *Proceedings of the annual meeting of the cognitive science society*, vol. 33, no. 33, 2011.

[11] A. Xie, D. P. Losey, R. Tolsma, C. Finn, and D. Sadigh, "Learning latent representations to influence multi-agent interaction," *arXiv preprint arXiv:2011.06619*, 2020.

[12] A. A. Sawant, R. Robbes, and A. Bacchelli, "To react, or not to react: Patterns of reaction to API deprecation," *Empirical Software Engineering*, vol. 24, no. 6, pp. 3824–3870, 2019.

[13] I. D. Loram, H. Gollee, M. Lakie, and P. J. Gawthrop, "Human control of an inverted pendulum: is continuous control necessary? is intermittent control effective? is intermittent control physiological?" *The Journal of physiology*, vol. 589, no. 2, pp. 307–324, 2011.

[14] Y. Asai, S. Tateyama, and T. Nomura, "Learning an intermittent control strategy for postural balancing using an emg-based human-computer interface," *PLoS One*, vol. 8, no. 5, p. e62956, 2013.

[15] I. Lubashevsky and H. Ando, "Intermittent control properties of car following: Theory and driving simulator experiments," *arXiv preprint arXiv:1609.01812*, 2016.

[16] P. Gawthrop, I. Loram, M. Lakie, and H. Gollee, "Intermittent control: a computational theory of human control," *Biological cybernetics*, vol. 104, no. 1, pp. 31–51, 2011.

[17] 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.

[18] 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, 2018.

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

[20] 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 (ICRA)*. IEEE, 2019, pp. 4325–4331.

[21] G. Markkula, E. Boer, R. Romano, and N. Merat, "Sustained sensorimotor control as intermittent decisions about prediction errors: Computational framework and application to ground vehicle steering," *Biological cybernetics*, vol. 112, no. 3, pp. 181–207, 2018.

[22] L. Sun, M. Cai, W. Zhan, and M. Tomizuka, "A game-theoretic strategy-aware interaction algorithm with validation on real traffic data," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 11 038–11 044.

[23] Y. Wang, Y. Ren, S. Elliott, and W. Zhang, "Enabling courteous vehicle interactions through game-based and dynamics-aware intent inference," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 217–228, 2020.

[24] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[25] P. Christodoulou, "Soft actor-critic for discrete action settings," *arXiv preprint arXiv:1910.07207*, 2019.