# How do kinematic and visual cues influence cyclist overtaking? A comparison of test environments

Alexander Rasch[1*], Prateek Thalya[2], Jesper Sandin[3] & Marco Dozza[1]

[1]Chalmers University of Technology, Gothenburg, Sweden

[2]Magna Electronics Sweden AB, Vårgårda, Sweden

[3]Swedish National Road and Transport Research Institute (VTI), Gothenburg, Sweden

Artificial test environments have become 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 (SIM: N = 25), a test track (TT: N = 18), and a hybrid “driver-vehicle-in-the-loop” environment (DVIL: N = 33) in which the participants drove a real vehicle on a test track while wearing a virtual-reality headset, so they could only see the virtual environment. Overall, the results verified that the direction of the main factors affecting overtaking strategy and performance was similar across all environments. In presence of a close oncoming vehicle, drivers in all environments preferred the more cautious accelerative (waiting behind the cyclist) over flying (passing the cyclist while facing the oncoming vehicle) overtaking strategy, and this effect was strongest in SIM. Metrics related to safety margins to the cyclist followed a similar trend in all environments in the presence of oncoming traffic; however, the TT environment had the smallest overall safety 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.

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

## 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 (World Health 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., 2020).

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

*Send correspondence to: Alexander Rasch, alexander.rasch@chalmers.se.

This preprint has been submitted to a journal.

not taken into account properly (Denk et al., 2023). Furthermore, 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, Himmels, Andreev, Lindner, Syed, Riener, Huber, & Kates, 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, Himmels, Andreev, Lindner, Syed, Riener, Huber, & Kates, 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. 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, and those 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. 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 experiment 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, Schneider, & Rüger (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. 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, Dozza, Bohman, Thalya, Larsson, & Lubbe (2018); Kovaceva et al. (2020); and Denk, Himmels, Andreev, Lindner, Syed, Riener, Huber, & Kates (2023) are among the few examples of such comparisons. Boda, Dozza, Bohman, Thalya, Larsson, & Lubbe (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, Bärgman, & Dozza (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 the models fitted on naturalistic data showed superior performance. Denk, Himmels, Andreev, Lindner, Syed, Riener, Huber, & Kates (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 in both cases produced a substantial number of dropouts due to motion sickness.

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. 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, using the scenario of cyclist overtaking.

## 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 (Figure 1).

### Experimental protocol

In all experiments, participants mainly drove on a straight road (3.75 m wide) resembling a rural road with a speed limit of 70 km/h, and encountered a simulated cyclist riding at a constant 20 km/h in the same lane, 1 m from the lane edge). Each participant completed three trials, one for each of the three different configurations (in random order): 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. In all environments, participants were instructed to drive as they would in real traffic.

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.

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

### Environments

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

#### *Test track*

The TT data were obtained in an experiment at Vårgårda airfield, Vårgårda, Sweden, in 2018 (Figure 1, Panel a). A robot cyclist, based on a Euro NCAP-validated Guided Soft Target, and an oncoming balloon vehicle served as the other road users. 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 (its physical maximum). In the no oncoming condition, the vehicle was on the test track and 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 (former 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. (2020).

#### *Driving simulator*

The SIM data were collected in VTI's advanced moving-base driving simulator Sim-IV in Gothenburg, Sweden, in 2021 (Figure 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.

#### *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 (Figure 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.

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 |

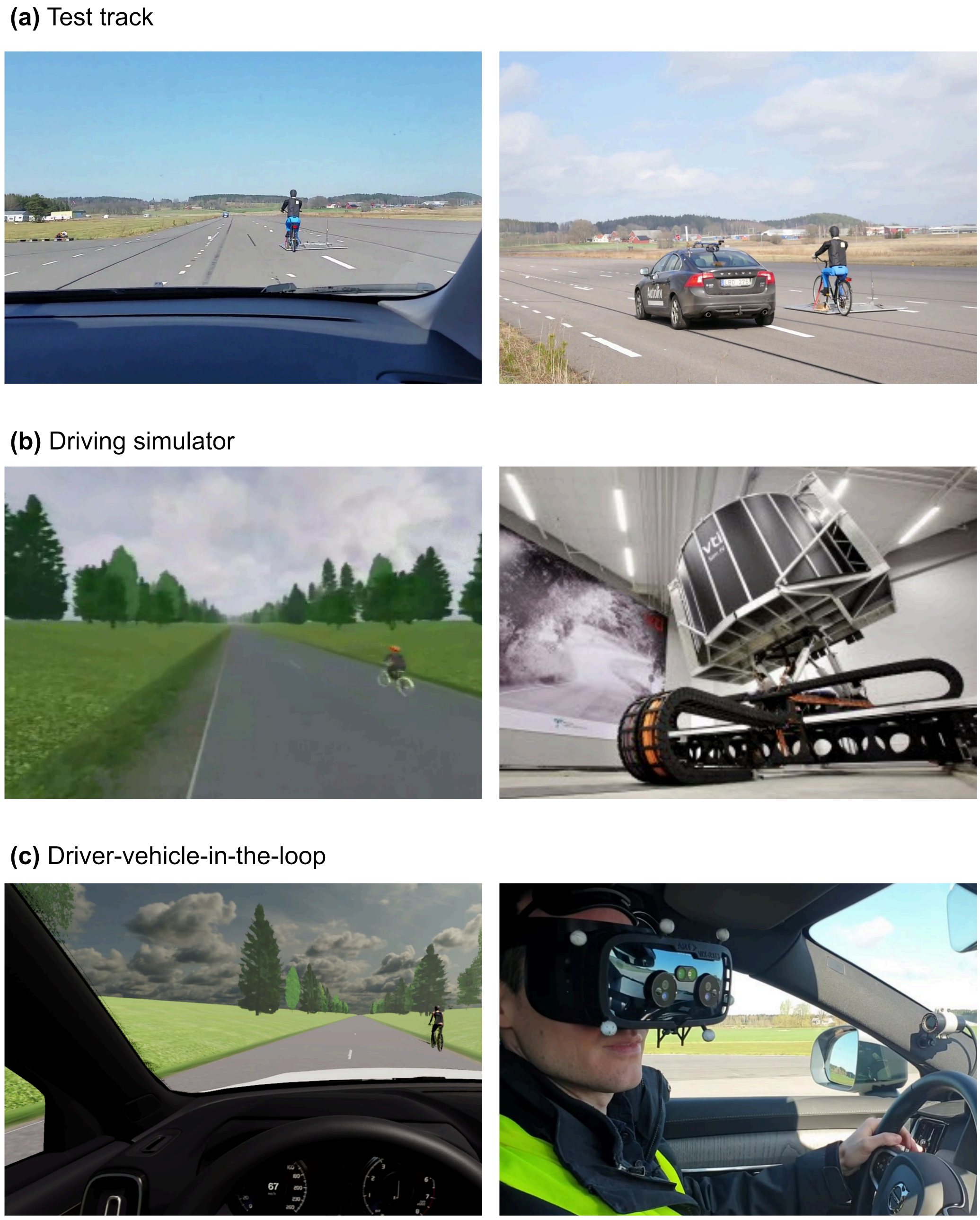


Figure 1: Data-collection environments; Pure test track (a; left panel shows the cockpit view, right panel the robot cyclist), driving simulator (b; left panel shows cockpit view, right panel the simulator used, Sim-IV), and the driver-vehicle-in-the-loop setup used on a test track (c; left panel shows the cockpit view, right panel the driver wearing a virtual-reality headset).

### 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).

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

## 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, Schindler, Bianchi-Piccinini, & Karlsson, 2016; Kovaceva, Nero, Bärgman, & Dozza, 2019; Rasch, Boda, Thalya, Aderum, Knauss, & Dozza, 2020). Overall, drivers' maneuvering was divided into two distinct strategies: *flying* maneuvers, in which drivers overtook the cyclist before an 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. Figure 2 depicts the phases and metrics for a flying overtaking maneuver in the presence of an oncoming vehicle.

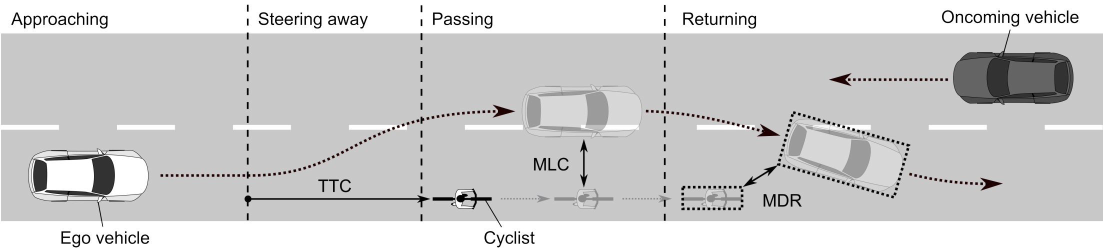


Figure 2: Cyclist-overtaking scenario (flying strategy in the presence of an oncoming vehicle). Phases: approaching, steering away, passing, and returning. Metrics: TTC (time-to-collision of the ego vehicle to the cyclist at the moment of steering onset), MLC (minimum lateral clearance), and MDR (minimum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

## 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 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 0.2 m. 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 0.2 m of the maximum lateral clearance during the maneuver, and the end of the passing phase when the lateral clearance decreased again by 0.2 m from its maximum.

## Overtaking metrics

We also defined a set of metrics to quantify driver behavior in the different phases of the overtaking maneuver (Figure 2). These metrics characterize the drivers' comfort zone as they circumvent the cyclist, in accordance with previous work (Dozza, Schindler, Bianchi-Piccinini, & Karlsson, 2016; Kovaceva, Nero, Bärgman, & Dozza, 2019; Rasch, Boda, Thalya, Aderum, Knauss, & Dozza, 2020). In the approaching phase, we calculated the time-to-collision (TTC) of the 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 returning: MDR). Table 3 summarizes the metrics and their explanations.

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 |

## Statistical models

To quantify the differences (including uncertainties) among the environments, we employed Bayesian regression models. Following the work by Klüver, Herrigel, Heinrich, Schöner, & Hecht (2016), the models were set up to include the data from all environments, 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 uses Markov-Chain Monte Carlo 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).

### *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:

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

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 $\beta_0$), 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 $\mathbf{env}$ 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 $\beta$ 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.

***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, Boda, Thalya, Aderum, Knauss, & Dozza, 2020) 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 any oncoming vehicle has already passed). The time gap to the oncoming vehicle (gap) was coded 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 $\beta$ 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.

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

***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), the probability that the parameter's effect is non-zero (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 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 (Equation 3, Equation 4); for strategy the predictor was the probability of a flying maneuver, while for all other metrics the value of the (linear) predictor ($\mu$) 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 (Equation 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 (Equation 4). As for the qualitative analyses, we identified clear contrasts from the PD of the contrast distribution (when greater than 95% in our case).

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

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

## Results

### Overtaking strategy

Participants in SIM performed more accelerative maneuvers than those in TT and DVIL (Figure 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 7).

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 (Figure 3). The model confirmed the clear reducing influence of time gap on the probability of a flying strategy for all three environments (Figure 4), without clear differences between them (Table 7).

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 7).

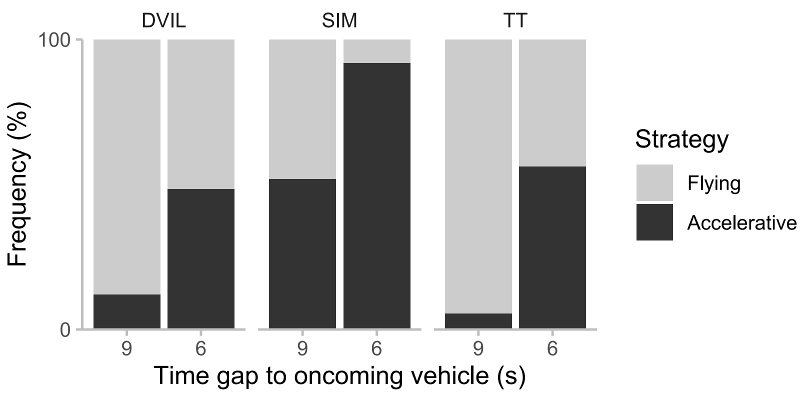


Figure 3: Strategy distribution (frequency of flying and accelerative strategies) for the three environments TT (test track), SIM (driving simulator), and DVIL (driver-vehicle-in-the-loop), grouped by time gap to the oncoming vehicle.

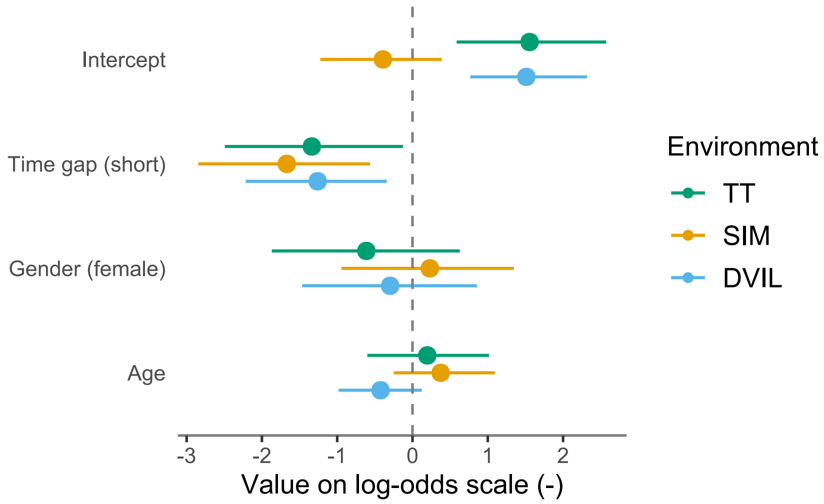


Figure 4: Strategy model; each parameter distribution is summarized as a median (dot) and bar (95% highest-density interval). The x-axis shows the parameter value on the log-odds scale; negative values indicate a higher probability of performing an accelerative maneuver while positive values indicate a higher probability of performing a flying maneuver. The numerical values of all parameters can be found in Table 5.

## Overtaking metrics

The lateral clearance profiles suggest that participants' lateral positioning differed during the approaching phase in the different environments (Figure 5). Participants in SIM kept the closest lateral clearance to the cyclist until about 25 m behind the cyclist in accelerative strategies and about 100 m 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 (Figure 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 10 m ahead of the cyclist), they kept smallest clearances to the cyclist, followed by DVIL and then SIM.

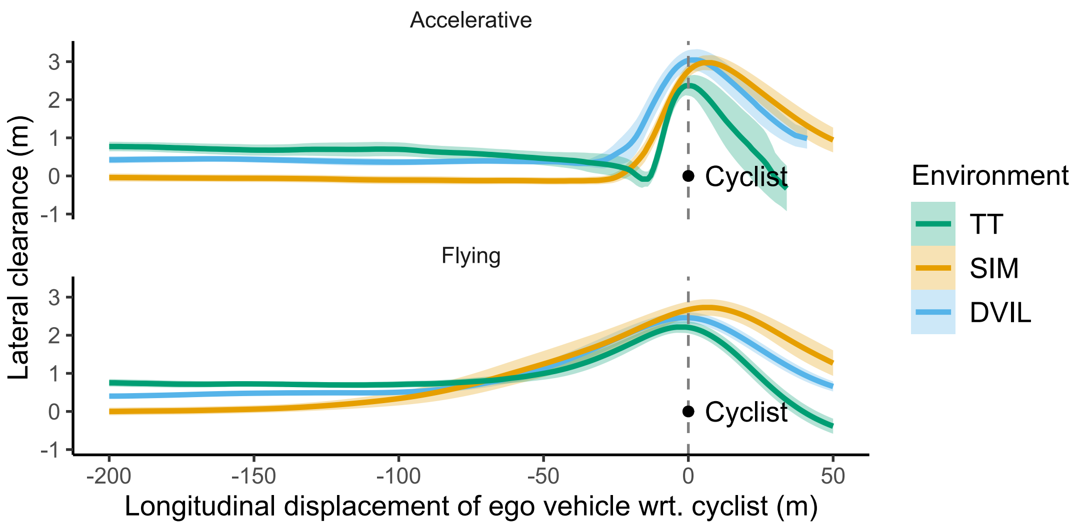


Figure 5: Profiles of lateral clearances to the cyclist in different environments under different strategies. The thick lines show the means and the shaded areas the 95% confidence intervals.

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 (Figure 6).

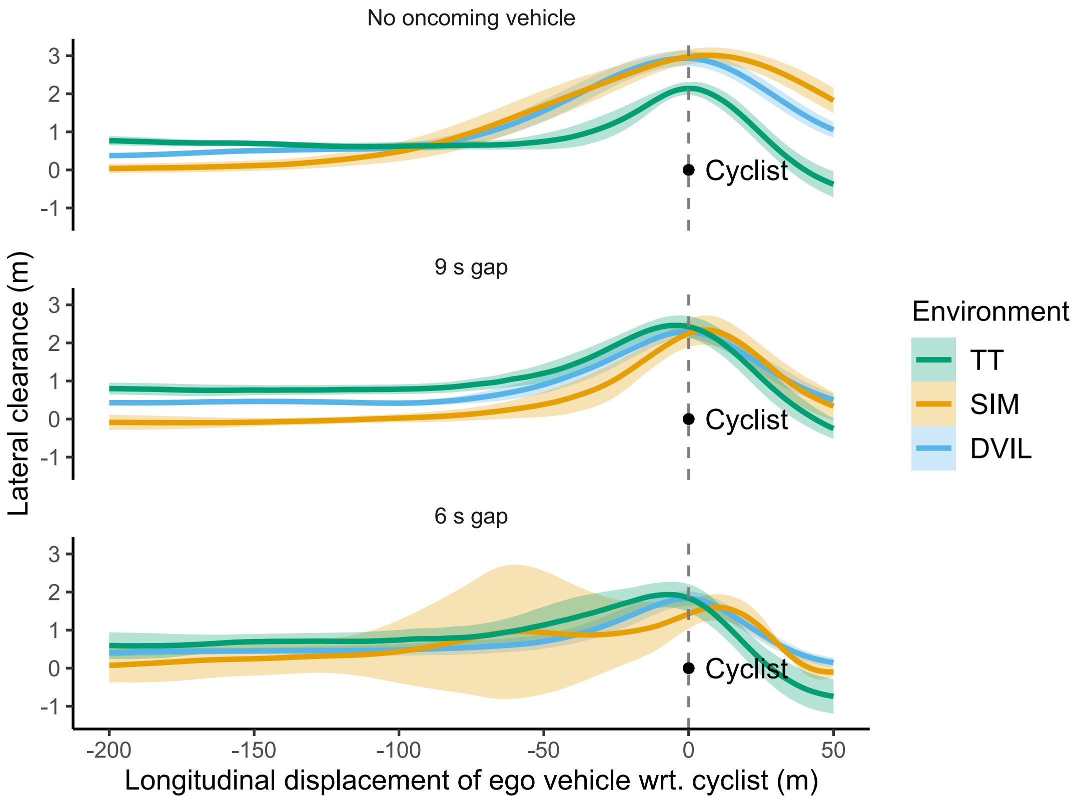


Figure 6: Profiles of lateral clearance to the cyclist in the different environments for flying strategies, grouped by the different oncoming vehicle configurations (no oncoming, 9 s gap, and 6 s gap). The thick lines show the mean and the shaded areas the 95% confidence intervals.

A deeper look into the individual comfort metrics revealed that most of them were distributed reasonably similarly between the environments (Figure 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)).

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 SIM. The effect of strategy on TTC was −1.23 ([−3.25, 0.83] 95% HDI) s greater in TT than in SIM, and the effect of strategy on MLC was −0.04 ([−0.34, 0.29] 95% HDI) m greater in TT than in SIM. On the other hand, no notable differences between TT and DVIL could be identified for the effect of strategy on TTC or MLC (Table 8). For MDR, participants in SIM differed from participants in the other environments, maintaining greater safety margins to the cyclist in flying maneuvers (5.13 ([−0.88, 12.46] 95% HDI) m, SIM-DVIL: 0.44 ([−2.71, 3.61] 95% HDI) m; Table 8).

The presence and timing of the oncoming vehicle influenced participants in all three environments, particularly for MLC (Figure 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 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 8). 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 8).

With respect to demographics, only age had an effect. It appeared to influence MDR in DVIL: older participants maintained lower MDRs (Figure 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 (Figure 8, Table 8).

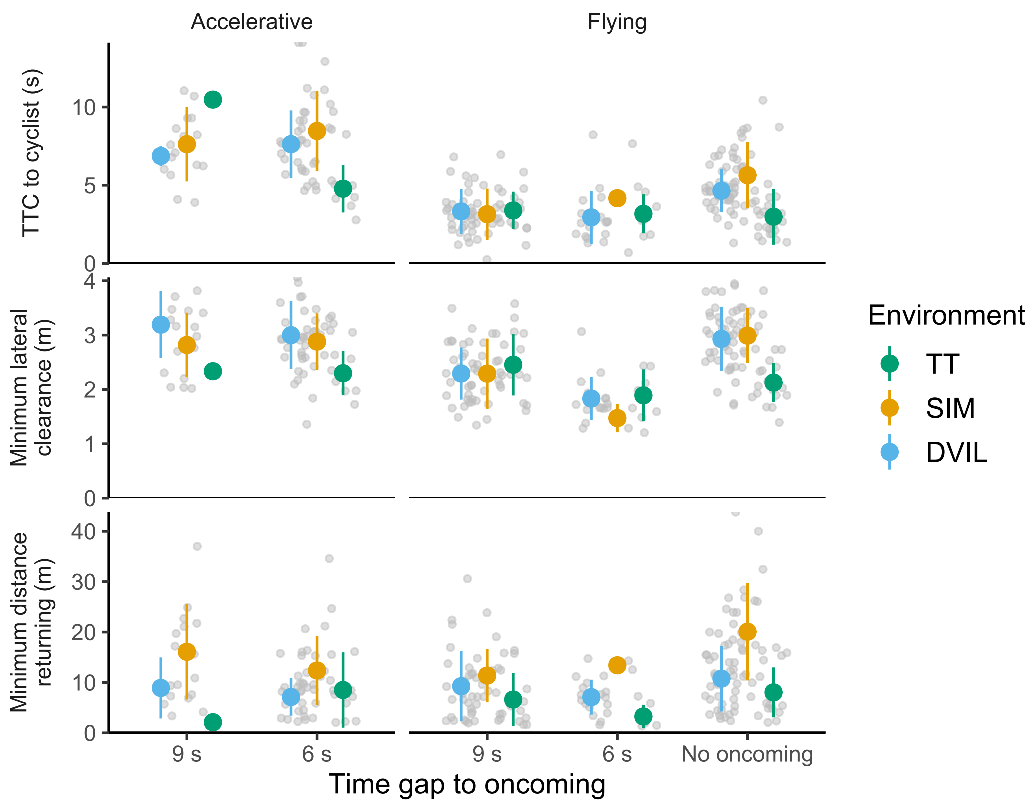


Figure 7: Comfort-zone metrics data for each environment, displayed as individual observations (grey dots), mean (colored dot), and +/- one standard deviation (vertical bars).

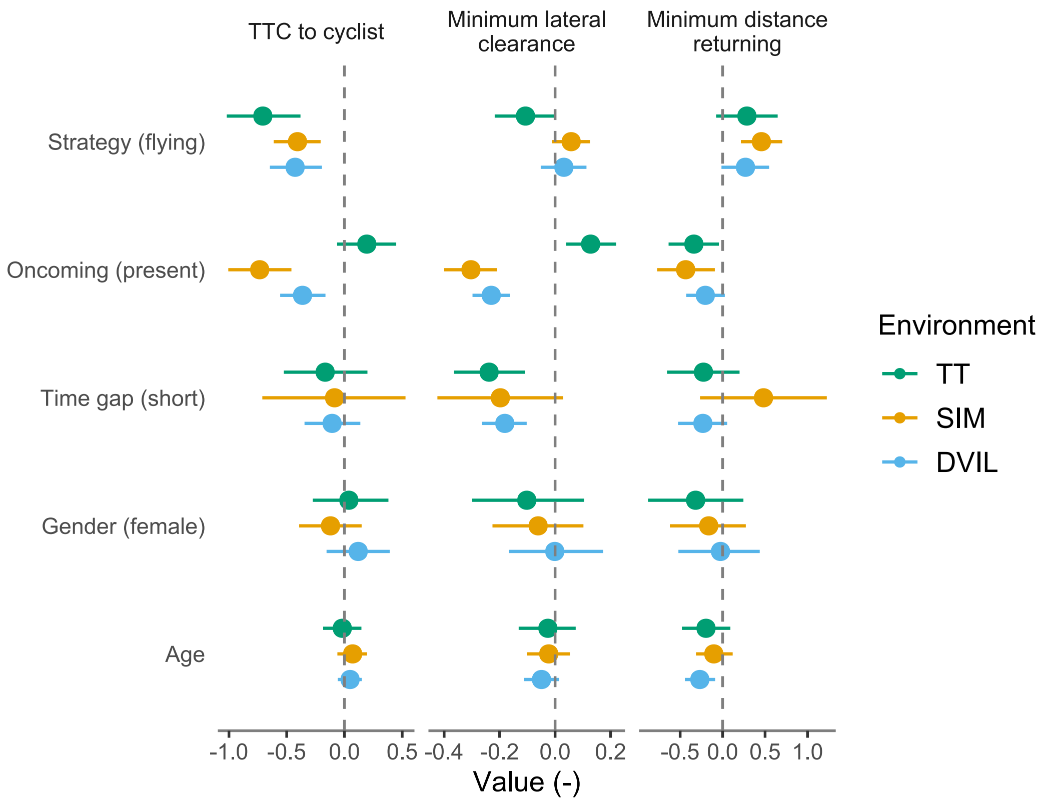


Figure 8: Comfort-zone metrics model; each parameter distribution is summarized as a median (dot) and bar (95% highest-density interval). The intercepts, not displayed due to their larger values, can be found in Table 6.

# Discussion

## How does the data-collection environment affect driver behavior?

### *Overtaking strategy*

When the time gap to the oncoming vehicle was short instead of long, 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, Nero, Bärgman, & Dozza, 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 than the HMD in DVIL (see Table 1), which may have affected the participants' depth perception. Further, in the post-drive questionnaire the SIM participants indicated that they had difficulty correctly estimating the distance to the oncoming vehicle (Figure 9). As a result, they might not have been able to assess the oncoming vehicle's distance with confidence, reverting to the safer strategy as a fallback.

### *Overtaking metrics*

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 cyclists were overtaken by car drivers Dozza, Schindler, Bianchi-Piccinini, & Karlsson (2016) or truck drivers

(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, Moretto, Zhou, & Itoh, 2018) and naturalistic studies (Rasch, Tarakanov, Tellwe, & Dozza, 2023).

Without oncoming traffic, participants in TT—in contrast to the other environments—maintained lower safety margins in the approaching and passing phases, suggesting that with virtual visuals induce safer behaviors. This result is in line with the findings of (Purucker, Schneider, & Rüger, 2018), that perceived criticality was higher in simulated environments than on a pure test track with real visual cues. An alternative explanation is that drivers in the virtual environments who chose a flying maneuver in the presence of the oncoming vehicle knew that the vehicle was virtual and feared the risk of a head-on collision less and, as a result, left larger safety margins to the cyclist. Further, in TT, another reason that the drivers might minimize safety margins to the robot cyclist is its artificial appearance. 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, Nero, Bärgman, & Dozza, 2019; Rasch, Tarakanov, Tellwe, & Dozza, 2023).

### Practical implications: which environment to use for what?

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.

Table 4 condenses the results and lessons learned from comparing three test environments in this paper. It should be noted that the generalizability of this figure may be limited, as there can be significant differences within the same kind of environment (i.e., other test-tracks, simulators, and DVIL setups may be different from the ones we utilized). The reader should also be cautioned that the minimalist bullet lists report only the two main pros and cons for each environment and may have sacrificed some details for the sake of conciseness.

Table 4: Comparison of the three data-collection environments.

| Environment | Visual information | Motion information | Pros | Cons |
|---|---|---|---|---|
| Driving simulator | Artificial | Artificial | Most economical.<br>Not dependent on the weather. | Depth and longitudinal speed estimation may be challenging.<br>Lowest ecological validity. |
| Driver-vehicle-in-the-loop | Artificial | Real | Virtual collisions may be acceptable.<br>Interactions with other road users do not require expensive robots that may have limited kinematics. | Most complicated and expensive setup.<br>Still in the trial phase as a test environment. |
| Test track | Real | Real | Most realistic visual and motion cues.<br>State-of-the-art for the safety assessment of ADAS. | Not all kinematics are possible because of the limits of the other road users (robots).<br>The risk of critical events may not be ethically (or financially) acceptable. |

### Limitations and future work

Participants were not the same across our data-collection environments for organizational reasons; for instance, the driver simulator and the test track were in two different cities, and traveling in between would have been impractical for the participants. Nevertheless, the number of participants was similar to those of previous studies and participants' demographics were comparable among environments. Therefore, we have no reason to believe that this limitation substantially impacted our results. On the contrary, the between-subject assessments in this paper may demonstrate the generalizability of the results. 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, Himmels, Andreev, Lindner, Syed, Riener, Huber, & Kates, 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; Figure 9 ), 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 eliminating the insurgence of motion sickness.

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, Tarakanov, Tellwe, & Dozza, 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, Bruzelius, & Eriksson, 2019).

## 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, the visual resolution 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 a safer behavior, at least when drivers overtake cyclists.

## 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.

## Declarations

### Funding

This work was supported by the project Modelling 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.

### Conflicts of interest/Competing interests

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

### Ethics approval

Approval for the pure test-track experiment was obtained from the local ethical board in Gothenburg, Sweden (under registration number 600–17). The driving simulator and driver-vehicle-in-the-loop experiments were approved by the national Swedish ethical review authority (registration number 2020-05419).

### Consent to participate

All participants provided written informed consent to participate in the experiments.

### Consent for publication

All participants provided written consent for the publication of anonymized data.

### Availability of data and materials

Materials and analysis code are available at https://osf.io/kpa5x/.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.

### Code availability

Analysis code is available at https://osf.io/kpa5x/.

# Appendix

## Model parameters

Table 5: 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 |
|---|---|---|---|---|
| $\beta_0^{\mathrm{DVIL}}$ | 1.51 | 0.73 | 2.28 | 1 |
| $\beta_{\mathrm{age}}^{\mathrm{DVIL}}$ | −0.42 | −0.97 | 0.14 | 0.94 |
| $\beta_{\mathrm{gap}}^{\mathrm{DVIL}}$ | −1.26 | −2.19 | −0.33 | 1 |
| $\beta_{\mathrm{gender}}^{\mathrm{DVIL}}$ | −0.3 | −1.57 | 0.76 | 0.7 |
| $\beta_0^{\mathrm{SIM}}$ | −0.39 | −1.2 | 0.4 | 0.82 |
| $\beta_{\mathrm{age}}^{\mathrm{SIM}}$ | 0.37 | −0.32 | 1.01 | 0.87 |
| $\beta_{\mathrm{gap}}^{\mathrm{SIM}}$ | −1.67 | −2.78 | −0.51 | 1 |
| $\beta_{\mathrm{gender}}^{\mathrm{SIM}}$ | 0.23 | −0.94 | 1.35 | 0.65 |
| $\beta_0^{\mathrm{TT}}$ | 1.56 | 0.58 | 2.57 | 1 |
| $\beta_{\mathrm{age}}^{\mathrm{TT}}$ | 0.2 | −0.63 | 0.98 | 0.69 |
| $\beta_{\mathrm{gap}}^{\mathrm{TT}}$ | −1.34 | −2.5 | −0.14 | 0.98 |
| $\beta_{\mathrm{gender}}^{\mathrm{TT}}$ | −0.61 | −1.91 | 0.58 | 0.83 |

Table 6: 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 |
| $\beta_0^{\mathrm{DVIL}}$ | 1.9 | 1.69 | 2.1 | 1 | 1.02 | 0.91 | 1.11 | 1 | 1.89 | 1.61 | 2.21 | 1 |
| $\beta_{\mathrm{age}}^{\mathrm{DVIL}}$ | 0.05 | −0.06 | 0.15 | 0.82 | −0.05 | −0.11 | 0.01 | 0.94 | −0.27 | −0.45 | −0.1 | 1 |
| $\beta_{\mathrm{St}}^{\mathrm{DVIL}}$ | −0.43 | −0.65 | −0.21 | 1 | 0.03 | −0.05 | 0.11 | 0.77 | 0.27 | −0.02 | 0.55 | 0.97 |
| $\beta_{\mathrm{OP}}^{\mathrm{DVIL}}$ | −0.36 | −0.56 | −0.17 | 1 | −0.23 | −0.3 | −0.16 | 1 | −0.2 | −0.43 | 0.02 | 0.96 |
| $\beta_{\mathrm{gap}}^{\mathrm{DVIL}}$ | −0.11 | −0.34 | 0.14 | 0.8 | −0.18 | −0.26 | −0.1 | 1 | −0.23 | −0.54 | 0.03 | 0.95 |
| $\beta_0^{\mathrm{SIM}}$ | 2.11 | 1.92 | 2.29 | 1 | 1.04 | 0.94 | 1.15 | 1 | 2.48 | 2.2 | 2.78 | 1 |
| $\beta_{\mathrm{age}}^{\mathrm{SIM}}$ | 0.07 | −0.06 | 0.19 | 0.86 | −0.02 | −0.1 | 0.06 | 0.72 | −0.1 | −0.33 | 0.1 | 0.83 |
| $\beta_{\mathrm{St}}^{\mathrm{SIM}}$ | −0.41 | −0.62 | −0.21 | 1 | 0.06 | −0.01 | 0.13 | 0.95 | 0.46 | 0.21 | 0.7 | 1 |
| $\beta_{\mathrm{OP}}^{\mathrm{SIM}}$ | −0.73 | −1.02 | −0.48 | 1 | −0.3 | −0.41 | −0.22 | 1 | −0.44 | −0.77 | −0.09 | 0.99 |
| $\beta_{\mathrm{gap}}^{\mathrm{SIM}}$ | −0.09 | −0.74 | 0.49 | 0.61 | −0.2 | −0.43 | 0.03 | 0.96 | 0.48 | −0.24 | 1.25 | 0.89 |
| $\beta_0^{\mathrm{TT}}$ | 1.66 | 1.35 | 1.98 | 1 | 0.88 | 0.74 | 1.02 | 1 | 1.73 | 1.31 | 2.16 | 1 |
| $\beta_{\mathrm{age}}^{\mathrm{TT}}$ | −0.02 | −0.19 | 0.14 | 0.58 | −0.03 | −0.13 | 0.07 | 0.7 | −0.2 | −0.49 | 0.08 | 0.91 |
| $\beta_{\mathrm{St}}^{\mathrm{TT}}$ | −0.71 | −1.03 | −0.4 | 1 | −0.11 | −0.22 | 0 | 0.97 | 0.29 | −0.06 | 0.66 | 0.94 |
| $\beta_{\mathrm{OP}}^{\mathrm{TT}}$ | 0.19 | −0.06 | 0.45 | 0.93 | 0.13 | 0.04 | 0.22 | 1 | −0.34 | −0.64 | −0.05 | 0.99 |
| $\beta_{\mathrm{gap}}^{\mathrm{TT}}$ | −0.17 | −0.51 | 0.22 | 0.81 | −0.24 | −0.37 | −0.11 | 1 | −0.23 | −0.64 | 0.21 | 0.86 |
| $\sigma_{\mathrm{ID}}$ | 0.22 | 0.13 | 0.31 | 1 | 0.17 | 0.14 | 0.21 | 1 | 0.46 | 0.37 | 0.58 | 1 |
| $\sigma$ | 0.39 | 0.34 | 0.43 | 1 | 0.13 | 0.12 | 0.15 | 1 | 0.45 | 0.4 | 0.5 | 1 |

## Contrasts

Table 7: 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 | −0.01 | −0.19 | 0.15 | 0.56 | −0.38 | −0.52 | −0.23 | 1 |
| 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 | 0.3 | 0.54 |
| Gender (female vs. male) | 0.1 | −0.19 | 0.4 | 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 | 0.9 |

Table 8: Contrasts between environments for 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 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metric | 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 | −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 | 0.98 | −0.95 | −3.2 | 1.2 | 0.8 |
| | Oncoming presence (present vs. absent) | −1.46 | −2.9 | −0.15 | 0.98 | −3.45 | −4.88 | −2.1 | 1 | −2.01 | −3.09 | −1 | 1 |
| | Time gap (short vs. long) | 0.13 | −1.59 | 2 | 0.56 | 0.3 | −1.61 | 2.33 | 0.62 | 0.2 | −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 | 2.4 | 0.68 |
| | Age (continuous) | 0.23 | −0.77 | 1.2 | 0.68 | 0.52 | −0.53 | 1.54 | 0.83 | 0.29 | −0.5 | 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 | 0.6 | −0.27 | −0.61 | 0.07 | 0.94 | −0.23 | −0.57 | 0.11 | 0.9 |
| | Oncoming presence (present vs. absent) | −0.18 | −0.47 | 0.12 | 0.89 | −1.05 | −1.34 | −0.71 | 1 | −0.86 | −1.12 | −0.62 | 1 |
| | Time gap (short vs. long) | −0.01 | −0.45 | 0.46 | 0.53 | 0.12 | −0.38 | 0.6 | 0.69 | 0.13 | −0.19 | 0.43 | 0.81 |
| | Gender (female vs. male) | −0.16 | −0.8 | 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 | −0.3 | 0.31 | 0.51 | −0.06 | −0.35 | 0.2 | 0.68 |
| MDR | Overall | −10.16 | −14.85 | −5.8 | 1 | −1.86 | −4.22 | 0.81 | 0.93 | 8.28 | 4.18 | 13.17 | 1 |
| | 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 | −8.3 | 6.67 | 0.58 | 1.6 | −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 | 1.4 | 0.81 |

## Post-drive survey results

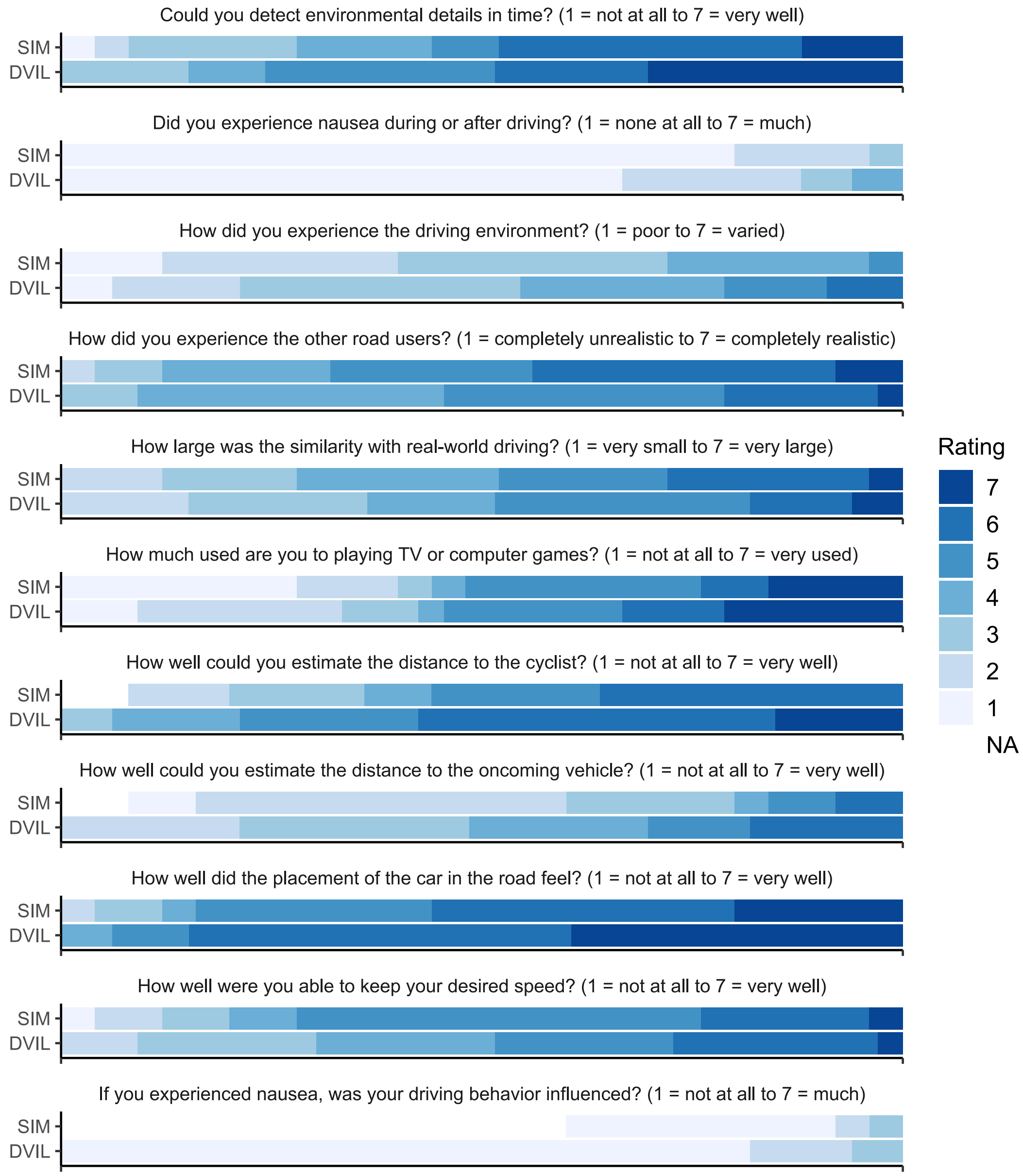


Figure 9: Post-drive questionnaire results for participants in driving simulator (SIM) and driver-vehicle-in-the-loop (DVIL). For the test-track (TT) experiment, these data were not collected.

## 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 \texttt{marginaleffects} in \texttt{R} and \texttt{Python}. *Journal of Statistical Software*. https://marginaleffects.com/

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*. 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. *2020 IEEE Intelligent Vehicles Symposium (IV)*, *Iv*, 1929–1934. 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

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

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

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

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. *Driving Simulation Conference*, 128–138.

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/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. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2118–2125. https://doi.org/10.1109/ITSC.2018.8569552

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), 1541–1542. https://doi.org/10.21105/joss.01541

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*(October2022), 391–404. https://doi.org/10.1016/j.trf.2023.05.011

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

Rasch, A., Boda, C.-N., Thalya, P., Aderum, T., Knauss, A., & Dozza, M. (2020). How do oncoming traffic and cyclist lane position influence cyclist overtaking by drivers?. *Accident Analysis & Prevention*, *142*, 105569–105570. https://doi.org/https://doi.org/10.1016/j.aap.2020.105569

Rasch, A., Panero, G., Boda, C.-N., & Dozza, M. (2020). How do drivers overtake pedestrians? Evidence from field test and naturalistic driving data. *Accident Analysis & Prevention*, *139*. 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/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 meter trucks on real roads. *Advances in Human Factors and Ergonomics 2012- 14 Volume Set : Proceedings of the 4th AHFE Conference 21-25 July 2012*.

World Health Organization. (2023). *Global status report on road safety*.