`HMIway-env`: A Framework for Simulating Behaviors and Preferences to Support Human-AI Teaming in Driving

Deepak Gopinath* Jonathan DeCastro* Guy Rosman
Emily Sumner Allison Morgan Shabnam Hakimi Simon Stent

Toyota Research Institute
USA

firstname.lastname@tri.global

Abstract

We introduce a lightweight simulation and modeling framework, `HMIway-env`, for studying human-machine teaming in the context of driving. The goal of the framework is to accelerate the development of adaptive AI systems which can respond to individual driver states, traits, and preferences, by serving as a data-generation engine and training environment for learning personalized human-AI teaming policies. We extend `highway-env`, an OpenAI Gym-based simulator environment, to enable specification of human driver behavior, and design of vehicle-driver interactions and outcomes. We describe one instance of our framework incorporating models for distracted and cautious driving, which we validate through crowd-sourced feedback, and show early experimental results toward the training of better intervention policies.

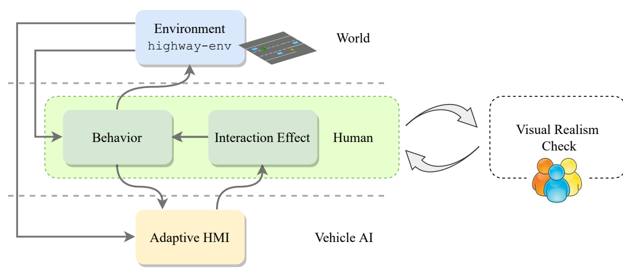


Figure 1. **Overview of `HMIway-env`.** We interface with `highway-env` [13] via a model of human behavior and intervention effects. We use this augmented driving simulator to support the rapid training and testing of adaptive intervention models, to support imperfect human drivers in various conditions. In this paper we describe one possible use case—studying intervention strategies for distracted drivers—but other models of behavior and intervention may be explored using the same framework.

1. Introduction

To build effective human-AI teams in safety-critical settings, it is important to construct ways of interacting with and assisting users as they engage in a task. Doing so is challenging because it encompasses both understanding the overall context and any cognitive deficiencies that affect the user's behaviors. It must strive to strike a balance between each user's preferences for receiving help, and the effectiveness of the help itself with respect to their situation.

In the driving domain, various situational and cognitive factors can make the driving task particularly challenging. For instance, imagine a driver that is easily distracted. As they engage in heavy phone use, they might fail to notice that a car next to them is trying to merge into their lane. In critical moments such as these, drivers may benefit from receiving timely alerts. However, interacting with the driver at the wrong moment or without discretion could lead to alert fatigue and compromise their safety, enjoyment, and trust in the vehicle over time, to varying extents depending on the individual. Hence, personalized interventions are needed to provide effective assistance that both helps the user and does not counteract their intents.

In this paper, we introduce a simulation and modeling framework, dubbed `HMIway-env`, that allows researchers to study human-machine interaction models that are better adapted to individuals, supporting improved human-AI teaming in driving. `HMIway-env` combines: (1) the behavior of the human in various safety-critical driving scenarios, (2) a mechanism for an AI system to interact with or intervene a driver, and (3) the preferences of the human to receive AI help thereby enabling the design of personalized vehicle assistance. Various existing simulation frameworks exist, e.g. [10, 16], however none of these frameworks try to compose the human and vehicle in a general form to account for drivers' internal states, traits and preferences for receiving AI assistance.

We build on the `highway-env` OpenAI Gym environment [7, 13], by incorporating human characteristics and human-AI interaction into the environment's models. Our work expands on `highway-env` in three main respects:

- We introduce a personalized driver model that captures certain *traits* and *preferences* of the individual.
- We introduce a mechanism for external *intervention* with the driver.
- We augment the human model with an *intervention efficacy* model that governs the overall effect of the intervention on current driver state.

Figure 1 shows our overall framework, while Fig. 2 shows these models interacting for a simulated distracted driving scenario.

To validate the modeling elements as well as solicit feedback on aspects for further improvement of the model, we leverage a human experimental study. Experimental feedback can be used, e.g., to calibrate whether or not, under the parameters chosen, a third-party would feel that the models produce safe or risky behavior to the desired degree, or whether any counter-intuitive behaviors have been unintentionally introduced.

Our initial study focuses on modeling momentary distractions in a population of drivers exhibiting a variety of different levels of driving caution. Distracted driving is a leading cause of risk in driving [27] and can be modelled relatively simply yet faithfully [21]. Since distracted driving leads to danger, many researchers are actively developing in-vehicle human-machine interface (HMI) techniques to try to mitigate it [6, 25].

2. Related Work

Data Simulation. In recent years, simulation environments have proven transformative for the study of sequential decision-making in many settings, not only in driving [5, 9, 13] but also gaming [18, 40] and general control environments [7, 38]. While significant effort has been allocated to simulators for the purpose of understanding a human's reaction to HMIs [33, 42], there is a growing need for simulation environments that lend well to quick iteration of AI systems without a human in the loop. Past work has also encompassed learning for human partners without gathering expensive human data by leveraging the sheer diversity in the styles of training agents [34]. This approach ignores specific characteristics of human behavior which are key in many safety-critical applications of HMIs such as driving. Finally, simulated and digital environments have proven to be important tools for psychological research in general [32], both by themselves and to augment more traditional experiments [3, 20, 31]. While these afford more efficient and controlled data collection, they become even more important in the context of studying interactive scenarios that involve road risk and cannot be readily tested in the wild. An additional recent effort involves the verification of realism of simulated results [36], which is part of a larger effort to check the human-plausibility of computer-generated outputs [30, 37].

`Highway-env` [13] is a lightweight model and processed-perception simulator tool that has been used to explore different driver factors such as aggressiveness [16], as well as gauging explainability for human subjects [36]. This makes it a strong candidate tool for exploring HMIs that relate to driver characteristics; ours is the first known work to explore it for this purpose.

Driver Behavioral Models. One important aspect of such simulators is the incorporation of driver behavioral models that capture phenomena of interest. These include rational driving behaviors such as [41], as well as phenomena such as distraction [24] and secondary visual attention tasks [17], bounded rationality [2, 14, 22], personal characteristics [26] such as risk propensity [11] and social behavior [29], as well as other factors often addressed in human factors, safety, and HMI research [1, 23, 35].

Human-AI Interaction Models. Finally, the design of such simulation environments is strongly tied to the kind of approaches we wish to test on them. Current AI systems that interact with the driver are diverse, and range from a variety of driver safety systems [4] to shared autonomy approaches [15, 29]. In particular, in this paper we focus on approaches that address distraction [12] and personalized human-AI teaming approaches. These interaction models also seek to endow the AI with human-aware decision making capabilities and can also potentially support continual adaptation to evolving individual preferences via personalized lifelong learning.

3. Models for Shared Human-AI Driving Teams

In this paper, we focus on human-AI driving teams in which the AI system interacts with the driver via sensory information at key moments in the driving task. In such a scenario, we assume that the human is always in full control of the vehicle, and the AI system merely provides suggestions, warning alerts or other measures to help the driver. We contrast this with more autonomy-driven teaming setups, where the nature of interaction is to override the driver by taking control of the vehicle. Most such setups are agnostic to the inherent preferences and abilities of the human partner. We aim to demonstrate the flexibility of our framework to represent various characteristics of human drivers

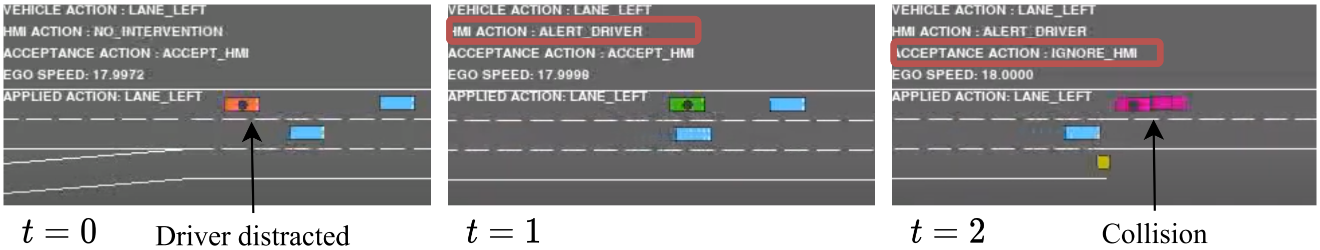


Figure 2. **Illustration of a roll-out from a simulation of distracted driving.** A driver is simulated to become distracted and ignore the attempted interaction by a safety HMI, resulting in a collision. Our framework allows the rapid exploration of Human-AI driving team strategies through different models of driver behaviors and preferences within a collection of minimalist driving environments including merge scenarios, intersections and roundabouts.

such as distraction level, cautiousness, and preferences toward AI-based interventions.

We frame the model as a reinforcement learning (RL) problem, which solves for a *policy*—a function mapping observations by the agent to actions taken on the part of the agent. Since we express both the human and AI system in terms of their higher-level goals and constraints in terms of a system of rewards, we may take the agent (the policy) to represent both the human and AI system operating together or as two policies that interact with one another. To generate our results, we adopt the former (i.e., a joint policy) approach, but our framework supports both options.

The policy is learned from observations of the world based on collected experience, and is gradually improved through maximization of a total reward for each roll-out generated by the policy as it is trained. In our human-AI teaming setting, the state consists of a simulated world: other vehicles, the road environment, and the state of the human. Observations are potentially corrupted measurements of state. The model of the human is expressed as a Markov Decision Process (MDP), which specifies the probability of transitioning from one state to another, along with any rewards (e.g. speed preferences) or penalties (e.g. collisions) collected along the way. To train policies, we adopt a policy gradient method: the proximal policy optimization (PPO) algorithm.[1]

3.1. Highway-env → HMIway-env

In order to augment the existing environments in `highway-env` to capture human factors, we introduce additional parameters into the environment model to capture: (a) the cautiousness exhibited by the driver, (b) the likelihood of the driver becoming distracted and attentive, and (c) and the willingness of a driver to be influenced by an external alert issued by a vehicle AI system.

Similar to the approach taken by Morton et. al. [19], the observations fed into the policy encode the position and velocities of nearby vehicles as lidar observations. Lidar-like representations resemble a human's perception system in a coarse sense (and therefore it is sensible to use such representations for training a driver policy) and moreover it is also true to the vehicle's on-board perception system (justifying its use for training the vehicle AI's intervention policy). The parameter that encodes the driver's cautiousness level, referred to as *obstacle inflation factor*, ω, inflates the spatial footprint of the surrounding vehicles and as a result the distance to the surrounding vehicles will be reduced proportionally. This models the tendency of drivers to maintain different levels of space with other cars depending on their individual preferences.

We seek to train a *joint* policy for the human and AI system as it is one of the most straightforward options to pick from a multitude of possibilities. The policy's actions consist of both human-initiated vehicle actions (v_t^H) and the AI system's intervention actions on the driver (v_t^A). For v_t^H, the action space consists of a discrete set of semantic actions comprising {*speed_up, slow_down, keep_speed, move_left, move_right*}, where the first two actions result in changes in speed, while the last two bring about lane changes. The action space for v_t^A is binary and consists of a discrete set of intervention actions given by {*alert, no_alert*}. The mechanism by which v_t^A affects the vehicle actions applied to the vehicle is facilitated by the intermediate distraction and alert-acceptance model described in detail in Sec. 3.2.

We introduce a new vehicle class in `highway-env` called `PilotedMDPVehicle` to encapsulate a distraction model and an intervention acceptance model. The ego-vehicle that is controlled by the RL agent is modelled as a `PilotedMDPVehicle`. `PilotedMDPVehicle` takes the joint policy action as input and updates its state in a three-step process. First, the intervention action affects the acceptance state of the vehicle. Second, the distraction state is updated using the subsumed distraction model. Third, the vehicle action conditioned on the distracted state is applied to the vehicle. Finally, the behavior of the other vehicles on the road are modelled using the intelligent driver

[1] https://stable-baselines3.readthedocs.io/en/master/modules/ppo.html

model [39] (the default approach in all `highway-env` environments). The initial speeds are uniformaly distributed in a range within the prescribed speed limits of the road on which they are spawned.

3.2. Vehicle AI–Mediated Distraction Model

To showcase the utility of our framework, we aim to develop a distraction model that captures a variety of phenomena observed in driving. Our model can capture momentary lapses in driving, such as when a driver is engaged in a secondary task (e.g. cellphone use) or is otherwise inattentive to the road. It can also encompass delays in action that might result, for example, from the driver being engaged in conversation or in a state of high cognitive load [8].

In our framework, the non-rational behavior exhibited by the driver is modeled primarily by the interaction of two binary random variables: the distraction state, d_t, and the acceptance state, i_t. In Fig. 3, d_t, encodes whether the driver is distracted or attentive, and the acceptance state, i_t, encodes the driver's inclination to be influenced by the vehicle AI's alert signal. The downstream effect of responding to an alert is that it indirectly affects the transition dynamics of the distraction state.

In Fig. 3 (right), the true action v_t consists of *two* components, namely: the vehicle action v_t^H, and the AI's intervention action $v_t^A \in \{alert, no\_alert\}$. We introduce a counter variable c_t to encode how long a *successful intervention by the vehicle AI* remains in effect. That is, when the vehicle AI issues an alert the effect of the alert persists for a fixed time window, referred to as the *intervention effectiveness window* of size N, during which the acceptance state remains 1.

In our model, the transitions for the intervention effect variable i_t and c_t are determined by the following equations:

$$\text{if} \quad i_{t-1} = 0 \quad \text{then} \quad \begin{cases} i_t = 1, c_t = 0 & \text{if } v_t^A = alert \\ i_t = 0, c_t = 0 & \text{if } v_t^A = no\_alert \end{cases} \tag{1}$$

$$\text{if} \quad i_{t-1} = 1, \quad v_t^A = alert, \quad \text{then} \quad c_t = 0 \quad \text{and} \quad i_t = 1 \tag{2}$$

$$\text{if} \quad i_{t-1} = 1, \quad v_t^A = no\_alert, \quad \text{then}$$
$$c_t = (c_{t-1}+1) \bmod N, \begin{cases} i_t = 1 & \text{if } c_{t-1} < N-1 \\ i_t = 0 & \text{if } c_{t-1} = N-1 \end{cases} \tag{3}$$

Equations (1) and (2) capture the behavior that, when a driver complies with an alert from the vehicle AI, their acceptance state is always set to be 1, regardless of the acceptance state at the previous time step. Additionally, c_t is set to be 0, indicating that the vehicle AI's intervention has been re-triggered. If the driver's acceptance state is 0, then it continues to be 0 if no alert has been received. In (3), if the acceptance state is already 1 (which implies that the vehicle AI's alert was already accepted by the driver at an earlier time step), then the acceptance state continues to be 1 (that is the alert continues to have an effect on the driver) as long as c_t remains within the *intervention effectiveness window*. Once c_t is greater than the window size, the acceptance state is reset to 0, if no more alerts are accepted.

In the current implementation, the distraction variable d_t evolves as a controlled Markov chain whose transition probabilities are modulated according to the *acceptance state* i_t of the driver. The amount of modulation is controlled by a parameter γ, *which captures the driver's willingness to be influenced by the alert issued by the vehicle AI*. Upon accepting the alert from the vehicle AI, the baseline transition probabilities of the distraction state Markov chain are modulated in such a way the probability of becoming distracted is *reduced* and of becoming more attentive is *increased*. Specifically, if $\beta \in (0, 1)$ is the baseline probability of transitioning from an attentive state ($d_{t-1} = 0$) to a distracted state ($d_t = 1$), the conditional (modified according to the acceptance state) transition probabilities for the Markov model is given by

$$p(d_t = 1 | d_{t-1} = 0) = \max(0, \beta - \gamma \mathbb{1}(i_t = 1)) \tag{4}$$

where γ is the *intervention effectiveness factor.* and $\mathbb{1}(\cdot)$ is the indicator function. Likewise, the probability of transitioning from a distracted to an attentive state becomes higher when the driver is willing to accept the AI agent's alert. Therefore,

$$p(d_t = 0 | d_{t-1} = 1) = \min(1, \alpha + \gamma \mathbb{1}(i_t = 1)) \tag{5}$$

where α is the baseline probability of transitioning from a distracted to an attentive state.

From the above equations, we can see that if the driver accepts an alert from the vehicle's AI, the acceptance state will be set to 1, and as a result the transition probabilities in (4) and (5) are modulated and this modulation remains in effect for at least N time steps.

In our framework, at every timestep t, first c_t is updated, followed by the acceptance state i_t and then finally the distraction variable d_t.

For $t > 0$, the value of d_t, the distraction state, affects the applied vehicle action a_t as follows,

$$a_t = \begin{cases} a_{t-1} & \text{if } d_t = 1 \\ v_t^H & \text{if } d_t = 0 \end{cases} \tag{6}$$

with $a_0 = v_0^H$.

3.3. Reward Structure

Table 1 shows the full list of driving-related rewards used by the joint human-vehicle AI model capturing a joint policy trained for the actions of both the human and AI system.

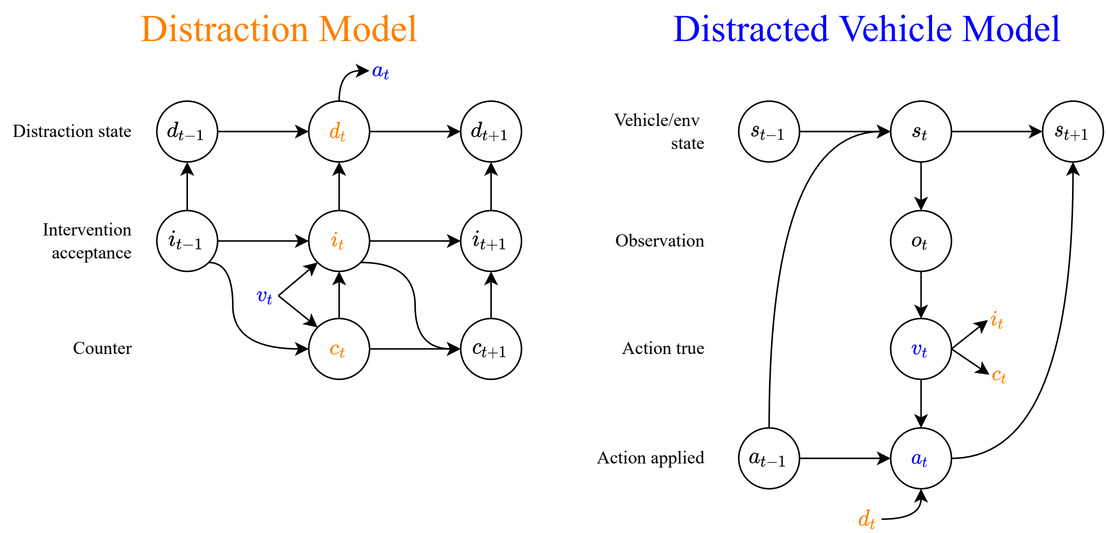


Figure 3. **Our implementation of a model of distracted driving.** `PilotedMDPVehicle` consists of two transition systems: a distraction model (left), which governs the tendency of a driver to become distracted and comply with or ignore an intervention, and a driving model (right) that models the vehicle's actions based on both the physical environment and distraction state.

Table 1. Driving related rewards for the joint human-vehicle AI system.

R_{coll}	C_{coll} if crashed, 0 otherwise
R_{speed}	$C_{speed} \frac{vehicle\_speed}{max\_speed}$
$R_{right-lane}$	$C_{right-lane}$, if on right lane, 0 otherwise
$R_{merging}$	$C_{merging} \frac{target\_speed - current\_speed}{target\_speed}$ if on the merging lane, 0 otherwise
$R_{lane-change}$	$C_{lane-change}$ if v_t^H is $move\_left$ or $move\_right$, 0 otherwise
$R_{distraction}$	$C_{distraction}$ if driver is in distracted state, 0 otherwise
R_{alert}	C_{alert} if $v_t^A = no\_alert$ and $d_{t-1} = 0$, 0 otherwise
$R_{accept-alert}$	$C_{accept-alert}$ if $v_t^A = alert$ with $d_{t-1} = 1$ and $d_t = 0$, 0 otherwise

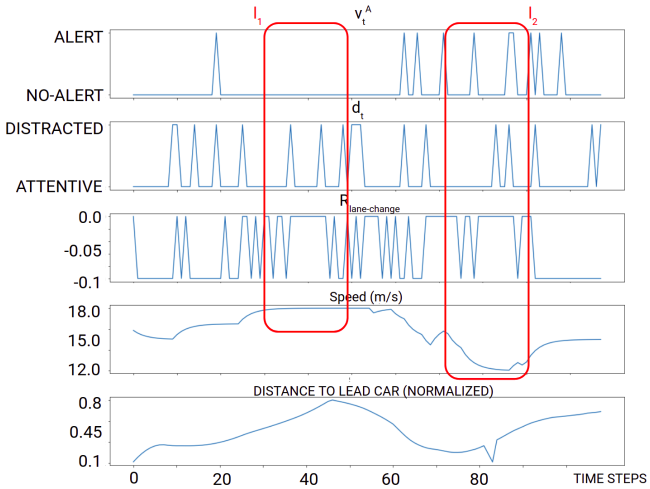
In Table 1, R_{coll}, R_{speed}, $R_{right-lane}$, $R_{merging}$, $R_{lane-change}$ are the reward components pertaining to driving performance. $R_{distraction}$ pertains to the human tendency to be distracted and R_{alert} capture the rewards the vehicle AI receives for issuing sparse alerts. $R_{accept-alert}$ is the reward term that connects the vehicle AI to the human.

For training our models, we set the coefficients related to driving rewards to values such that the vehicle favors being safe on the right lane, at higher speeds and seeks to minimize lane changes.

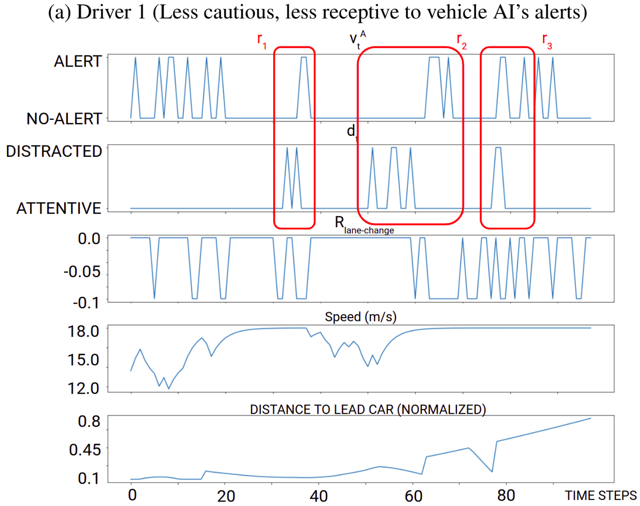
4. Model Behavior Analysis

In this section, we present an initial analysis of the roll-outs generated from models trained with different model parameters. These different model parameters represent two driver types who differ in terms of how cautious they are (risk-taking vs. risk-averse) and their willingness to be influenced by the vehicle AI's alert actions. We also experiment with different settings of C_{alert} to vary the alert sparsity and in doing so we indirectly encode a driver's preference to be alerted in the first place. The models were trained in a modified merge environment from `highway-env` in which the ego-vehicle is modeled as an instance of `PilotedMDPVehicle`.

Figure 4 shows the data traces for roll-outs for two different driver types: (a) Driver 1 who is less cautious and unwilling to accept the vehicle AI's alerts and (b) Driver 2 who is more cautious and willing to accept the vehicle AI's alerts fully. From the data traces, we can see that Driver 2 is distracted less often compared to Driver 1 due to their higher willingness (higher value of γ) to accept the vehicle AI's alerts despite the baseline distraction probability β being the same for both. We also observe a trend where the vehicle AI's interventions are more active when driving related rewards decrease (red boxes l_1 and l_2 in Figure 4a). For example, although Driver 1 gets distracted during the initial part of the trajectory, the vehicle AI's interventions remain sparse, likely due to the fact that the vehicle speed is still fairly high. However, immediately after the half-way mark, there is a decrease in speed and a corresponding increase in the number of AI interventions. We also observe,



(a) Driver 1 (Less cautious, less receptive to vehicle AI's alerts)



(b) Driver 2 (More cautious, more receptive to vehicle AI's alerts)

Figure 4. Data traces from two different driver types. For both models, (α, β)=$(0.8, 0.2)$ and C_{alert}=6.0. $(\omega, \gamma) = (3.0, 0.1)$ for Driver 1 and $(9.0, 1.0)$ for Driver 2. Regions of interest discussed in the text are circled in red.

especially for Driver 2, that when the driver becomes distracted, the vehicle AI issues a slightly delayed alert, particularly when driving-related rewards decrease (red boxes r_1, r_2 and r_3 in Figure 4b).

Table 2 presents different driving performance related metrics for two driver types for models trained with different C_{alert} coefficients. Overall, we can see that the vehicle AI is more effective in making Driver 2 less distracted indicating that the learnt vehicle AI policy is able to effectively intervene and make the driver more attentive. We also observe that the average high speed reward tends to be lower for Driver 2, when $C_{alert} = 6.0$ and 9.0. Visual inspection of the roll-outs also reveal that Driver 2 exhibits more variations in speed, and is able to regulate the speed depending on how close the nearby vehicles are. The vehicle AI is able to help the Driver 2 to be more cautious, despite the fact that there is a high probability of being distracted. Lane change rewards are also comparable between the driver type (except for $C_{alert} = 6.0$), indicating that the alerts are not inadvertently causing fishtailing behavior. Another interesting observation is with respect to the minimum Time-To-Collision (TTC) for the different driver types. Here, we observe that Driver 2 is able to achieve lower TTC compared to Driver 1 suggesting that if the driver is more receptive to the vehicle AI's alerts, the joint human-AI team is likely more confident in following the lead car more closely.

Our pilot analysis suggests that even with a naive joint policy training approach, the vehicle AI is able to assist the distracted driver in achieving consistently high levels of driving performance. Our future work will include exploring multi-stage training (which `HMIway-env` already has support for) with a constrained version of PPO akin to the approach presented in [28], with the exception that the 'shared autonomy' in our scenario is facilitated via the intermediate distraction and acceptance model.

5. Human Subjects Study

To validate whether the model we created matched people's intuitions of what risky (distracted) driving looks like, we ran a study with human subjects. Participants were shown clips of videos generated with varying levels of cautiousness and distraction. They were asked to assess how distracted the driver is, how risky the driver is, and how safe they would feel in the vehicle.

5.1. Methods

We created eight different driving behavior videos to show subjects.[2] The video set consisted of videos varying in cautiousness (high or low ω values) and level of distraction ($\beta \in [0.0, 0.6]$). The order of the conditions was randomized each time to control for repeated exposure. Subjects were instructed to pay attention to the driving behavior of the "green rectangle" which represented the car of focus. A gray dot was overlaid on the green rectangle to orient the subject's attention (see Fig. 5).

After viewing each video, subjects were instructed to answer a series of questions about the green driver's behavior using a 4-point Likert response scale. These questions were related to how safe, distracted, risky, or similar to their own driving behavior they perceived the green driver's behavior to be. Subjects were then prompted to provide open ended feedback about how the driver could improve their driving. Participants were paid $3.00 for their participation.

We recruited 500 random subjects with US driver licenses to participate through Prolific. Our participants

[2] https://www.youtube.com/playlist?list=PLgkyRHe_bn13hcbG2VJNIIYU3ZnV8O24m

Table 2. Metrics for models trained for two different driver types. For all models, $(\alpha, \beta) = (0.6, 0.4)$. Driver parameters are same as in Figure 4.

	$C_{alert} = 3.0$		$C_{alert} = 6.0$		$C_{alert} = 9.0$	
	Driver 1	Driver 2	Driver 1	Driver 2	Driver 1	Driver 2
High Speed Reward	438	444	424	382	430	372
Distraction Reward	-316	-39	-330	-89	-349	-98
Lane Change Reward	-2.90	-2.97	-3.2	-4.6	-4.0	-3.7
Minimum TTC to Lead Vehicle (s)	0.48	0.19	0.49	0.22	0.52	0.26
Number of Alert Acceptances	7.0	4.6	5.7	6.8	5.4	7.2

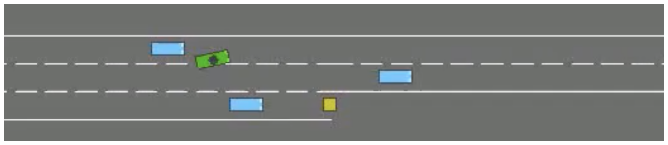


Figure 5. **Example of what participants were shown.** Participants were instructed: "You will now be presented with a series of short videos. The rectangles shown represent cars. Please watch carefully and make sure to focus on the driving behavior of the green car. There is a gray dot in the middle of the rectangle to help you identify the car. Feel free to re-watch the video as many times as needed." All videos can be found here.

Table 3. **Perceived Safety.** Means (standard deviations) of how people answered, "How safe would you feel if you were a passenger in this car? (1=very safe, 4=not at all safe)". This question was asked to elicit how safe one would feel in the target vehicle. We note that the perceived safety decreases with an increase in distraction parameter β.

ω	$\beta = 0.0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
Low	2.69 (0.98)	1.91 (0.93)	3.49 (0.96)	3.82 (0.72)
High	2.28 (1.04)	3.40 (0.80)	2.24 (0.99)	3.54 (0.92)

Table 4. **Perceived Riskiness.** Means (standard deviations) of how people answered, "How risky does the driver's driving appear to be? (1=very risky, 4=not at all risky)". This question was asked to see if people's intuitions about riskiness matched the different inflation values we used (low for risk-seeking and high for risk-aversion). We note that perceived riskiness increases as the distraction parameter β increases.

ω	$\beta = 0.0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
Low	2.59 (0.94)	3.43 (0.76)	2.41 (0.96)	1.61 (0.80)
High	3.04 (0.97)	1.76 (0.83)	3.14 (0.91)	2.28 (0.95)

Table 5. **Perceived Distraction.** Means (standard deviations) of how people answered, "How distracted does the driver appear to be? (1=not at all distracted, 4=very distracted)". This question was asked to see if people's intuitions about distraction matched the varying β levels we used. We note that the perceived distraction level increases with an increase in the distraction parameter β.

ω	$\beta = 0.0$	$\beta = 0.2$	$\beta = 0.4$	$\beta = 0.6$
Low	2.07 (1.20)	1.50 (0.97)	3.01 (1.15)	3.37 (1.00)
High	1.67 (1.09)	3.23 (1.09)	3.14 (1.12)	3.14 (1.16)

predominantly identified as white or of European descent (71%). 10.02% identified as being of Asian descent, 7.56% identified as black or African American, 1.70% identified as American Indian or Alaskan Native, 1.13% identified as Middle Eastern or North African, and 0.19% identified as Native Hawaiian or Pacific Islander, 27.4% were ages 26-35, 24.6% were ages 36-45, 17.6% were 46-55, 15.6% were 56-65, 7.6% were ages 66 and older, and 6.8% were ages 18-25. 41% of our sample identified as women. 91.43% of participants reported as not identifying as ethnically Hispanic or Latino/x while 7.37% of participants did.

5.2. Results

Human subject study data can be useful in evaluating how well our driver models portray realistic behavior. In Table 3, subjects generally found the simulated driving behavior to be less safe as distraction level increased. This was generally consistent with perceived risk in Table 4 and perceived distraction level in Table 5. However, low obstacle inflation factor was not necessarily related to perceived riskiness. For example, the video that was generated with a high inflation factor and $\beta = 0.2$ (video here) had one of the highest perceived level of risk. While the simulated driver left a lot of space between cars, its behavior during a lane change gave the perception of high risk. This finding is consistent with the objective of building driver models that indeed exhibit different risks, one that could help us update such models in a principled way.

Additionally, there was a wide range of individual differences in how people perceived the videos. This suggests that features of the individual annotators could influence the perception of safety, risk and distractedness. For example, if somebody is a very distracted driver, they may be less sensitive to distracted behavior in a video, whereas if they are generally risk-averse they may be more sensitive to a lower inflation factor. In follow-up studies, we plan to explore this relationship in more detail.

6. Conclusion

In this paper, we introduce `HMIway-env`, a simulation and modeling framework for encoding human behaviors and preferences in a driving environment. The framework supports modeling behaviors of a human and their preferences for receiving AI assistance in order to support various tasks such as data generation, algorithm prototyping, and learning interaction policies. Though we demonstrate the approach using a model for human distraction, this is extensible to other driver-specific behaviors, such as attention shifts, the effects of age, alarm fatigue, and other physical and cognitive impairments. We plan to develop the framework further with additional behavioral models and interaction strategies and incorporating human subject study results to refine their realism.

References

[1] L Aberg and P A Rimmö. Dimensions of aberrant driver behaviour. *Ergonomics*, 41(1):39–56, Jan. 1998. 2

[2] Arwa Alanqary, Gloria Z Lin, Joie Le, Tan Zhi-Xuan, Vikash K Mansinghka, and Joshua B Tenenbaum. Modeling the mistakes of boundedly rational agents within a Bayesian theory of mind. *arXiv preprint arXiv:2106.13249*, 2021. 2

[3] Antonio A Arechar, Simon Gächter, and Lucas Molleman. Conducting interactive experiments online. *Exp. Econ.*, 21(1):99–131, 2018. 2

[4] Klaus Bengler, Klaus Dietmayer, Berthold Farber, Markus Maurer, Christoph Stiller, and Hermann Winner. Three decades of driver assistance systems: Review and future perspectives. *IEEE Intell. Transp. Syst. Mag.*, 6(4):6–22, 2014. 2

[5] Luca Bergamini, Yawei Ye, Oliver Scheel, Long Chen, Chih Hu, Luca Del Pero, Błażej Osiński, Hugo Grimmett, and Peter Ondruska. Simnet: Learning reactive self-driving simulations from real-world observations. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5119–5125. IEEE, 2021. 2

[6] Stewart Birrell, Mark Young, Neville Stanton, and Paul Jennings. Using adaptive interfaces to encourage smart driving and their effect on driver workload. In *Advances in Human Aspects of Transportation*, pages 31–43. Springer, 2017. 2

[7] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. *arXiv preprint arXiv:1606.01540*, 2016. 2

[8] Michelle Chan and Anthony Singhal. The emotional side of cognitive distraction: Implications for road safety. *Accident Analysis Prevention*, 50:147–154, 2013. 4

[9] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on Robot Learning*, pages 1–16. PMLR, 2017. 2

[10] Salma Elmalaki, Huey-Ru Tsai, and Mani Srivastava. Sentio: Driver-in-the-Loop forward collision warning using multisample reinforcement learning. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, SenSys '18, pages 28–40, New York, NY, USA, Nov. 2018. Association for Computing Machinery. 1

[11] Julie Hatfield and Ralston Fernandes. The role of risk-propensity in the risky driving of younger drivers. *Accid. Anal. Prev.*, 41(1):25–35, Jan. 2009. 2

[12] Katja Kircher and Christer Ahlström. The driver distraction detection algorithm attend. In *Driver Distraction and Inattention*, pages 327–348. CRC Press, 2017. 2

[13] Edouard Leurent. An environment for autonomous driving decision-making. https://github.com/eleurent/highway-env, 2018. 1, 2

[14] I Lubashevsky, P Wagner, and R Mahnke. Bounded rational driver models. *The European Physical Journal B - Condensed Matter and Complex Systems*, 32(2):243–247, Mar. 2003. 2

[15] Mauricio Marcano, Sergio Díaz, Joshué Pérez, and Eloy Irigoyen. A review of shared control for automated vehicles: Theory and applications. *IEEE Transactions on Human-Machine Systems*, 50(6):475–491, 2020. 2

[16] Angelos Mavrogiannis, Rohan Chandra, and Dinesh Manocha. B-GAP: Behavior-guided action prediction for autonomous navigation. *arXiv preprint arXiv:2011.03748*, 2020. 1, 2

[17] Barbara Metz, Nadja Schömig, and Hans-Peter Krüger. Attention during visual secondary tasks in driving: Adaptation to the demands of the driving task. *Transp. Res. Part F Traffic Psychol. Behav.*, 14(5):369–380, Sept. 2011. 2

[18] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb. 2015. 2

[19] Jeremy Morton and Mykel J Kochenderfer. Simultaneous policy learning and latent state inference for imitating driver behavior. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6, Oct. 2017. 3

[20] Kate Nussenbaum, Maximilian Scheuplein, Camille V Phaneuf, Michael D Evans, and Catherine A Hartley. Moving developmental research online: Comparing in-lab and web-based studies of model-based reinforcement learning. *Collabra Psychol.*, 6(1), Nov. 2020. 2

[21] Jami Pekkanen, Otto Lappi, Paavo Rinkkala, Samuel Tuhkanen, Roosa Frantsi, and Heikki Summala. A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task. *R. Soc. Open Sci.*, 5(9):180194, Sept. 2018. 2

[22] Alexander Peysakhovich. Reinforcement learning and inverse reinforcement learning with system 1 and system 2. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, pages 409–415, New York, NY, USA, Jan. 2019. Association for Computing Machinery. 2

[23] J Reason, A Manstead, S Stradling, J Baxter, and K Campbell. Errors and violations on the roads: a real distinction? *Ergonomics*, 33(10-11):1315–1332, Oct. 1990. 2

[24] Michael A Regan, John D Lee, and Kristie Young. *Driver distraction: Theory, effects, and mitigation*. CRC press, 2008. 2

[25] Katharina Reinmueller, Linda Koehler, and Marco Steinhauser. Adaptive warning signals adjusted to driver passenger conversation: Impact of system awareness on behavioral adaptations. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58:242–252, 2018. 2

[26] Guillermo Reyes. An adaptive and personalized In-Vehicle Human-Machine-Interface for an improved user experience. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, IUI '20, pages 35–36, New York, NY, USA, Mar. 2020. Association for Computing Machinery. 2

[27] Jonathan J. Rolison, Shirley Regev, Salissou Moutari, and Aidan Feeney. What are the factors that contribute to road accidents? an assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accident Analysis Prevention*, 115:11–24, 2018. 2

[28] Charles Schaff and Matthew R Walter. Residual policy learning for shared autonomy. In *Proceedings of Robotics: Science and Systems (RSS)*, 2020. 6

[29] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proc. Natl. Acad. Sci. U. S. A.*, 116(50):24972–24978, Dec. 2019. 2

[30] Konstantin Shmelkov, Cordelia Schmid, and Karteek Alahari. How good is my GAN? In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 213–229, 2018. 2

[31] Patrick J Smith, Anna C Need, Elizabeth T Cirulli, Ornit Chiba-Falek, and Deborah K Attix. A comparison of the cambridge automated neuropsychological test battery (CANTAB) with "traditional" neuropsychological testing instruments. *J. Clin. Exp. Neuropsychol.*, 35(3):319–328, Feb. 2013. 2

[32] Gijsbert Stoet. PsyToolkit: A novel Web-Based method for running online questionnaires and Reaction-Time experiments. *Teach. Psychol.*, 44(1):24–31, Jan. 2017. 2

[33] Olaf Stroosma, M M (rené) van Paassen, and Max Mulder. Using the SIMONA research simulator for human-machine interaction research. In *AIAA Modeling and Simulation Technologies Conference and Exhibit*, Guidance, Navigation, and Control and Co-located Conferences. American Institute of Aeronautics and Astronautics, Aug. 2003. 2

[34] DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. Collaborating with humans without human data. *Advances in Neural Information Processing Systems*, 34, 2021. 2

[35] Matus Sucha, Lenka Sramkova, and Ralf Risser. The manchester driver behaviour questionnaire: self-reports of aberrant behaviour among czech drivers. *European Transport Research Review*, 6(4):493–502, Sept. 2014. 2

[36] Pradyumna Tambwekar, Andrew Silva, Nakul Gopalan, and Matthew Gombolay. Specifying and interpreting reinforcement learning policies through simulatable machine learning. *arXiv preprint arXiv:2101.07140*, 2021. 2

[37] Yu Tian, Zhangkai Ni, Baoliang Chen, Shiqi Wang, Hanli Wang, and Sam Kwong. Generalized visual quality assessment of GAN-generated face images. *arXiv preprint arXiv:2201.11975*, 2022. 2

[38] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033, Oct. 2012. 2

[39] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Phys. Rev. E*, 62:1805–1824, Aug 2000. 4

[40] Pedro A Tsividis, Joao Loula, Jake Burga, Nathan Foss, Andres Campero, Thomas Pouncy, Samuel J Gershman, and Joshua B Tenenbaum. Human-level reinforcement learning through theory-based modeling, exploration, and planning. *arXiv preprint arXiv:2107.12544*, 2021. 2

[41] Markus Wulfmeier, Dushyant Rao, Dominic Zeng Wang, Peter Ondruska, and Ingmar Posner. Large-scale cost function learning for path planning using deep inverse reinforcement learning:. *Int. J. Rob. Res.*, Aug. 2017. 2

[42] Rachael A Wynne, Vanessa Beanland, and Paul M Salmon. Systematic review of driving simulator validation studies. *Saf. Sci.*, 117:138–151, Aug. 2019. 2