2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

Active Probing and Influencing Human Behaviors Via Autonomous Agents

Shuangge Wang[1], Yiwei Lyu[2], John M. Dolan[3]

***Abstract*— Autonomous agents (robots) face tremendous challenges while interacting with heterogeneous human agents in close proximity. One of these challenges is that the autonomous agent does not have an accurate model tailored to the specific human that the autonomous agent is interacting with, which could sometimes result in inefficient human-robot interaction and suboptimal system dynamics. Developing an online method to enable the autonomous agent to learn information about the human model is therefore an ongoing research goal. Existing approaches position the robot as a passive learner in the environment to observe the physical states and the associated human response. This passive design, however, only allows the robot to obtain information that the human chooses to exhibit, which sometimes doesn't capture the human's full intention. In this work, we present an online optimization-based probing procedure for the autonomous agent to clarify its belief about the human model in an active manner. By optimizing an information radius, the autonomous agent chooses the action that most challenges its current conviction. This procedure allows the autonomous agent to actively probe the human agents to reveal information that's previously unavailable to the autonomous agent. With this gathered information, the autonomous agent can interactively influence the human agent for some designated objectives. Our main contributions include a coherent theoretical framework that unifies the probing and influence procedures and two case studies in autonomous driving that show how active probing can help to create better participant experience during influence, like higher efficiency or less perturbations.**

I. Introduction

It is imperative for robots to behave reactively in a human-present environment because all safety specifications ought to be met. An autonomous vehicle, for instance, should yield to a human vehicle trying to nudge in front of it [1], [2]; a reconnaissance drone should avoid adversarial behaviors. Robots, however, are usually not designed to behave purely in a reactive manner because it makes them too conservative. Consider a scenario of autonomous driving (Fig. 1) where the human vehicle is traveling in the outer lane (lower), but at a fast enough speed that it's more efficient to switch to the inner lane (upper). Many human drivers don't have the awareness to switch lane because they are usually egocentric, even subconsciously, in that they would rather remain in their current lane unless blocked by some other vehicles. Such human egocentricity and strict infrastructure preconditions render purely communication-based approaches, like vehicle signaling or V2X [3]–[7], fruitless in addressing these inefficiencies. Some works, therefore, proposed interaction-based approaches, like game-theoretical influence [8], [9], wave stabilization [10]–[13], and herding [14]–[19], that use autonomous agents to influence human agents physically. In Fig. 1, for instance, the autonomous vehicle would block the fast human vehicle, influencing it to drive in the inner lane.

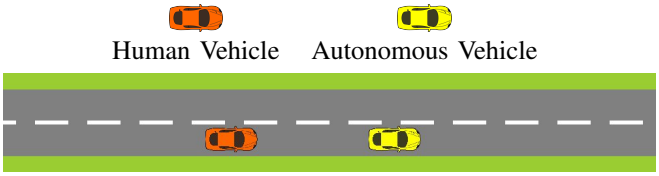


Fig. 1: Both vehicles currently travel to the right in the outer lane (lower). Autonomous vehicle (yellow) intends to influence human vehicle (orange) with intention to drive fast to inner lane (upper).

Since such influence is exerted in close proximity, the autonomous agent needs an accurate human model. Although generally reasonable, models produced from offline techniques may not capture characteristics specific to the human agent with whom the robot interacts closely. For instance, in Fig. 1, the autonomous vehicle is interested in, rather precisely, the human vehicle's desired travel velocity, in which each human differs from another.

Existing online approaches tackle this problem by positioning the autonomous agent as a passive observer, in which it observes the environmental states and their associated human response and then chooses the model that best explains this correlation. The issue with this design is that the autonomous agent is passive, so it only has access to information that the human agent chooses to exhibit. Hence, the autonomous agent can only make decisions based on the human information that's readily available. In Fig. 1 for instance, a passive autonomous vehicle would presume the human vehicle intends to travel at most as fast as itself, whereas in reality the human could want to drive faster, only to be blocked by the autonomous vehicle.

In this work, we enable autonomous agents to leverage their actions to estimate the human internal model by actively interacting with the human to reveal more information. Rather than relying on passive observations, the autonomous agent can account for the fact that the human will react to its actions, so the autonomous agent can "probe", i.e., select actions that will trigger human reactions that will best challenge its initial belief. By probing iteratively, the autonomous agent converges to an increasingly accurate

[1]Shuangge Wang is with the Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, CA, 90089 USA. Email: larrywan@usc.edu

[2]Yiwei Lyu is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA. Email: yiweilyu@andrew.cmu.edu

[3]John M. Dolan is with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213 USA. Email: jdolan@andrew.cmu.edu

human model. Then, based on the probed information, the autonomous agent can actively influence other agents for some designated objectives, like higher efficiency or better driving experience. We propose our approach under some very mild assumptions, making it transferable to various human-robot interaction scenarios. Our key contributions include: 1) a coherent theoretical framework that unifies the probing and influencing procedures; 2) a proven solvable trajectory-planning optimization; 3) two case studies as application examples in the domain of autonomous driving with numerical simulations used to demonstrate the precision of probing results and efficacy in creating better participant driving experience during influence.

II. Related Work

To exert influence on humans, an autonomous agent would have to interact with different human agents in close proximity, who are heterogeneous agents that differ significantly in their internal models, to which the autonomous agent does not have direct access [20]–[22]. Such an internal model might characterize human's intentions, preferences, objectives, strategies, etc. Works in robotics and perception have focused on estimating these internal models using algorithms based on observations of human's actions, such as intent-driven behavior prediction [23]–[30], inverse reinforcement learning (IRL) [31]–[35], hidden model prediction [36], [37], affective state estimation [38], and activity recognition [39]. Although the human model derived from the above methods performs generally reasonably, it might not capture specific characteristics of the human agent that the autonomous agent is interacting with. The autonomous agent, therefore, needs an online procedure to learn the model specially tailored to the human agent that the autonomous agent is interacting with.

Some online approaches frame this problem as a Partially Observable Markov Decision Process (POMDP) [40]–[42], in which the autonomous agent parameterize the human's intent through a model, inferred through Markovian or Bayesian estimation of the hidden parameters of the internal models from observations of the physical states of the world [43]–[46]. In this paradigm, the autonomous agent is mainly a reactive agent in the environment to observe, which hugely sacrifices the robot's agency to initiate action to actively reveal information about the human.

Some existing works enable active probing for interactive motion planning by incorporating a heuristic active information gathering objective, e.g., information entropy, into the autonomous agent's trajectory optimization framework for human value function parameter estimation [47], [48]. Building upon this work, we allow the autonomous agent to optimize the information radius, i.e., the cohesion between two beliefs, relative to its latest belief of the human model, so instead of having a fixed reference belief like in [48] the autonomous agent aims to maximize the information radius relative to a dynamic reference, its current belief, at every time iteration.

III. Theory

A. Human-Robot Joint Dynamics

For all notations below, we use subscripts to denote the time step and superscripts to capture the attributes' ownership (human or robot). In a human-robot joint system, we define the state vector as $s_t \in \mathbb{R}^n$, the robot's input vector as $u_t^{\mathcal{R}} \in \mathbb{U}^{\mathcal{R}} \subseteq \mathbb{R}^{m^{\mathcal{R}}}$, confined to admissible control space $\mathbb{U}^{\mathcal{R}}$, the human's input vector as $u_t^{\mathcal{H}} \in \mathbb{U}^{\mathcal{H}} \subseteq \mathbb{R}^{m^{\mathcal{H}}}$, confined to admissible control space $\mathbb{U}^{\mathcal{H}}$, and finally the discrete-time control-affine dynamics of the joint system as

$$s_{t+1} = f(s_t) + M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}} + M^{\mathcal{H}}(s_t)u_t^{\mathcal{H}} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ captures the non-linear autonomous dynamics and $M^{\mathcal{R}} : \mathbb{R}^n \to \mathbb{R}^{n\times m^{\mathcal{R}}}$ and $M^{\mathcal{H}} : \mathbb{R}^n \to \mathbb{R}^{n\times m^{\mathcal{H}}}$ are state-dependent input transformation matrices for robot and human respectively [47].

B. Belief Update

The autonomous agent possesses a belief of φ that characterizes the human agent's utility function $r_\varphi^{\mathcal{H}} : \mathbb{R}^n \to \mathbb{R}$. For driving scenarios, a typical φ could characterize the desired velocity of the human vehicle, and a typical $r_\varphi^{\mathcal{H}}$ would include features like safety and speed. We generalize the autonomous agent's belief by proposing a non-parametric representation which can approximate a wider range of distributions. At time t, belief of φ is defined as bel_t with finite domain space Φ. The autonomous agent updates this belief via a particle-filtering recursion [49]

$$bel_{t+1}(\varphi) \propto bel_t(\varphi) \cdot p(u_t^{\mathcal{H}}|s_t, u_t^{\mathcal{R}}, \varphi),\ \forall\varphi \in \Phi \tag{2}$$

where the conditional probability is obtained through a softmax operation based on the Boltzmann model of exponential likeliness of human actions with greater utility [29], [34]

$$p(u_t^{\mathcal{H}}|s_t, u_t^{\mathcal{R}}, \varphi) = \frac{e^{r_\varphi^{\mathcal{H}}\left(f(s_t)+M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}}+M^{\mathcal{H}}(s_t)u_t^{\mathcal{H}}\right)}}{\sum_{\tilde{u}_t^{\mathcal{H}}\in\mathbb{U}^{\mathcal{H}}} e^{r_\varphi^{\mathcal{H}}\left(f(s_t)+M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}}+M^{\mathcal{H}}(s_t)\tilde{u}_t^{\mathcal{H}}\right)}} \tag{3}$$

in which $\mathbb{U}^{\mathcal{H}}$ is discretized for softmax normalization. The complete belief update algorithm is shown in algorithm 1.

Algorithm 1 Belief Update

Input: $bel_t, s_t, u_t^{\mathcal{R}}, u_t^{\mathcal{H}}$

$\eta \leftarrow 0$
for all $\varphi \in \Phi$ **do**
  $r \leftarrow e^{r_\varphi^{\mathcal{H}}\left(f(s_t)+M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}}+M^{\mathcal{H}}(s_t)u_t^{\mathcal{H}}\right)}$ {Boltzmann}
  $\tilde{r} \leftarrow \sum_{\tilde{u}_t^{\mathcal{H}}\in\mathbb{U}^{\mathcal{H}}} e^{r_\varphi^{\mathcal{H}}\left(f(s_t)+M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}}+M^{\mathcal{H}}(s_t)\tilde{u}_t^{\mathcal{H}}\right)}$
  $bel_{t+1}(\varphi) \leftarrow bel_t(\varphi) \cdot \frac{r}{\tilde{r}}$ {belief update}
  $\eta \leftarrow \eta + bel_{t+1}(\varphi)$
end for
for all $\varphi \in \Phi$ **do**
  $bel_{t+1}(\varphi) \leftarrow \frac{bel_{t+1}(\varphi)}{\eta}$ {belief normalization}
end for
return bel^{t+1}

C. Probing

The motivation behind probing is to allow the autonomous agent to actively interact with the human agent to reveal more information that was previously unavailable, meaning that the autonomous agent should choose actions that best challenge its current belief at every time step. Quantitatively, the autonomous agent chooses actions that maximize the information radius between its current belief and the projected belief if such actions are to be executed.

We introduce the Jensen-Shannon divergence (JSD) as a measure of information radius to quantify the cohesion between two beliefs, bel_a and bel_b [50], [51]

$$D_{\mathrm{JS}}[bel_a, bel_b] = \frac{D_{\mathrm{KL}}\left[bel_a : \overline{bel}_{a,b}\right] + D_{\mathrm{KL}}\left[bel_b : \overline{bel}_{a,b}\right]}{2} \tag{4}$$

where D_{KL} is the Kullback–Leibler divergence (KLD) [52], [53] and $\overline{bel}_{a,b}$ is the arithmetic mixture of bel_a and bel_b

$$D_{\mathrm{KL}}[bel_a : \overline{bel}_{a,b}] = \mathop{\mathbb{E}}_{\varphi\sim bel_a} \log\left(\frac{2\cdot bel_a(\varphi)}{bel_a(\varphi)+bel_b(\varphi)}\right) \tag{5}$$

At state s_t, the autonomous agent predicts how the human agent, characterized by φ, will react to its action $u_t^{\mathcal{R}}$ using

$$Q(s_t, u_t^{\mathcal{R}}, \varphi) = \mathop{\arg\max}_{\tilde{u}_t^{\mathcal{H}}\in\mathbb{U}^{\mathcal{R}}} r_\varphi^{\mathcal{H}}(f(s_t)+M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}}+M^{\mathcal{H}}(s_t)\tilde{u}_t^{\mathcal{H}}) \tag{6}$$

We solve the probing problem using Model Predictive Control (MPC), in which the autonomous agent chooses a sequence of actions that optimizes the JSD between the current belief and the projected belief at finite horizon T

$$\max_{u_{0:T-1}^{\mathcal{R}}} \mathop{\mathbb{E}}_{\varphi\sim bel_0} \sum_{t=0}^{T-1} D_{\mathrm{JS}}[bel_0, bel_{t+1}] - D_{\mathrm{JS}}[bel_0, bel_t] \tag{7a}$$

$$\text{s.t.} \quad s_0 = s_t, bel_0 = bel_t \tag{7b}$$

$$u_t^{\mathcal{H}} = Q(s_t, u_t^{\mathcal{R}}, \varphi) \tag{7c}$$

$$s_{t+1} = f(s_t) + M^{\mathcal{R}}(s_t)u_t^{\mathcal{R}} + M^{\mathcal{H}}(s_t)u_t^{\mathcal{H}} \tag{7d}$$

$$bel_{t+1}(\varphi) \propto bel_t(\varphi)\cdot p(u_t^{\mathcal{H}}|s_t, u_t^{\mathcal{R}}, \varphi) \tag{7e}$$

To ensure solvability, we prove that D_{JS}, which maps to $[0,\infty)$ in theory, is upper bounded in optimization (7).

Proof. **Boundedness**:

We first make a slight assumption that bel_0 is bounded and has compact support, hence

$$\sup_{\varphi\in\Phi} bel_0(\varphi) < \infty \wedge \inf_{\varphi\in\Phi} bel_0(\varphi) > 0 \tag{8}$$

which helps to substantiate the boundedness of KLD [54]. We will initialize the belief such that condition (8) is satisfied in section IV.

We hypothesize inductively that $\forall a \in \{0,\dots,T-1\}$, $\sup_{\varphi\in\Phi} bel_a(\varphi) < \infty$. Since $p(u_t^{\mathcal{H}}|s_t, u_t^{\mathcal{R}}, \varphi)$ maps to an image of $(0,1)$, using condition (8) as base case, we have

$$\sup_{\varphi\in\Phi} bel_a(\varphi) < 1 < \infty,\ \forall a \in \{0,\dots,T\} \tag{9}$$

By similar induction technique, we have

$$\inf_{\varphi\in\Phi} bel_a(\varphi) > 0,\ \forall a \in \{0,\dots,T\} \tag{10}$$

Hence, we have extended condition (8) to

$$\sup_{\varphi\in\Phi} bel_a(\varphi) < \infty \wedge \inf_{\varphi\in\Phi} bel_a(\varphi) > 0,\ \forall a \in \{0,\dots,T\} \tag{11}$$

Therefore, $\forall a \in \{0,\dots,T\}$, $\exists \bar{s} = \sup_{\varphi\in\Phi} bel_a(\varphi)$ such that $0 < \bar{s} < \infty$. Similarly, $\forall a, b \in \{0,\dots,T\}$, $\exists \underline{\mathrm{i}} = \inf_{\varphi\in\Phi} bel_a(\varphi) + bel_b(\varphi)$ such that $0 < \underline{\mathrm{i}} < \infty$.

Therefore, by equation (5), we have $\forall a, b \in \{0,\dots,T\}$

$$\begin{aligned} D_{\mathrm{KL}}[bel_a : \overline{bel}_{a,b}] &= \mathop{\mathbb{E}}_{\varphi\sim bel_a} \log\left(\frac{2\cdot bel_a(\varphi)}{bel_a(\varphi)+bel_b(\varphi)}\right) \\ &\le \mathop{\mathbb{E}}_{\varphi\sim bel_a} \sup_{\varphi\in\Phi} \log\left(\frac{2\cdot bel_a(\varphi)}{bel_a(\varphi)+bel_b(\varphi)}\right) \\ &\le \log(2\cdot\bar{s}) - \log(\underline{\mathrm{i}}) < \infty \end{aligned} \tag{12}$$

By symmetry, $D_{\mathrm{KL}}[bel_b : \overline{bel}_{a,b}] < \infty$ can be easily proved using the same technique, which together concludes the boundedness of JSD. □

We adopt a dynamic-programming-based approach to optimize equation (7), while other quasi-Newton methods, like the BFGS algorithm [55]–[58], are also applicable. Although the computational complexity grows exponentially with respect to the state dimension, the high parallelizability of equation (7d) and (7e) can attenuate the curse of dimensionality. Moreover, we argue that successfully reasoning about human-robot interactions over a short horizon does not require a full-fidelity model of the joint dynamics, so highly informative insights can still be obtained tractably via approximation. We define the value function of executing n consecutive controls starting from time k as

$$V(k,n) = \mathop{\mathbb{E}}_{\varphi\sim bel_0} \sum_{t=k}^{k+n-1} D_{\mathrm{JS}}[bel_0, bel_{t+1}] - D_{\mathrm{JS}}[bel_0, bel_t] \tag{13}$$

The value function on the horizon therefore satisfies

$$\begin{aligned} V(0,T) &= \mathop{\mathbb{E}}_{\varphi\sim bel_0} \sum_{t=0}^{k-1} D_{\mathrm{JS}}[bel_0, bel_{t+1}] - D_{\mathrm{JS}}[bel_0, bel_t] \\ &\quad + \mathop{\mathbb{E}}_{\varphi\sim bel_0} \sum_{t=k}^{T-1} D_{\mathrm{JS}}[bel_0, bel_{t+1}] - D_{\mathrm{JS}}[bel_0, bel_t] \\ &= V(0,k) + V(k,T-k),\ \forall k \in \{0,\dots,T\} \end{aligned} \tag{14}$$

which shows that the path-dependency fits a Bellman optimality equation [59]. The optimal value function and control policy can therefore be obtained in polynomial time by backtracking the Hamilton–Jacobi–Bellman (HJB) equation [60]

$$V(t,T-t) = \max_{u_t^{\mathcal{R}}\in\mathbb{U}^{\mathcal{R}}} \Big\{ V(t,1) + V(t+1,T-t-1) \Big\} \tag{15}$$

Following this policy, the autonomous agent interactively probes the human agent and gradually converges its belief until the change of JSD is too small. The autonomous agent then chooses $\hat{\varphi}$, which could be a linear combination of all $\varphi \in \Phi$ weighted by their $bel(\varphi)$ or simply the most likely $\varphi \in \Phi$, as the human model parameter.

D. Influence

We characterize an influence as a sequence of atomic objectives, each with a utility function, that accounts for a major influence if all executed in order, and we delegate the responsibility of planning these atomic objectives to some high-level planner. For each objective, we incorporate $\hat{\varphi}$ into the utility function for both robot and human.

$$\max_{u^{\mathcal{R}}_{0:T-1}} \sum_{t=0}^{T-1} r^{\mathcal{R}}_{\hat{\varphi}(s_{t+1})} \tag{16a}$$

$$\text{s.t.} \quad s_0 = s_t \tag{16b}$$

$$u^{\mathcal{H}}_t = Q(s_t, u^{\mathcal{R}}_t, \hat{\varphi}) \tag{16c}$$

$$s_{t+1} = f(s_t) + M^{\mathcal{R}}(s_t)u^{\mathcal{R}}_t + M^{\mathcal{H}}(s_t)u^{\mathcal{H}}_t \tag{16d}$$

Similarly, this optimization problem can be solved using HJB recursion in polynomial time.

IV. Simulation

In this section, we present two car-following-based scenarios in which probing and influencing can be used to facilitate better participant experience and optimality for human drivers. Both scenarios start with the human vehicle following the autonomous vehicle.

A. Ground Truth

To generate the ground truth trajectories for the human-driven vehicle, we use the intelligent driver model (IDM) [61]–[63], which is known to accurately imitate human driving behaviors.

$$u^{\mathcal{H}} = u_{\max}\left[1 - \left(\frac{v^{\mathcal{H}}}{v_{\text{des}}}\right)^4 - \left(\frac{d_{\text{des}}}{x^{\mathcal{R}} - x^{\mathcal{H}}}\right)^2\right] \tag{17}$$

in which

$$d_{\text{des}} = d_{\min} + \tau_{\text{gap}} \cdot v^{\mathcal{H}} - \frac{v^{\mathcal{H}} \cdot (v^{\mathcal{H}} - v^{\mathcal{R}})}{2\sqrt{a_{\max} \cdot b_{\text{pref}}}} \tag{18}$$

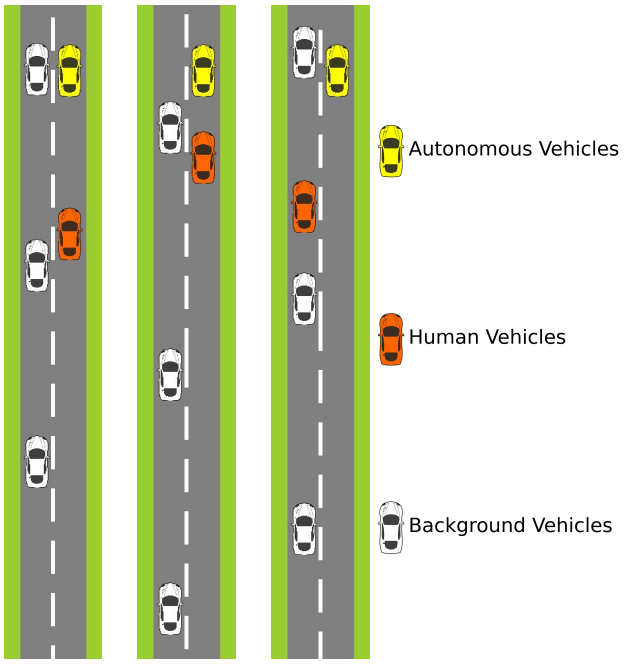
where superscripted notations are system dynamics and subscripted notations are constant parameters. We assume that humans will maintain their driving style, so the constant parameters are static over time. Without loss of generality, we also use IDM to model other background vehicles in the environment. We simulate all vehicles as mass points.

B. Exploitation and Exploration

To balance exploitation and exploration, the autonomous vehicle alternates between $5\,\text{s}$ of passive observation and $5\,\text{s}$ of active probing. We also set the MPC horizon to $5\,\text{s}$. Thanks to the boundedness of JSD, we can add a safety objective, $\lambda \cdot r^{\mathcal{R}}_{safe}(s_{t+1})$, on the autonomous agent's optimization to enforce some safety features, and we choose λ empirically.

C. Human Model

The autonomous vehicle models the human underlying utility using a combination of features, namely desired headway and desired velocity. For each scenario, we choose $|\Phi| = 30$ such that each $\varphi \in \Phi$ maps to a distinct desired velocity or desired headway, and we initialize them to a uniform distribution, which satisfies condition (8).



(a) Phase 1 (b) Phase 2 (c) Phase 3

Fig. 2: Phase 1: Autonomous vehicle maintains velocity. Phase 2: Autonomous vehicle brakes to block the human vehicle. Phase 3: Human vehicle merges due to blocking. All vehicles travel upwards.

D. Scenario 1: Influence fast drivers to switch lane

Consider a two-lane highway (Fig. 2a) with an inner lane (left) and an outer lane (right). Here, we cause the autonomous vehicle to actively probe the desired velocity of the human vehicle. If the human vehicle exhibits the intention to travel at a high velocity, the autonomous vehicle will perform a series of maneuvers to help the human vehicle merge to the inner lane in the widest gap between the background vehicles. While approaching the widest gap, the autonomous vehicle slows down to block the human vehicle (Fig. 2b), and the human vehicle switches lanes shortly after that (Fig. 2c).

We choose the IDM parameters as $u_{\max} = 0.73\,\text{m/s}^2$, $b_{\text{pref}} = 1.67\,\text{m/s}^2$, $v_{\text{des}} = 25\,\text{m/s}$, $\tau_{\text{gap}} = 1.5\,\text{s}$, and $d_{\min} = 2\,\text{m}$. We start the car-following scenario with relative headway of $100\,\text{m}$, the autonomous vehicle and the human vehicle both traveling at $20\,\text{m/s}$. We also included a passive observing approach to compare as a baseline. Fig. 3 is a snapshot of the belief from two approaches taken every $10\,\text{s}$.

By $50\,\text{s}$, the active approach's peak happens at φ_{19}, which maps to a desired velocity of $23.56\,\text{m/s}$, which is close to the IDM parameter, v_{des}, of $25\,\text{m/s}$. In comparison, the passive observation baseline peaks at φ_{16} that maps to $19.86\,\text{m/s}$, which is very far from the ground truth. This is because the passive approach suffers from no exploration to trigger human reaction, so the autonomous vehicle will assume the human vehicle intends to travel only as fast as itself.

Leveraging this probed information, the autonomous vehicle can set a cutoff, $23\,\text{m/s}$ in our simulation for instance, to influence the humans with high desired velocity to drive in

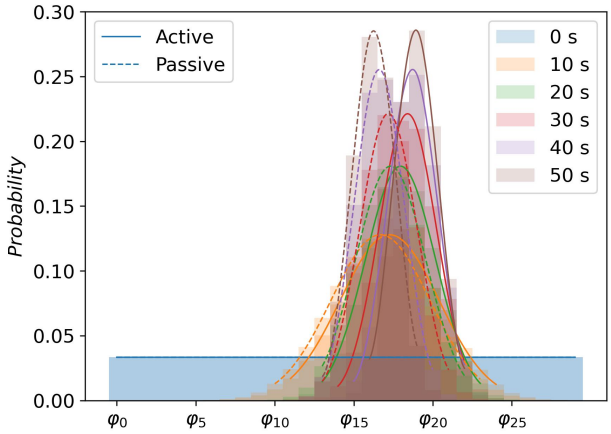


Fig. 3: Belief Snapshot

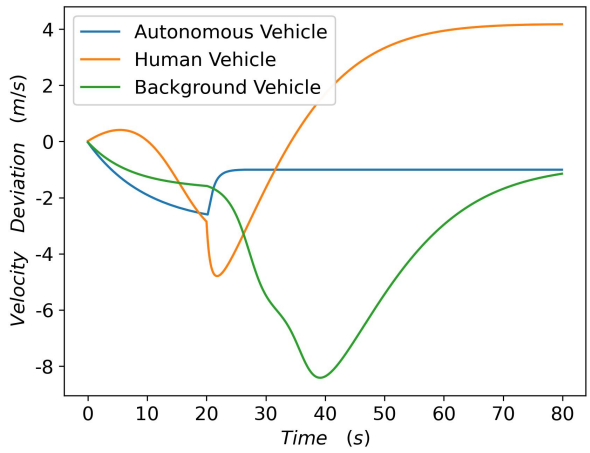


Fig. 4: Velocity Deviation

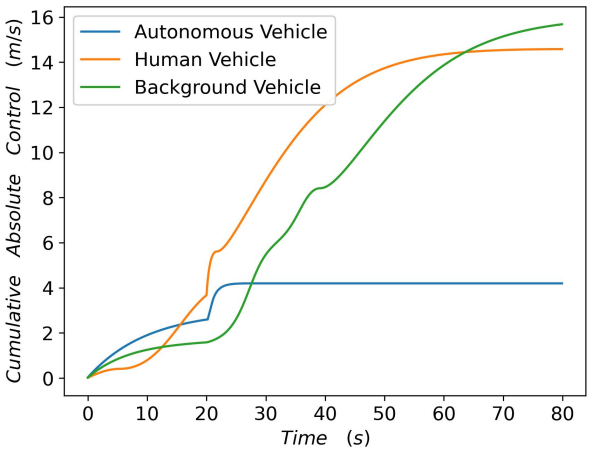
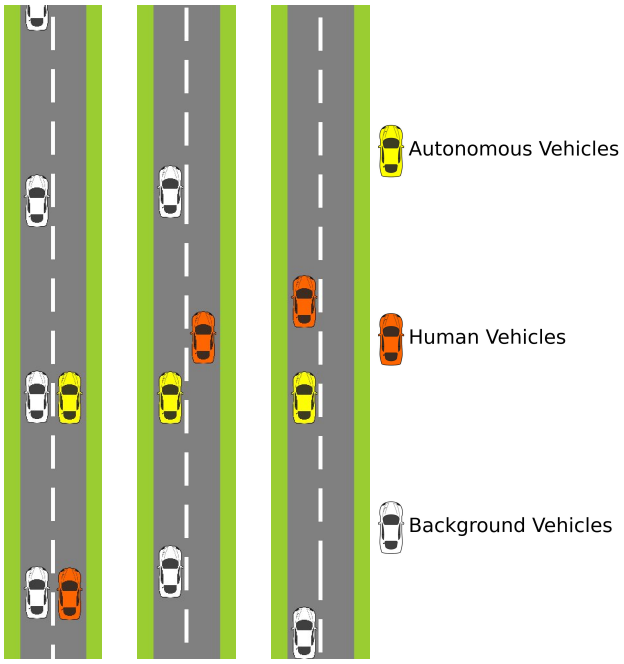


Fig. 5: Cumulative Absolute Control



(a) Phase 1 (b) Phase 2 (c) Phase 3

Fig. 6: Phase 1: Autonomous vehicle merges first. Phase 2: Autonomous vehicle slows down to create gap for human vehicle. Phase 3: Human vehicle merges. All vehicles travel upwards.

the inner lane. According to Fig. 4, the influence brought about 20.04% increase in the human vehicle's velocity, whereas the passive approach wouldn't be able to initiate the influence procedure at all because it does not try to reveal information that the human is not showing, hence the autonomous vehicle becomes more and more wrongly convinced that the human vehicle intends to travel only as fast as $19.86\,\mathrm{m/s}$. According to Fig. 5, the influence introduces a bounded perturbation, about $15.68\,\mathrm{m/s}$ of cumulative absolute control, on average background vehicles, which could be easily attenuated with autonomous vehicles using flow stopper techniques [10], [64].

E. Scenario 2: Helping human to switch lane

Consider a scenario like Fig. 6a, in which the lane the autonomous and human vehicle currently occupy is about to end, either due to traffic, construction, or lane merge. Both vehicles, therefore, have to switch to the left lane, which is occupied by some background vehicles. Assume the headway gaps between the background vehicles are too narrow for humans while traveling at such a high speed. Fortunately, autonomous vehicles are capable of performing the switching. The autonomous vehicle, therefore, helps the human vehicle to switch lanes by first probing the desired headway of the human vehicle around a specific velocity, in this case $20\,\mathrm{m/s}$. The autonomous vehicle will then switch lanes and slow down to create enough gap based on the probed headway (Fig. 6b). Finally, the human vehicle can merge into the lane with ease (Fig. 6c).

We choose the IDM parameters as $u_{\max} = 0.73\,\mathrm{m/s^2}$, $b_{\mathrm{pref}} = 1.67\,\mathrm{m/s^2}$, $v_{\mathrm{des}} = 20\,\mathrm{m/s}$, $\tau_{\mathrm{gap}} = 1.5\,\mathrm{s}$, and $d_{\min} = 2\,\mathrm{m}$. Similarly, we initialize the road condition to the same

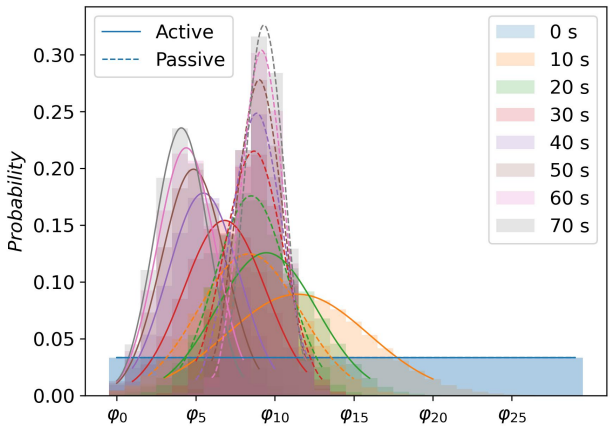


Fig. 7: Belief Snapshot

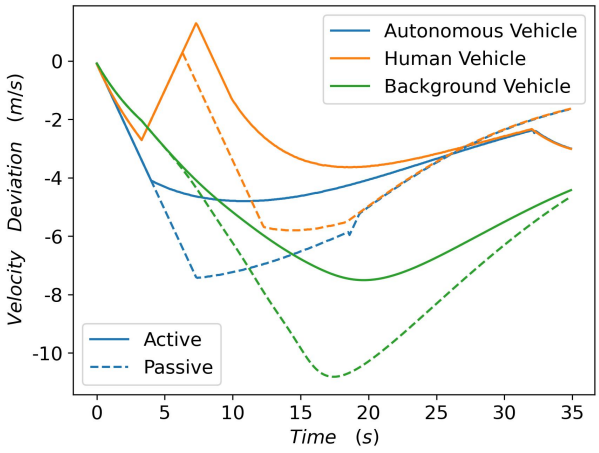


Fig. 8: Velocity Deviation

condition as the previous scenario, and we include a passive observing approach to compare as a baseline.

Fig. 7 is a snapshot of the belief from two approaches taken every $10\,\mathrm{s}$. By $70\,\mathrm{s}$, the probability for the active approach peaks at φ_4, which maps to a desirable headway around $48.27\,\mathrm{m}$, whereas that of passive approach peaks at φ_9, which maps to a desirable headway around $108.62\,\mathrm{m}$. For reference, according to data from the Next Generation Simulation for US Highway 101 [65], the average headway for cars traveling around $20\,\mathrm{m/s}$ is about $42.18\,\mathrm{m}$. Although not absolutely precise, the active approach generates a much more accurate profile than the passive approach does.

Based on the probed information, the autonomous vehicle can proceed to create a gap for the human vehicle. For comparison, we simulated a baseline where the autonomous vehicle is passive during the information gathering process, so the autonomous vehicles would have to slow down to create a wider gap, inducing larger perturbations on the background vehicles. According to Fig. 9, the cumulative absolute control for all three types of vehicles in the active approach is significantly lower than that in the passive approach. The reductions in perturbation are respectively

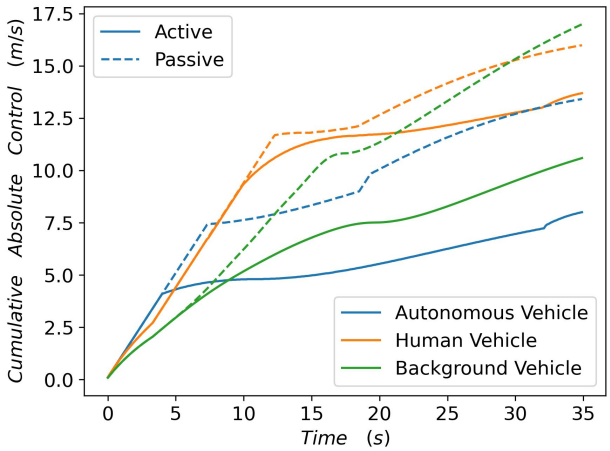


Fig. 9: Cumulative Absolute Control

40.36%, 14.33%, and 37.66% for autonomous, human, and background vehicle. According to Fig. 8, the active approach generates less extreme velocity deviation for all three types of vehicles in general, which helps to reduce the intensity and propagation of traffic wave [66].

Moreover, our baseline is under the assumption that the autonomous vehicle would overtake under this scenario. Without active probing, the autonomous vehicle is more likely to behave quite conservatively, so it will most likely wait until all of the background vehicles have passed to switch lanes. This subjects the autonomous and human vehicles to almost a complete stop and a wait time that depends on the number of consecutive closely spaced background vehicles behind, meaning that the deviation will continue to increase if there is no large gap. Our active probing and influencing approach, on the other hand, is agnostic to this condition because the autonomous vehicle creates its own lane-change opportunity.

V. Conclusions

In this work, we present an active probing approach for an autonomous agent to actively interact with a human agent to reveal information about a human's underlying utility and internal model. Our simulation results in autonomous driving demonstrate how the gathered information can be leveraged to increase driver experience and overall optimality compared to a passive learning baseline method. Future work could adopt learning-based methods to replace the heuristic probing objective with a more efficient and scenario-specific objective. It could also be worthwhile to relax the static human model assumption, hence empowering the autonomous agent to actively learn the human's adaptation policy.

VI. Acknowledgement

This work was supported by the CMU Argo AI Center for Autonomous Vehicle Research.

Special thanks to Mrinal Verghese and Bhaskar Krishnamachari for their insightful suggestions, Rachel Burcin for her hospitality, and the 2022 Robotics Institute Summer Scholars for their company.

REFERENCES

[1] Y. Lyu, W. Luo, and J. M. Dolan, "Probabilistic safety-assured adaptive merging control for autonomous vehicles," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 10 764–10 770.

[2] S. Van Koevering, Y. Lyu, W. Luo, and J. Dolan, "Provable probabilistic safety and feasibility-assured control for autonomous vehicles using exponential control barrier functions," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 952–957.

[3] S. Kato, S. Tsugawa, K. Tokuda, T. Matsui, and H. Fujii, "Vehicle control algorithms for cooperative driving with automated vehicles and intervehicle communications," *IEEE Transactions on intelligent transportation systems*, vol. 3, no. 3, pp. 155–161, 2002.

[4] J. Baber, J. Kolodko, T. Noel, M. Parent, and L. Vlacic, "Cooperative autonomous driving: intelligent vehicles sharing city roads," *IEEE Robotics & Automation Magazine*, vol. 12, no. 1, pp. 44–49, 2005.

[5] H. Stübing, M. Bechler, D. Heussner, T. May, I. Radusch, H. Rechner, and P. Vogel, "simtd: a car-to-x system architecture for field operational tests [topics in automotive networking]," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 148–154, 2010.

[6] A. De La Fortelle, X. Qian, S. Diemer, J. Grégoire, F. Moutarde, S. Bonnabel, A. Marjovi, A. Martinoli, I. Llatser, A. Festag *et al.*, "Network of automated vehicles: the autonet 2030 vision," in *ITS World Congress*, 2014.

[7] L. Hobert, A. Festag, I. Llatser, L. Altomare, F. Visintainer, and A. Kovacs, "Enhancements of v2x communication in support of cooperative autonomous driving," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 64–70, 2015.

[8] B. Toghi, R. Valiente, D. Sadigh, R. Pedarsani, and Y. P. Fallah, "Cooperative autonomous vehicles that sympathize with human drivers," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4517–4524.

[9] 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.

[10] C. Wu, A. M. Bayen, and A. Mehta, "Stabilizing traffic with autonomous vehicles," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 6012–6018.

[11] A. R. Kreidieh, C. Wu, and A. M. Bayen, "Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 1475–1480.

[12] C. Wu, A. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: A modular learning framework for autonomy in traffic," *arXiv preprint arXiv:1710.05465*, 2017.

[13] N. Kheterpal, K. Parvate, C. Wu, A. Kreidieh, E. Vinitsky, and A. Bayen, "Flow: Deep reinforcement learning for control in sumo," *EPiC Series in Engineering*, vol. 2, pp. 134–151, 2018.

[14] J. Grover, N. Mohanty, C. Liu, W. Luo, and K. Sycara, "Noncooperative herding with control barrier functions: Theory and experiments," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 80–86.

[15] N. Mohanty, J. Grover, C. Liu, and K. Sycara, "Distributed multirobot control for non-cooperative herding," *arXiv preprint arXiv:2301.03293*, 2023.

[16] W. Lee and D. Kim, "Autonomous shepherding behaviors of multiple target steering robots," *Sensors*, vol. 17, no. 12, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/12/2729

[17] A. Pierson and M. Schwager, "Controlling noncooperative herds with robotic herders," *IEEE Transactions on Robotics*, vol. 34, no. 2, pp. 517–525, 2017.

[18] R. Vaughan, N. Sumpter, J. Henderson, A. Frost, and S. Cameron, "Robot control of animal flocks," in *Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC) held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA) Intell*. IEEE, 1998, pp. 277–282.

[19] ——, "Experiments in automatic flock control," *Robotics and autonomous systems*, vol. 31, no. 1-2, pp. 109–117, 2000.

[20] H. G. Stassen, G. Johannsen, and N. Moray, "Internal representation, internal model, human performance model and mental workload," *Automatica*, vol. 26, no. 4, pp. 811–820, 1990.

[21] R. J. Jagacinski and R. A. Miller, "Describing the human operator's internal model of a dynamic system," *Human Factors*, vol. 20, no. 4, pp. 425–433, 1978.

[22] Y. Lyu, W. Luo, and J. M. Dolan, "Responsibility-associated multi-agent collision avoidance with social preferences," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3645–3651.

[23] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, "Planning-based prediction for pedestrians," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3931–3936.

[24] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *2012 IEEE intelligent vehicles symposium*. IEEE, 2012, pp. 1162–1167.

[25] A. D. Dragan and S. S. Srinivasa, *Formalizing assistive teleoperation*. MIT Press, July, 2012.

[26] M. Awais and D. Henrich, "Human-robot collaboration by intention recognition using probabilistic state machines," in *19th International Workshop on Robotics in Alpe-Adria-Danube Region (RAAD 2010)*. IEEE, 2010, pp. 75–80.

[27] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.

[28] T.-H. D. Nguyen, D. Hsu, W.-S. Lee, T.-Y. Leong, L. P. Kaelbling, T. Lozano-Perez, and A. H. Grant, "Capir: Collaborative action planning with intention recognition," in *Seventh Artificial Intelligence and Interactive Digital Entertainment Conference*, 2011.

[29] R. D. Luce, *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.

[30] Y. Lyu, C. Dong, and J. M. Dolan, "Fg-gmm-based interactive behavior estimation for autonomous driving vehicles in ramp merging control," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1250–1255.

[31] P. Abbeel and A. Y. Ng, "Exploration and apprenticeship learning in reinforcement learning," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 1–8.

[32] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," *arXiv preprint arXiv:1206.4617*, 2012.

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

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

[35] D. Ramachandran and E. Amir, "Bayesian inverse reinforcement learning." in *IJCAI*, vol. 7, 2007, pp. 2586–2591.

[36] C.-P. Lam, A. Y. Yang, and S. S. Sastry, "An efficient algorithm for discrete-time hidden mode stochastic hybrid systems," in *2015 European Control Conference (ECC)*. IEEE, 2015, pp. 1212–1218.

[37] Y. Lyu, W. Luo, and J. M. Dolan, "Adaptive safe merging control for heterogeneous autonomous vehicles using parametric control barrier functions," in *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, pp. 542–547.

[38] D. Kulic and E. A. Croft, "Affective state estimation for human–robot interaction," *IEEE transactions on robotics*, vol. 23, no. 5, pp. 991–1000, 2007.

[39] T. Van Kasteren, A. Noulas, G. Englebienne, and B. Kröse, "Accurate activity recognition in a home setting," in *Proceedings of the 10th international conference on Ubiquitous computing*, 2008, pp. 1–9.

[40] S. Javdani, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization," *Robotics science and systems: online proceedings*, vol. 2015, 2015.

[41] A. Fern, S. Natarajan, K. Judah, and P. Tadepalli, "A decision-theoretic model of assistance," *Journal of Artificial Intelligence Research*, vol. 50, pp. 71–104, 2014.

[42] T. Bandyopadhyay, K. S. Won, E. Frazzoli, D. Hsu, W. S. Lee, and D. Rus, "Intention-aware motion planning," in *Algorithmic foundations of robotics X*. Springer, 2013, pp. 475–491.

[43] J. F. Fisac, A. Bajcsy, S. L. Herbert, D. Fridovich-Keil, S. Wang, C. J. Tomlin, and A. D. Dragan, "Probabilistically safe robot planning with confidence-based human predictions," *arXiv preprint arXiv:1806.00109*, 2018.

[44] H. Hu, K. Nakamura, and J. F. Fisac, "Sharp: Shielding-aware robust planning for safe and efficient human-robot interaction," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 5591–5598, 2022.

[45] R. Tian, L. Sun, A. Bajcsy, M. Tomizuka, and A. D. Dragan, "Safety assurances for human-robot interaction via confidence-aware game-theoretic human models," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 11 229–11 235.

[46] R. P. Bhattacharyya, R. Senanayake, K. Brown, and M. J. Kochenderfer, "Online parameter estimation for human driver behavior prediction," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 301–306.

[47] H. Hu and J. F. Fisac, "Active uncertainty learning for human-robot interaction: An implicit dual control approach," *arXiv preprint arXiv:2202.07720*, 2022.

[48] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 66–73.

[49] S. Thrun, "Probabilistic robotics," *Communications of the ACM*, vol. 45, no. 3, pp. 52–57, 2002.

[50] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information theory*, vol. 37, no. 1, pp. 145–151, 1991.

[51] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969.

[52] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[53] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[54] F. Nielsen, "On a generalization of the jensen–shannon divergence and the jensen–shannon centroid," *Entropy*, vol. 22, no. 2, p. 221, 2020.

[55] C. G. Broyden, "The convergence of a class of double-rank minimization algorithms 1. general considerations," *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.

[56] R. Fletcher, "A new approach to variable metric algorithms," *The computer journal*, vol. 13, no. 3, pp. 317–322, 1970.

[57] D. Goldfarb, "A family of variable-metric methods derived by variational means," *Mathematics of computation*, vol. 24, no. 109, pp. 23–26, 1970.

[58] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of computation*, vol. 24, no. 111, pp. 647–656, 1970.

[59] R. E. Bellman and S. E. Dreyfus, *Applied dynamic programming*. Princeton university press, 2015, vol. 2050.

[60] R. Bellman, "Dynamic programming," *Science*, vol. 153, no. 3731, pp. 34–37, 1966.

[61] M. Treiber and D. Helbing, "Explanation of observed features of self-organization in traffic flow," *arXiv preprint cond-mat/9901239*, 1999.

[62] M. Treiber and A. Kesting, "Traffic flow dynamics," *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*, pp. 983–1000, 2013.

[63] ——, "The intelligent driver model with stochasticity-new insights into traffic flow oscillations," *Transportation research procedia*, vol. 23, pp. 174–187, 2017.

[64] Y. Sugiyama, M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S.-i. Tadaki, and S. Yukawa, "Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam," *New journal of physics*, vol. 10, no. 3, p. 033001, 2008.

[65] J. Colyar and J. Halkias, "Us highway 101 dataset," *Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-07-030*, pp. 27–69, 2007.

[66] A. R. Kreidieh, C. Wu, and A. M. Bayen, "Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 1475–1480.