

Driver behavior during cyclist overtaking: A comparison of test environments

Downloaded from: https://research.chalmers.se, 2026-11-23 13:15 UTC

Citation for the original published paper (version of record):

Rasch, A., Thalya, P., Sandin, J. et al (2027). Driver behavior during cyclist overtaking: A comparison of test environments. Transportation Research Part F: Traffic Psychology and Behaviour, 124. http://dx.doi.org/10.1016/j.trf.2026.103860

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Transportation Research Part F: Psychology and Behaviour 124 (2027) 103860

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

Driver behavior during cyclist overtaking: A comparison of test environments

Alexander Rasch [a,*], Prateek Thalya [b], Jesper Sandin [c], Marco Dozza [a]

[a] Chalmers University of Technology, Hörsalsvägen 7B, Göteborg, Sweden
[b] Magna Electronics Sweden AB, Wallentinsvägen 22, Vårgårda, Sweden
[c] Swedish National Road and Transport Research Institute (VTI), Regnbågsgatan 1, Göteborg, Sweden

ARTICLE INFO

Keywords:
Driver behavior
Traffic safety
Test track
Driving simulator
Virtual reality
Bayesian modeling

ABSTRACT

Artificial test environments are indispensable for modeling human behavior in traffic because they provide more control of experimental factors and fewer ethical concerns than a real traffic environment. These advantages are particularly critical for research on interactions involving vulnerable road users such as cyclists and drivers of motorized vehicles. However, previous research comparing test environments has predominantly focused on car-to-car interaction scenarios. This study investigated a cyclist-overtaking scenario in three different environments: a driving simulator (N = 25), a test track (N = 18), and a hybrid "driver-vehicle-in-the-loop" environment (N = 33) in which the participants drove a real vehicle on a test track while wearing a virtual-reality headset. Results indicated that the main factors influencing overtaking strategies and performance were consistent across all environments. Drivers tended to favor a cautious approach—waiting behind the cyclist—when facing a close oncoming vehicle, with this tendency being most pronounced in the driving simulator. Safety margins to the cyclist followed a similar trend in all environments; however, the test-track environment had the smallest margins, suggesting that virtual visual cues induce safer behavior. Because overtaking maneuvers depend on depth estimation, the visual resolution of the virtual environment may be more important than the environment itself in explaining the results.

1. Introduction

Drivers' behavior must be understood and predicted in order to proactively address traffic safety issues. Models simulating their behavior can play a crucial role, providing a reference for traffic planning and regulation, infrastructure design, and advanced driver assistant systems (ADAS) (AbuAli & Abou-zeid, 2016). In particular, it has become more important to simulate drivers' interactions with vulnerable road users (VRUs: mainly cyclists and pedestrians), due to their increased popularity and presence in traffic and the fact that they often need to share the roads with motorized traffic (Organization, 2023).

Collecting data from individual drivers is essential for capturing and analyzing their behavior. Such data may be collected in different *environments*, with or without real traffic. *Naturalistic studies* and *field tests* both take place in real traffic. Naturalistic studies measure drivers' everyday driving behavior in an unobtrusive manner, using sensor-equipped vehicles (Kovaceva et al., 2019), site-based sensors (Rasch et al., 2023), or drones (Bock et al., 2020; Krajewski et al., 2018), for instance. Field tests involve sensor-equipped road users in more controlled experiments, such as a specific scenario or intersection (Dozza et al., 2016; Rasch et al., 2020b).

* Corresponding author.
E-mail address: alexander.rasch@chalmers.se (A. Rasch).

https://doi.org/10.1016/j.trf.2026.103860
Received 24 March 2026; Received in revised form 28 October 2026; Accepted 28 October 2026
Available online 4 November 2026
1369-8478/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Studies conducted in real traffic typically result in data with the highest ecological validity (Bärgman, 2016). However, the applicability of such data may be compromised by the necessary ethical and practical restrictions on the experimental setup. For instance, critical interactions with real humans may not be repeatable and the risk is not morally justifiable. Furthermore, data collected on real roads typically contain many confounders that may impact statistical analyses, leading to false conclusions if not taken into account properly (Denk et al., 2023). Additionally, naturalistic studies may need to utilize lower sampling rates and measurement resolution/accuracy than other types of studies, since they need to run over longer time horizons to capture enough events of interest and therefore have more limitations on sensing equipment and data storage. These characteristics may limit the suitability of real-traffic data for the complex driver models that systems like ADAS require, and may weaken the conclusiveness of analyses over short spatial or temporal intervals. For instance, measures like passing distance or time-to-collision (TTC) need to be measured accurately at the exact moments in time because they may determine critical interactions between drivers and VRUs.

Studies carried out in environments without real traffic typically take the form of randomized, controlled trials with volunteer participants on a *test track* or in a *driving simulator* (see, for example, Boda et al., 2018). These artificial environments have less ecological validity than real-traffic studies, but, on the other hand, they offer greater safety and better repeatability. Test-track experiments are carried out with real vehicles driven, but for ethical and repeatability reasons any other road users, such as cyclists or other vehicles, are usually soft dummy targets, such as balloon forms, which do not cause damage when struck. Simulator experiments are typically run indoors in mock-up vehicles with or without a motion base that generates motion cues for the participants. While kinematics are more realistic in those with a motion base, in both types of simulators kinematics and visuals are artificial. Both types typically use static screens around the vehicle to provide participants with a view of the simulated environment. Previous studies have explored hybrid/virtual-reality arrangements, with head-mounted displays (HMDs) instead of static screens, both in simulators (Denk et al., 2023) and on test tracks (Blissing et al., 2019; Purucker et al., 2018). Such displays may offer better visual cues to participants compared to static screens through a more immersive driving experience; however, they may come at the expense of increased simulator sickness (Denk et al., 2023).

Previous research efforts have focused on comparing aspects of driver behavior between simulator and real-world studies to assess the simulator validity. Those studies have generally agreed that driving simulators can at least provide relative validity, i.e., reproducing behavioral trends found in real traffic—while sometimes even providing absolute validity, i.e., reproducing the actual behavioral effects found in real traffic (Fisher et al., 2011). For instance, Klüver et al. (2016) compared driver behavior in a real-world field test and in five different simulator environments. They showed that all five simulators could at least be regarded as relatively valid, as indicated either by a non-significant effect or low effect size of the test environments on metrics such as the standard deviation of lane position and time headway. The simulators with a motion base yielded the closest similarities to real-world data. A similar study by Himmels et al. (2024) compared six different driving simulators with a real-world field test and showed that all simulators induced behavior similar to that seen in real-world driving, as determined via Bayesian ANOVA for similar positioning metrics as used by Klüver et al. (2016) and speed. They found that their environments generally provided absolute validity for metrics related more to vehicle positioning than speed control. Further, less simulator sickness was reported in the higher-fidelity simulators. In addition, Moll et al. (2023) showed that a simulator experiment and a field test yielded similar results regarding driver behavior when overtaking cyclists, both in terms of overtaking speeds and lateral clearance to the cyclist.

A study by Purucker et al. (2018) compared drivers' *perceived* criticality in different car-following scenarios for four different artificial traffic environments: a simulator without a motion base, one with a motion base, a test track, and a hybrid setup in which participants drove the vehicle on a test track but viewed the scenario virtually with an HMD. They assessed differences between environments in terms of self-rated perceived criticality by fitting ordinal logistic regression models and investigating the significance of the model parameters. Their results indicate that drivers judge scenarios as more critical in environments with artificial visuals than in a test-track environment.

Test-track experiments may represent a valid compromise between simulator and real-world studies; however, they have rarely been included in environment comparisons, particularly for VRU-interaction scenarios. Boda et al. (2018), Kovaceva et al. (2020), and Denk et al. (2023) are among the few examples of such comparisons. Boda et al. (2018) compared driver behavior in a simulator and on a test track for straight-crossing interactions with cyclists, showing that brake-onset behavior was similar, while drivers braked harder in the simulator. Kovaceva et al. (2020) used cyclist-overtaking data from a naturalistic study and a test-track study to show that drivers in both datasets used similar visual cues to avoid rear-end crashes by steering, while behavioral models fitted on naturalistic data showed superior performance. Denk et al. (2023) analyzed right-turn interactions between a driver and a cyclist in two simulator environments without a motion base (using static screens vs. an HMD) and on a test track. While participants subjectively rated the realism of both simulator environments higher than that of the test track, the simulator produced a substantial number of dropouts due to motion sickness.

Demographic factors such as age, gender, and driving experience have consistently been shown to influence driver behavior, as evidenced by empirical studies based on both real world driving (test track or on road) and driving simulators. Results from real world data indicate that older drivers and drivers with lower annual mileage generally adopt more conservative longitudinal strategies, characterized by lower speeds, larger following distances, and larger time margins (e.g. Joshi et al., 2025). Simulator studies largely reproduce these relative differences between demographic groups, showing similar trends in car following behavior and risk regulation (Michaels et al., 2017; Ramezani-Khansari et al., 2021). However, absolute values of safety metrics such as time headway and TTC often differ between simulator and real world driving, reflecting differences in sensory feedback, perceived control, and risk. Thus, while simulators capture demographic effects on driver behavior reliably in relative terms, real world driving data remain essential for accurately quantifying absolute safety margins.

Previous work has mainly focused on the verification and validation of test environments, most often driving simulators. However, test-track environments, particularly hybrid environments that present a compromise between simulation and reality, have rarely been included in such comparisons. Furthermore, most studies comparing driver behavior in different environments have focused on interactions with other cars, while interactions with VRUs have rarely been addressed.

Cyclist overtaking is a particularly challenging VRU scenario for drivers as it requires subtle and accurate lateral control to reach a suitable clearance from the cyclist, as well as carefully planned longitudinal control and decision-making to plan whether to overtake in the presence of oncoming traffic (Dozza et al., 2016; Kovaceva et al., 2019). For cyclists, being overtaken is the most common and most dangerous scenario on rural roads where speed limits are high and dedicated cyclist infrastructure absent (Moll et al., 2023), resulting in high injury severity in the event of a collision (Kircher & Sandin, 2025). Swedish crash data indicate that fatal rear-end collisions involving cyclists occur more frequently on rural than on urban roads (Kullgren et al., 2019). Empirical studies show that cyclists are more likely to perceive an overtake as uncomfortable when clearance is small, when the overtaking vehicle is traveling at a higher speed, when the vehicle is a heavy vehicle, or when the driver does not reduce speed before passing (Llorca et al., 2017; López et al., 2020). Overall, both objective safety and perceived safety during overtaking depend on a combination of lateral clearance, vehicle speed, vehicle type, and driver behavior.

Previous research has suggested that drivers tend to compensate for the risk of a head-on collision with oncoming traffic with an increase in collision risks with the cyclist (Rasch et al., 2020a). However, no prior study has investigated the role of different artificial data-collection environments. Therefore, the objective of this study was to perform an in-depth comparison of different artificial, state-of-the-art, data-collection environments for traffic safety research and development, using the scenario of cyclist overtaking.

2. Method

We obtained data from drivers in an experimental cyclist-overtaking scenario conducted in three different environments: 1) a test track (TT), 2) a driving simulator (SIM), and 3) a hybrid, driver-vehicle-in-the-loop (DVIL) setup (Fig. 1).

2.1. Experimental protocol

In all experiments, participants completed a cyclist overtaking on a straight road stretch (3.75m wide), resembling a rural road with a speed limit of 70 km/h. The simulated cyclist was riding at a constant 20 km/h in the same lane, 1m from the lane edge. Each participant completed three trials, one for each of the three different configurations: no oncoming vehicle (oncoming absent), 9 s (long gap), or 6 s (short gap). The gap times refer to the TTC to the oncoming vehicle, measured at the moment when the ego vehicle's TTC to the cyclist reached 2 s.

The gap times were chosen in accordance with those used in a previous simulator experiment on cyclist overtaking (Bianchi Piccinini et al., 2018). The values represented a compromise between short and long gaps that favor a distinct spread in overtaking strategies (execute vs. abort the attempt in presence of oncoming traffic), taken from the aforementioned study and assessed in pilot tests. While Bianchi Piccinini et al. (2018) is the best reference for this paper as it modulated TTC in cyclist overtaking, the practice of modulating TTC with oncoming traffic is well established for car-to-car overtaking. For example, field observations suggest that drivers typically accept gaps to oncoming traffic corresponding to approximately 10–12 s when overtaking another motorized vehicle (Hassein et al., 2017; Polus et al., 2000). Since cyclists travel much slower than motorized vehicles, the duration of the overtaking maneuver is considerably shorter, implying that drivers can be expected to accept shorter gaps in the order of the ones used in this study.

At the beginning, participants completed a training run in each environment to familiarize themselves with it. They were instructed to keep to the desired speed of 70 km/h to practice maintaining longitudinal control of the vehicle. Additionally, in the simulated environments (SIM and DVIL), cones were placed to let the participants practice maintaining lateral control of the vehicle when changing lanes. In all training runs, an oncoming vehicle was present in the adjacent lane to make participants aware of the possible presence of oncoming traffic. The order of the trials was randomized uniquely for each participant. In all environments, participants were instructed to drive as they would in real traffic.

After the experiment, participants answered a questionnaire capturing their demographics and their experience, including how realistic the test environment seemed Fig. C.1.

2.2. Environments

Table 1 gives an overview of the three environments and their properties.

2.2.1. Test track

The TT data were obtained in an experiment at Vårgårda airfield, Vårgårda, Sweden, in 2018 (Fig. 1, Panel a). Two road users were employed based on Euro NCAP-validated Guided Soft Targets: a robot cyclist mounted on a remote-controlled platform, and an oncoming balloon vehicle. They were both controlled by a CHRONOS server (Bjelkeflo et al., 2018). Participants drove a Volvo S60 fitted with data-collection equipment. In the two conditions with an oncoming vehicle, it traveled at a constant speed of 40 km/h, pulled by a motor-connected cable along a rail. The limited speed was a direct consequence of the maximum pull force of the motor, as well as the air resistance and safety risk of derailing. In the no-oncoming condition, the vehicle had to remain on the test track

Table 1
Overview of the three data-collection environments; test-track, driving simulator, and driver-vehicle-in-the-loop. The speed of the oncoming vehicle was, depending on the environment, either constant (TT), or adaptive (SIM, DVIL) in that it was changed dynamically to match the target time gaps more consistently.

Environment	TT	SIM	DVIL
Visuals	Real	Artificial	Artificial
Kinematics	Real	Artificial	Real
Cyclist speed	20 km/h constant	20 km/h constant	20 km/h constant
Cyclist lane position (m)	1.0	1.0	1.0
Oncoming speed	40 km/h constant	70 km/h adaptive	70 km/h adaptive
Road configuration	Single straight stretch	Continuous road	Single straight stretch
Lane width (m)	3.75	3.75	3.75
Screen resolution (arcmin/px)	NA	2.48 (horizontal), 1.06 (vertical)	0.86 (focus area), 2.00 (peripheral area)
Screen refresh rate (Hz)	NA	30	90
Year of collection	2018	2021	2022

(due to technical reasons) and was visible to participants but remained stationary at a far distance to impact driver behavior as little as possible.

Participants in the experiment were employees of Autoliv or Magna Electronics (formerly Veoneer Active Safety), two companies that produce vehicle safety systems. (None of the participants were involved in safety-system development.) Further details can be found in the work by (Rasch et al., 2020a).

2.2.2. Driving simulator

The SIM data were collected in VTI's advanced moving-base driving simulator Sim-IV in Gothenburg, Sweden, in 2021 (Fig. 1, Panel b). Its motion system is composed of a hexapod mounted on a sledge, providing eight degrees of freedom. A dome on the hexapod can host a passenger vehicle cabin, as in the present study, or a truck cabin. The visual system is independent of the cabin choice and provides the driver with a field of view (FoV) of almost 180 deg (Jansson et al., 2014). To ensure participants' immersion in the driving task and avoid the unnecessary braking and accelerating needed in the TT experiment (when returning to the starting position), the virtual world consisted of a continuous road made of straight stretches (where the overtaking maneuvers occurred) connected by curve elements. The oncoming vehicle was controlled to drive at 70 km/h initially, but its speed was adjusted while the driver was approaching the cyclist to ensure that the desired time gap (six or nine seconds, depending on the condition) was met at 2 s TTC to the cyclist. Participants were recruited through public advertisement.

2.2.3. Driver-vehicle-in-the-loop

The DVIL data were obtained by Volvo Cars and Magna on the same test track used for the TT data (Vårgårda airfield), in 2022 (Fig. 1, Panel c). The DVIL setup, developed by Volvo Cars, consisted of a real vehicle (Volvo V60) on the test track, so the drivers experienced real kinematic cues; however, they saw a simulated scenario through a virtual-reality HMD (Varjo XR-3, Varjo, Finland). The virtual surroundings were the same as in the SIM environment, and the other road users were controlled with the same algorithm. However, in this environment, participants drove on one straight stretch of road without curves, due to the physical constraints of the airfield. Participants were employees of Magna; however, as in the TT experiment, none of them were involved in safety-system development.

2.3. Participants

Participants in all environments had to have a valid driver's license, be between 18 and 60 years old, drive at least three days per week, and not drive professionally. Participants in the experiments were comparable between environments according to their age distribution, gender ratio, yearly mileage, and driver's-license age (Table 2). We excluded the data from participants who did not feel that the experiment was realistic at all (as determined from the post-drive questionnaire; three participants in TT, one in SIM, and four in DVIL).

2.4. Scenario definitions

To quantify driver behavior in the different environments, we characterized an overtaking maneuver with a set of phases and metrics, following previous research (Dozza et al., 2016; Kovaceva et al., 2019; Rasch et al., 2020a). Overall, drivers' maneuvering was divided into two distinct strategies: *flying* maneuvers, in which drivers overtook the cyclist before the oncoming vehicle and without slowing down, and *accelerative* maneuvers, in which drivers braked behind the cyclist to let the oncoming vehicle pass first, to then re-accelerate and pass the cyclist. Fig. 2 depicts the phases and metrics for a flying overtaking maneuver in the presence of an oncoming vehicle.

(a) Test track



(b) Driving simulator



(c) Driver-vehicle-in-the-loop



Fig. 1. Data-collection environments; Pure test track (a; left panel shows the cockpit view, right panel the robot cyclist), driving simulator (b; left panel shows cockpit view, right panel the simulator used, Sim-IV), and the driver-vehicle-in-the-loop setup used on a test track (c; left panel shows the cockpit view, right panel the driver wearing a virtual-reality headset). Video footage from the three environments can be seen at https://youtu.be/Fqzr1F7vP7g.

Table 2
Summary of participants in the experiments collected in the three different environments, test track (TT), driving simulator (SIM), and driver-vehicle-in-the-loop (DVIL).

Characteristic	TT (N = 18)	SIM (N = 25)	DVIL (N = 33)
Age (years)	43 (9)	41 (10)	40 (10)
Driver's license age (years)	25 (9)	22 (10)	20 (12)
Gender			
female	5 (28%)	9 (36%)	6 (18%)
male	13 (72%)	16 (64%)	27 (82%)
Yearly mileage (km)			
0-10.000	8 (44%)	5 (20%)	2 (6.1%)
10.000–20.000	7 (39%)	13 (52%)	18 (55%)
20.000–30.000	2 (11%)	5 (20%)	10 (30%)
more than 30.000	1 (5.6%)	2 (8.0%)	3 (9.1%)
Mean (SD); n (%)			

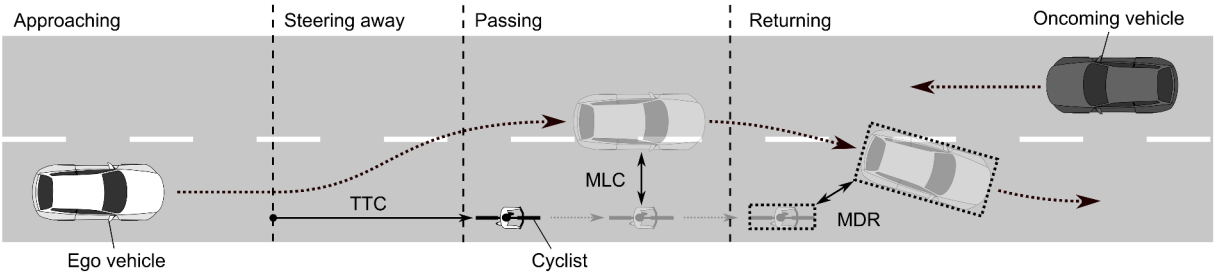

Fig. 2. Cyclist-overtaking scenario (flying strategy in the presence of an oncoming vehicle). Phases: approaching, steering away, passing, and returning. Metrics: TTC (time-to-collision of the ego vehicle to the cyclist at the moment of steering onset), MLC (minimum lateral clearance), and MDR (minimum distance returning).

Table 3
Overtaking metrics and explanations.

Phase	Acronym	Metric	Unit	Definition
Approaching	TTC	TTC to cyclist at action	s	TTC to cyclist at moment of action; brake onset for accelerative, steer onset for flying maneuvers
Passing	MLC	Minimum Lateral Clearance	m	Lateral distance between left side of cyclist and right side of ego vehicle at the moment of passing
Returning	MDR	Minimum Distance Returning	m	Minimum Euclidean distance during the returning phase

2.5. Overtaking phases

The different environments used different vehicles, which had different steering and pedal systems. Therefore, to make the transitions between the phases consistent and comparable, we relied mainly on position and speed measures, as opposed to steering and pedal signals. We further relied as much as possible on phase definitions from previous similar work (Rasch et al., 2020a). We defined steering onset (the end of the approaching phase in a flying maneuver) as the moment when the change in lateral position of the ego vehicle exceeded 0.2m. Braking onset (in an accelerative maneuver) was defined as the moment when the speed of the ego vehicle dropped by at least 10 km/h in 3 s. The start of the passing phase was determined as the first moment when the lateral clearance to the cyclist was within 0.2m of the maximum lateral clearance during the maneuver, and the end of the passing phase when the lateral clearance decreased again by 0.2m from its maximum.

2.6. Overtaking metrics

We also defined a set of metrics to quantify driver behavior in the different phases of the overtaking maneuver (Fig. 2). These metrics characterize the driver's comfort zone when circumventing the cyclist, in accordance with previous work (Dozza et al., 2016; Kovaceva et al., 2019; Rasch et al., 2020a). In the approaching phase, we calculated the TTC of the ego vehicle's front bumper to the rear end of the cyclist, at the moment of the first evasive action (steering onset for flying maneuvers, and braking onset for accelerative maneuvers). In the passing phase, we calculated the minimum lateral clearance (MLC) between the right side of the ego vehicle and the left side of the cyclist when both road users are next to each other. In the returning phase, we calculated the minimum (Euclidean) distance between the two rectangular bounding boxes of the ego vehicle and cyclist (minimum distance re7turning: MDR). Table 3 summarizes the metrics and their explanations.

2.7. Statistical models

To quantify the differences (including uncertainties) between the environments, we employed Bayesian regression models. Following the work by Klüver et al. (2016), the models were fitted on the data from all environments with a categorical variable for the environment and interactions with the other independent variables. All models were fitted in *brms* 2.21.0 (Bürkner, 2017), an R package for fitting Bayesian regression models that employs Markov-Chain Monte Carlo (MCMC) sampling via the No-U-Turn Sampler algorithm. In contrast to frequentist regression, Bayesian regression enables the quantification of the full probability distribution of the models' parameters and allows the specification of prior distributions that may help regularize the model in order to prevent overfitting. We generally summarized the uncertainty distribution with the 95% highest-density interval (HDI). In line with the Bayesian analysis reporting guidelines by Kruschke (2021), we inspected the convergence of the MCMC sampler via qualitative, visual inspections of the chains (Bürkner, 2017). We also determined quantitatively that the convergence was consistent within and between chains by verifying that the R-hat statistic was close to one (between 1.0 and 1.1). Furthermore, we verified the fit of the model to the data through posterior predictive checks (per Gelman et al. (2019); provided in the Supplementary Material).

2.7.1. Overtaking strategy

We modeled the binary choice of overtaking strategy (flying vs. accelerative) with a Bernoulli distribution, which samples zeros (accelerative) or ones (flying), depending on the probability of a flying strategy (p_i) as follows:

$$\begin{aligned} &\text{Strategy}_i \sim \text{Bernoulli}\ (p_i) \\ &\text{logit}\ p_i = \beta_0^T\ \mathbf{env}_i + \left(\beta_{\text{gap}}^T\ \mathbf{env}_i\right)\ \text{gap}_i + \left(\beta_{\text{gender}}^T\ \mathbf{env}_i\right)\ \text{gender}_i + \left(\beta_{\text{age}}^T\ \mathbf{env}_i\right)\ \text{age}_i \end{aligned} \tag{1}$$

The logit link function is used to transform the linear predictor to the probability scale ($p \in [0, 1]$). The model contains an intercept (parameter β_0), symbolizing the overall strategy tendency, the time gap to the oncoming vehicle (gap, a dummy binary variable; 0 = long gap, 1 = short gap), gender (gender, 0 = male, 1 = female), and age (age, the standardized, continuous, age in years). The vector **env** comprises three dummy variables ($\mathbf{env} = [\text{TT},\ \text{SIM},\ \text{DVIL}]^T$). Each is binary (binary; 0 = not the environment, 1 = the environment). For example, $[1, 0, 0]^T$ indicates the TT environment. Each parameter vector β contains a parameter for each environment ($\beta = \left[\beta^{\text{TT}},\ \beta^{\text{SIM}},\ \beta^{\text{DVIL}}\right]^T$); therefore, the environment variables are multiplied element-wise by their corresponding parameters ($\beta^T \mathbf{env}$), representing interactions with the other independent variables. We used weakly-informative prior distributions for the parameters; normal distributions with means of zero and a standard deviations of one. We fitted the model only on the trials with an oncoming vehicle present, because in the absence of an oncoming vehicle a driver would always perform a flying maneuver.

2.7.2. Comfort-zone metrics

Due to their physical nature, comfort-zone metrics are always greater than zero and generally skewed right, so we modeled them with log-normal distributions, as in previous work (Rasch et al., 2020a) We included the effect of strategy (flying vs. accelerative) as a dummy variable St (0 = accelerative, 1 = flying). Similarly, the presence of the oncoming vehicle was coded as OP (0 = absent, 1 = present). The variable was included through an interaction with strategy to ensure that it only added an effect to the model when the strategy was flying (an accelerative strategy indicates that the oncoming vehicle has already passed). The time gap to the oncoming vehicle (gap) was coded binary as in the strategy model (0 = long gap, 1 = short gap), through an interaction with strategy and oncoming presence. Gender and age were included in the same way as in the strategy model. We also included the effect of the environment through interactions with the dummy variable env as in the strategy model. We used non-informative default prior distributions: flat, uniform distributions for the β parameters, Cholesky Lewandowski-Kurowicka-Joe (LKJ) correlation distributions (shape parameter 1) for the correlation of random effects between the environments, and Student's t distributions (location zero, scale 2.5, three degrees of freedom) for all standard deviations. We fitted the model on all trials—except for one accelerative maneuver in SIM, in which the driver did not perform any braking at all apart from engine braking, which we could not tell apart from the target-speed regulation.

$$\begin{aligned} &\text{metric}_i \sim \text{LogNormal}\ (\mu_{\text{metric},i}, \sigma_{\text{metric}}) \\ &\mu_{\text{metric},i} = \beta_0^T\ \mathbf{env}_i + \left(\beta_{\text{St}}^T\ \mathbf{env}_i\right)\ \text{St}_i + \left(\beta_{\text{OP}}^T\ \mathbf{env}_i\right)\ \text{St}_i\ \text{OP}_i \\ &\qquad + \left(\beta_{\text{gap}}^T\ \mathbf{env}_i\right)\ \text{St}_i\ \text{OP}_i\ \text{gap}_i \\ &\qquad + \left(\beta_{\text{gender}}^T\ \mathbf{env}_i\right)\ \text{gender}_i + \left(\beta_{\text{age}}^T\ \mathbf{env}_i\right)\ \text{age}_i \\ &\qquad + u_{\text{ID}\times\text{env},i}, \\ &u_{\text{ID}\times\text{env},i} \sim N(0, \sigma_{\text{ID}\times\text{env}}), \end{aligned} \tag{2}$$

2.7.3. Inference

Our main goal for this study was to quantify the differences between the environments. Therefore, we developed comparative analyses on both qualitative and quantitative levels. For qualitative analyses, we first investigated the fitted model parameters from all environments. We identified a clear effect of a model parameter if its probability of direction (PD), i.e. the proportion of the parameter's posterior distribution sharing the same sign as its median (Makowski et al., 2019), is greater than a threshold (we used an arbitrary threshold of 95%; different stakeholders may use a different percentile). We focused on clear effects in order to keep the

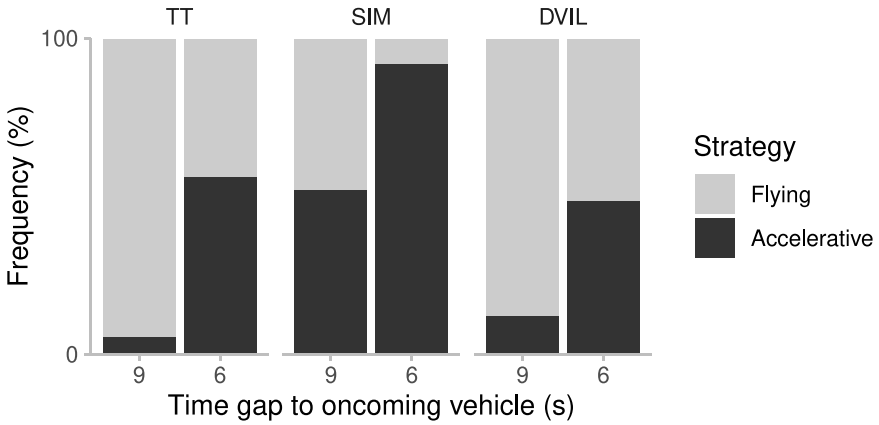

Fig. 3. Strategy distribution (frequency of flying and accelerative strategies) for the three environments TT (test track), SIM (driving simulator), and DVIL (driver-vehicle-in-the-loop), grouped by time gap to the oncoming vehicle.

results tractable and highlight the most important factors affecting driver behavior. The reader is cautioned not to discard effects that did not exceed our threshold of probability.

For quantitative analyses, we calculated the pairwise differences, or contrasts, between environments using the model predictions. All predictions were made based on the expected value of the predictor (Eqs. (3), (4)); for strategy the predictor was the probability of a flying maneuver, while for all other metrics the value of the (linear) predictor (μ) was transformed back to the data scale (e.g. meters). All contrasts were calculated with the *marginaleffects* package (Arel-Bundock et al., 2024), which can calculate marginal effects from the posterior samples of Bayesian models, thereby retaining the full probability distribution. For each model, we calculated overall contrasts, i.e., differences between the average predictions for all data for each environment (Eq. (3)). Then, for each variable in the models, we calculated contrasts of so-called *average marginal effects* (AMEs), which are slopes of the metric of interest. They capture the change in the outcome metric under a small change of the variable of interest (Eq. (4)). As for the qualitative analyses, we identified clear contrasts from the PD of the contrast distribution (when greater than 95% in our case).

$$\text{Contrast}(\text{metric},\ j-k) = \mathbb{E}\left[\text{metric} \mid \text{env} = j\right] - \mathbb{E}\left[\text{metric} \mid \text{env} = k\right] \tag{3}$$

$$\text{Contrast}(\text{metric},\ \text{var},\ j-k) = \mathbb{E}\left[\frac{\partial \text{metric}}{\partial \text{var}} \,\middle|\, \text{env} = j\right] - \mathbb{E}\left[\frac{\partial \text{metric}}{\partial \text{var}} \,\middle|\, \text{env} = k\right] \tag{4}$$

3. Results

3.1. Overtaking strategy

Participants in SIM performed more accelerative maneuvers than those in TT and DVIL (Fig. 3). The contrasts revealed that, overall, participants in SIM had a 0.37 ([0.19, 0.54] 95% HDI) lower probability of performing a flying maneuver than participants in TT, and a 0.38 ([0.23, 0.52] 95% HDI) lower probability than those in DVIL. TT and DVIL had similar overall strategy outcomes (Table B.3).

In all three environments, drivers' strategy choice appeared to be influenced by the time gap to the oncoming vehicle, resulting in fewer flying maneuvers when the time gap was shorter (Fig. 3). The model confirmed the clear reducing influence of time gap on the probability of a flying strategy for all three environments (Fig. 4), without clear differences between them (Table B.3).

Gender and age did not appear to have any clear influence on strategy choice across environments (fig-strategy-model). However, there was a notable difference in the effect of age between SIM and DVIL (-0.11 ([-0.29, 0.06] 95% HDI)): older participants in SIM tended to prefer more flying maneuvers while older participants in DVIL preferred more accelerative maneuvers (Table B.3).

3.2. Overtaking metrics

The lateral clearance profiles suggest that participants' lateral positioning differed during the approaching phase in the different environments (Fig. 5). Participants in SIM kept the closest lateral clearance to the cyclist until about 25m behind the cyclist in accelerative strategies and about 100m behind the cyclist in flying strategies. In contrast, participants in DVIL stayed further left, and participants in TT stayed yet further to the left (for both strategies). In TT, drivers performing an accelerative strategy steered shortly back just before reaching the cyclist (Fig. 5). They also had the smallest lateral clearance in the passing phase, followed by SIM and then DVIL. In addition, in the returning phase (starting about 10m ahead of the cyclist), they kept the smallest clearances to the cyclist, followed by DVIL and then SIM.

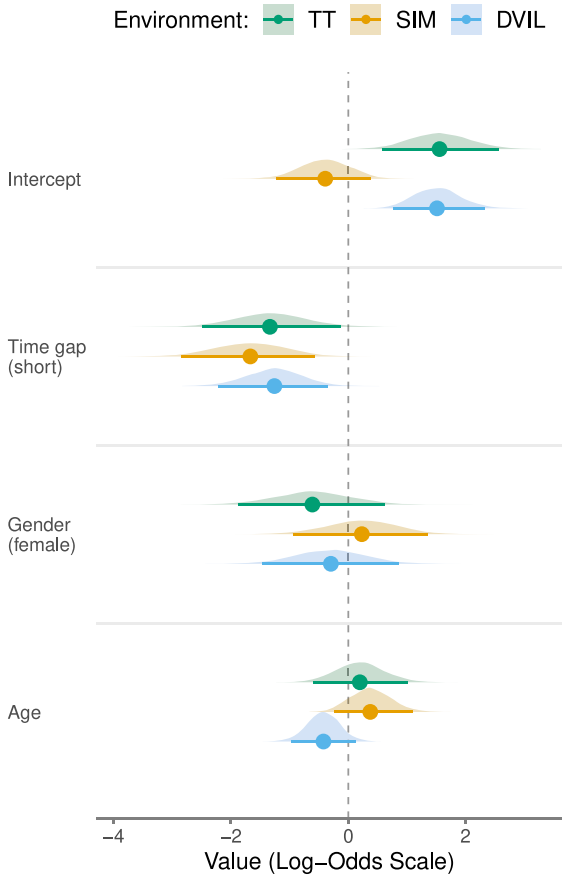

Fig. 4. Strategy model; each parameter posterior distribution is shown as density plot and summarized with median (dot) and 95% highest-density interval (horizontal bar). The x-axis shows the parameter value on the log-odds scale; negative values indicate a higher probability of performing an accelerative maneuver while positive values indicate a higher probability of performing a flying maneuver. The numerical values of all parameters can be found in Table A.1.

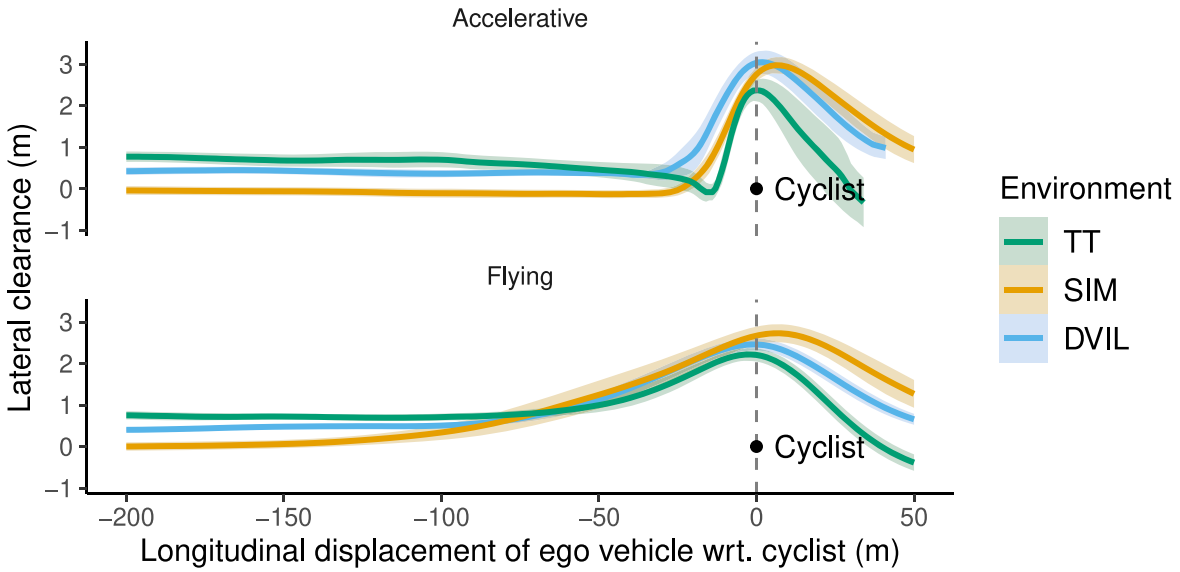

Fig. 5. Profiles of lateral clearances to the cyclist in different environments under different strategies. The thick lines show the means and the shaded areas the 95% confidence intervals.

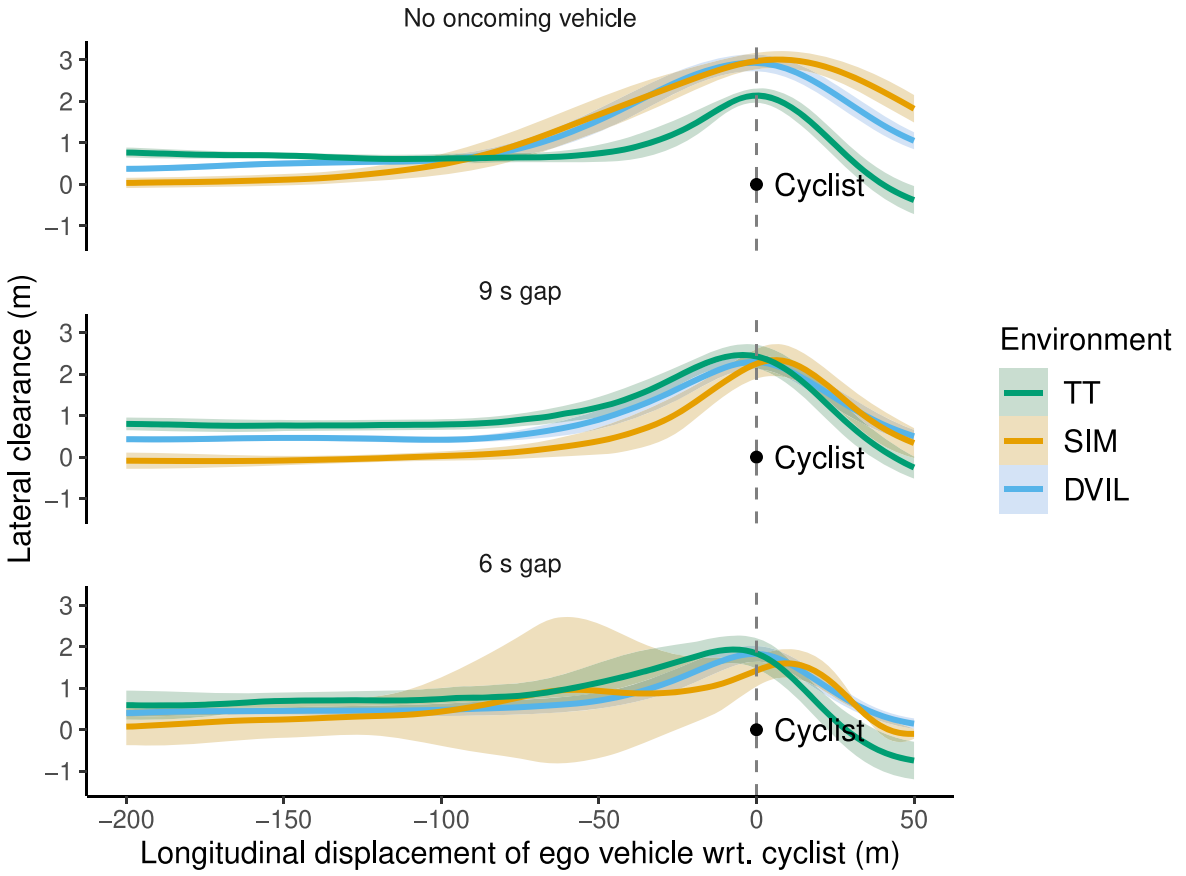

Fig. 6. Profiles of lateral clearance to the cyclist in the different environments for flying strategies, grouped by the different oncoming vehicle configurations (no oncoming, 9 s gap, and 6 s gap). The thick lines show the mean and the shaded areas the 95% confidence intervals.

The three environments also showed trajectory differences in flying strategies carried out under the different oncoming vehicle configurations. For instance, with no oncoming vehicle, participants in TT maintained a closer trajectory to the cyclist in both longitudinal and lateral directions than participants in the other environments during the whole maneuver, while those in SIM and DVIL were similar to each other (Fig. 6).

A deeper look into the individual comfort metrics revealed that most of them were distributed reasonably similarly between the environments (Fig. 7), under different strategies and oncoming vehicle configurations. With respect to means and standard deviations, distributions appeared generally similar between environments in scenarios with similar numbers of observations. However, the contrasts revealed that, overall, participants in TT kept shorter TTCs than participants in SIM (TT-SIM: -1.58 ([-2.58, -0.66] 95% HDI)) and DVIL (TT-DVIL: -1.17 ([-2.05, -0.21] 95% HDI)), while SIM and DVIL were more similarly distributed. Contrasts for MLC showed a similar trend: clearances were shorter in TT overall (TT-SIM: -0.35 ([-0.66, -0.06] 95% HDI), TT-DVIL: -0.34 ([-0.62, -0.07] 95% HDI)). In the returning phase, SIM accounted for clearly larger safety margins (MDRs) than the other environments (TT-SIM: -10.16 ([-14.85, -5.8] 95% HDI)), SIM-DVIL: 8.28 ([4.18, 13.17] 95% HDI)), while TT and DVIL were more similar to each other (TT-DVIL: -1.86 ([-4.22, 0.81] 95% HDI); PD = 0.93).

The model coefficients revealed that the strategy choice had a clear influence on the TTC to the cyclist at the first evasive action (steering for flying, braking for accelerative) for all environments: the TTC at this point was shorter under a flying strategy than an accelerative one. Clear differences between environments mainly occurred between TT and DVIL. The effect of strategy on TTC was -2.18 ([-4.23, 0] 95% HDI) s lower in TT than in DVIL, and the effect of strategy on MLC was -0.27 ([-0.61, 0.07] 95% HDI) m lower (PD = 0.94) in TT than in DVIL. On the other hand, no notable differences between TT and SIM could be identified for the effect of strategy on TTC or MLC (Table B.4). For MDR, the effect of strategy differed most between TT and the other environments, with a greater increase in MDR for flying maneuvers (5.13 ([-0.88, 12.46] 95% HDI) m, TT-DVIL: 5.52 ([-0.14, 13.12] 95% HDI) m; Table B.4).

The presence and timing of the oncoming vehicle influenced participants in all three environments, particularly for MLC (Fig. 8). However, while safety margins were generally reduced when the oncoming vehicle was closer, TT participants' behavior differed from the other environments in the approaching and passing phases in the mere presence of the oncoming vehicle. While in SIM and DVIL, both TTC and MLC were reduced in the presence of an oncoming vehicle, the opposite trend appeared for TT. The estimated contrast

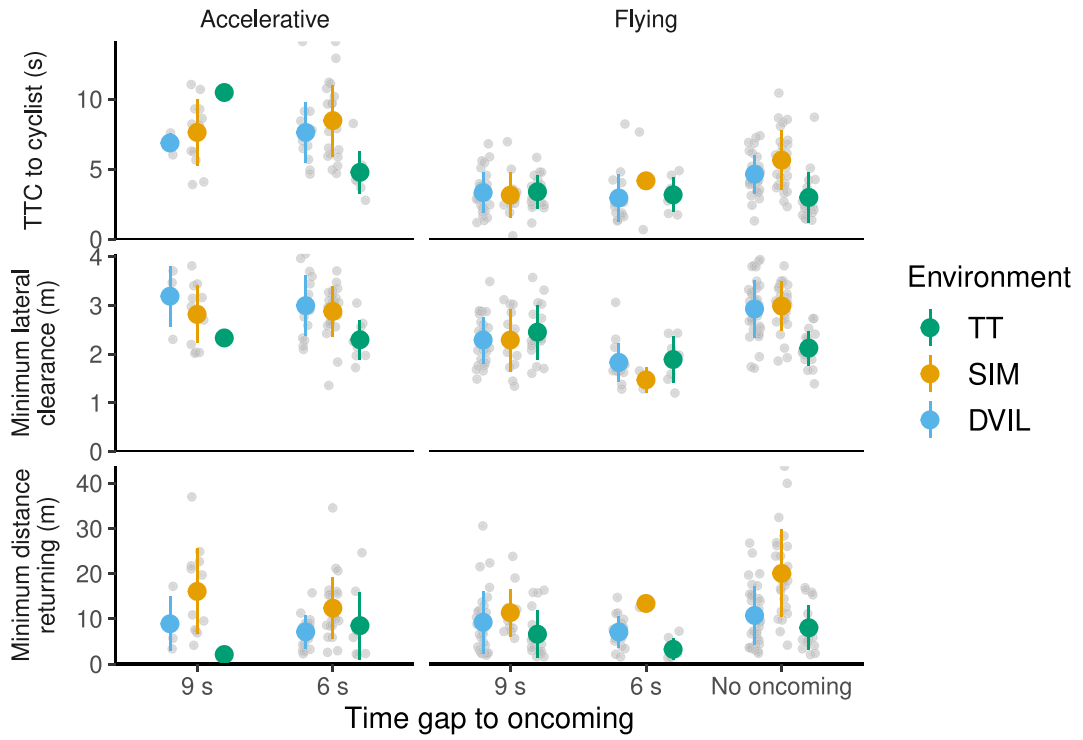

Fig. 7. Comfort-zone metrics data for each environment, displayed as individual observations (gray dots), mean (colored dot), and +/- one standard deviation (vertical bars).

in the effect of the presence of the oncoming vehicle was clear between all environments for TTC: -1.46 ([-2.9, -0.15] 95% HDI) s (TT-SIM), -3.45 ([-4.88, -2.1] 95% HDI) s (TT-DVIL), and -2.01 ([-3.09, -1] 95% HDI) s (SIM-DVIL). For MLC, the effect of the presence of oncoming traffic in TT differed clearly from the other environments: TT-SIM -0.18 ([-0.47, 0.12] 95% HDI) m, and TT-DVIL -1.05 ([-1.34, -0.71] 95% HDI) (Table B.4). For MDR, there were no clear differences between environments. However, participants in SIM returned with notably (PD $>$ 0.90) larger margins when the time gap was short compared to the other environments (TT-SIM: 9.75 ([-3.49, 29.26] 95% HDI) m, SIM-DVIL: -0.63 ([-3.52, 2.29] 95% HDI) m; Table B.4).

The presence and timing of the oncoming vehicle influenced participants in all three environments, particularly for MLC (Fig. 8). However, while safety margins were generally reduced when the oncoming vehicle was closer, TT participants' behavior differed from the other environments in the approaching and passing phases in the mere presence of the oncoming vehicle. While in SIM and DVIL, both TTC and MLC were reduced in the presence of an oncoming vehicle, the opposite trend appeared for TT. The estimated contrast in the effect of the presence of the oncoming vehicle was clear between all environments for TTC: -1.46 ([-2.9, -0.15] 95% HDI) s (TT-SIM), -3.45 ([-4.88, -2.1] 95% HDI) s (TT-DVIL), and -2.01 ([-3.09, -1] 95% HDI) s (SIM-DVIL). For MLC, the effect of the presence of oncoming traffic in DVIL differed clearly from the other environments (TT-DVIL: -1.05 ([-1.34, -0.71] 95% HDI) m, SIM-DVIL -0.86 ([-1.12, -0.62] 95% HDI); Table [tbl-metrics-contrasts]). For MDR, the effect of time gap differed between TT and the other environments (TT-SIM: 9.75 ([-3.49, 29.26] 95% HDI) m, TT-DVIL: 9.04 ([-4.48, 28.04] 95% HDI) m with PD = 0.93).

With respect to demographics, only age had an effect. It appeared to influence MDR in DVIL: older participants maintained lower MDRs (Fig. 8). However, there were no clear differences to TT and SIM, in which none of the demographics appeared to have a clear influence on the model (Fig. 8, Table B.4).

4. Discussion

When the time gap to the oncoming vehicle was short, drivers chose more accelerative maneuvers in all environments. This fact confirms one of the main results from previous research conducted in both artificial (Bianchi Piccinini et al., 2018) and naturalistic environments (Kovaceva et al., 2019). Thus, all three environments can be regarded as (at least relatively) valid in this aspect.

Strategy choice generally differed most between SIM and the other environments. In the presence of oncoming traffic, participants in SIM opted for the accelerative strategy more than twice as often as participants in the other environments. This finding may be related to the fact that the SIM screens had a lower resolution and refresh rate than the HMD in DVIL (see Table 1), which may have affected the participants' depth perception and ability to estimate the TTC to the oncoming vehicle. Indeed, in the post-drive questionnaire, the SIM participants indicated that they had difficulty correctly estimating the distance to the oncoming vehicle Fig. C.1. As a result, they might not have been able to assess the oncoming vehicle's distance with confidence, reverting to the safer

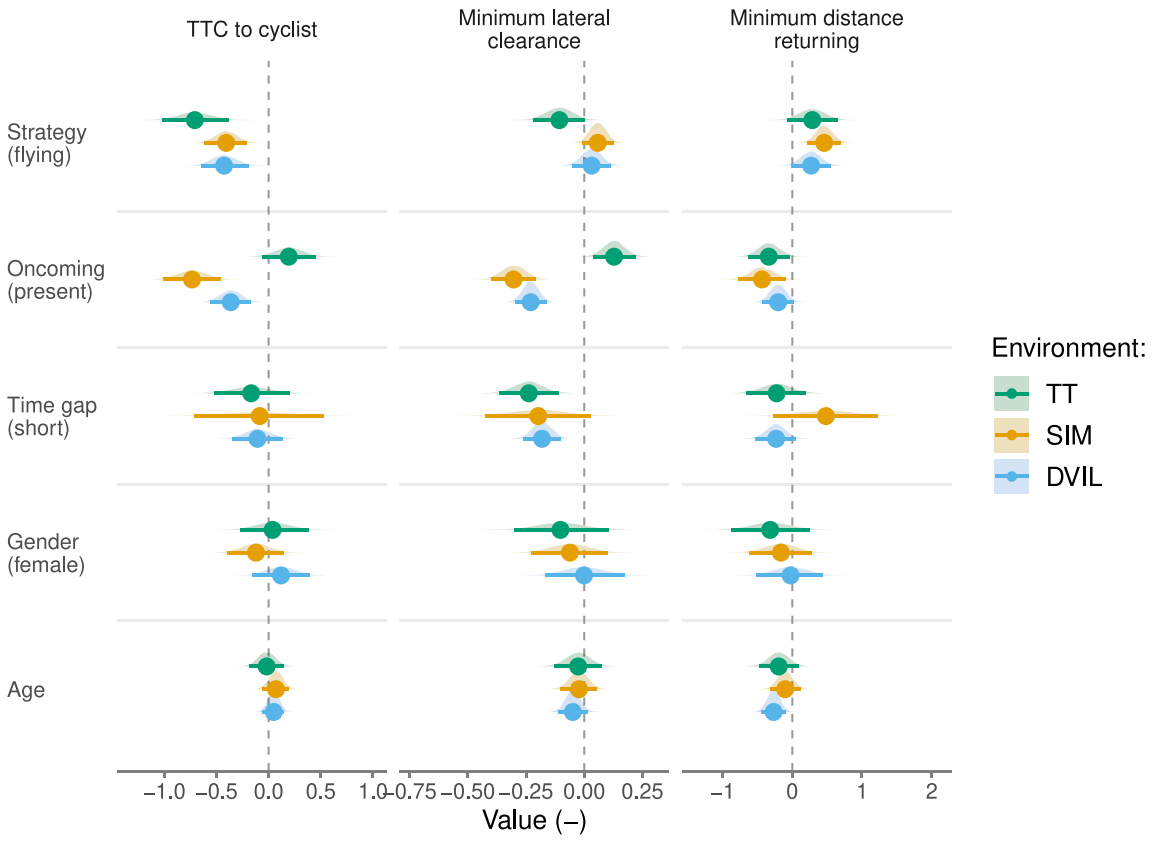

Fig. 8. Comfort-zone metrics model; each parameter distribution is displayed as density and summarized as a median (dot) and 95% highest-density interval (horizontal bar). The intercepts, not displayed due to their larger values compared to the slopes, can be found in Table A.2.

strategy as a fallback. Another possible explanation for the difference between SIM and TT may be the difference in road layouts: In TT, the artificial road on the airfield was visually not as clearly delimited from the surrounding environment compared to the simulated environments with trees lined up beside the road. This might have increased drivers' perceived lateral freedom. However, we did not measure participants' perception of the space available, nor did we record any attempts to use the adjacent space for evasive maneuvers. Therefore, this explanation should be regarded as a possible interpretation and not as a demonstrated behavioral mechanism.

In all environments, participants exhibited the same risk-compensatory behavior: in flying maneuvers when the oncoming vehicle was present, drivers generally reduced their risk of a head-on collision with the oncoming vehicle by decreasing the safety margins to the cyclist, increasing the risk of a side-swipe collision with the cyclist. The same behavior has been observed in previous field studies when car drivers overtake cyclists (Dozza et al., 2016) as well as trucks (Sandin et al., 2012). Further, with the shorter time gap to the oncoming vehicle, the safety margin to the cyclist was smaller; this finding has also been reported previously in simulator (Bianchi Piccinini et al., 2018) and naturalistic studies (Kovaceva et al., 2019; Rasch et al., 2023).

Without oncoming traffic, participants in TT—in contrast to the other environments—maintained lower safety margins in the approaching and passing phases. There can be several explanations for this. One is that the criticality is perceived as higher in simulated environments than on a pure test track with real visual cues (Purucker et al., 2018). Another related explanation is that the drivers in TT might minimize safety margins to the robot cyclist because of its artificial appearance. A third reason could be that in real vehicles, drivers regulate time margins differently than in simulated environments, reflecting a more accurate perception of vehicle dynamics and a stronger coupling between control inputs and motion feedback (Boda et al., 2018). However, the driving behavior in TT appears to be more conservative than in real-world driving, since absolute values for TTC and MLC were generally smaller in naturalistic data (Kovaceva et al., 2019; Rasch et al., 2023).

Our results may provide a reference for traffic-safety researchers deciding what type of environment is most appropriate for their data collection, depending on their individual needs and goals. A driving simulator, providing artificial visual and motion information to drivers, may be the most economical setup among those we tested. Furthermore, it is independent of the weather; however, it may compromise depth and longitudinal speed estimation, making decision-making more difficult for drivers and thereby resulting in more cautious behavior than expected or planned. A test-track setup with real visual and motion information has the most realistic visual and motion cues for drivers and is state-of-the-art for assessing the safety of ADAS. On the other hand, not all kinematics

scenarios are possible to set up because of the limits of the other road users (usually represented by robots) and the dimensions of the test track. Additionally, the risk of critical events may not be ethically (or financially) acceptable. A hybrid setup between a simulator and test track, like our DVIL setup, with artificial visual and real motion information, may allow for more realistic interactions with other road users, as they can easily be simulated, and collisions may be acceptable. The real motion feedback allows for better speed estimation and may support drivers' decision-making. A DVIL system, however, requires a technically more complicated setup to link visual and motion information without latency accurately. Furthermore, as a test environment, it is usually still in the trial phase.

Comparing test environments in a fair manner is extremely difficult, if not impossible. Ideally, the same participants would perform identical experiments in each environment on the same day, with the order randomized also across test environments. Furthermore, the sample should be large and appropriately balanced, potentially across multiple demographic factors beyond gender. Even under these ideal conditions, there may be inherent differences in how test environments (e.g. road layout) influence perception (e.g., visual cues), and behavior (e.g., vehicle control) cannot be fully controlled. While the challenge of a fair comparison across test environments is not insurmountable, it can only be addressed through a sufficient number of studies that collectively build evidence, gradually strengthening observed associations toward causal interpretations. Below, we outline some of the unavoidable limitations associated with the type of comparison conducted in this study.

The fact that the oncoming vehicle had to travel at a lower speed in TT (40 km/h) than in SIM and DVIL (70 km/h) might have introduced a confounding effect. Following the same practice as in previous car-to-cyclist and car-to-car overtaking studies, TTC was controlled as the main factor consistently across environments (Bianchi Piccinini et al., 2018; Hassein et al., 2017; Polus et al., 2000). TTC represents the time gap to the oncoming traffic and is an established surrogate safety measure with an intrinsic link to optical looming, which is a major factor in driver gap acceptance. However, we cannot rule out that other related factors, such as angular size and angular expansion rate of the oncoming vehicle as well as its metric distance and closing speed—which had to differ for TT to match TTC with the other environments—might have had an effect that confounded the differences we found between TT and the other environments.

In our experiment, participants were not the same across our data-collection environments for organizational reasons; for instance, the driving simulator and the test track were in two different cities, and traveling between them would have been impractical for the participants. Furthermore, on the test track, only Autoliv and Magna employees were allowed to drive, making the TT and DVIL populations more similar to each other than to the SIM population, which might have affected the results. Nevertheless, the number of participants was similar to those of previous studies, and participants' demographics were comparable among environments. Therefore, we have no strong reason to believe that this limitation substantially impacted our results. On the contrary, the between-subject assessments in this paper may help demonstrate the generalizability of the results. Of course, a larger number of participants would have helped mitigate all these limitations. It is also worth noting that involving the same subjects in all experiments also poses limitations not only for organizational but also for analytical reasons. For example, in this case, the environment would be a within-subject factor that might require further investigation to determine whether and how the order of environments impacts behavior.

Driving simulators are all different from each other—and the hybrid environment in this study was custom-made (as no commercial product exists today for rendering virtual worlds on test tracks). As a consequence, it is up for debate the extent to which the results from this comparison apply to other driving simulators or hybrid test environments.

Motion sickness is often reported as a crucial issue in virtual worlds, especially when the subject is fully immersed (Denk et al., 2023), and there may be a mismatch between visual and kinematic cues. In our experiment, motion sickness was not a critical issue (most participants rated the occurrence of nausea as non-existent or very low; Fig. C.1), possibly because of the specific scenario we tested, with limited lateral movement. In experiments that include more vigorous steering and larger head movements (e.g., negotiating intersections), motion sickness may negatively impact data collection. Our experiments indicated a slightly higher chance of motion sickness in the hybrid environment than in the driving simulator, which we did not expect. While motion sickness in the driving simulator may be due to the limited kinematic cues, in the hybrid environment, a lag between the vehicle controls and the rendering of the virtual environment may be the culprit. However, as technology evolves, this lag will become smaller, possibly preventing motion sickness. Finally, it is important to note that sensory perception, perceived safety, and vehicle operation vary across environments and may significantly influence the results.

Future work may compare the results from the artificial environments presented in this study with results from naturalistic data. Obtaining a naturalistic dataset with comparable conditions (e.g., road layout, time gaps, cyclists' lateral positions) will be challenging. It might be possible, however, with site-based observations on a similar stretch of road (Rasch et al., 2023). Our study focused on behavioral differences at a higher, tactical level of driving. Future work should also investigate the lower, operational levels of driver behavior (braking and steering) to determine how they may differ between environments. However, to do so, the same vehicle with the same controls (pedals, steering system) may be necessary in all environments.

We used an approach speed of only 70 km/h, which is a common speed limit on rural roads in Europe. Future studies may test different, possibly lower, speeds that are more common in urban areas, to assess the extent to which speed perception and optical flow influence driving behavior. Future studies should also investigate the effect of artificial-reality-based HMDs that project the real world through cameras on the outside of the display, overlaid with artificial road users (Blissing et al., 2019).

Finally, our rule-based methodology for determining the overtaking phases was a trade-off between relying on similar previous work, mainly the test-track experiment from Rasch et al. (2020a), and developing a methodology that worked reasonably robust for all participants in all environments. We verified this qualitatively by visually inspecting the trajectories and phase division. Future work should assess whether this simple methodology could be improved with alternative, e.g. data-driven methods, and how sensitive the results are to the chosen method.

5. Conclusion

This study compared driver behavior when overtaking cyclists in three different test environments: a driving simulator, a test track, and a hybrid environment where the participants drove a real vehicle on a test track while only seeing a virtual environment. Overall, the results verified that the directions of the main factors affecting overtaking strategy and performance were similar across all environments. Because overtaking maneuvers depend on depth estimation and accurate timing, the visual resolution and refresh rate of the virtual environment may be more important than the environment itself in explaining the results. Driving scenarios other than overtaking may not be equally affected by the (same) technical specifications of the individual virtual environments. It was surprising that the safety margins in the test track experiment were the smallest across all environments. It is possible that virtual visual cues induce safer behavior, at least when drivers overtake cyclists. While our primary objective was to compare across test environments, we are inevitably also comparing different populations of participants. Moreover, additional confounding factors—such as variations in perceived safety across environments—introduce limitations and suggest that the results should be interpreted as indicating possible associations rather than direct causality.

CRediT authorship contribution statement

Alexander Rasch: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Prateek Thalya:** Writing – review & editing, Writing – original draft, Resources, Investigation, Data curation, Conceptualization; **Jesper Sandin:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization; **Marco Dozza:** Writing – review & editing, Writing – original draft, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgments

The authors would like to thank Magna for carrying out the two test-track experiments (the TT environment in collaboration with Autoliv and the DVIL in collaboration with Volvo Cars). We further thank VTI for its support in conducting the driving-simulator experiment. We thank Carol Flannagan for valuable discussions about the statistical model and Kristina Mayberry for language revisions. This work was supported by the project Modeling Interaction between Cyclists and Automobiles (MICA2), funded jointly by Vinnova, the Swedish Energy Agency, the Swedish Transport Administration, and the Swedish vehicle industry through the Strategic Vehicle Research and Innovation (FFI) Program under Grant 2019–03082. The test-track data were collected in collaboration with the project Drivers in Interaction with Vulnerable Road Users (DIV), funded by Toyota Motor Europe, Autoliv, and Magna.

Data availability

Data and analysis code are available at https://git.chalmers.se/m2/vehicle-safety/mica2/environments. Due to data protection rules, we cannot make the demographics data available, but documented code for the analyses in this article is available and is illustrated with simulated demographics data.

Appendix A. Model parameters

Table A.1
Fitted parameters of the overtaking-strategy model related to time gap to oncoming (gap), gender, age, and intercept (0), for each test environment: driving simulator (SIM), test track (TT), and driver-vehicle-in-the-loop (DVIL). The posterior distributions are summarized with median and lower and upper 95% highest-density interval (HDI), and probability of direction (PD).

Parameter	Median	l-95%	u-95%	PD
β_0^{DVIL}	1.51	0.73	2.28	1.00
β_{age}^{DVIL}	-0.42	-0.97	0.14	0.94
β_{gap}^{DVIL}	-1.26	-2.19	-0.33	1.00
β_{gender}^{DVIL}	-0.30	-1.57	0.76	0.70
β_0^{SIM}	-0.39	-1.20	0.40	0.82
β_{age}^{SIM}	0.37	-0.32	1.01	0.87
β_{gap}^{SIM}	-1.67	-2.78	-0.51	1.00
β_{gender}^{SIM}	0.23	-0.94	1.35	0.65
β_0^{TT}	1.56	0.58	2.57	1.00
β_{age}^{TT}	0.20	-0.63	0.98	0.69
β_{gap}^{TT}	-1.34	-2.50	-0.14	0.98
β_{gender}^{TT}	-0.61	-1.91	0.58	0.83

Table A.2
Fitted parameters of the comfort-metrics model related to strategy (St), oncoming presence (OP), time gap to oncoming (gap), age, gender and intercept (0), for each test environment: driving simulator (SIM), test track (TT), and driver-vehicle-in-the-loop (DVIL). The posterior distributions are summarized with median and lower and upper 95% highest-density interval (HDI), and probability of direction (PD).

	TTC				MLC				MDR			
Param	Median	l-95%	u-95%	PD	Median	l-95%	u-95%	PD	Median	l-95%	u-95%	PD
β_0^{DVIL}	1.90	1.69	2.10	1.00	1.02	0.91	1.11	1.00	1.89	1.61	2.21	1.00
β_{age}^{DVIL}	0.05	-0.06	0.15	0.82	-0.05	-0.11	0.01	0.94	-0.27	-0.45	-0.10	1.00
β_{St}^{DVIL}	-0.43	-0.65	-0.21	1.00	0.03	-0.05	0.11	0.77	0.27	-0.02	0.55	0.97
β_{OP}^{DVIL}	-0.36	-0.56	-0.17	1.00	-0.23	-0.30	-0.16	1.00	-0.20	-0.43	0.02	0.96
β_{gap}^{DVIL}	-0.11	-0.34	0.14	0.80	-0.18	-0.26	-0.10	1.00	-0.23	-0.54	0.03	0.95
β_0^{SIM}	2.11	1.92	2.29	1.00	1.04	0.94	1.15	1.00	2.48	2.20	2.78	1.00
β_{age}^{SIM}	0.07	-0.06	0.19	0.86	-0.02	-0.10	0.06	0.72	-0.10	-0.33	0.10	0.83
β_{St}^{SIM}	-0.41	-0.62	-0.21	1.00	0.06	-0.01	0.13	0.95	0.46	0.21	0.70	1.00
β_{OP}^{SIM}	-0.73	-1.02	-0.48	1.00	-0.30	-0.41	-0.22	1.00	-0.44	-0.77	-0.09	0.99
β_{gap}^{SIM}	-0.09	-0.74	0.49	0.61	-0.20	-0.43	0.03	0.96	0.48	-0.24	1.25	0.89
β_0^{TT}	1.66	1.35	1.98	1.00	0.88	0.74	1.02	1.00	1.73	1.31	2.16	1.00
β_{age}^{TT}	-0.02	-0.19	0.14	0.58	-0.03	-0.13	0.07	0.70	-0.20	-0.49	0.08	0.91
β_{St}^{TT}	-0.71	-1.03	-0.40	1.00	-0.11	-0.22	0.00	0.97	0.29	-0.06	0.66	0.94
β_{OP}^{TT}	0.19	-0.06	0.45	0.93	0.13	0.04	0.22	1.00	-0.34	-0.64	-0.05	0.99
β_{gap}^{TT}	-0.17	-0.51	0.22	0.81	-0.24	-0.37	-0.11	1.00	-0.23	-0.64	0.21	0.86
σ_{ID}	0.22	0.13	0.31	1.00	0.17	0.14	0.21	1.00	0.46	0.37	0.58	1.00
σ	0.39	0.34	0.43	1.00	0.13	0.12	0.15	1.00	0.45	0.40	0.50	1.00

Appendix B. Contrasts

Table B.3
Contrasts between environments for strategy model (TT = test track, SIM = driving simulator, DVIL = driver-vehicle-in-the-loop). The distributions represent the expected values predicted from the model, summarized by median and lower and upper 95% highest-density interval (HDI), and probability of direction (PD). The contrasts for the intercept are calculated from the estimated marginal means, while the contrasts for the other parameters are calculated from the estimated marginal slopes/trends.

	TT - SIM				TT - DVIL				SIM - DVIL			
Variable	Median	l-95%	u-95%	PD	Median	l-95%	u-95%	PD	Median	l-95%	u-95%	PD
Overall	0.37	0.19	0.54	1.00	-0.01	-0.19	0.15	0.56	-0.38	-0.52	-0.23	1.00
Time gap (short vs. long)	-0.04	-0.29	0.22	0.63	-0.02	-0.32	0.27	0.57	0.01	-0.27	0.30	0.54
Gender (female vs. male)	0.10	-0.19	0.40	0.75	0.16	-0.14	0.49	0.85	0.06	-0.28	0.38	0.63
Age (continuous)	0.14	-0.01	0.27	0.97	0.03	-0.14	0.22	0.61	-0.11	-0.29	0.06	0.90

Table B.4
Contrasts between environments for the comfort-zone model (TT = test track, SIM = driving simulator, DVIL = driver-vehicle-in-the-loop, TTC = time-to-collision, MLC = minimum lateral clearance, MDR = minimum distance returning). The distributions represent the expected values predicted from the model, summarized by median and lower and upper 95% highest-density interval (HDI), and probability of direction (PD). The contrasts for the intercept are calculated from the estimated marginal means, while the contrasts for the other variables are calculated from the estimated marginal slopes/trends.

	TT - SIM				TT - DVIL				SIM - DVIL			
Variable	Median	l-95%	u-95%	PD	Median	l-95%	u-95%	PD	Median	l-95%	u-95%	PD
TTC												
Overall	-1.58	-2.58	-0.66	1.00	-1.17	-2.05	-0.21	0.99	0.42	-0.48	1.36	0.81
Strategy (flying vs. accelerative)	-1.23	-3.25	0.83	0.89	-2.18	-4.23	0.00	0.98	-0.95	-3.20	1.20	0.80
Oncoming presence (present vs. absent)	-1.46	-2.90	-0.15	0.98	-3.45	-4.88	-2.10	1.00	-2.01	-3.09	-1.00	1.00
Time gap (short vs. long)	0.13	-1.59	2.00	0.56	0.30	-1.61	2.33	0.62	0.20	-1.26	1.46	0.61
Gender (female vs. male)	-1.37	-3.46	0.93	0.89	-0.91	-3.06	1.13	0.81	0.42	-1.35	2.40	0.68
Age (continuous)	0.23	-0.77	1.20	0.68	0.52	-0.53	1.54	0.83	0.29	-0.50	1.08	0.78
MLC												
Overall	-0.35	-0.66	-0.06	0.99	-0.34	-0.62	-0.07	0.99	0.01	-0.26	0.29	0.53
Strategy (flying vs. accelerative)	-0.04	-0.34	0.29	0.60	-0.27	-0.61	0.07	0.94	-0.23	-0.57	0.11	0.90
Oncoming presence (present vs. absent)	-0.18	-0.47	0.12	0.89	-1.05	-1.34	-0.71	1.00	-0.86	-1.12	-0.62	1.00
Time gap (short vs. long)	-0.01	-0.45	0.46	0.53	0.12	-0.38	0.60	0.69	0.13	-0.19	0.43	0.81
Gender (female vs. male)	-0.16	-0.80	0.42	0.71	0.05	-0.55	0.67	0.57	0.22	-0.39	0.84	0.76
Age (continuous)	0.06	-0.21	0.32	0.68	0.00	-0.30	0.31	0.51	-0.06	-0.35	0.20	0.68
MDR												
Overall	-10.16	-14.85	-5.80	1.00	-1.86	-4.22	0.81	0.93	8.28	4.18	13.17	1.00
Strategy (flying vs. accelerative)	5.13	-0.88	12.46	0.97	5.52	-0.14	13.12	0.98	0.44	-2.71	3.61	0.61
Oncoming presence (present vs. absent)	-1.53	-9.09	6.77	0.66	-1.48	-8.75	6.82	0.66	0.05	-2.63	2.93	0.52
Time gap (short vs. long)	9.75	-3.49	29.26	0.95	9.04	-4.48	28.04	0.93	-0.63	-3.52	2.29	0.67
Gender (female vs. male)	-2.36	-10.18	5.55	0.73	-0.72	-8.30	6.67	0.58	1.60	-3.21	7.17	0.74
Age (continuous)	0.73	-3.09	4.76	0.65	-0.34	-4.51	3.54	0.57	-1.09	-3.71	1.40	0.81

Appendix C. Post-drive survey results

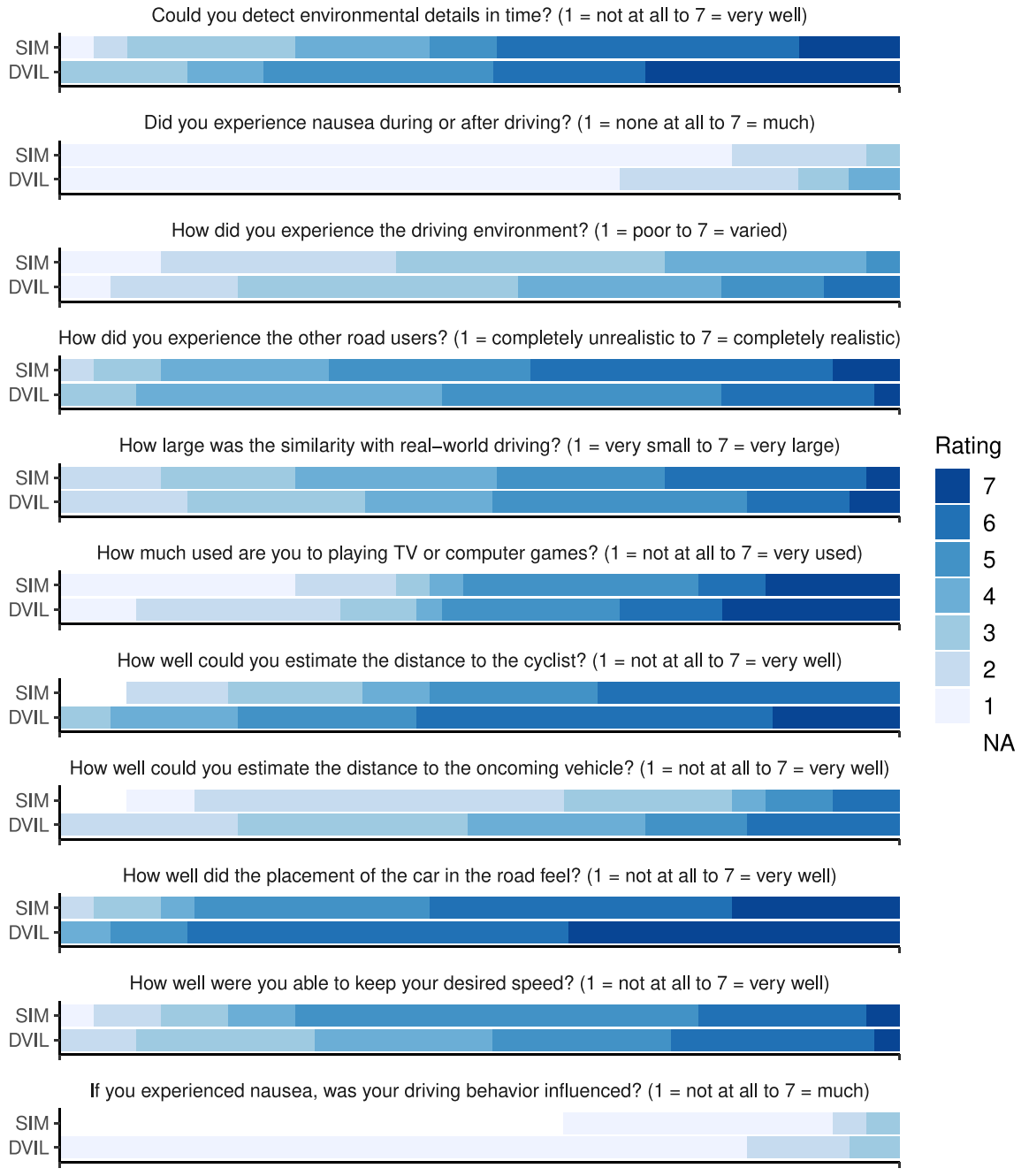

Fig. C.1. Post-drive questionnaire results for participants in driving simulator (SIM) and driver-vehicle-in-the-loop (DVIL). For the test-track (TT) experiment, these data were not collected.

References

AbuAli, N., & Abou-zeid, H. (2016). Driver behavior modeling: Developments and future directions. *International Journal of Vehicular Technology*, *2016*, 1–12. https://doi.org/10.1155/2016/6952791

Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginaleffects in r and python. *Journal of Statistical Software*, . https://marginaleffects.com.

Bärgman, J. (2016). Methods for Analysis of Naturalistic Driving Data in Driver Behavior Research. Ph.D. thesis. Chalmers University of Technology. http://publications.lib.chalmers.se/records/fulltext/244575/244575.pdf.

Bianchi Piccinini, G. F., Moretto, C., Zhou, H., & Itoh, M. (2018). Influence of oncoming traffic on drivers' overtaking of cyclists. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59*, 378–388. https://doi.org/10.1016/j.trf.2018.09.009

Bjelkeflo, L., Laxing, R., Djup, M., Gustafsson, P., Lundin, N., Johansson, V., Naterjee, V., Pohl, J., Schiller, E., & Svensson, B. J. (2018). CHRONOS Part 1. Technical Report FFI - Vinnova. https://www.vinnova.se/contentassets/3f1ea323948d43ce90732c74dc10fd19/2016-02573.pdf.

Blissing, B., Bruzelius, F., & Eriksson, O. (2019). Driver behavior in mixed and virtual reality – a comparative study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *61*, 229–237. https://doi.org/10.1016/j.trf.2017.08.005

Bock, J., Krajewski, R., Moers, T., Runde, S., Vater, L., & Eckstein, L. (2020). The inD dataset: A drone dataset of naturalistic road user trajectories at german intersections. In *2020 IEEE Intelligent vehicles symposium (IV)* Iv (pp. 1929–1934). IEEE. https://doi.org/10.1109/IV47402.2020.9304839

Boda, C.-N., Dozza, M., Bohman, K., Thalya, P., Larsson, A., & Lubbe, N. (2018). Modelling how drivers respond to a bicyclist crossing their path at an intersection: How do test track and driving simulator compare? *Accident Analysis & Prevention*, *111*, 238–250. https://doi.org/10.1016/j.aap.2017.11.032

Bürkner, P.-C. (2017). Brms : An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Denk, F., Himmels, C., Andreev, V., Lindner, J., Syed, A. A., Riener, A., Huber, W., & Kates, R. (2023). Studying interactions of motorists and vulnerable road users: Empirical comparison of test track and simulator experiments. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, (September), 992–999. https://doi.org/10.1109/ITSC57777.2023.10421865

Dozza, M., Schindler, R., Bianchi-Piccinini, G., & Karlsson, J. (2016). How do drivers overtake cyclists? *Accident Analysis & Prevention*, *88*, 29–36. https://doi.org/10.1016/j.aap.2015.12.008

Fisher, D. L., Rizzo, M., Caird, J. K., & Lee, J. D., et al. (2011). Handbook of Driving Simulation for Engineering, Medicine, and Psychology. CRC Press. https://doi.org/10.1201/b10836

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-Squared for Bayesian regression models. *The American Statistician*, *73*(3), 307–309. https://doi.org/10.1080/00031305.2018.1549100

Hassein, U., Diachuk, M., & Easa, S. (2017). Evaluation of dynamic passing gap acceptance on two-lane highways using field data. *Canadian Journal of Civil Engineering*, *44*(11), 871–880. https://doi.org/10.1139/cjce-2016-0572

Himmels, C., Venrooij, J., Parduzi, A., Peller, M., & Riener, A. (2024). The bigger the better? investigating the effects of driving simulator fidelity on driving behavior and perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, *101*(January), 250–266. https://doi.org/10.1016/j.trf.2024.01.007

Jansson, J., Sandin, J., Augusto, B., Fischer, M., Blissing, B., & Källgren, L. (2014). Design and performance of the VTI sim IV. In *Driving simulation conference* (pp. 128–138).

Joshi, A., Adugyamfi, K., Merickel, J., Gunaratne, P., & Sharma, A., et al. (2025). Driver Age and Its Effect on Key Driving Metrics: Insights from Dynamic Vehicle Data. https://doi.org/10.48550/arXiv.2501.06918

Kircher, K., & Sandin, J. (2025). Long combination vehicles overtaking cyclists: A field study in regular traffic. *Traffic Safety Research*, *9*, e000105. https://doi.org/10.55329/xvqs2323

Klüver, M., Herrigel, C., Heinrich, C., Schöner, H. P., & Hecht, H. (2016). The behavioral validity of dual-task driving performance in fixed and moving base driving simulators. *Transportation Research Part F: Traffic Psychology and Behaviour*, *37*, 78–96. https://doi.org/10.1016/j.trf.2015.12.005

Kovaceva, J., Bärgman, J., & Dozza, M. (2020). A comparison of computational driver models using naturalistic and test-track data from cyclist-overtaking manoeuvres. *Transportation Research Part F: Traffic Psychology and Behaviour*, *75*, 87–105. https://doi.org/10.1016/j.trf.2020.09.020

Kovaceva, J., Nero, G., Bärgman, J., & Dozza, M. (2019). Drivers overtaking cyclists in the real-world: evidence from a naturalistic driving study. *Safety Science*, *119*, 199–206. https://doi.org/10.1016/j.ssci.2018.08.022

Krajewski, R., Bock, J., Kloeker, L., & Eckstein, L. (2018). The highD dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In *2018 21st international conference on intelligent transportation systems (ITSC)* (pp. 2118–2125). IEEE (*vol. 2018-Novem*). https://doi.org/10.1109/ITSC.2018.8569552

Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour*, *5*(10), 1282–1291. https://doi.org/10.1038/s41562-021-01177-7

Kullgren, A., Stigson, H., Rizzi, M., & Ydenius, A. (2019). The potential of vehicle and road infrastructure interventions in fatal bicyclist accidents on swedish roads—what can in-depth studies tell us? *Traffic Injury Prevention*, *20*(sup1), S7–S12. https://doi.org/10.1080/15389588.2019.1610171

Llorca, C., Angel-Domenech, A., Agustin-Gomez, F., & García, A. (2017). Motor vehicles overtaking cyclists on two-lane rural roads: Analysis on speed and lateral clearance. *Safety Science*, *92*, 302–311. https://doi.org/10.1016/j.ssci.2015.11.005

López, G., Pérez-Zuriaga, A. M., Moll, S., & García, A. (2020). Analysis of overtaking maneuvers to cycling groups on two-lane rural roads using objective and subjective risk. *Transportation Research Record*, *2674*(7), 343–355. https://doi.org/10.1177/0361198120921169

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). Bayestestr: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Michaels, J., Chaumillon, R., Nguyen-Tri, D., Watanabe, D., Hirsch, P., Bellavance, F., Giraudet, G., Bernardin, D., & Faubert, J., et al. (2017). Driving simulator scenarios and measures to faithfully evaluate risky driving behavior: A comparative study of different driver age groups. *PloS One*, *12*(10), e0185909. https://doi.org/10.1371/journal.pone.0185909

Moll, S., López, G., Llopis-Castelló, D., Dols, J. F., & García, A. (2023). Drivers' behaviour when overtaking cyclists on rural roads: Driving simulator validation using naturalistic data. *Transportation Research Part F: Traffic Psychology and Behaviour*, *95*(October 2022), 391–404. https://doi.org/10.1016/j.trf.2023.05.011

Organization, W. H. (2023). Global status report on road safety. Technical Report World Health Organization Geneva.

Polus, A., Livneh, M., & Frischer, J. (2000). Evaluation of the passing process on two-lane rural highways. *Transportation Research Record*, *1701*(1), 53–60.

Purucker, C., Schneider, N., & Rüger, F. (2018). UR:BAN human factors in traffic. *UR:BAN Human Factors in Traffic*, (pp. 423–446). https://doi.org/10.1007/978-3-658-15418-9

Ramezani-Khansari, E., Tabibi, M., Moghadas Nejad, F., & Mesbah, M. (2021). Comparing the effect of age, gender, and desired speed on car-following behavior by using driving simulator. *Journal of Advanced Transportation*, *2021*(1), 9922321. https://doi.org/10.1155/2021/9922321

Rasch, A., Boda, C.-N., Thalya, P., Aderum, T., Knauss, A., & Dozza, M. (2020a). How do oncoming traffic and cyclist lane position influence cyclist overtaking by drivers? *Accident Analysis & Prevention*, *142*, 105569. https://doi.org/10.1016/j.aap.2020.105569

Rasch, A., Panero, G., Boda, C.-N., & Dozza, M. (2020b). How do drivers overtake pedestrians? evidence from field test and naturalistic driving data. *Accident Analysis & Prevention*, *139*, 105494. https://doi.org/10.1016/j.aap.2020.105494

Rasch, A., Tarakanov, Y., Tellwe, G., & Dozza, M. (2023). Drivers passing cyclists: How does sight distance affect safety? results from a naturalistic study. *Journal of Safety Research*, *87*, 76–85. https://doi.org/10.1016/j.jsr.2023.09.006

Sandin, J., Renner, L., & Andersson, J. (2012). The effect on passenger cars' meeting margins when overtaking 30 m trucks on real roads. In *Advances in human factors and ergonomics 2012- 14 vol set : Proceedings of the 4th AHFE conference 21–25 july 2012* Advances in Human Factors and Ergonomics Series.