

# Editing Driver Character: Socially-Controllable Behavior Generation for Interactive Traffic Simulation

Wei-Jer Chang\*, Chen Tang\*, Chenran Li, Yeping Hu, Masayoshi Tomizuka, and Wei Zhan

**Abstract**—Traffic simulation plays a crucial role in evaluating and improving autonomous driving planning systems. After being deployed on public roads, autonomous vehicles need to interact with human road participants with different *social preferences* (e.g., selfish or courteous human drivers). To ensure that autonomous vehicles take safe and efficient maneuvers in different interactive traffic scenarios, we should be able to evaluate autonomous vehicles against reactive agents with different social characteristics in the simulation environment. We propose a socially-controllable behavior generation (SCBG) model for this purpose, which allows the users to specify the *level of courtesy* of the generated trajectory while ensuring *realistic and human-like* trajectory generation through learning from real-world driving data. Specifically, we define a novel and differentiable measure to quantify the level of courtesy of driving behavior, leveraging marginal and conditional behavior prediction models trained from real-world driving data. The proposed courtesy measure allows us to auto-label the courtesy levels of trajectories from real-world driving data and conveniently train an SCBG model generating trajectories based on the input courtesy values. We examine SCBG on the Waymo Open Motion Dataset (WOMD) and show that we are able to control the SCBG model to generate realistic driving behaviors with desired courtesy levels. In particular, we find that SCBG is able to identify different motion patterns of courteous behaviors according to the scenarios.

**Index Terms**—Traffic Simulation, Intelligent Transportation Systems.

## I. INTRODUCTION

**S**IMULATION plays a critical role in the development and evaluation of autonomous vehicles (AVs). It accelerates the development cycle by enabling efficient evaluation [1] and closed-loop training [2], which are costly and risky in real-world scenarios. One key component to design for the simulator is the behavior model of the simulated road participants. In conventional driving simulation software, typical methods to synthesize the simulated agents include replaying driving logs [3] and heuristics-based approaches [4], [5]. Replay driving logs collected in real-world traffic scenarios ensures *realistic* driving behavior. However, this method does not

Manuscript received: March, 31, 2023; Accepted June, 8, 2023.

This paper was recommended for publication by Editor Aniket Bera upon evaluation of the Associate Editor and Reviewers' comments.

This work is supported by the InnoHK of the Government of the Hong Kong Special Administrative Region via the Hong Kong Centre for Logistics Robotics. W.J. Chang, C. Tang, C. Li, Y. Hu, M. Tomizuka, and W. Zhan are with the Department of Mechanical Engineering, University of California, Berkeley, CA 94720 USA {weijer\_chang, chen\_tang, chenran\_li, yeping\_hu, tomizuka, wzhan}@berkeley.edu

\* Equal contribution.

Digital Object Identifier (DOI): see top of this page.

## Socially-Controllable Behavior Generation (SCBG)

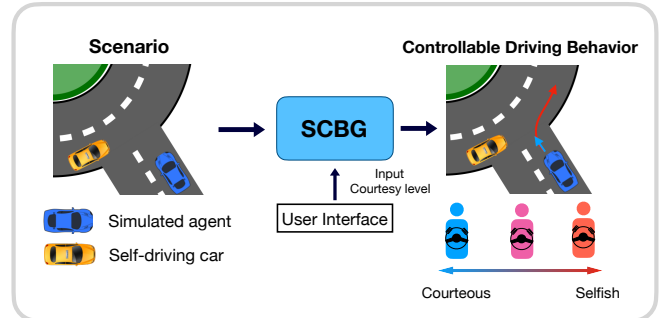


Fig. 1. Illustration of the proposed SCBG model. The SCBG model aims to control the driving behavior of a simulated agent based on the input courtesy level. By generating driving behaviors with controlled courtesy levels, the SCBG model could potentially enable more efficient evaluation and training of autonomous driving algorithms in interactive traffic scenarios.

account for the *reaction* of the simulated agents to the tested AVs. Thus, it is only reliable if the behavior of the AVs stays close to the recorded driving logs, which limits its capability to evaluate AV algorithms in highly interactive scenarios. In contrast, we may leverage domain knowledge to synthesize heuristic-based reactive agents. While the synthesized reactive behavior is sensible, it lacks sufficient realism because these heuristic-based models lack the modeling capacity to simulate sophisticated human-like behavior.

To synthesize *realistic* reactive agents, deep learning methods have been proposed to learn reactive driving behavior from large-scale real-world data [2], [6], [7]. However, such approaches can only synthesize reactive behavior following the training data distribution without a mechanism to *control* the behavior of the reactive agents [8]. Recently, several works have attempted to fill in this gap and proposed *controllable* learning-based reactive behavior generation frameworks which allow users to control different aspects of the simulation (e.g., level of safety criticality [9], [10], [11], complying user-specified rules [8]). The control mechanisms allow users to specify the simulated scenarios, which could further accelerate the development pipeline. For example, it has been demonstrated that we could evaluate the planner's performance more efficiently by deliberately generating more safety-critical testing scenarios [9].

In this work, we investigate such a controllable learning-based behavior generation framework. Different from prior

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

works, we aim to explicitly control the *social preferences* of the reactive agents. After being deployed on public roads, AVs will encounter human drivers with different driving styles. By specifying the social preferences of the simulated reactive agents, we may have a more comprehensive evaluation of the AVs' capability to handle different human drivers. For instance, an ideal AV should yield to *selfish* drivers to ensure safety but behave less conservatively when encountering *courteous* drivers to ensure efficiency. Also, we may leverage the controllable simulation to design training curricula to accelerate policy training or synthesize driving policies with robust performance for AV as it interacts with human drivers with diverse social characteristics.

The main challenge lies in *how to quantify and label the social preference of human driving behavior*. Human driving behavior involves sophisticated reasoning procedures which may be characterized from various different aspects. Inspired by prior works [12], [13], we choose to control the level of *courtesy* of the reactive agents. Courtesy can be formalized as the change in the expected *reward* of the other agents due to the simulated agent's actions. Intuitively, selfish agents may negatively influence the utilities of other agents, whereas courteous agents take actions that maximize the utilities of other agents. Courtesy has been shown as an important factor describing social interaction in driving [13]. Furthermore, we propose a novel data-driven auto-labeling framework that leverages a marginal behavior predictor and a conditional behavior predictor [14], [15] to estimate the courtesy values of driving behaviors in real-world data. With the auto-labeling framework, we can avoid costly and biased manual efforts in labeling the social characteristics of driving data.

Given real-world driving data with courtesy value labels, we train a *socially-controllable* behavior generation (SCBG) model synthesizing a simulated agent's future trajectory given an input courtesy value. During training, we leverage the auto-labeling framework to augment the training data with synthesized trajectories. We also utilize the differentiable nature of the auto-labeling operation to design a *courtesy loss*, which directly matches the courtesy values of the generated trajectories with the input courtesy values. Furthermore, we introduce a range predictor to estimate the range of feasible courtesy values in a given scenario, which ensures that the input courtesy value of the SCBG model is feasible during inference time. We examine the proposed method on the Waymo Open Motion Dataset (WOMD) [16] and show that we are able to control the SCBG model to generate realistic driving behaviors with desired courtesy levels. We find that the SCBG model is able to identify the different motion patterns of courteous behaviors according to the scenarios. This work is a crucial step toward developing a scalable traffic simulator with socially-controllable and realistic reactive agent models.

## II. RELATED WORKS

### A. Behavior Generation for Traffic Simulation

The methods for traffic simulations can be broadly categorized into two main groups: heuristic-based and learning-based approaches. In heuristic-based simulations, the reactive

agents are controlled with human-specified rules, such as Intelligent Driver Model (IDM) [4], [17]. However, these methods lack the modeling capacity required to simulate realistic and human-like behavior. On the other hand, learning-based simulations leverage deep learning-based methods to mimic driving behavior from large-scale driving data [6], [7], [18]. For instance, TrafficSim [6] trained a variational autoencoder-based model to jointly simulate the agents interacting with each other. Symphony [18] combined parallel beam search and goal conditioning to improve the realism and diversity of the simulated driving behavior.

### B. Controllable Behavior Generation

Recently, several methods have been developed to train controllable and realistic driving behavior generation models from large-scale driving datasets [10], [11], [9], [8]. Some of them aim to generate safety-critical scenarios for efficient evaluation. For instance, [10] and [11] generate challenging near-collision scenarios using adversarial optimization. [9] proposes a generative model that can produce interactions with varying safety levels controlled by a style coefficient in the latent space. Unlike them, [8] proposes a controllable behavior generation framework that allows users to specify the desired properties of the trajectories. Our work differs from existing works in that we focus on controlling the *social* characteristics of the generated driving behaviors.

Our work is also related to the literature on modeling social interaction in human driving behavior. For example, [13] formalizes the courtesy level of human drivers as the effect of a driver's behavior on the other drivers' utilities. Based on this, [19] further characterizes human drivers' social preferences. They then formulate pairwise interaction as a Stackelberg game where each driver's utility is a weighted sum of three reward terms for these perspectives. These approaches can be seen as model-based methods for controllable behavior generation. However, they have limited modeling capacity and scalability beyond pairwise interactions. Instead, we aim to explore a data-driven framework that can generate realistic and socially-controllable driving behavior in multi-agent scenarios.

## III. SOCIALLY-CONTROLLABLE BEHAVIOR GENERATION

In this section, we introduce the Socially-Controllable Behavior Generation (SCBG) framework. In Sec. III-A, we introduce the problem setting considered in this work. In Sec. III-B, we define a quantitative data-driven measure of courtesy, which is the core element of our SCBG framework. In Sec. III-C, we introduce the data-driven framework to auto-label the courtesy values of vehicle trajectories from real-world data. In Sec. III-D, we explain how we train an SCBG model from real-world data. In Sec. III-E, we introduce a courtesy range predictor which is used to control the SCBG model during inference.

### A. Problem Formulation

We consider a simulated interactive traffic scenario consisting of two vehicles denoted by Vehicle A and Vehicle

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

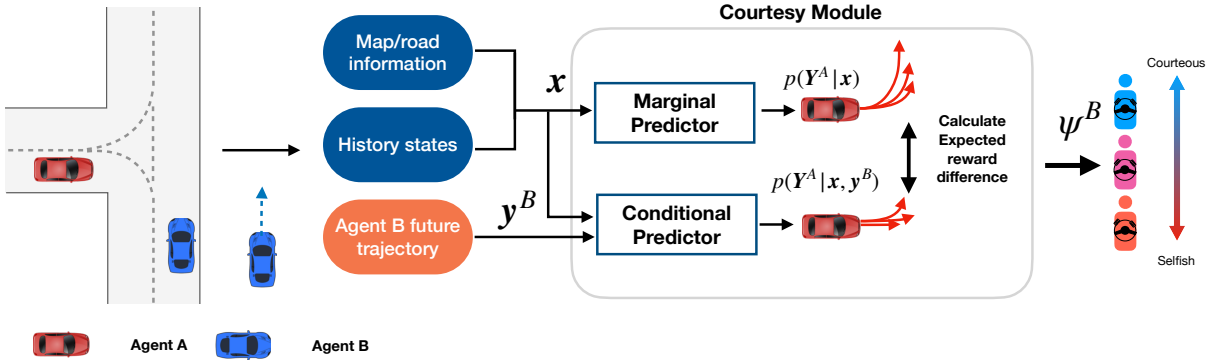


Fig. 2. Our proposed auto-labeling framework. The proposed auto-labeling framework estimates and labels the courtesy values of agent B in a sample of interactive pairs by leveraging predictors trained on real-world data. Marginal and conditional trajectory distributions of agent A are estimated by querying the respective predictors, and the difference in the expected reward values under the two distributions is used to quantify the courtesy level. This approach provides an estimate of the courtesy value without requiring humans to manually label the courtesy values.

B, where Vehicle A is controlled by the tested autonomous driving software and Vehicle B is controlled by the behavior generation model we develop. We aim to design a behavior generation model that allows the users to control *how courteous the behavior of Vehicle B is to Vehicle A*. Concretely, we denote the past observation by  $\mathbf{x}$ , which collects the historical trajectories of the two vehicles and other scene information (e.g., map, static or non-interacting objects). Given  $\mathbf{x}$ , the behavior generation model outputs a future trajectory for Vehicle B to follow, denoted by  $\mathbf{y}^B$ . In addition, the model takes a coefficient  $\psi$  as input, which controls the level of courtesy of the generated trajectory. Formally, the model is defined as:

$$\mathbf{y}^B = g_{\theta}(\mathbf{x}, \psi), \quad (1)$$

where  $\theta$  denotes the model parameters. We aim to train the model from real-world driving data to ensure the realism of the generated trajectories. However, the courtesy level  $\psi$  is a latent variable that is not recorded in the dataset. Thus, we need to define a quantitative measure of courtesy that allows convenient auto-labeling without manual effort.

### B. Quantifying Courtesy

Inspired by [12], [13], we formalize courtesy as the change in the expected *reward* of Vehicle A due to the actions of Vehicle B. Formally, given a prospective future trajectory of Vehicle B,  $\mathbf{y}^B$ , we define its level of courtesy as:

$$\psi(\mathbf{y}^B) = \mathbb{E}[r(\mathbf{Y}^A)|\mathbf{x}, \mathbf{y}^B] - \mathbb{E}[r(\mathbf{Y}^A)|\mathbf{x}], \quad (2)$$

where

$$\mathbb{E}[r(\mathbf{Y}^A)|\mathbf{x}, \mathbf{y}^B] = \int_{\mathbf{y}^A} r(\mathbf{y}^A) p(\mathbf{y}^A|\mathbf{x}, \mathbf{y}^B) d\mathbf{y}^A, \quad (3)$$

$$\mathbb{E}[r(\mathbf{Y}^A)|\mathbf{x}] = \int_{\mathbf{y}^A} r(\mathbf{y}^A) p(\mathbf{y}^A|\mathbf{x}) d\mathbf{y}^A. \quad (4)$$

The variable  $\mathbf{y}^A$  denotes the future trajectory of Vehicle A. The distribution  $p(\mathbf{y}^A|\mathbf{x}, \mathbf{y}^B)$  is the *conditional probability* of Vehicle A's future trajectory, given the observation and Vehicle B's future trajectory. The distribution  $p(\mathbf{y}^A|\mathbf{x})$  is the *marginal*

*probability* of Vehicle A's future trajectory. The reward function  $r(\cdot)$  gives the cumulative reward of a given trajectory. In practice, we may specify different reward functions according to the scenarios and the user's needs.

The term  $\mathbb{E}[r(\mathbf{Y}^A)|\mathbf{x}]$  indicates the expected reward of Vehicle A under all the possible interactions between the two vehicles, whereas  $\mathbb{E}[r(\mathbf{Y}^A)|\mathbf{x}, \mathbf{y}^B]$  indicates the expected reward of Vehicle A if Vehicle B executes a particular trajectory  $\mathbf{y}^B$ . The value  $\psi$  then indicates the effect of a given future trajectory of Vehicle B on the expected reward of Vehicle A. A positive  $\psi$  indicates a positive effect on the reward and thus implies a courteous and cooperative agent. Conversely, a negative  $\psi$  indicates a negative effect on the reward and thus implies a selfish agent.

This formulation provides flexibility to accommodate the behavior generation model to the particular aspects of social interaction we want to consider, by adopting different reward functions for courtesy quantification. For instance, to capture the behavioral impact of one vehicle on another, we can define the reward as the change in the trajectory distribution of the other vehicle, which can be quantified by KL-divergence [15]. In this work, we utilize a basic reward function, specifically average speed, to illustrate the proposed framework. In future work, we will conduct comprehensive evaluation to identify the most suitable rewards for different real-world scenarios.

### C. Auto-Labeling Courtesy

The courtesy value defined above depends on the trajectory distribution of the interacting vehicle. However, we do not have access to the ground-truth trajectory distribution of the vehicles appearing in real-world data. Instead, we propose to estimate and auto-label the courtesy values leveraging trajectory predictors trained from real-world data. The auto-labeling framework is illustrated in Fig. 2. Given a sample from the dataset, i.e.,  $\mathbf{x}, \mathbf{y}^B \sim \mathcal{D}$ , we query a marginal predictor for an estimated marginal distribution  $p(\mathbf{y}^A|\mathbf{x})$ , and query a conditional predictor for an estimated conditional distribution  $p(\mathbf{y}^A|\mathbf{x}, \mathbf{y}^B)$ . With the estimated distributions, we can then estimate the courtesy value of  $\mathbf{y}^B$  using Eqn. (2). We

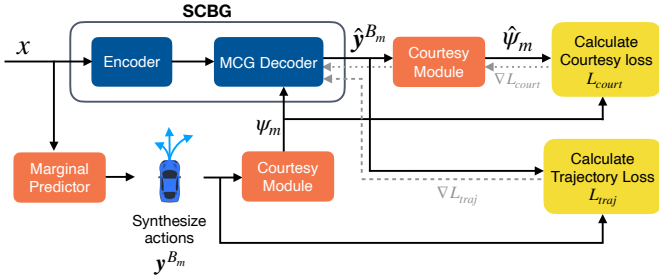


Fig. 3. The SCBG model architecture and training pipeline. The SCBG model takes as input the scene observation  $\mathbf{x}$  and conditions on the courtesy level to produce a trajectory  $\hat{\mathbf{y}}^{B,m}$ . The model is trained by matching the generated trajectories and their courtesy values with the labeled ones. To enhance the model performance, we leverage a marginal predictor to augment the training data with predicted trajectories.

denote the overall courtesy computation operation by  $J_\phi(\cdot, \cdot)$ , where  $\phi$  denotes the parameters of the predictors:

$$\psi = J_\phi(\mathbf{x}, \mathbf{y}^B). \quad (5)$$

Note that the auto-labeling procedure does not impose any restrictions on the prediction models used. In our experiment, we adopted the state-of-the-art Multipath++ [20] as the backbone prediction model, where the predicted trajectory distribution is represented as a Gaussian Mixture Model (GMM). The original Multipath++ model is designed for marginal prediction. We accommodate it to support both marginal and conditional predictions following the practice in [15]. Specifically, we add an additional future encoder to encode  $\mathbf{y}^B$ . Under the conditional prediction mode, the encoded embedding is fused with the embedding of the other input features. Under the marginal prediction mode, we turn off the future encoder so that the model does not take  $\mathbf{y}^B$  as input. We follow the training scheme in [15] to train the model for both inference modes simultaneously.

#### D. Socially-Controllable Behavior Generation Model

We now present the SCBG model architecture and its training pipeline, which are summarized in Fig. 3. We implement the SCBG model based on a Multipath++ backbone, which consists of a scene encoder and a multi-context gating (MCG) trajectory decoder. The original MCG decoder in Multipath++ utilizes a set of learned anchor embeddings for multimodal trajectory prediction. Since our SCBG model is designed for closed-loop simulation tasks, it is crucially important to ensure the long-term stability of the closed-loop behavior [6], [21], [22]. Modeling the multimodality at the trajectory level could be troublesome for closed-loop simulation because we need to ensure the consistency in modality across nearby timesteps, which requires non-trivial adaptation of the MCG decoder. For simplification, we remove the anchor embeddings and let the model output a single-modal trajectory given an input courtesy value in this work. In the future, we plan to introduce a high-level goal inference module, such as the one in [22], to model the multimodality when extending the current framework for socially-controllable closed-loop simulation. In particular, we want to highlight two key components of the training process.

1) *Data Augmentation*: Real-world data can only provide one sample of the future trajectory per scenario, resulting in one sampled behavior that corresponds to a single courtesy value. However, to achieve controllable behavior generation, the model needs to be reliable under a diverse set of input courtesy values. To address this challenge, we augment the training dataset with synthesized trajectories sampled from the marginal predictor. While in auto-labeling, we query the marginal predictor for  $p(\mathbf{y}^A|\mathbf{x})$ , in this context, we query the marginal predictor for  $p(\mathbf{y}^B|\mathbf{x})$  and sample  $M$  trajectories from the predicted distribution. By sampling from the marginal distribution, we augment the dataset with plausible trajectories of diverse courtesy levels. Our experiments demonstrate that data augmentation plays a critical role in ensuring the generated trajectories indeed correspond to the input courtesy value.

2) *Loss Function*: During training, we use a loss function that consists of two parts. The first part computes the error between the generated and labeled trajectories. Specifically, given a sample from the dataset,  $\mathbf{x}, \mathbf{y}^{B,0} \sim \mathcal{D}$ , we denote the  $M$  synthesized trajectories as  $\mathbf{y}^{B,m}, m = 1, \dots, M$ . And we denote the labeled courtesy values for  $\mathbf{y}^{B,m}$  as  $\psi_m$ . The trajectory loss,  $L_{traj}$ , is then defined as:

$$L_{traj} = l(\mathbf{y}^{B,0}, \hat{\mathbf{y}}^{B,0}) + \alpha \sum_{m=1}^M l(\mathbf{y}^{B,m}, \hat{\mathbf{y}}^{B,m}), \quad (6)$$

where  $\hat{\mathbf{y}}^{B,m}$  is the trajectory generated by the SCBG model given the courtesy value  $\psi_m$ , and  $l$  is a differentiable loss function, such as mean squared error. The coefficient  $\alpha$  is a hyperparameter used to balance the losses between the ground truth and synthesized trajectories. And we use Huber loss as the loss function  $l(\cdot, \cdot)$ .

The second part of the loss function compares the input courtesy values against the courtesy values of the generated trajectories, which ensures the generated trajectories indeed match the input courtesy values. The courtesy loss  $L_{court}$  is defined as:

$$L_{court} = \sum_{m=0}^M \|\psi_m - J_\phi(\mathbf{x}, \mathbf{y}^{B,m})\|^2. \quad (7)$$

Since the courtesy computation module  $J_\phi(\cdot, \cdot)$  is differentiable, we can directly leverage it to compute the courtesy loss and backpropagate through it for gradient computation. The overall loss function is then defined as:

$$L = L_{traj} + \beta L_{court}, \quad (8)$$

where the coefficient  $\beta$  is a hyperparameter balancing the trajectory loss and the courtesy loss.

#### E. Courtesy Range Predictor

Since the reward distribution varies across scenarios, so does the range of feasible courtesy values. As a result, feeding arbitrary courtesy values to the SCBG model may lead to out-of-distribution input which results in unreliable generated behavior. To this end, we train a courtesy range predictor to predict the interval of feasible courtesy values for a given scenario. Since we do not have access to the ground-truth

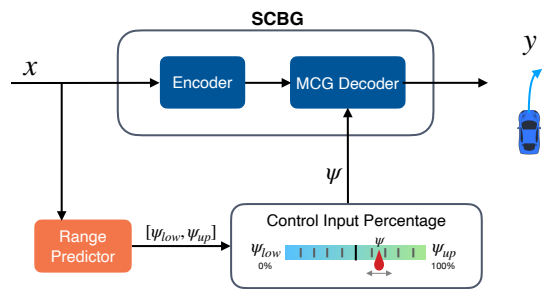


Fig. 4. The paradigm of behavior generation at inference time. The range predictor predicts the feasible range of courtesy values for a given scenario. This range is then used to scale the user input percentage to the appropriate input value, which is fed into the SCBG model.

distributions, we use quantile regression [23] to estimate the statistics of the courtesy value from data. As shown in Fig. 4, the estimated range is used to normalize the courtesy values across scenarios so that the user only needs to specify the level of courtesy as a value from the unit interval. The range predictor shares the encoder of the SCBG model. During training, we freeze the parameters of the encoder and train a new decoder to predict the 0.1 and 0.9 quantiles of  $\psi$ . We follow the practice in [24] and use the pinball loss function to train the range predictor:

$$L_{\tau} = \begin{cases} (\tau - 1) \cdot (\psi - \hat{\psi}_{\tau}), & \text{if } \psi < \hat{\psi}_{\tau} \\ \tau \cdot (\psi - \hat{\psi}_{\tau}) & \text{otherwise,} \end{cases} \quad (9)$$

where  $\tau$  denotes the target quantile (i.e., 0.1 and 0.9) and  $\hat{\psi}_{\tau}$  denotes the predicted courtesy value at the  $\tau$ -quantile.

#### IV. EXPERIMENTS

In this section, we conduct experiments to validate our proposed SCBG framework on real-world driving data.

##### A. Dataset

We use the Waymo Open Motion Dataset (WOMD) [16]. The dataset provides a subset with labels identifying a pair of interacting agents in the scenarios. We refer to this subset as the interactive subset. In our work, we focus on vehicle-to-vehicle interaction and vehicle behavior generation while we still include the historical trajectories of the other types of agents (i.e., pedestrians and cyclists) in the model input.

##### B. Training Marginal and Conditional Predictors

We build our prediction model on the open-source implementation of Multipath++ [25]. We follow the practice in [15] and train a prediction model supporting both marginal and conditional predictions during inference. We report the prediction errors (measurement time of 8s) on the interactive subset of the validation set in Table I. The evaluation metrics are defined as in [26] and computed with respect to six predicted trajectory samples, except for the ADE metric. The ADE metric measures the error between the ground-truth trajectory and the predicted sample with the highest probability. Since the SCBG model generates a single trajectory, we list the ADE values

TABLE I  
PREDICTION MODEL PERFORMANCE

Mode	ADE (m)	minADE (m)	minFDE (m)	MR (%)	mAP
Marginal	3.18	1.16	2.52	20.6	0.262
Conditional	3.13	1.11	2.35	19.5	0.274

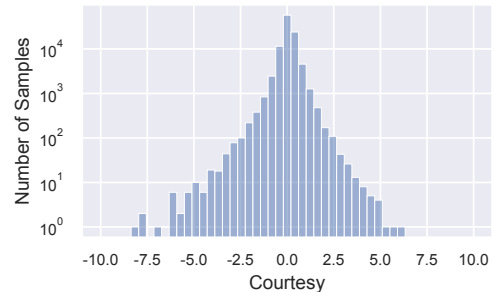


Fig. 5. Histogram of courtesy values on the interactive subset of WOMD.

here as references to validate the realism of the trajectories generated by the SCBG model, which will be discussed later in Sec. IV-D. Notably, the prediction errors are comparable with the results reported in [25], which achieved a top-10 performance in the Waymo Motion Prediction Challenge 2022.

##### C. Auto-labeling Courtesy Value

Using the trained predictor, we auto-labeled the courtesy values of the data, following the method discussed in Sec. III-B. Since the courtesy values are non-trivial only in interactive scenarios, we auto-labeled the courtesy values and trained the SCBG model on the interactive subset. In our experiment, we use average speed as the reward function when quantifying the level of courtesy in Eqn. (2). The average speed indicates the agent's progress along its route, which is the primary driving target for on-road driving. Fig. 5 shows the histogram of the extracted courtesy values on the interactive subset. It is worth noting that the histogram concentrates at zero. One reason is that there exist many non-interactive cases even in the interactive subset. Besides, the logged trajectory in the dataset does not necessarily affect the expected utility of the interacting agent.

##### D. Socially-Controllable Behavior Generation

1) *Implementation Details*: Since a large portion of the data has relatively low absolute courtesy values (Fig. 5), we split the data into two subsets with a threshold absolute courtesy value of 2. During training, we sample 50% of the batch data from each subset to avoid overwhelming the training data with trivial samples. To train the SCBG model, we load the encoder parameters from the Multipath++ model and only train the decoder for SCBG.

2) *Evaluation Metrics*: We want to validate that the proposed SCBG model can achieve *socially-controllable* trajectory generation and that the generated trajectories are *realistic*. To evaluate the model's controllability, we compare the courtesy values of the generated trajectories with the input courtesy

TABLE II  
SOCIALY-CONTROLLABLE BEHAVIOR GENERATION EVALUATION RESULTS

model	CourtesyMSE			TrajADE
	<i>data</i>	<i>range</i>	<i>arbitrary</i>	
①	0.062 ± 0.321	0.154 ± 0.470	10.35 ± 9.40	3.27 ± 2.60
①+②	0.045 ± 0.252	0.137 ± 0.467	9.97 ± 9.26	3.22 ± 2.91
①+②+③	0.034 ± 0.172	0.120 ± 0.376	9.77 ± 9.09	3.02 ± 2.72

①: Baseline Model    ②: Data Augmentation    ③: Courtesy Loss

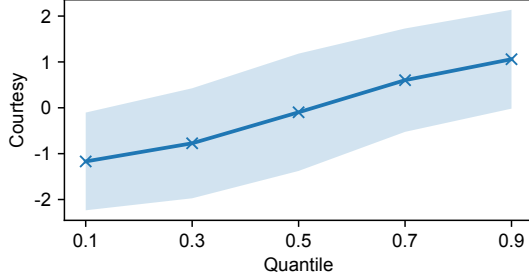


Fig. 6. The relationship between input quantile and the courtesy value of the generated trajectory in the high-courtesy subset. The courtesy value increases as the input quantile increases, demonstrating the controllability of SCBG.

values. Since we do not have access to the ground-truth trajectory distributions, we leverage the auto-labeling method to define a data-driven evaluation metric for controllability, denoted by *CourtesyMSE*, which is the mean squared error (MSE) between the input and auto-labeled courtesy values of the generated trajectory:

$$\text{CourtesyMSE} = \frac{1}{N} \sum_{n=1}^N \frac{1}{|\Psi_n|} \sum_{\psi_{n,i} \in \Psi_n} (\psi_{n,i} - \hat{\psi}_{n,i})^2, \quad (10)$$

where  $\hat{\psi}_{n,i} = J_\phi(\mathbf{x}_n, g_\theta(\mathbf{x}_n, \psi_{n,i}))$  and  $N$  is the number of samples in the dataset.  $\Psi_n$  is a set of selected courtesy values of interest. We consider three strategies to define  $\Psi_n$  for different perspectives of evaluation:

- *Data*: We define the set  $\Psi_n$  with the courtesy values of the ground-truth and augmented trajectories:

$$\Psi_{n,data} = \{J_\phi(\mathbf{x}_n, \mathbf{y}_n^{B,m})\}_{m=0}^M.$$

The resulting metric quantifies controllability over input courtesy values following the data distribution, eliminating the influence of infeasible input courtesy values.

- *Range*: We use the range predictor to predict the 0.1 and 0.9 quantiles of the courtesy values and interpolate between the predicted quantiles with a fixed interval:

$$\Psi_{n,range} = \{\hat{\psi}_{0.1}(\mathbf{x}_n), \hat{\psi}_{0.1+\delta\tau}(\mathbf{x}_n), \dots, \hat{\psi}_{0.9}(\mathbf{x}_n)\}.$$

The resulting metric reflects the controllability of the SCBG model in a practical setting where the range of feasible courtesy values is unknown.

- *Arbitrary*: The input courtesy values are selected by interpolating between the minimum and maximum courtesy values of the entire dataset, denoted by  $\psi_{\min}$  and  $\psi_{\max}$ , in a sample-agnostic way:

$$\Psi_{arbitrary} = \{\psi_{\min}, \psi_{\min} + \delta\psi, \dots, \psi_{\max}\}.$$

The resulting metric serves as a baseline showing the model performance without the range predictor.

To evaluate realism, we follow the common practice [7], [8] to compare the generated trajectory against the ground truth from data. Specifically, we compare the trajectory generated with the courtesy value of the ground-truth trajectory against the ground-truth trajectory. We use ADE to quantify the trajectory error and define the metric as follows:

$$\text{TrajADE} = \frac{1}{TN} \sum_{n=1}^N \sum_{t=1}^T \|\mathbf{y}_{n,t}^{B,0} - \hat{\mathbf{y}}_t^{B,0}\|_2. \quad (11)$$

We can then evaluate the realism of the generated trajectories by comparing TrajADE with ADE reported in Table I.

### E. Quantitative Results

Table II summarizes the results of three model variants: 1) the baseline variant without data augmentation or courtesy loss; 2) the baseline model with data augmentation; and 3) the complete SCBG model with both data augmentation and courtesy loss. The results validate the effectiveness of the proposed data augmentation scheme and the courtesy loss. Both modules significantly reduce courtesy and trajectory errors, implying improved controllability and realism. Interestingly, the courtesy loss helps reduce TrajADE even though it does not directly penalize trajectory errors, which shows that matching the courtesy values provides informative supervision signals that guide the model to better capture the driving behaviors during training.

Table II also highlights the important role of the range predictor by comparing the courtesy errors under different strategies for constructing  $\Psi_n$ . Without the range predictor, the courtesy error significantly increases (i.e.,  $\Psi_n = \Psi_{arbitrary}$ ) because of the out-of-distribution input courtesy values. Meanwhile, the courtesy error with the range predictor (i.e.,  $\Psi_n = \Psi_{n,range}$ ) is comparable to the courtesy error when the courtesy values from the data are used (i.e.,  $\Psi_n = \Psi_{n,data}$ ).

Overall, the proposed SCBG model enables socially controllable and realistic driving behavior generation, which is validated by the low *CourtesyMSE* value of 0.120 and a *TrajADE* value smaller than the ADEs of the trajectory prediction models reported in Table I. We further illustrate the model's controllability in the practical setting where users give quantile commands as in Fig. 4. As shown in Fig. 6, the courtesy value of the generated trajectory effectively increases with the quantile input. The correlation coefficient between these two values is 0.60 on the entire interactive validation subset and 0.79 on the high-courtesy subset (i.e., data with absolute courtesy values larger than 2).

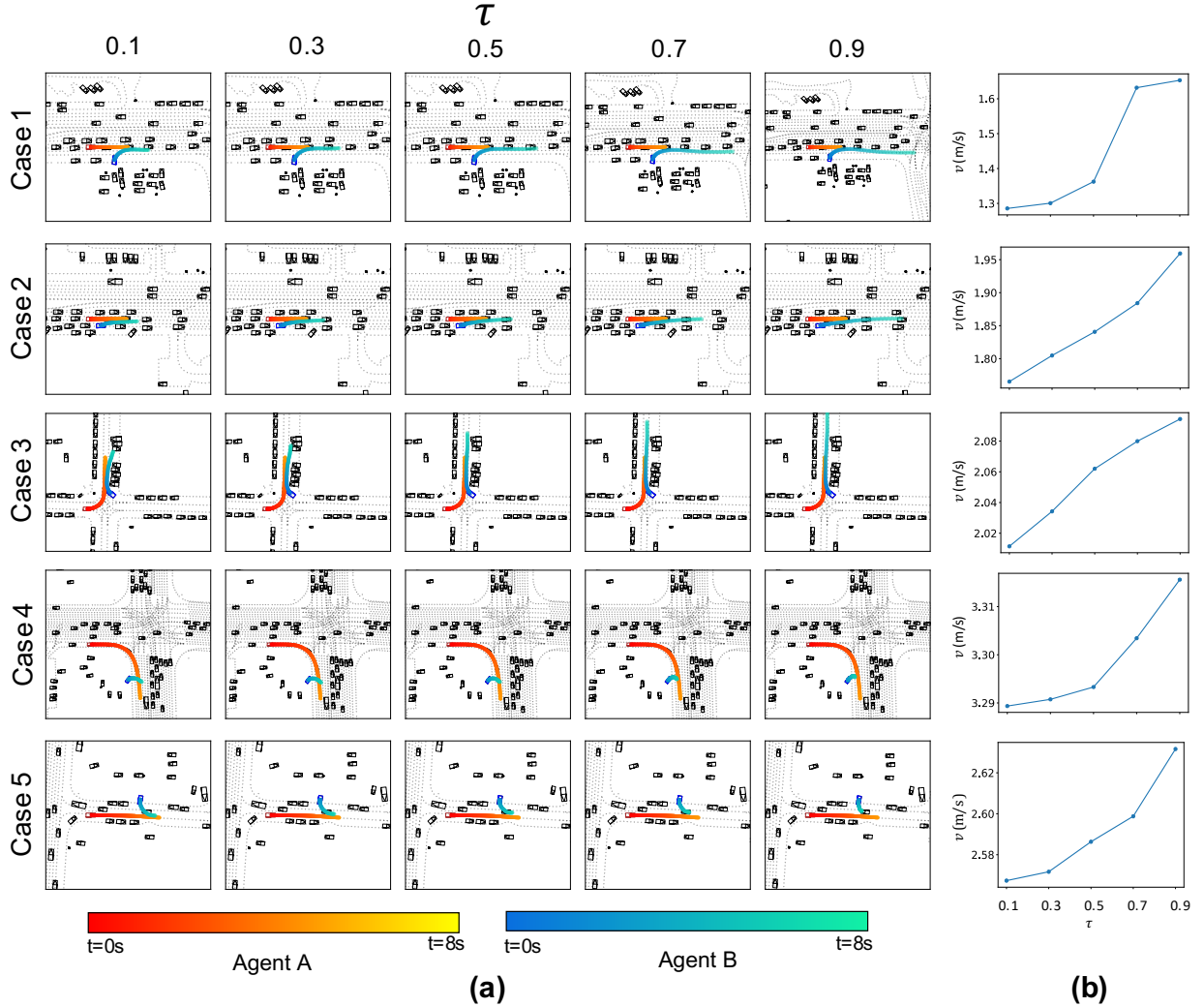


Fig. 7. (a) Visualization of the generated trajectories by control courtesy value. The blue line indicates the trajectory of the controlled agent B, while the red line represents the ground truth trajectory of agent A. This visualization shows that as the input courtesy value increases, agent B increases agent A's reward by either changing lanes (Case 1), increasing its speed (Cases 2-3), or yielding (Cases 4-5). (b) Relationship between the input courtesy quantiles and the predicted average speed of agent A, given agent B's trajectory.

### F. Qualitative Analysis

In this section, we visualize some representative examples showing that SCBG is able to generate different courteous behaviors according to the scenarios. In Fig 7, we plot the trajectories generated for the controlled agent B with different input courtesy quantiles. We also plot Agent A's ground-truth trajectories observed in the dataset to visualize the nominal behavior of Agent A. For example, the controlled agent attempts to merge into the lane where Agent A is driving in Cases 1-3. The controlled agent accelerates to prevent blocking Agent A when the courtesy level increases. Interestingly, when the input quantiles are 0.7 and 0.9 in Case 1, the controlled agent further switches its lane to the right, creating more space for Agent A. In contrast, the controlled agent slows down and yields when the courtesy level increases in Cases 4-5. Despite the diverse behavior patterns observed in the visualization, all types of courteous behavior generated by the SCBG model result in a higher reward for the other agent (Fig 7(b)).

### V. DISCUSSION AND LIMITATION

We demonstrate that the proposed SCBG framework can control the courtesy level of the generated driving behavior, which is a crucial first step toward achieving socially controllable traffic simulation. The proposed SCBG framework offers a key advantage in that it can be easily scaled to handle multi-agent interactions by calculating the joint influence of the controlled agent on all the other agents. One limitation of the current framework is that the auto-labeling method is affected by the causality issue of conditional behavior prediction [15], [27]. Specifically, a conditional prediction model cannot differentiate between the correlation and causation of two agents' trajectories. One potential solution is incorporating prior knowledge of causal relations when designing the closed-loop simulation [14]. In addition, the current framework does not account for situations where different interactive behaviors can lead to the same courtesy value, as it relies on generating a deterministic trajectory based on a courtesy value. In future

**IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.**

work, we will account for this fact by enabling multi-modal behavior generation. Our ultimate goal is to extend SCBG to a scalable socially-controllable closed-loop simulation framework that enables multiple simulated agents with specified social preferences to interact with the test AVs. One key step is to condition the SCBG model on a goal or reference path to enhance long-horizon closed-loop stability [18], [22]. Afterward, we are interested in exploring how to leverage a socially-controllable traffic simulator to accelerate policy training and robustify the trained policy.

## VI. CONCLUSION

In this study, we introduce socially-controllable behavior generation (SCBG), a model for generating realistic driving behavior based on desired courtesy levels. We propose a novel data-driven quantification of courtesy for socially interactive traffic scenarios, which enables auto-labeling of the latent courtesy values. We present a novel training algorithm to train the SCBG model from large-scale real-world driving data. In particular, we introduce a data augmentation scheme and a novel courtesy loss to improve the controllability and realism of the trained model. Our experimental results show that the SCBG model can learn to generate realistic driving behaviors with desired courtesy levels from real-world driving data. In future work, we will extend SCBG for closed-loop simulation and explore its application in closed-loop policy training and evaluation. By utilizing this framework, we believe that AVs will be able to interact with human drivers with diverse social characteristics in a safe and smooth manner.

## ACKNOWLEDGMENT

The authors would like to thank Prof. Anca Dragan, Lingfeng Sun, and Hengbo Ma for their insightful suggestions.

## REFERENCES

- [1] D. J. Fremont, E. Kim, Y. V. Pant, S. A. Seshia, A. Acharya, X. Bruso, P. Wells, S. Lemke, Q. Lu, and S. Mehta, "Formal scenario-based testing of autonomous vehicles: From simulation to the real world," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–8.
- [2] Q. Li, Z. Peng, L. Feng, Q. Zhang, Z. Xue, and B. Zhou, "Metadrive: Composing diverse driving scenarios for generalizable reinforcement learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3461–3475, 2023.
- [3] O. Scheel, L. Bergamini, M. Wolczyk, B. Osinski, and P. Ondruska, "Urban driver: Learning to drive from real-world demonstrations using policy gradients," in *Conference on Robot Learning*. PMLR, 2022, pp. 718–728.
- [4] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Conference on robot learning*. PMLR, 2017, pp. 1–16.
- [5] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, "Microscopic traffic simulation using sumo," in *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE, 2018, pp. 2575–2582.
- [6] S. Suo, S. Regalado, S. Casas, and R. Urtasun, "Trafficsim: Learning to simulate realistic multi-agent behaviors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 400–10 409.
- [7] L. Bergamini, Y. Ye, O. Scheel, L. Chen, C. Hu, L. Del Pero, B. Osinski, H. Grimmer, and P. Ondruska, "Simnet: Learning reactive self-driving simulations from real-world observations," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 5119–5125.
- [8] Z. Zhong, D. Rempe, D. Xu, Y. Chen, S. Veer, T. Che, B. Ray, and M. Pavone, "Guided conditional diffusion for controllable traffic simulation," *arXiv preprint arXiv:2210.17366*, 2022.
- [9] Z.-H. Yin, L. Sun, L. Sun, M. Tomizuka, and W. Zhan, "Diverse critical interaction generation for planning and planner evaluation," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 7036–7043.
- [10] D. Rempe, J. Phillion, L. J. Guibas, S. Fidler, and O. Litany, "Generating useful accident-prone driving scenarios via a learned traffic prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 305–17 315.
- [11] J. Wang, A. Pun, J. Tu, S. Manivasagam, A. Sadat, S. Casas, M. Ren, and R. Urtasun, "Advsim: Generating safety-critical scenarios for self-driving vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9909–9918.
- [12] 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.
- [13] L. Sun, W. Zhan, M. Tomizuka, and A. D. Dragan, "Courteous autonomous cars," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 663–670.
- [14] Q. Sun, X. Huang, J. Gu, B. C. Williams, and H. Zhao, "M2i: From factored marginal trajectory prediction to interactive prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6543–6552.
- [15] E. Tolstaya, R. Mahjourian, C. Downey, B. Vadarajan, B. Sapp, and D. Anguelov, "Identifying driver interactions via conditional behavior prediction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 3473–3479.
- [16] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou *et al.*, "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9710–9719.
- [17] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.
- [18] M. Igl, D. Kim, A. Kuefler, P. Mougou, P. Shah, K. Shiarlis, D. Anguelov, M. Palatucci, B. White, and S. Whiteson, "Symphony: Learning realistic and diverse agents for autonomous driving simulation," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2445–2451.
- [19] L. Wang, L. Sun, M. Tomizuka, and W. Zhan, "Socially-compatible behavior design of autonomous vehicles with verification on real human data," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3421–3428, 2021.
- [20] B. Varadarajan, A. Hefny, A. Srivastava, K. S. Refaat, N. Nayakanti, A. Comman, K. Chen, B. Douillard, C. P. Lam, D. Anguelov *et al.*, "Multipath++: Efficient information fusion and trajectory aggregation for behavior prediction," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 7814–7821.
- [21] W.-J. Chang, Y. Hu, C. Li, W. Zhan, and M. Tomizuka, "Analyzing and enhancing closed-loop stability in reactive simulation," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3665–3672.
- [22] D. Xu, Y. Chen, B. Ivanovic, and M. Pavone, "Bits: Bi-level imitation for traffic simulation," *arXiv preprint arXiv:2208.12403*, 2022.
- [23] R. Koener and G. Bassett Jr, "Regression quantiles," *Econometrica: journal of the Econometric Society*, pp. 33–50, 1978.
- [24] J. W. Taylor, "A quantile regression neural network approach to estimating the conditional density of multiperiod returns," *Journal of forecasting*, vol. 19, no. 4, pp. 299–311, 2000.
- [25] S. Konev, "Mpa: Multipath++ based architecture for motion prediction," 2022. [Online]. Available: <https://arxiv.org/abs/2206.10041>
- [26] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou *et al.*, "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9710–9719.
- [27] C. Tang, W. Zhan, and M. Tomizuka, "Interventional behavior prediction: Avoiding overly confident anticipation in interactive prediction," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 11 409–11 415.