

Does Unpredictability Influence Driving Behavior?

Sepehr Samavi¹, Florian Shkurti², and Angela P. Schoellig³

Abstract—In this paper we investigate the effect of the unpredictability of surrounding cars on an ego-car performing a driving maneuver. We use Maximum Entropy Inverse Reinforcement Learning to model reward functions for an ego-car conducting a lane change in a highway setting. We define a new feature based on the unpredictability of surrounding cars and use it in the reward function. We learn two reward functions from human data: a baseline and one that incorporates our defined unpredictability feature, then compare their performance with a quantitative and qualitative evaluation. Our evaluation demonstrates that incorporating the unpredictability feature leads to a better fit of human-generated test data. These results encourage further investigation of the effect of unpredictability on driving behavior.

I. INTRODUCTION

An increasing number of autonomous vehicles (AVs) have started to be deployed in mixed traffic settings, where the AV shares the same road with human-driven vehicles. To be successful in these settings, the AV needs to understand and predict the continuous chain of interactions between drivers. Specifically, the AV needs to be able to predict the behavior of other vehicles to seamlessly interact and arrive at its goal.

In a typical AV autonomy stack, a prediction model generates information about the future of the surrounding environment (e.g. location of adjacent vehicles, pedestrians, road geometry) and supplies that information into a planning module to produce collision-free trajectories [1]–[4]. Therefore, the accuracy of the prediction model is critical for safe AV operation. To this end, there has been much progress in enabling accurate predictions [5]–[7]. However, any model comes with uncertainty, and thus a number of recent methods have focused on quantifying the uncertainty and/or confidence of prediction models [8], [9]. The goal of this uncertainty quantification is to avoid dangerous planning caused by inaccurate predictions by switching to another prediction model [9], or switching to a more conservative planner [10]. While most methods investigate how to incorporate uncertainty into a robot’s autonomy stack, we explore whether predictive uncertainty can provide insight on the trajectories taken by human drivers in scenarios where an adjacent driver is behaving erratically, and thus unpredictably.

Computing resources used in preparing this research were provided, in part, by the Province of Ontario, the Government of Canada through CIFAR, and companies sponsoring the Vector Institute for Artificial Intelligence.

¹University of Toronto Institute for Aerospace Studies (UTIAS) and the Vector Institute. sepehr@robotics.utias.utoronto.ca.

²Department of Computer Science at the University of Toronto florian@cs.toronto.edu.

³Technical University of Munich, UTIAS and the Vector Institute. angela.schoellig@tum.de.

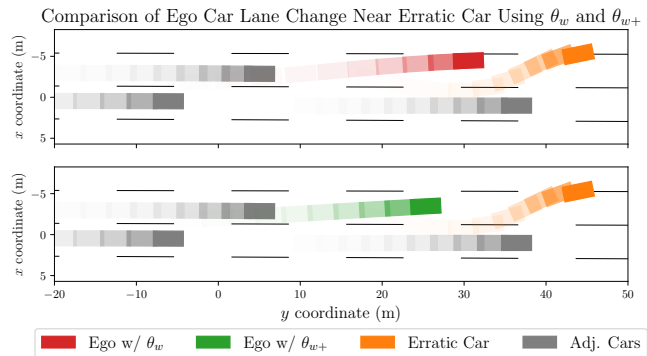


Fig. 1: We compare a snapshot (at $t = 3s$) of ego car lane change trajectories generated using the baseline (red) and the unpredictability-aware (green) models, for identical adjacent cars (gray and orange) and lanes. The history of each vehicle is illustrated with decreasing opacity. The erratic adjacent vehicle is perturbed to zigzag. We observe that the trajectory from the unpredictability-aware model (green) delays changing into the target lane and maintains a higher distance from the erratic car (orange) than the baseline (red). Video at tiny.cc/unpredi.

In this paper we propose *unpredictability* as a feature for a trajectory planning reward function. For an ego-vehicle performing a trajectory, we measure the unpredictability of each interacting adjacent car by using the performance metrics of an off-the-shelf vehicle trajectory prediction model [6]. We assume that if the off-the-shelf model performs poorly on a particular adjacent car at a particular time, then that car is behaving unpredictably and an ego-car should keep its distance. Our intuition is captured in Fig. 1, where we illustrate a snapshot of ego cars (red and green) performing a simulated lane change in an identical scenario involving an erratic adjacent car (orange). In the top figure, we can see that in a trajectory generated without accounting for unpredictability, the ego-car (red) maintains a lower distance from the erratic car compared to the one in the bottom figure where unpredictability is accounted for (green).

The main contribution of this paper is defining a reward feature incorporating unpredictability and analyzing human lane change behavior using an Inverse Reinforcement Learning (IRL) approach. Using IRL, we learn two reward functions for generating lane change trajectories: an unpredictability-aware reward that includes the defined feature and a baseline reward that does not. We demonstrate that, compared to the baseline, the unpredictability-aware reward generates trajectories that are more similar to those in a human-generated dataset. We also analyze the qualitative similarities between trajectories generated by the two rewards and human trajectories in test scenarios with variable unpredictable behavior by adjacent cars. This analysis provides insights into the effect of the unpredictability feature.

II. RELATED WORK

Modelling human driving behavior is explored from different perspectives in AV research.

Prediction and Forecasting: In the classical modular AV design, the output of a prediction model is used in motion planning [1]–[3]. In the field of prediction models for AVs, many works use camera fed directly into Deep Neural Network (DNN) architectures such as Long Short-Term memory (LSTM) recurrent neural network architectures [11]–[13] or Variational Auto Encoders (VAEs) [14]–[16] to forecast the evolution of a scene. Other proposals use similar network architectures with a variety of sensors, such as cameras or LIDAR to represent the surrounding world as a Dynamic Occupancy Grid Map (DOGMa) in 2D [17]–[19] or 3D [20]. A third group of prediction models focus on predicting time-series trajectories (future positions and velocity) [5], [6], [21]. In order to produce more accurate predictions, some works explicitly account for interactions between the agents [5], [6], and even account for the effect of an ego AV’s plans on the predictions of future behavior by other agents [21]. We propose tracking the predictive accuracy of the model proposed in [6], which predicts trajectories of vehicles driving on a highway, as an unpredictability measure that we incorporate in a reward feature we describe in Section III-B.

Imitation Learning Approaches: Other methods focus on imitating human driving styles by learning AV planning models that imitate human behavior from data. A common approach is to learn a parametric reward function represented a weighted linear combination of nonlinear features [22], [23] or a DNN [24] using a Maximum Entropy Inverse Reinforcement Learning (MaxEnt IRL) learning approach [25], then use the reward function in a separate planner to optimize trajectories that imitate human behavior. In contrast, end-to-end methods forgo explicitly learning a reward function and learn policies directly from human data [26], [27].

Feature-design Approaches: In contrast to the proposals presented above that focus on *how* to generate driving behavior, our work focuses on investigating *what* features can be used to generate driving behavior. Among methods that focus on feature design, the authors in [28] propose maximizing the visibility of the ego-car in highway and blind intersection scenarios. In a driving simulation, they demonstrate that a visibility-aware AV slows down before entering another vehicle’s blind spots. In [29], the authors formalize a notion of being courteous to other interacting drivers. They hypothesize that human drivers are in fact also courteous. The authors test their hypothesis using an IRL analysis, in the same way as this paper intends to do. Rather than courtesy, we define the unpredictability of surrounding traffic as a feature to incorporate in trajectory planning.

Uncertainty and Confidence: The concept of unpredictability has also been studied in the context of estimating model uncertainty and confidence for prediction models in AVs and other robotics applications. Uncertainty estimation methods include using Bayesian techniques to track a parameter that governs the variance of predictions [8], comparing the output from an ensemble models [9], and tracking the

difference between predictions and actual outcomes [10], [30]. The motivation behind these methods is to use the estimate to enable downstream planning and control of the AV to avoid dangerous actions caused by low-confidence predictions.

In this paper, however, we investigate whether incorporating an estimate of unpredictability into a trajectory planner could have the additional benefit of generating plans that are more similar to human ones by generating more conservative behavior when faced with an unpredictable adjacent vehicle. The authors in [31] have shown that it is possible to learn a driving policy by maximizing only the predictive accuracy of the behavior of adjacent vehicles as the reward in a Reinforcement Learning (RL) approach. However, their model is exclusively trained and tested in a simulator, and does not use any human data. Therefore, the notion of predictability in their proposal is limited to the driver models that coded in the simulator. We focus on real human trajectory data and compare trajectories generated by using a baseline reward and an unpredictability-aware reward with trajectories from a dataset of human lane changes.

III. PROBLEM FORMULATION

A. Lane Change Planning Problem

We formulate a lane change maneuver as trajectory optimization. Concretely, we model the maneuver as a deterministic discrete-time optimal control task over a finite horizon, K , and continuous state and action spaces. Starting from an initial state, \mathbf{x}_0 , that is known a priori, the optimization problem can be written as,

$$\max_{\mathbf{x}, \mathbf{u}} \sum_{k=0}^{K-1} \mathcal{R}(\mathbf{x}_k, \mathbf{u}_k) \quad (1)$$

$$\text{subject to } \mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k), \quad k = 0 \dots K - 1$$

where $(\mathbf{x}, \mathbf{u}) = ([\mathbf{x}_1^\top, \dots, \mathbf{x}_K^\top]^\top, [\mathbf{u}_0^\top, \dots, \mathbf{u}_{K-1}^\top]^\top)$ are the sequences of states and actions in lifted notation, \mathcal{R} is the reward function, and f is the ego-vehicle dynamics.

System and Dynamics: We model the ego-vehicle with a kinematic unicycle model. The state is $\mathbf{x}_k = [x_k, y_k, \psi_k]^\top \in \mathbb{R}^3$, where (x_k, y_k) is the two-dimensional position of the ego-vehicle on a 2D map and ψ_k is its heading at time step k . The control action is defined as $\mathbf{u}_k = [v_k, \omega_k]^\top \in \mathbb{R}^2$, where v_k is the longitudinal velocity input and ω_k is the steering input of the car. We discretize time using the forward Euler method with a sampling rate of $\delta t = 0.1s$ to obtain the dynamics model,

$$f(\mathbf{x}_k, \mathbf{u}_k) = \mathbf{x}_k + \delta t [v_k \cos(\psi_k), v_k \sin(\psi_k), \omega_k]^\top. \quad (2)$$

Reward Function: The reward function is formulated as a linear combination of nonlinear reward features,

$$\mathcal{R}_k(\mathbf{x}_k, \mathbf{u}_k) = \boldsymbol{\theta}^\top \boldsymbol{\phi}(\mathbf{x}_k, \mathbf{u}_k), \quad (3)$$

where $\boldsymbol{\phi} := [\phi_1, \dots, \phi_p]^\top$ are a set of p features and each feature is defined as a nonlinear function that incorporates elements of the ego-vehicle’s surroundings that are relevant

to performing a lane change maneuver. The parameters $\theta := [\theta_1, \dots, \theta_p]^\top \in \mathbb{R}_{\geq 0}^p$ provide the weight of each feature in ϕ relative to the others.

We reviewed features commonly used for trajectory generation in highway driving. The authors in [32] present a detailed survey of features. The features that compose our baseline reward function are enumerated below,

- 1) *Lateral deviation from target lane* [29] is represented as,

$$\phi_d := -\exp\left(\frac{d_k}{w}\right), \quad (4)$$

where d_k is the distance from the position of the car, \mathbf{x}_k , to the centerline of the target lane. We parametrize the target-lane as a line and w is the average distance between the centerlines of the current-lane and target-lane in the vicinity of the lane change maneuver.

- 2) *Deviation from the mean speed of traffic* is represented as,

$$\phi_v := -(v_k - v_d)^2, \quad (5)$$

where v_d is the mean speed of the 4 closest adjacent vehicles over the entire trajectory.

- 3) *High angular speed* is represented as,

$$\phi_a := -\omega_k^2. \quad (6)$$

- 4) *Time To Collision (TTC) with preceding vehicles* is represented as,

$$\phi_p := -\sum_i h_1(\alpha_{i,k}) \exp\left(-\frac{1}{t_p^2} \frac{\|\mathbf{x}_k - \mathbf{x}_{i,k}\|^2}{v_k^2}\right), \quad (7)$$

where i is the index of the two preceding vehicles, and $\alpha_{i,k}$ is the angle between the heading of the ego-car and the vector $\mathbf{x}_{i,k} - \mathbf{x}_k$. The function $h_1(\alpha_{i,k}) = \exp(-c\alpha_{i,k})$ when $-\frac{\pi}{2} \leq \alpha_{i,k} \leq \frac{\pi}{2}$, and 0 otherwise, restricts penalizing TTC to preceding vehicles that are *in front* of the ego-vehicle. The constants c and t_p are hyperparameters.

- 5) *Time To Collision (TTC) with following vehicle* is represented as,

$$\phi_f := -h_2(\mathbf{x}_k) \exp\left(-\frac{1}{t_f^2} \frac{\|\mathbf{x}_{f,k} - \mathbf{x}_k\|^2}{v_{f,k}^2}\right), \quad (8)$$

where f is the index of the vehicle that follows behind the ego-car in the target lane, and $v_{f,k}$ is its speed. The function $h_2(\mathbf{x}_k) = \frac{d_k^2}{w^2}$ (with d_k as in (4)) restricts penalizing the TTC to the duration when the ego-vehicle is *crossing into* the target lane (i.e. $h_2(\cdot) > 0$), but not after being safely *in front* of the following vehicle.

B. Unpredictability Metric and Reward Feature

Most AVs that operate in human environments, use a prediction model of the other agents' behavior [33]. Our approach to measuring predictability – which we call *unpredictability* – uses the tracked performance of such prediction models throughout a maneuver. We propose to take an off-the-shelf prediction model that has been learned from human

data [6], then to use the performance of that model as a heuristic for measuring predictability.

We calculate the mean Euclidean predictive error of the prediction with respect to the ground truth and use this value as the unpredictability metric. For adjacent vehicle i at time k , we evaluate the predictive error of the prediction made by the model at time $k - t_n$. Then take the average of the error from time $k - t_n$ to k as the unpredictability measure. Concretely, for adjacent car i at time step k the mean Euclidean predictive error is,

$$z_{i,k} := \frac{1}{t_n} \sum_{j=k-t_n+1}^k \sqrt{(x_{i,j} - \hat{x}_{i,j}^{(k-t_n)})^2 + (y_{i,j} - \hat{y}_{i,j}^{(k-t_n)})^2}, \quad (9)$$

where t_n is the number of time steps we look to the past, 0.2s in our implementation, $(\hat{x}_{i,j}^{(k-t_n)}, \hat{y}_{i,j}^{(k-t_n)})$ is the prediction made at time $k - t_n$ by the model [6] of where car i would be at time j , and $(x_{i,j}, y_{i,j})$ is the observed position of car i at time j .

In order to incorporate the measure of unpredictability into our baseline lane change reward function, we augment the reward function with two additional features analogous to ϕ_p and ϕ_f ,

- 7) *Unpredictability-weighted TTC with preceding vehicles.*

$$\phi_{pz} := -\sum_i h_1(\alpha_{i,k}) \exp\left(-\frac{1}{t_p^2} \frac{\|\mathbf{x}_k - \mathbf{x}_{i,k}\|^2 - c_p z_{i,k}^2}{v_k^2}\right), \quad (10)$$

- 8) *Unpredictability-weighted TTC with following vehicle.* Analogously, we define

$$\phi_{fz} := -h_2(\mathbf{x}_k) \exp\left(-\frac{1}{t_f^2} \frac{\|\mathbf{x}_{f,k} - \mathbf{x}_k\|^2 - c_f z_{f,k}^2}{v_{f,k}^2}\right), \quad (11)$$

The intuition behind our formulation is that, if an adjacent vehicle is behaving unpredictably, we expect the driver of the ego-car to weigh a collision with them more highly. As such, we use the unpredictability error (9) multiplied by a constant (c_p and c_f , respectively) as an additional dynamic weight on the baseline features.

IV. EXPERIMENTAL METHODOLOGY

The primary focus of this paper is to investigate the effect of incorporating unpredictability on generated lane change trajectories. We hypothesize that incorporating the unpredictability feature in a lane change planner produces trajectories that better fit to dataset of human lane changes. We use an IRL approach to demonstrate this effect.

Independent Variable: We learn two reward functions using the IRL algorithm that we will describe in Section IV-B, on the same human lane change training data, $\{(\mathbf{x}_e^{(i)}, \mathbf{u}_e^{(i)})\}_{i=0}^N$. We learn the vector $\theta_w \in \mathbb{R}_{\geq 0}^5$ that parametrizes the baseline reward function described in Section III-A, and the vector $\theta_{w+} \in \mathbb{R}_{\geq 0}^7$ that parametrizes the reward function that includes the additional unpredictability feature, described in Section III-B.

Dependent Variable: We generate two sets of trajectories that are optimal under the respective learned reward functions, $\{(\mathbf{x}_w^{(i)}, \mathbf{u}_w^{(i)})\}_{i=0}^{N_{\text{test}}}$ and $\{(\mathbf{x}_{w+}^{(i)}, \mathbf{u}_{w+}^{(i)})\}_{i=0}^{N_{\text{test}}}$, in identical test scenarios. Then, for each test scenario, we compare the two generated trajectories from the learned models with the human expert trajectory from that scenario by calculating Mean Euclidean Error (MEE). For the test dataset, we calculate the Average MEE, $\overline{MEE}_w = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \sum_{k=1}^K \|\mathbf{x}_w^{(i)} - \mathbf{x}_e^{(i)}\|$ for the baseline, and \overline{MEE}_{w+} for the unpredictability-aware model.

A. Lane Change Dataset Preprocessing

We use the human driving datasets, NGSIM US-101 [34] and I-80 [35] from the USA and the highD [36] from Germany, which contain position, time, and lane label information for each vehicle. The I-80 and US-101 datasets are each split into three sections with increasing traffic congestion, $t0$, $t1$, and $t2$.

Extracting Lane Changes: We denoise the dataset using a symmetric exponential moving filter to smooth the paths [37]. For each path in each dataset, we use finite differences to calculate heading and linear and angular velocity. We search the dataset for changes in the lane label of a vehicle, t_{lc} , and extract the trajectory of the ego-car conducting the lane change on the interval $t \in [t_{lc} - 2.0, t_{lc} + 5.0]$, which is the typical duration of lane changes on the datasets [37]. We also extract the paths of four adjacent vehicles: two vehicles preceding the ego-car before and after crossing into the target lane, and two vehicles following the ego-vehicle before and after crossing into the target lane. We characterize the *current-lane* and the *target-lane* by performing linear least squares regression on the paths of all vehicles assigned to each lane across the entire dataset in the vicinity of the lane change. We normalize all position values such that every trajectory begins at $\mathbf{x}_0 = [0, 0, \psi_0]^\top$.

Training, Validation, Testing Split: We perform training, validation, and testing set splitting for the lane change IRL models such that the sets match the split used to train, validate and test the off-the-shelf prediction model [6] that we use for the unpredictability feature described in Section III-B. Only the lane changes used in the training set for the off-the-shelf prediction model are included in our lane change IRL training set.

B. Learning Lane Change Reward Function from Humans

IRL Algorithm: Given training set of N human lane change trajectories, $\{(\mathbf{x}_e^{(i)}, \mathbf{u}_e^{(i)})\}_{i=0}^N$, we obtain a maximum likelihood estimate of reward parameters, $\theta_{(\cdot)}$ [22].

We define a probability distribution over lane change trajectories as,

$$p(\mathbf{u}|\mathbf{x}_0) = \frac{\exp\left(\sum_{k=0}^{K-1} \mathcal{R}_k(\mathbf{x}_k, \mathbf{u}_k)\right)}{\int \exp\left(\sum_{k=0}^{K-1} \mathcal{R}_k(\tilde{\mathbf{x}}_k, \tilde{\mathbf{u}}_k)\right) d\tilde{\mathbf{u}}}. \quad (12)$$

The denominator is the sum of every dynamically feasible trajectory and intractable to calculate due to the continuous state and action spaces. Since the dynamics are deterministic, we can find \mathbf{x} given \mathbf{u} and \mathbf{x}_0 , thus the distribution can

TABLE I: Model Performance on Training Datasets

Highway	# Traj.	$\overline{MEE}_w \downarrow$	$\overline{MEE}_{w+} \downarrow$	% Imp \uparrow
i80 t0	188	3.77 \pm 2.53	3.73 \pm 2.58	1.17%
i80 t1	152	3.87 \pm 2.34	3.73 \pm 2.32	3.68%
i80 t2	180	3.50 \pm 2.37	3.46 \pm 2.32	1.00%
us101 t0	116	4.46 \pm 3.05	3.93 \pm 2.82	11.83%
us101 t1	93	5.13 \pm 2.92	3.54 \pm 1.96	31.03%
us101 t2	133	5.11 \pm 3.15	3.84 \pm 2.89	24.94%
highD	55	3.40 \pm 2.38	3.36 \pm 2.34	1.18%

TABLE II: Model Performance Comparison on Testing Datasets

Highway	# Traj.	$\overline{MEE}_w \downarrow$	$\overline{MEE}_{w+} \downarrow$	% Imp \uparrow
i80 t0	23	3.21 \pm 2.47	3.08 \pm 2.46	4.11%
i80 t1	40	4.43 \pm 2.78	4.38 \pm 2.83	1.05%
i80 t2	27	3.63 \pm 2.21	3.54 \pm 2.12	2.48%
us101 t0	20	5.50 \pm 2.82	4.85 \pm 2.13	11.78%
us101 t1	22	6.12 \pm 4.93	5.27 \pm 4.06	13.92%
us101 t2	23	3.81 \pm 1.56	3.51 \pm 2.63	8.03%
highD	8	4.39 \pm 2.31	4.13 \pm 2.25	6.05%

be written as a function of \mathbf{u} . Following [22] we use the Laplace approximation to find an approximate solution to the denominator. We arrive at a closed-form estimate of the probability,

$$p(\mathbf{u}|\mathbf{x}_0) \approx e^{\frac{1}{2}\mathbf{g}^\top \mathbf{H}^{-1} \mathbf{g}} |\mathbf{H}|^{-\frac{1}{2}} (2\pi)^{-\frac{d_{\mathbf{u}}}{2}}, \quad (13)$$

where $\mathbf{g} := \frac{\partial \mathcal{R}}{\partial \mathbf{u}}$ is the gradient of the reward with respect to the actions, $\mathbf{H} := \frac{\partial^2 \mathcal{R}}{\partial \mathbf{u}^2}$ is the Hessian with respect to the actions, $|\cdot|$ denotes the determinant, and $d_{\mathbf{u}}$ denotes the dimension of the lifted vector $\mathbf{u} \in \mathbb{R}^{Kd}$. We finally obtain the approximate log likelihood,

$$\mathcal{L} = \frac{1}{2}\mathbf{g}^\top \mathbf{H}^{-1} \mathbf{g} + \frac{1}{2} \log |\mathbf{H}| - \frac{d_{\mathbf{u}}}{2} \log(2\pi). \quad (14)$$

Assuming the N trajectories in a dataset are independent and identically distributed (i.i.d), we can sum the likelihood of each trajectory to get the likelihood over the entire dataset then optimize the reward parameters θ that maximize the likelihood of the dataset under the model.

IRL Implementation Details: We use the Sequential Least-Squares Quadratic Program (SLSQP) algorithm [38] from the NLOpt package (v2.6.2) [39]. We calculate the \mathbf{g} and \mathbf{H} in addition to the gradient of (14) automatically using JAX [40]. We perform a sweep over hyperparameters defined in Section III, and use the value that performs best on the training set. For numerical stability, we use min-max normalization on the features, exploit the identity $\log |\mathbf{H}| = 2\text{tr}(\mathbf{L})$ where \mathbf{L} is the Cholesky factor, and regularize \mathbf{H} to be negative definite [22].

V. RESULTS AND DISCUSSION

Tab. I summarizes trajectory similarity results on the training sets and Tab. II on the test sets in addition to the number of trajectories used for training and testing, respectively. For each dataset, the third column summarizes the average and standard deviation of the MEE, in meters, between trajectories that are optimal with respect to the baseline reward and the corresponding expert trajectories in the dataset. Likewise, the fourth column summarizes the same values for trajectories optimal with respect to the unpredictability-aware reward. The last column presents the improvement in average MEE of the unpredictability-aware

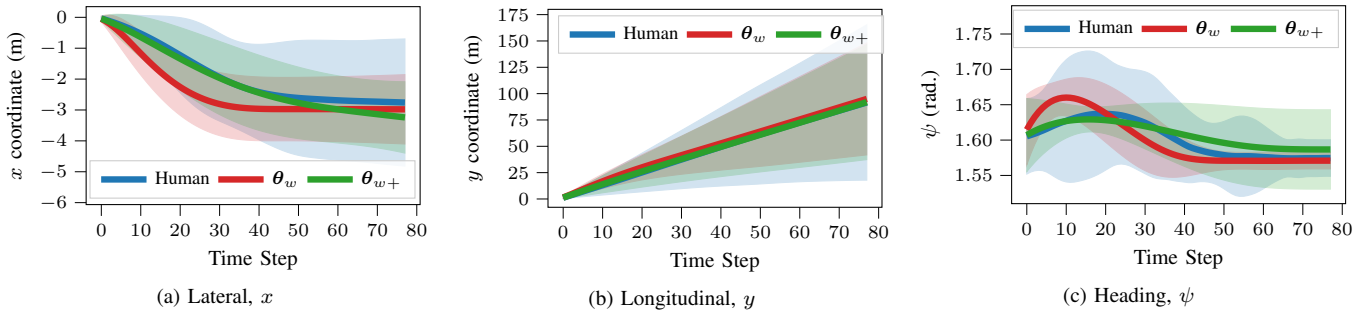


Fig. 2: State values compared to time. We compare the human ego-car lane change trajectory (blue) with trajectories generated by optimizing rewards with baseline reward, θ_w (red), and unpredictability-aware reward, θ_{w+} (green) on the US-101 $t1$ test dataset. We plot the average (solid lines) and $3\text{-}\sigma$ bounds (translucent fill) of the values. In the lateral direction (a) and heading (c), we observe more similarity between the θ_{w+} model (green) and the human data (blue) compared to the θ_w model (red).

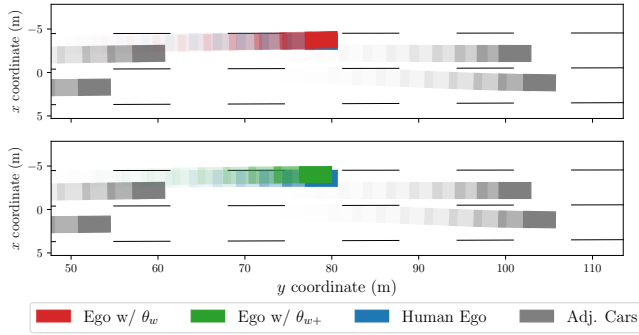


Fig. 3: Snapshot at $t = 5.5s$ for baseline (red) unpredictability-aware (green) trajectories, compared to the human in the test set (blue). In this scenario none of the adjacent vehicles are unpredictable. The unpredictability-aware model (green) is very similar the baseline (red).

model over the baseline, as a percentage of \overline{MEE}_w . On the test sets, we observe an MEE improvement between 1 to 13% in the models incorporating the unpredictability feature and a weighted average improvement of 5.86% across all test datasets.

We observe that the trajectories generated using the reward function with unpredictability are consistently closer to the human expert trajectories compared to the baseline both on the training datasets (Tab. I) and on the test datasets (Tab. II). Note that the results are generated using diverse trajectory datasets from North America and Europe, and contain trajectories with variable levels of traffic congestion, capturing the diversity of human highway driving behavior.

In Fig. 2, we plot the states of ego-car lane change trajectories from the humans in the US-101 $t1$ test set, as well as the trajectories generated using the baseline reward, θ_w , and the reward that incorporate unpredictability, θ_{w+} . For each state value, we plot the mean and $3\text{-}\sigma$ bounds at each time step for all trajectories in the test dataset. We observe that for all state values, both models succeed in fitting the human trajectory data reasonably well, and both models succeed in traveling to the target lane. We note that incorporating unpredictability results in a better match with the lateral x coordinate information indicating that in this dataset, the model with unpredictability is more hesitant to move into the target lane, similar to the human trajectory data. This result is consistent with the numerical improvement of 13.9% listed

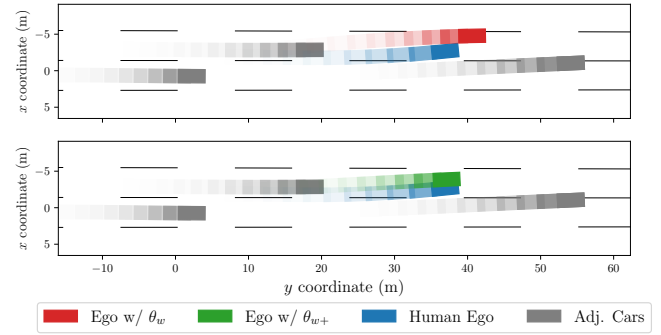


Fig. 4: Snapshot at $t = 3.2s$. In this scenario, the adjacent vehicle preceding the ego car is behaving unpredictably by changing lanes in front of the ego car, who is also changing lanes. The unpredictability-aware model (green) delays changing into the target lane and maintains a higher distance from the preceding vehicle than the baseline (red), similar to the human (blue).

in Tab. II. We believe the variation in improvement between the datasets can be attributed to the uneven distribution of unpredictable scenarios.

A. Qualitative Results

We also conduct a qualitative analysis of particular lane change scenarios from the test datasets with variations in the unpredictability of adjacent vehicles.¹ As described in the motivating example in Fig. 1, we expect trajectories from the model incorporating unpredictability to keep a greater distance from an erratic or unpredictable adjacent vehicle compared to the baseline and a similar distance in the absence of unpredictability. Fig. 3 demonstrates a lane change scenario with low unpredictability for all the gray adjacent vehicles. We can observe that both models are very similar to the human trajectory, the green (θ_{w+}) and red (θ_w) vehicles are both very close to the blue (human).

Figs. 4-6 illustrate trajectories with an unpredictable adjacent car. In Fig. 4, the preceding car is also moving to the target-lane simultaneously to the ego-car's lane change. We can see that the green vehicle (θ_{w+}) delays entering the target lane, similar to the blue human car, while the red vehicle (θ_w) does not. We make a similar observation even when the unpredictable adjacent vehicle is behind the ego car in the target lane (as in Fig. 5). In this scenario, the adjacent

¹ Animations of all scenarios can be found at tiny.cc/unpredi.

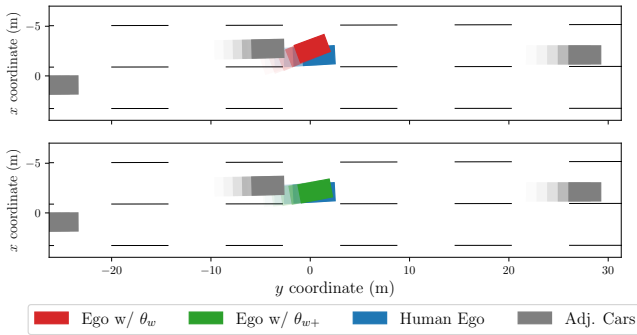


Fig. 5: Snapshot at $t = 0.5s$. In this scenario, the adjacent vehicle in the target lane speeds up to not allow the ego vehicle in the target lane. The unpredictability-aware model (green) delays entering the target lane, similar to the human (blue), while the baseline model (red) cuts in front of the adjacent car.

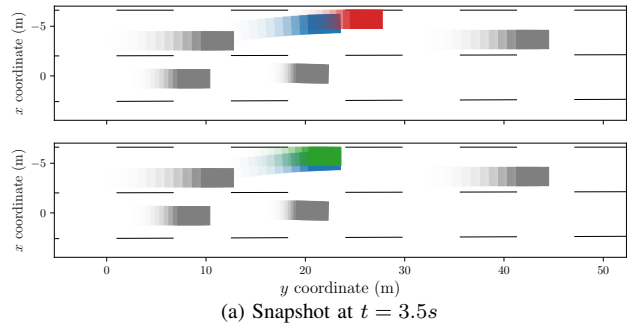
vehicle accelerates just as the human is changing into the target lane. While the green car (θ_{w+}) delays entering the target lane, similar to the blue human car, the red car (θ_w) quickly enters the target lane. Like these two examples, on average, θ_{w+} trajectories are more similar to human ones compared to θ_w . However, we also observed some scenarios where this is not the case for the whole trajectory. In Fig. 6, early in the lane change (6a) we observe that the green car (θ_{w+}) delays entering the target, consistent with the other two scenarios. Here, however, the preceding car in the target lane behaves unpredictably by increasing its speed. The headway of the ego-car is increasing (6b), yet the green vehicle (θ_{w+}) maintains its cautious distance: the increasing headway is counteracted by the unpredictability of the preceding car’s acceleration.

In all the presented scenarios we observe that the trajectories generated using the unpredictability-aware behave more cautiously by staying farther away compared to the baseline. Note that on average across each test dataset, the unpredictability-aware trajectories are also more similar to the human data than the less-cautious baselines (recall Tab. II); the instances of over-cautiousness are a minority.

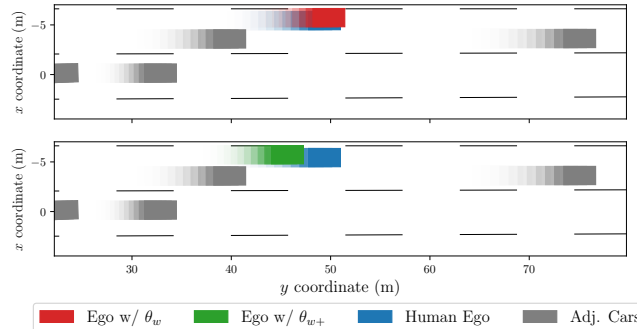
B. Limitations and Future Work

Although we can demonstrate an improvement in fitting the human data by using the unpredictability-aware reward, the scope of our study has limitations. First, our metric of unpredictability relies on the output of an off-the-shelf prediction model. Other prediction models or notions of unpredictability (e.g. covariance [8], [9]) may have variable success in generating values that are effective in the reward that we present. An evaluation of such different notions is one direction for future work.

Second, our formulation structures reward functions as linear combinations of non-linear features, and the unpredictability-aware reward contains an extra feature versus the baseline. With this structure, one could argue that the additional degree of freedom bestowed by the extra feature means that *any* feature that is sufficiently uncorrelated from those in the baseline could result in an improved fit. Although our quantitative results and qualitative analysis on simulated



(a) Snapshot at $t = 3.5s$



(b) Snapshot at $t = 7.0s$

Fig. 6: Snapshots at $t = 3.5s$ (a) and $t = 7.0s$ (b) of the same scenario. The vehicle preceding the ego car behaves unpredictably by increasing its speed. Early in the trajectory (a) The unpredictability-aware model (green) delays entering the target lane, similar to the human car (blue) and unlike the baseline (red). Later in the trajectory (b), it continues to maintain a higher distance from the preceding vehicle despite the preceding car getting farther away, unlike the human (blue).

(Fig. 1) and real scenarios (Figs. 3-6) serve as an encouraging indication that unpredictability may be a cause, our study is limited in showing a definitive link. A further step is to also evaluate more diverse reward structures, for example, with a nonlinear combination of features [41] or neural net representation [24].

Third, we found that erratic or unpredictable behavior was rare in the highway lane changes that we used. A direction of future investigation is to extract driving scenarios where more unpredictable behavior can be found (e.g. from interactive driving datasets [42]) and quantitatively analyze the effect of unpredictability in those particular occurrences, in addition to the average across all scenarios.

VI. CONCLUSION

In this paper, we proposed *unpredictability* as a feature in a reward function that can generate lane change trajectories. We used a performance metric of an off-the-shelf trajectory prediction model as a measure of the unpredictability of an adjacent driver. We analyzed human lane change behavior using an Inverse Reinforcement Learning (IRL) approach and show that incorporating this unpredictability measure can produce a better fit of human trajectories. We also provide qualitative insights into how the unpredictability feature may have an influence on the generated trajectories. We believe our results encourage further investigation on the role of unpredictability in generating driving behavior.

REFERENCES

- [1] P. Falcone, F. Borrelli, J. Asgari, H. E. Tseng, and D. Hrovat, "Predictive Active Steering Control for Autonomous Vehicle Systems," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 3, pp. 566–580, 2007. [Online]. Available: <http://ieeexplore.ieee.org/document/4162483/>
- [2] G. Schildbach and F. Borrelli, "Scenario model predictive control for lane change assistance on highways," in *IEEE Intelligent Vehicles Symposium*. IEEE, 2015, pp. 611–616.
- [3] A. Liniger, A. Domahidi, and M. Morari, "Optimization-based autonomous racing of 1:43 scale RC cars," *Optimal Control Applications and Methods*, vol. 36, no. 5, pp. 628–647, 2015. [Online]. Available: <http://doi.wiley.com/10.1002/oca.2123>
- [4] K. Burnett, J. Qian, X. Du, L. Liu, D. J. Yoon, T. Shen, S. Sun, S. Samavi, M. J. Sorocky, M. Bianchi, K. Zhang, A. Arkhangorodsky, Q. Sykora, S. Lu, Y. Huang, A. P. Schoellig, and T. D. Barfoot, "Zeus: A system description of the two-time winner of the collegiate SAE autodrive competition," *Journal of Field Robotics*, no. April, 2020.
- [5] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [6] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018, pp. 1549–1557.
- [7] B. Ivanovic and M. Pavone, "The Trajectron: Probabilistic Multi-Agent Trajectory Modeling With Dynamic Spatiotemporal Graphs," in *IEEE International Conference on Computer Vision (ICCV)*, 2019, pp. 2375–2384. [Online]. Available: <http://arxiv.org/abs/1810.05993>
- [8] 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.
- [9] L. Sun, X. Jia, and A. D. Dragan, "On complementing end-to-end human behavior predictors with planning," *Robotics: Science and Systems XVII*, 2021.
- [10] K. D. Katyal, G. D. Hager, and C.-M. Huang, "Intent-aware pedestrian prediction for adaptive crowd navigation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3277–3283.
- [11] W. Lotter, G. Kreiman, and D. Cox, "Deep predictive coding networks for video prediction and unsupervised learning," in *International Conference on Learning Representations*, 2016.
- [12] R. Villegas, J. Yang, Y. Zou, S. Sohn, X. Lin, and H. Lee, "Learning to generate long-term future via hierarchical prediction," in *International Conference on Machine Learning (ICML)*, 2017, pp. 5429–5449.
- [13] W. Byeon, Q. Wang, R. K. Srivastava, and P. Koumoutsakos, "ContextVP: Fully Context-Aware Video Prediction," *European Conference on Computer Vision (ECCV)*, 2018.
- [14] M. Babaeizadeh, C. Finn, D. Erhan, R. H. Campbell, and S. Levine, "Stochastic variational video prediction," in *International Conference on Learning Representations*, 2018. [Online]. Available: <https://openreview.net/forum?id=rk49Mg-CW>
- [15] N. Rhinehart, R. McAllister, K. Kitani, and S. Levine, "PRECOG: PREDiction Conditioned On Goals in Visual Multi-Agent Settings," in *International Conference on Computer Vision*, 2019, pp. 2821–2830. [Online]. Available: <http://arxiv.org/abs/1905.01296>
- [16] T. Xue, J. Wu, K. L. Bouman, and W. T. Freeman, "Visual dynamics: Probabilistic future frame synthesis via cross convolutional networks," *Advances in Neural Information Processing Systems*, pp. 91–99, 2016.
- [17] S. Hoermann, M. Bach, and K. Dietmayer, "Dynamic Occupancy Grid Prediction for Urban Autonomous Driving: A Deep Learning Approach with Fully Automatic Labeling," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2056–2063, 2018.
- [18] N. Mohajerin and M. Rohani, "Multi-Step Prediction of Occupancy Grid Maps with Recurrent Neural Networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10 600–10 608. [Online]. Available: <http://arxiv.org/abs/1812.09395>
- [19] M. Itkina, K. Driggs-Campbell, and M. J. Kochenderfer, "Dynamic environment prediction in urban scenes using recurrent representation learning," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 2052–2059.
- [20] R. Luo, Wenjie and Yang, Bin and Urtasun, "Fast and Furious: Real Time End-to-End 3D Detection, Tracking and Motion Forecasting With a Single Convolutional Net," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3569–3577.
- [21] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 2020, pp. 683–700.
- [22] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," in *International Conference on Machine Learning (ICML)*, 2012, pp. 41–48.
- [23] M. Kuderer, S. Gulati, and W. Burgard, "Learning driving styles for autonomous vehicles from demonstration," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 2641–2646.
- [24] M. Wulfmeier, P. Ondruska, and I. Posner, "Maximum Entropy Deep Inverse Reinforcement Learning," University of Oxford, Tech. Rep., 2015. [Online]. Available: <http://arxiv.org/abs/1507.04888>
- [25] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, "Maximum Entropy Inverse Reinforcement Learning," in *Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 3*, 2008, pp. 1433–1438. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1620270.1620297>
- [26] J. Hawke, R. Shen, C. Gurau, S. Sharma, D. Reda, N. Nikolov, P. Mazur, S. Micklethwaite, N. Griffiths, A. Shah, and A. Kendall, "Urban Driving with Conditional Imitation Learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 251–257.
- [27] J. Huang, S. Xie, J. Sun, Q. Ma, C. Liu, D. Lin, and B. Zhou, "Learning a Decision Module by Imitating Driver's Control Behaviors," in *Conference on Robot Learning (CoRL)*, 2020.
- [28] N. Buckman, A. Pierson, S. Karaman, and D. Rus, "Generating Visibility-Aware Trajectories for Cooperative and Proactive Motion Planning," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3220–3226. [Online]. Available: <https://ieeexplore.ieee.org/document/9196809/>
- [29] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, "Courteous Autonomous Cars," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 663–670.
- [30] A. Sripathy, A. Bobu, D. S. Brown, and A. D. Dragan, "Dynamically switching human prediction models for efficient planning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 3495–3501.
- [31] D. Hayashi, Y. Xu, T. Bando, and K. Takeda, "A Predictive Reward Function for Human-like Driving based on a Transition Model of Surrounding Environment," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 7618–7624.
- [32] M. Naumann, L. Sun, W. Zhan, and M. Tomizuka, "Analyzing the Suitability of Cost Functions for Explaining and Imitating Human Driving Behavior based on Inverse Reinforcement Learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 5481–5487.
- [33] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and Decision-Making for Autonomous Vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 187–210, 2018. [Online]. Available: <http://www.annualreviews.org/doi/10.1146/annurev-control-060117-105157>
- [34] J. Colyar and J. Halkia, "USA 101 Freeway Dataset Federal Highway Administration," U.S. Department of Transportation, Tech. Rep., 2007, <https://www.fhwa.dot.gov/publications/research/operations/07030/index.cfm>.
- [35] —, "Interstate 80 Freeway Dataset Federal Highway Administration," U.S. Department of Transportation, Tech. Rep., 2006. [Online]. Available: <https://www.fhwa.dot.gov/publications/research/operations/06137/>
- [36] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems," *IEEE Conference on Intelligent Transportation Systems, Proceedings (ITSC)*, pp. 2118–2125, 2018.
- [37] C. Thiemann, M. Treiber, and A. Kesting, "Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data," *Transportation Research Record*, no. 2088, pp. 90–101, 2008.
- [38] D. Kraft, "Algorithm 733: TOMP—Fortran modules for optimal control calculations," *ACM Transactions on Mathematical Software*,

- vol. 20, no. 3, pp. 262–281, 1994. [Online]. Available: <https://dl.acm.org/doi/10.1145/192115.192124>
- [39] S. G. Johnson, “The NLOpt nonlinear-optimization package,” 2020. [Online]. Available: <http://github.com/stevenj/nlopt>
- [40] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, “JAX: composable transformations of Python+NumPy programs,” 2018. [Online]. Available: <http://github.com/google/jax>
- [41] V. K. Sergey Levine, Zoran Popovic, “Feature Construction for Inverse Reinforcement,” in *Advances in Neural Information Processing Systems*, 2010, pp. 1342–1350.
- [42] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. Qi, Y. Zhou, Z. Yang, A. Chouard, P. Sun, J. Ngiam, V. Vasudevan, A. McCauley, J. Shlens, and D. Anguelov, “Large scale interactive motion forecasting for autonomous driving : The waymo open motion dataset,” *arXiv*, 2021.