Noname manuscript No.
(will be inserted by the editor)

Test-track and On-road studies: Methodological Insights on the Assessment of Carsickness and its Modulating Factors

Rebecca Pham Xuan, Adrian Brietzke, Hanna Koerber,
Myriam Metzulat, Aaron Edelmann, Riender Happee,
Georgios Papaioannou*

Received: date / Accepted: date

Abstract The research on motion sickness and the work on effective countermeasures has gained additional attention due to the ongoing development of highly automated driving vehicles. In order to understand the phenomenon of motion sickness better and support the testing, this work sheds lights on motion sickness and multiple factors connected to it: The relation between the development of carsickness and testing environment itself is explored as well as their connection to psychological factor. Simultaneously, carsickness is often stimulated by non-driving-related tasks (NDRTs), while publications have shown different effects on motion sickness by various NDRTs. Therefore, it is also investigated whether the optical flow influences carsickness. Lastly, it will also be investigated whether there is an effect on performance caused by carsickness across different testing environments. 47 participants participated in an within-study in a vehicle either driving on a public road or on a test track. While administering comparable accelerations and two randomly assigned NDRTs, carsickness, psychological factors and performance was measured within these environments. The results indicate that it was possible to set-up rides in both environments which led to comparable motion sickness stimuli based on the vehicle accelerations. The optical flow did not lead to different carsickness. At the same time, few differences could be found among psychological factors, although they do not indicate that the transfer of results between testing environments would not be valid. Lastly, cognitive performance was affected by the two testing environments with regard to the reaction time. This work elucidates on the one hand that multiple factors influence motion sickness, though much remains, yet, to be understood. On the other hand, the feasibility of conducting comparable studies in different testing environments is shown. If the transfer of such is fully realized, motion sickness research can proceed in an effective and intertwined way.

Keywords Motion Sickness, MISC, Optical Flow, Psychological Factors, Performance, Testing Environment

---

Address(es) of author(s) should be given

Highlights

- Average and individual carsickness were comparable across both environments.
- Psychological factors were comparable across environments except trust, which was reduced on test-track
- Carsickness correlated similarly with pleasure and perceived safety across both environments, but not with acceptance and trust.
- Video optical flow did not significantly impact carsickness.
- Carsickness effect on cognitive performance was comparable across environments.

1 Introduction

Automated vehicles (AV) are set to be revolutionary, shifting how the average vehicle is utilised and offering great benefits to the society (Bansal et al., 2016; Fagnant and Kockelman, 2015). The possibility to comfortably engage in non-driving related tasks (NDRT) is expected to be at the forefront for their wide acceptance. Yet, major challenges remain to address carsickness in AVs. Extensive research with human experiments is being carried out in different settings to understand the occurrence of carsickness and eventually develop countermeasures to mitigate it in the context of AVs (Emond et al., 2024). However, standardization efforts for carsickness research are scarce (Bos et al., 2022), making it difficult to compare results between studies. This study therefore focuses on deriving methodological considerations about different settings (testing environment, psychological factors, engagement in NDRTs and others) and advancing the comparability between carsickness studies.

Testing environments for carsickness need to balance two requirements: replicability and realism. The motion stimulus shall be maintained as constant as possible across participants and conditions, especially when comparing different countermeasures or observing passenger behavior across multiple trips. At the same time, the findings should be applicable to real-world driving situations. Common testing environments for carsickness research are driving simulators, test-tracks, or public roads, which achieve different compromises among these requirements. However, simulators even employed with advanced motion cueing algorithms (Khusro et al., 2020), have been shown to provoke lower MS levels compared to on-road driving (Dam et al., 2024; Mühlbacher et al., 2020; Talsma et al., 2023). This indicates that in-vehicle tests, either on test tracks or public roads, more realistically assess the occurrence of carsickness. Despite securing the most realistic study environment, on-road testing raises the risk of accidents and requires costly resources, especially when testing AV prototypes. To achieve replicability in manual car rides, training drivers is important to ensure consistent driving behavior (Bengler et al., 2019). Even with trained drivers or fully automated vehicles, driving behavior in public road studies is not consistently reproducible due to unexpected dynamic events (vulnerable road users, other vehicles, traffic lights etc.). These elements cause a lack of replicability (internal validity), while, interestingly, they are indispensable components of reality, and therefore increase the external validity. In contrast, studies on test tracks lack the external validity of varying driving behavior while allowing high replicability. This makes large test tracks a feasible solution for testing (Jones et al., 2019; Brietzke et al., 2021). However, not all researchers have access to test tracks, they are relatively cost-intensive and the complexity of on-road driving is missing. In order to bridge the gap between studies on test tracks and public roads, this study aims to explore the effects of the testing environment on carsickness occurrence.

Besides the effect of the testing environment on the occurrence of carsickness, the participants' psychological state is critical. Comfort while being driven by AVs is a highly complex concept (Peng et al. (2024a)) and is affected by environmental, physical (Papaioannou et al. (2025)) and psychological factors (privacy, trust (Elbanhawi et al. (2015); Paddeu et al. (2020)), perceived safety (He et al. (2022)), naturalness (Peng et al. (2024b)), engagement in NDRTs (Metzulat et al. (2024)) and situation awareness, arousal, personality and others). The impact of these factors is well-studied with regards to comfort.

However, taking part in an experiment that involves automated driving could be associated with some kind of excitement due to the novelty of the technology and the high expectations that go along with it. To our knowledge, no study has yet investigated the effect of psychological factors on the occurrence of carsickness as well as the effect these factors have depending on the testing environment in on-road experiments. The current study puts a specific emphasis on the latter.

Non-driving related tasks are also of special importance when conducting carsickness studies. Many researchers use different non-driving related tasks to help induce higher levels of carsickness in experimental studies. Common naturalistic tasks include reading (e.g., Karjanto et al. (2021); Kremer et al. (2022); Tomzig et al. (2023)), video-watching (e.g., Brietzke et al. (2021); Mühlbacher et al. (2020); Tajdari et al. (2025)), gaming (Talsma et al. (2023)) and quizzes (Jain et al., 2023). Although more seldom, standardized tasks, i.e. static visual or auditory tasks have been implemented, too (Kuiper et al. (2018), Metzulat et al. (2024)). Research findings focusing on the relationship between carsickness and NDRTs underline the different influences on motion sickness between the various tasks (Isu et al., 2014; Morimoto et al., 2008; Metzulat et al., 2024). Nevertheless, to make a further step towards clarifying the importance of different elements in NDRTs on motion sickness, the current study concentrates on the effect of visual tasks with varying optical flow.

Carsickness might not only compromise AVs utility, but might also cause safety issues by impairing cognitive performance ((Diels and Bos, 2016; Emond and Zare, 2024)). In conditional automation, the driver still needs to be able to take over the car safely, e.g. in an emergency situation. If the driver develops symptoms of carsickness during the automated ride, e.g. due to NDRT engagement, the ability to react adequately and quickly might be impaired. Such a reduction in reaction capability could lead to safety-critical situations. Simulation sickness, sea sickness or cybersickness have shown performance decreases in visual (Bos, 2004; Bos et al., 2008; Golding and Kerguelen, 1992; Kaplan et al., 2017) and physical performance (Smyth et al., 2019a) as well as prolonged reaction times (Nalivaiko et al., 2015; Nesbitt et al., 2017; Bos, 2015; Smyth et al., 2019a). A test-track study, which investigated the effect of two different driving modes resulting in two carsickness levels on a simple tactile reaction task found higher reaction times in the mode with higher carsickness compared to the other mode with less carsickness, however there was no significant correlation of carsickness level and reaction times (Kantusch, 2023). As the underlying mechanisms and symptoms differ between different types of motion sickness, it should be investigated whether these effects are also present for carsickness. Metzulat et al. (2025) showed no significant negative effects of carsickness on performance in a visual search task, but significant negative effects of carsickness on reaction times in a simple reaction task. The performance data was partly retrieved from the here presented study and was pooled with data from another study, which was identical regarding the performance tasks, to strengthen the robustness. Additionally, hand-eye coordination was significantly impaired with increasing carsickness, while there was only a tendency for impairment of mental rotation with carsickness (Metzulat et al., 2025). However, that analysis did not take into account the study setting, being test-track or on-road testing. Therefore, in this paper, we will examine whether performance and the effect of carsickness on it differs between on-road and test-track conditions. We assume that differences in performance level could exist regardless of carsickness, as psychological states such as arousal may differ. However, the effect of carsickness on performance should be consistent in both environments.

To sum up, this study aims to explore which methodological aspects affect the assessment of carsickness across different testing environments (on-road and test-track), when horizontal accelerations as a function of time are replicated and occupants have internal vision. Driven by this objective, the following research questions (RQ) were defined:

RQ1. How differently does carsickness occur across different testing environments?
RQ2. What is the variation of various psychological factors (before and after the experimental procedure) across different testing environments?

RQ3. To what extent is the relation different between various psychological factors (before and after the experimental procedure) with carsickness across different testing environments?
RQ4. What is the effect of the optical flow of visually engaging non-driving related tasks on carsickness across different testing environments?
RQ5. What is the variation of the cognitive performance across different testing environments?
RQ6. To what extent is the effect of carsickness on cognitive performance different across different testing environments?

This paper will focus on RQ2-RQ5. RQ1 is also briefly addressed in (Harmankaya et al., 2024), which developed and validated the methodology around replicating on-road carsickness exposure. However, the current paper explores the influence of the testing environment in detail. RQ6 is also partially addressed in Metzulat et al. (2025), where the correlation between carsickness and cognitive performance is explored, but the effect across testing environments is presented here.

2 Methods

The experiment followed a 2x2x2 within-subject design. Testing environment, NDRT and performance tasks were the independent variables. The sequence of the testing environment, NDRT and performance tasks was randomized to counteract possible order effects.

2.1 Ethics statement

The experiment was performed in accordance with the Declaration of Helsinki. The study was approved by the Human Research Ethics Council of Delft University of Technology (Delft, The Netherlands; application number 3598). All participants gave their written informed consent prior to participation in the study. Participants received a compensation of 50 euros.

2.2 Participants

Participants were selected based on a screening survey [1] of more than 300 respondents. The questionnaire aimed at exploring the users' background on: motion sickness (demographic characteristics and motion sickness experience), non-driving activities while being driven, socio-demographics, driving experience and mobility, automation experience, general attitudes towards technology, personality. The survey results are not part of this work, and were collected in the scope of the Hi-Drive project. Fourty-seven participants (17 males, 29 females, 1 non-binary) took part in the experiment. They were preselected from the survey pool based on age, gender and motion sickness susceptibility. The participants' age ranged from 17 to 68 years (M = 29.30 years, SD = 13.01).

To assess susceptibility to MS, the Motion sickness susceptibility Questionnaire short-form (MSSQ-Short) was administered. The overall MSSQ-Short Score as well as the item regarding the experience of carsickness in the past ten years were weighted equally to assess the current theoretical susceptibility to carsickness. Depending on this score, participants were assigned to one of five categories, as described in Pham Xuan (2023). In order to prevent high drop-outs due to severe carsickness symptoms and at the same time to have participants that at least experienced carsickness to a certain extent, participants with very high susceptibility (category E) as well as participants with very low susceptibility (category A) were not invited to take part. It was aspired to mainly invite participants that met the criteria of categories C and D but in order to fill the participant pool, six additional participants of category B were invited.

---

[1] `https://tudelft.fra1.qualtrics.com/jfe/form/SV_5o7eClwKlADhR7E`

Across all invited participants, the mean MSSQ-Short value was 12.97 (*SD* = 5.99). This translates to a 55.90% percentile or rather a range of 44.96% to 65.57%. So, in general, susceptibility was slightly above average and within a limited range.

2.3 Experimental procedure

Prior to the first experimental session, the participants were informed about the general aim of the study, and gave their written consent. Then, participants filled out a pre-questionnaire [2] consisting of: Part 1: Anthropometrics, Part 2: Self Assessment Manikin (SAM) questionnaire (more information in 2.6.2), and Part 3: Motion sickness assessment questionnaire (MSAQ). Anthropometric data was measured for a few participants who were selected, aiming for gender balanced measurements to record human body dynamics while being driven by wearing the XSENS Motion Capture Suit. These participants were introduced to the XSENS system and calibrated it together with the experimenter. Finally, participants were made familiar with the performance task (see Section 2.6.3). They could practice it for two minutes, and then their pre-drive performance was collected in four minutes.

Thereafter, the first experimental session commenced. Participants either experienced the on-road or the test-track condition first. As stated, they gave their carsickness level on the MISC every minute. After reaching the end of the path in both testing environments or after stating a MISC level of 6 or higher, the carsickness accumulation phase terminated. Participants were then required to fulfill the post-drive performance task for four minutes directly after their latest MISC measure to ensure that the performance task was conducted at the highest possible level of carsickness. Thereafter, participants filled out a post-questionnaire [3], which consisted of Part 1: SAM questionnaire, Part 2: Comfort (2.1 MSAQ, 2.2 Acceptance, 2.3 ARCA), Part 3: Trust and Part 4: Perceived Safety. More details about the questionnaires are in Section 2.6.2.

Participants were invited to the second session with a gap of four to five days after the first session to avoid habituation effects. However, due to time limitations, the last five participants were invited with a smaller gap (approx. 1-2 days). The average gap across all participants was four days. The second session followed the same procedure as the first session.

2.4 Testing environment and vehicle

The on-road condition was carried out on a predefined urban and interurban route through and around the city of Delft, the Netherlands, with medium to high traffic density (see Fig. 1, left). The route contained sections with velocities of up to 80 km/h, but also slower sections due to school districts and residential areas with ample longitudinal and lateral acceleration phases due to traffic lights, turns and bends. The route took approximately 25 minutes to complete. Thus, there was enough time for the accumulation of carsickness.

In the on-road condition, due to regulatory constraints, the vehicle was manually driven by the safety driver. In total, three safety drivers performed the on-road conditions ($\sim$ 15 participants per driver). Prior to the experiment, the safety drivers completed an intensive training on the on-road course to assure driving style comparability across the safety drivers and to reduce variability between participant rides. Moreover, they followed a specific set of driving instructions, e.g., specifying speed limits for certain sections of the route to further ensure similarity between rides.

To replicate the on-road sickness exposure, the trajectory planning algorithm developed by (Harmankaya et al., 2024) was used to track the on-road driving dynamics and transfer them to a compact

---

[2] `https://tudelft.fra1.qualtrics.com/jfe/form/SV_4MBauHACisyBUmG`

[3] `https://tudelft.fra1.qualtrics.com/jfe/form/SV_b2RogXUDKppE1xA`

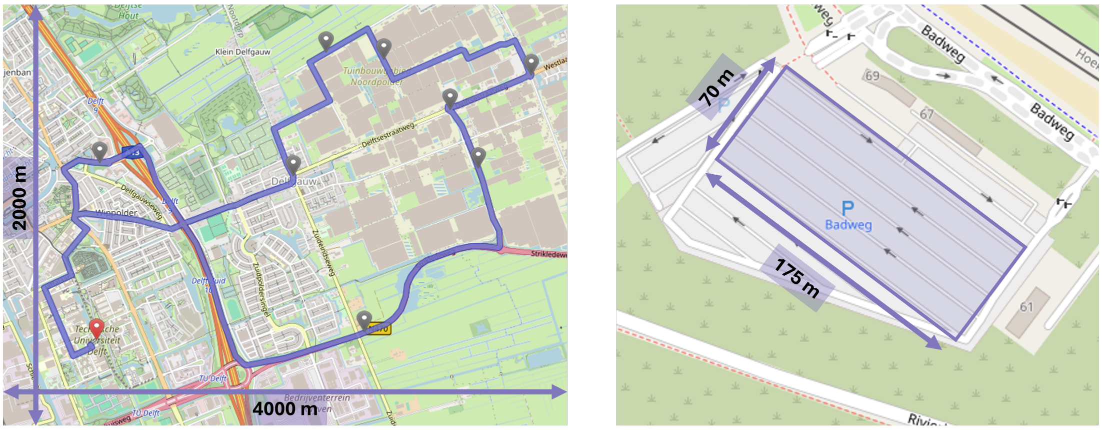


Fig. 1: (a) On-road and (b) test-track condition.

test-track. This was done with Nonlinear Model Predictive Control, which focused on tracking the longitudinal and lateral accelerations from driving dynamics data gathered from multiple drives by the three safety drivers. The algorithm also considered various constraints about the feasibility of the design path and the required dimensions of the compact test-track.

The test-track condition took place at a small parking lot (70 by 175 meters) at Hoek van Holland, the Netherlands (see Fig. 1, right). For the period of the experiment, the parking lot was fenced off. Therefore, there was no other traffic on the parking lot and thus, it represented a test-track surrounding. An E-Golf equipped with a SAE level 3 automated driving function was used as a test vehicle. The participants were seated on the front passenger seat, an experimenter was seated at the back to record the participants' feedback while a safety driver was on the driver's seat. In the test-track condition, the SAE level 3 system was activated and the safety drivers only acted as a back-up if the system requested a take-over. Whenever this occurred, the drivers briefly took over and then immediately transferred the control to the vehicle in order to ensure a high ratio of automated driving.

The general driving situation of the experimental conditions resulted in 94 (47 on-track, 47 on-road) sessions, lasting between 4 and 33 minutes. The driven distance in the on-road condition was up to 14.5 km. The planned trajectory on the test-track had a maximum distance of only 4.5 km (Figure 2), since it was designed to generate a stimulus of the same duration (about 26 minutes) but with lower velocities. The average speed in the on-road condition was 45 km/h (SD = 17) with a maximum of 86 km/h. In comparison, in the test-track condition, the average speed was 16 km/h (SD = 8), and the maximum was about 40 km/h. Additional characteristics are given in Appendix A.1.

Figure 2 shows a minor speed variability between runs (see the grey shaded area), indicating good replicability within conditions. As expected, replicability was best on the test-track using automation, whereas replicability was also quite good on-road where test drivers realised a rather consistent driving style.

The equivalence of the two conditions in terms of vehicle dynamics was evaluated using the motion sickness dosage value (MSDV) as described in ISO-2631:1997 Organization (1997). More specifically, we present the MSDV of vehicle accelerations for longitudinal ($\text{MSDV}_{un,X}$) and lateral ($\text{MSDV}_{un,Y}$) motion without the motion sickness frequency weighting (unweighted). This is selected to emphasize the vehicle dynamics, without assuming the validity of the common approaches in MS filter selection (Papaioannou et al., 2025; Harmankaya et al., 2024). On-road the accumulated $\text{MSDV}_{un,X}$ was 30.2 (SD = 7.4) and $\text{MSDV}_{un,Y}$ was 31.8 (SD = 8.5). On the contrary, for test-track, $\text{MSDV}_{un,X}$ was 16.8 (SD = 4.6)

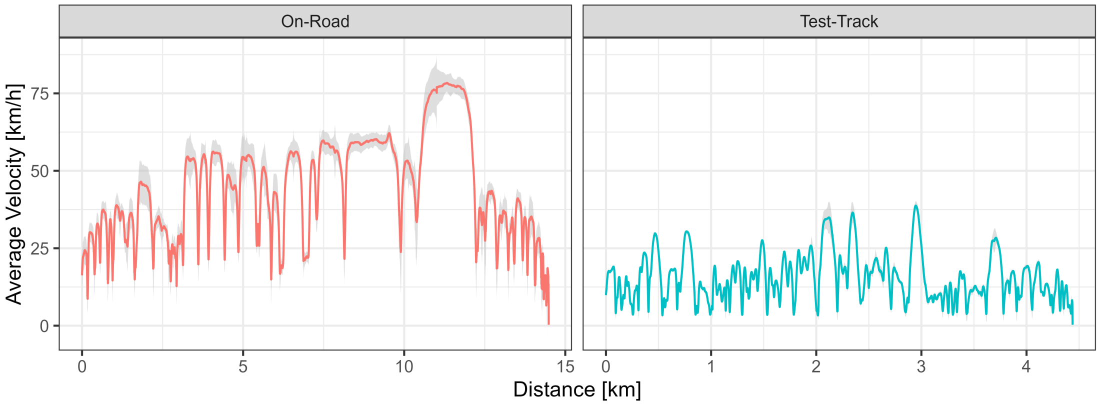


Fig. 2: Velocity as a function of distance for (a) on-road and (b) test-track. The grey shaded area represents the SD per distance interval. Horizontal axis limits are scaled to the maximum distance per condition to improve visibility.

and $\text{MSDV}_{un,Y}$ was 28.0 (*SD* = 8.6). In order to present a generalizable base for the dynamic stimulus, we calculated the average MSDV increase per second (Figure 3). This allows comparing any given natural or artificial ride regarding the given accelerations. However, this approach cannot capture the temporal motion sickness symptom development especially for longer stimulation's (larger than 5 minutes) that are characterized by varying phases within a stimulation.

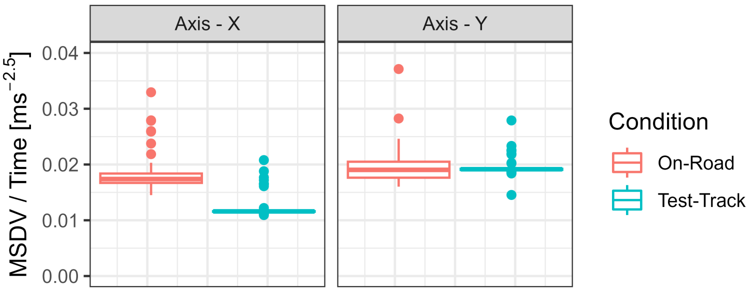


Fig. 3: Trip wise relative MSDV per second, clustered per axis of measurement and condition. *Note:* Maximum unweighted MSDV per session is divided by the total duration per session resulting in a representation of a comparable dynamic stimulus. Outliers occur due to large MSDV increase in the first half of the trajectory and occasionally early termination, which results in a higher relative MSDV per second.

2.5 Non-driving related task

We explored two video watching NDRTs, where participants watched Two videos from the same genre on a tablet held with the hands on the lap. Both videos depicted sporting events (a tennis and an ice-hockey game) to arouse similar emotional levels. To validate the differences in the visual dynamics (see Figure 4), the videos were analyzed regarding their respective optical flow using the Computer Vision Toolbox

(MathWorks 2022) prior to the experiment based on the Farnebäck method (Farnebäck (2003)). The difference in optical flow between the more dynamic ice hockey video ($M = 2.5 * 10^6$, $SD = 2.12 * 10^6$) and the more static tennis video ($M = 1.04 * 10^6$, $SD = 2.54 * 10^6$) was statistically significant when calculating a one-sided t-test with a medium to large effect size, $t(104400) = 110.39$, $p < .001$, $d = -0.63$.

To ensure and assess the participants engagement on the NDRTs, the videos were edited and included unexpected events. More specifically, a ball overlapped the tennis ball or hockey puck at random moments. The participants were required to count these events to verify their engagement in the NDRT. In total, sixteen events took place in both videos until the 25th minute. Their focus level during the ride was similar in both conditions and relatively high. More specifically, the participants counted on average 15.81 and 14.70 out of 20 events in total during the on-road and test-track condition. The differences between the counts could also have been caused by shorter driving times due to an early reaching of the termination criterion.

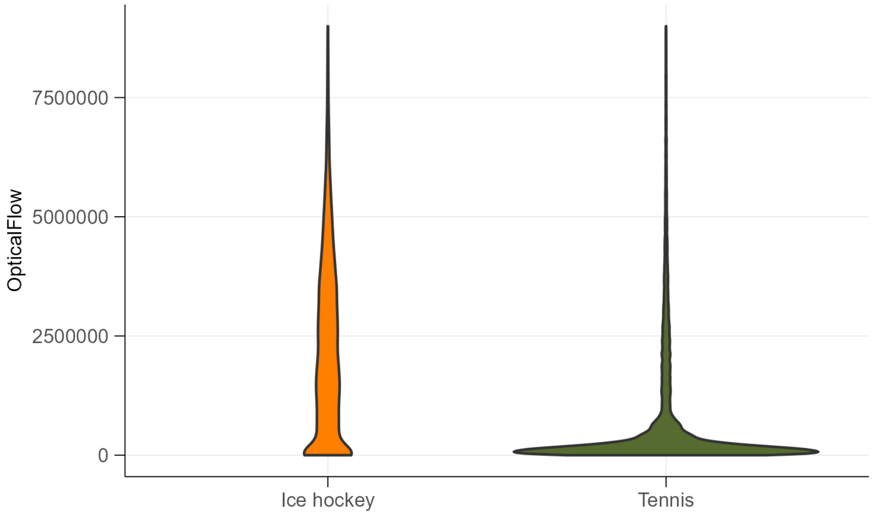


Fig. 4: Optical flow of the ice hockey and the tennis video

2.6 Dependent variables

2.6.1 Carsickness

To measure carsickness during the actual experimental rides, the misery scale (MISC; (Wertheim et al., 2001; Bos et al., 2005)) was used. Here, ratings from 0 - no problems to 10 - vomiting can be given. Participants were asked to verbally give their subjective motion sickness rating on the MISC scale once every minute. Additionally, participants were motivated in the general briefing to report changes in MISC at any time in between the one minute intervals. If participants reached a value equal to 6 (little nausea) or higher, the experimental session was terminated immediately. This was done due to ethical concerns. To analyze the MISC ratings between the conditions, the following procedure was implemented: If multiple ratings were recorded within a minute, only the first rating was evaluated. If no measurement was recorded in a given minute, the last MISC rating prior to that minute was used. As stated, if participants reached a

MISC rating of 6, the experimental session was terminated. In such cases, the rating of 6 was sustained and continued until the 25th minute, corresponding to the average ride duration. Furthermore, motion sickness prior to and after each session were assessed with the motion sickness assessment questionnaire (MSAQ), which includes sixteen items regarding motion sickness symptoms that are rated on a 9-point scale ranging from "not at all" to "severely" (Gianaros et al., 2001). Due to a mistake, one item was excluded (I felt queasy (G)) hence we calculated MSAQ as a percentage from the fifteen items.

2.6.2 Psychological factors

To investigate the psychological factors between the two testing environments, we focused on aspects of emotion (pleasure, arousal and sense of control) (Part 1), acceptance (Part 2.2), comfort (Part 2.3), trust (Part 3) and perceived safety (Part 4). In Part 1, the self-assessment manikin (SAM) was completed by the participants prior to and after each experimental session (Bradley and Lang, 1994). The SAM is a non-verbal questionnaire that uses pictograms to measure affect and feelings to an exposure. There are three dimensions: *Pleasure*, *Arousal* and *Sense of control*. Each dimension was measured in nine stages.

Acceptance, comfort, trust and perceived safety were measured after each session. In Part 2.2, acceptance was conceptualized as attitude towards the systems. The questionnaire clarified to the participants that as system they should consider either the SAE Level 3 automated system or the human driver. Hence, the van-der-Laan questionnaire was implemented (Van Der Laan et al., 1997). The scales are split into two subscales: usefulness and satisfaction. Thereby, scores, differing from the original publication, ranged on a 7-point Likert scale from -3 (low) to +3 (high). In Part 2.3, comfort was examined with selected questions from the automated comfort assessment questionnaire (ARCA, (Marberger et al., 2022)). The items were selected to fit the remaining questionnaire and both environments, the automated ride as well as the manual ride. Altogether twelve items regarding psychological and physical aspects were used from the ARCA. These are rated on a 7-point Likert scale. In Part 3 and 4, trust and perceived safety were assessed using five and seven respective items ranging from strongly disagree (1) to strongly agree (5) on a Likert scale. The items were partly derived from a survey by Nordhoff et al. (2021) developed for passengers of partly automated cars. In the analysis of these items, if necessary, scales were reversed so that an increase of any metric would indicate an improvement. Part 3 and 4 were aligned with the post-questionnaire generated to assess trust and perceived safety in the scope of the Hi-Drive project (Madigan et al., 2023).

2.6.3 Cognitive Performance

To measure cognitive performance, a visual search task and a simple reaction task were performed pre- and post-motion exposure. The Surrogate Reference Task (SuRT; (Mattes and Hallén, 2009)) was used to assess visual performance. The simple reaction task, based on a go/no-go paradigm (Donders (1969)), measured reaction times to sudden events. For both tasks, mean reaction times as well as the accuracy rate per block were analyzed. There were four blocks for each task of 1 minute, two for each difficulty. In each session, only one of the two tasks was completed pre-drive without carsickness and post-drive with carsickness. The task per test condition was randomized and balanced across all participants. The tasks were used according to Metzulat et al. (2025).

3 Results

3.1 Carsickness occurrence (RQ1)

The paper explored the variation of carsickness across different testing environments (RQ1). The overall MISC over both environments was 1.44 (*SD* = 1.52). The descriptive results of the MISC ratings over the duration of the rides are presented in Figure 5a.

With regard to the environmental conditions the average MISC in the on-road condition (*M* = 1.43, (*SD* = 1.47) was similar to the test-track condition (*M* = 1.45, *SD* = 1.58). In both conditions, on average, an increase of the MISC ratings could be observed throughout the respective sessions. Additionally, the number of drop-outs was almost evenly distributed between the on-road (*N* = 8) and the test-track (N = 6) condition and the number of participants reporting no carsickness during the whole ride in the on-road condition (*N* = 4) was similar to the test-track condition (*N* = 3). To confirm the descriptive analysis of the MISC data, a rmANOVA with session duration and condition as main factors was completed: No significant difference between the environments was revealed ($F(2,47) = 0.79, p = .75$). Furthermore, the MSAQ analysis (post-questionnaire, Part 2.1) also illustrated no significant difference between the two conditions (Figure 6). Among the four domains, the sopite-related (S) symptoms have higher percentages (irritation/annoyance, drowsiness, fatigue, uneasiness). The lack of significant difference was also captured in three out of four symptom domains in MSAQ. Significant differences were only captured on the peripheral symptoms ($W(2,47) = 78, p = .013$).

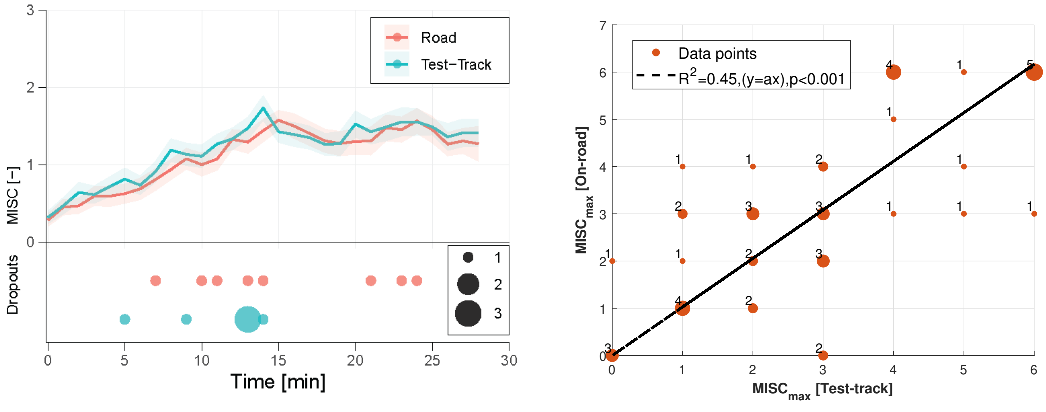
Considering the duration of rides, those which were not aborted due to sickness, took an average of 27 minutes (*SD* = 1.77, 39 rides) on the road and 26 minutes (*SD* = 0, 41 rides) on the test-track. Here, a significant effect of session duration on the accumulation of carsickness could be found when calculating the rmANOVA ($F(2,47) = 70.0, p < .001$). The replication of the individual motion sickness occurrence was also explored through Figure 5b, showing strong agreement at both the extreme ends and lower levels of MISC. The figure depicts the correlation between individual MISC_{max} values across two conditions. The data was fitted using a first-order polynomial model of the form (y = a * x). The best fit was obtained with a slope of a = 1.028, yielding an adjusted $R^2 = 0.45$ and a statistically significant result ($p < .001$).

The analysis of the maximum MISC and the corresponding MSDV (Figure 7), for example in longitudinal direction, underlines limitations regarding the interpretability of the MSDV in the given experiment. Sessions with a maximum MISC of zero lead to the highest MSDV. In comparison the MISC Level of six was in some cases already reached after a stimulation of about 14 $\text{ms}^{-1.5}$. Further trend analysis were not considered as the goal of the motion stimulation was rather to achieve a high reproducibility, which cannot be used to analyze the effect of MSDV on motion sickness symptoms.

3.2 Psychological factors (RQ2)

The paper addresses the variation of various participants' psychological factors before and after the experiment across different testing environments (RQ2). As psychological factors, we explore pleasure, arousal, sense of control, acceptance (defined by usefulness and satisfaction), Automated Ride Comfort Assessment (ARCA), trust and perceived safety.

The results of the three dimensions of the SAM (Pre- and Post-Questionnaire, Part 1) for both conditions are given in Figure 8. On average, participants experienced high levels of *Pleasure* in both conditions. Regarding, their *Arousal*, participants felt rather calm in both conditions, yet with slightly higher excitement in the test-track condition, albeit with a large variance in the responses. Finally, with respect to their *Sense of control*, participants reported a medium sense of control with slightly higher controllability in the on-road condition and again with a large spread in the answers. No significant difference was identified between the testing environments during pre- and post-experiment. The ensuing statistical analysis, only revealed a significant effect on arousal ($F(1,45) = 7.09, p = .011$).



(a) Pre-condition

(b) Post-condition

Fig. 5: (a) Subjective Motion Sickness Rating and Dropouts over Time for the different environments and (b) Correlation of individual MISC_{max} between the two conditions. The numbers of duplicates per coordinate are illustrated in the figure, while the size of the points changes accordingly. The data are fitted in first order polynomial functions (y = a * x). This figure was extracted from Harmankaya et al. (2024) for completeness in the analysis.

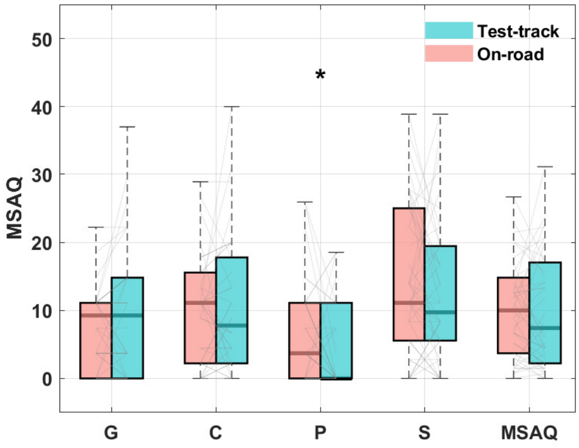


Fig. 6: Comparison of the four domains of MSAQ Gastrointestinal (G), Central (C), Peripheral (P), Sopite (S) and total MSAQ. Thin gray lines indicate individual changes between the on-road and test-track condition. Paired non-parametric Mann-Whitney test significance is presented by $^{*}p \leq .05$,$^{**}p \leq .01$, and $^{***}p \leq .001$.

The results of the acceptance measures (Post-Questionnaire, Part 2.2) on the van-der-Laan-scales were adjusted to a 7-point Likert scale ranging from -3 to 3. *Usefulness* ratings were slightly positive and above the scale average in both the on-road (*M* = 0.73, *SD* = 0.93) and test-track (*M* = 0.77, *SD* = 0.83) condition, with no significant difference between them. *Satisfaction* ratings also did not differ significantly between the conditions [$t(45) = -0.05, p = .957$], with on-road (*M* = 0.53, *SD* = 1.34) and

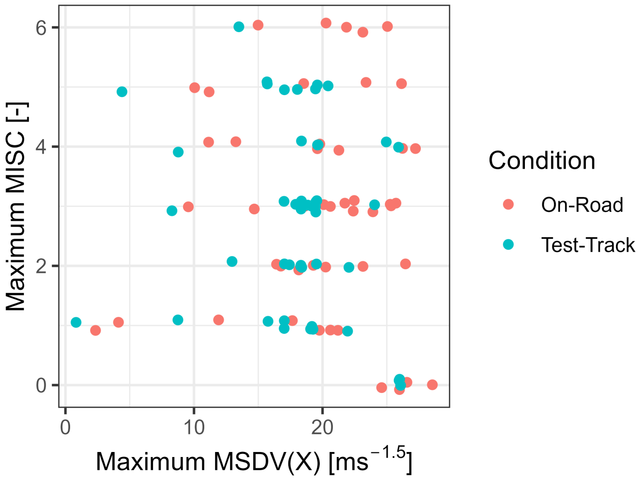
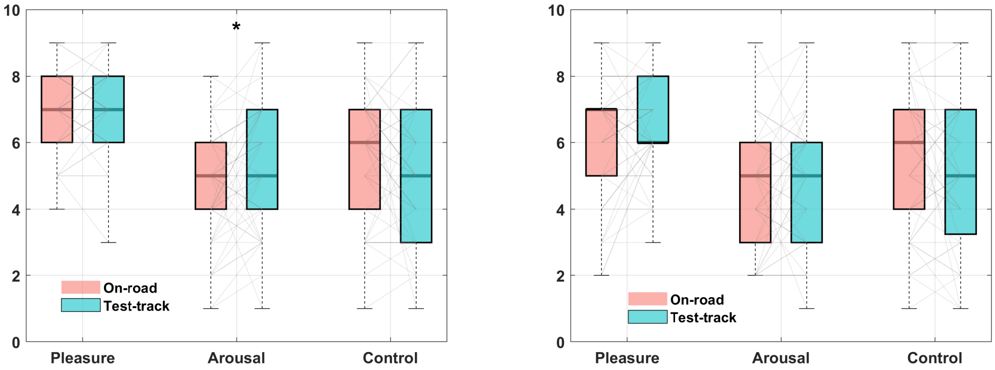


Fig. 7: Maximum MISC per trip vs. corresponding MSDV at the time of first occurrence. At the maximum MISC of 6 the ride was aborted and the maximum MSDV represents the end of the stimulation. In cases below MISC 6, the overall MSDV per trip can be higher than the displayed maximum MSDV as the first occurrence of maximum MISC is presented. Points are slightly jittered around the ordinate to improve visibility.



(a) Pre-condition (b) Post-condition

Fig. 8: Comparison of *Pleasure*, *Arousal* and *Sense of control* (SAM questionnaire) across testing environments of pre- (a) and post-ride (b). Thin gray lines indicate individual changes between the on-road and test-track condition. Paired non-parametric Mann-Whitney test significance is presented by $^{*}p \leq 0.05$ and $^{**}p \leq 0.01$

.

test-track (M = 0.54, SD = 1.22) ratings both being positive and above the scale average. Overall, both rides were assessed positively.

The ratings on the automated ride and comfort assessment items (Post-Questionnaire, Part 2.3) are shown in Figure 9. Most items were rated positively in both conditions with only the *Predictability* (Item 6) on-road and the *Fatigue* (Item 12) in both conditions being below the neutral rating (4). The *Feeling*

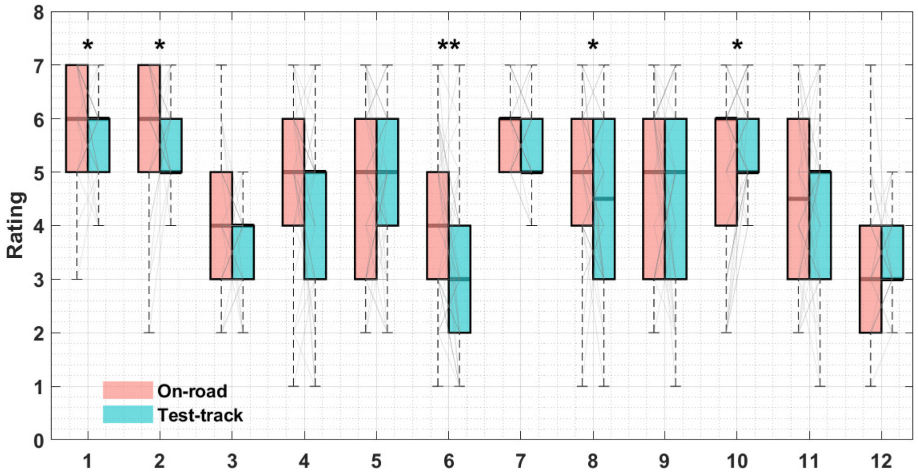


Fig. 9: Comparison of the specific comfort assessment items through the ARCA questionnaire across testing environments. The items refer to: 1: Sense of Safety; 2: Naturalness; 3: Feeling of control, 4: Travel progress; 5: Workload, 6: Predictability; 7: System Trust; 8: Interference with NDRT; 9: G-Forces Braking; 10: G-Forces Acceleration; 11: G-Forces Curves; 12: Fatigue. Values are from 1 - Negative to 7 - positive. Thin gray lines indicate individual changes between the on-road and test-track condition. Paired non-parametric Mann-Whitney test significance is presented by $^{*}p \leq .05$,$^{**}p \leq .01$, and $^{***}p \leq .001$.

of control in both conditions and the *Predictability* on-road were (close to) neutral (4). Significant differences between conditions were only identified for the *Sense of Safety* (Item 1, $W(2,47) = 290.0, p = .039$), the *Naturalness* (Item 2, $W(2,47) = 351.0, p = .041$), the *Predictability* (Item 6, $W(2,47) = 444.5, p = .003$), the *Interference with the NDRT* (Item 8, $W(2,47) = 432.0, p = .047$), and *G-Forces Acceleration* (Item 10, $W(2,47) = 225.5, p = .030$).

In order to get a general idea of the relationship between *Trust* and *Perceived Safety*, a mean value per dimension per participant and testing environment was calculated. This approach does not aim to introduce a new metric, but allows for a general estimation of the influence of the two dimensions on carsickness across different testing environments. *Trust* was higher in the on-road condition (*M* = 4.34, *SD* = 0.54) compared to the test-track condition (*M* = 3.87, *SD* = 0.64), although not significantly ($z = 1534.5, p < .001$). Additionally, *Perceived Safety* (Part 4) did not differ between conditions ($z = 1132.5, p = 0.694$; on-road: *M* = 4.07, *SD* = 0.44; test track: *M* = 4.02, *SD* = 0.46).

3.3 Psychological factors & carsickness (RQ3)

Following the analysis of the psychological factors by RQ2, the following sections presents the effect of carsickness on those psychological factors pre- and post-experiment (RQ3).

Regarding the SAM questionnaire (Part 1), it has been measured how the rating in the SAM dimensions changed between pre- and post-experiment with the subjective motion sickness (Figure 10). Therein, the change in *Pleasure* had a strong significant negative correlation with the maximum ratings on the MISC scale in both, the on-road ($r(186) = 0.40$, $p < .001$, 95% CI [.27, .51]) as well as the test-track condition ($r(171) = 0.34$, $p < .001$, 95% CI [.20, .47]). Hence, participants that reached higher maximum MISC ratings rated their mood as sadder compared to participants with lower maximum sickness ratings.

Additionally, the *Feeling of being in control* was significantly related to motion sickness on the test-track ($r(171) = 0.32$, $p < .001$, 95%CI [.18, .45]).

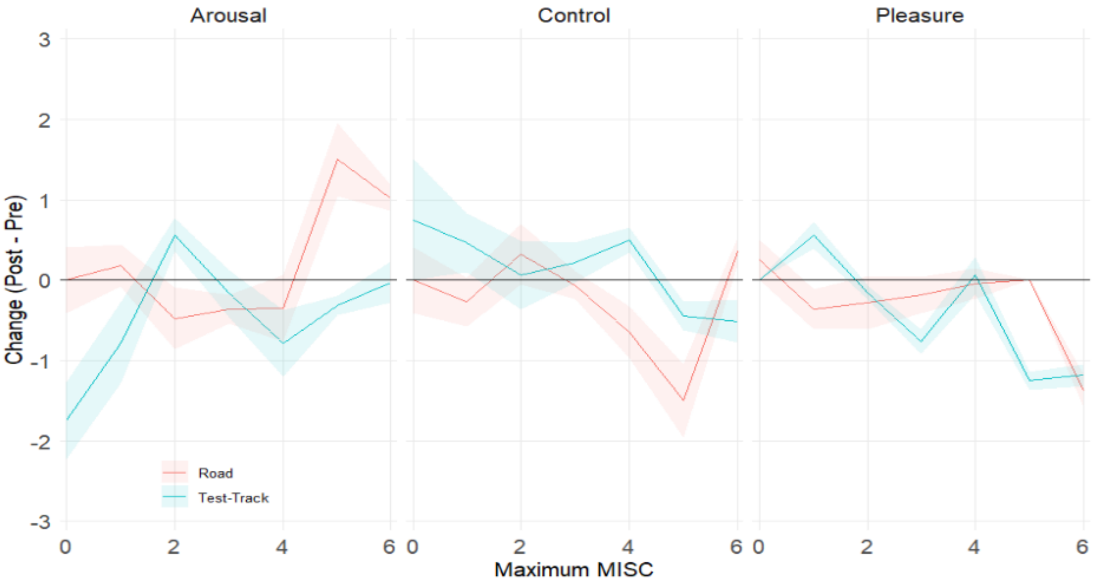
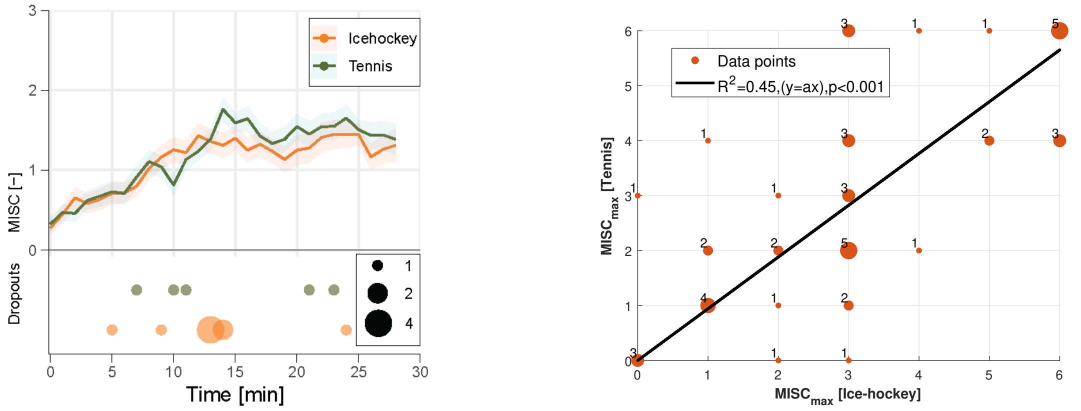


Fig. 10: The changes of SAM Dimensions over the maximum MISC

In order to estimate the relationship between *Usefulness* and *Satisfaction* (Part 2), *Trust* (Part 3) and *Safety* (Part 4) the correlations of the given answers after the ride with the maximum carsickness values per person were calculated. With regard to the van-der-Laan dimensions *Usefulness* and *Satisfaction*, only *Satisfaction* in the on-road condition yielded significant correlations. *Trust* and *Perceived Safety* were collected with several items, reflecting the multi-dimensionality and complexity of these quantities. Next to a question about general trust in the system, *Trust* was also measured by asking for trust in the system's ability to control the car well (keeping it lane centered, maintaining speed and distance to the car ahead), hesitation to use the system as well as general comfort to use the system. *Perceived Safety* estimated by asking for the feeling of safety, comfort, anxiety over the ride. Additionally, items about specific situations were asked: whether the respondent felt at risk and whether one felt being in danger at the worst moment. Lastly, it was asked whether the ride felt safer than expected and whether one would recommend the experience because of its safety. The calculated correlations of all of the described items regarding *Trust* and *Perceived Safety* with the maximum motion sickness can be found in the appendix A.3. Although the direction of the correlations is mostly negatively related to the experienced maximum motion sickness, the environments in which these relationships become significant are not consistent. This inconsistency could indicate that the different aspects of *Trust* and *Perceived Safety* are conflicted with regard to carsickness.

For the analysis of *Trust* and *Perceived Safety* with regard to motion sickness, the mean values of the grouped answers were used (as in Section 3.2). In terms of *Perceived Safety*, the Kendall-Tau-correlation between the subjective maximum Motion Sickness and a mean of all the items, showed a significant negative relationship in both environments (on-road: $\tau = -0.37$, $p = .008$; test track: $\tau = -0.36$, $p = .001$). The same relationship was observed with regard to *Trust*, although only significant in the test-track condition (on-road: $\tau = -0.12$, $p = .439$; test track: $\tau = -0.44$, $p = .002$). In both aspects, *Perceived Safety*



(a)

(b)

Fig. 11: (a) Subjective motion sickness rating and dropouts over time for the different environments and (b) Correlation of individual MISC$_{max}$ between the two non-driving related tasks. The numbers of duplicates per coordinate are illustrated in the figure, while the size of the points changes accordingly. The data are fitted in first order polynomial functions $(y = a * x)$.

and *Trust*, higher levels of motion sickness were, mostly significantly, associated with lower *Perceived Safety* and *Trust*. The results of the correlations for all items in both aspects can be found in the Appendix (A.4, A.3).

Twelve items of the ARCA were used to assess the passengers' *comfort*. Across all items, it could be observed that a higher maximum subjective motion sickness leads to a worse rating of the system. However, across the two testing environments, only five items were significantly influenced (see Section 3.2). Within these values, *Perceived Safety* and *Interference with NDRT* showed a significant negative relationship in the test-track condition ($\tau = -0.29$, $p = .020$ and $\tau = -0.29$, $p = .014$, respectively). The remaining seven items, which did not become significant in 3.2, are assumed to be disconnected from the testing environment, although they partly showed a significant relationship to the maximum motion sickness value. The full table of the calculated correlations is given in the appendix A.2.

3.4 Optical flow of non-driving related tasks (RQ4)

In this section, the effect of the optical flow of visually engaging non-driving related tasks on the occurrence of carsickness across different testing environments is explored (RQ4).

The subjective motion sickness ratings for the ice hockey condition averaged 1.43 (*SD* = 1.54), while it was 1.44 (*SD* = 1.51) for the tennis condition. Dropouts almost doubled (from 5 to 9) when participants watched the ice hockey video with the higher optical flow. The results for the development of the MISC over time with respect to the experienced NDRT is depicted in Fig. 11a. From a descriptive point of view, the development over time is comparable up until the 11th minute mark. Thereafter, the tennis condition had slightly higher MISC values until the end of the experiment. To further analyze the data, a rmANOVA was conducted. In both NDRT conditions, carsickness accumulated over time at the same rate, with a highly significant effect of time on the response $(F(2,25) = 71.8, p < .001)$. NDRT as factor, showed no significant effect of NDRT on the MISC response $(F(2,25) = 0.98, p = .47)$.

Additionally, the motion sickness occurrence between the different NDRTs needs to be observed on an individual level. Hence, we opt to explore this through Figure 11b, where we see a good correspondence at the maximum and at low levels of MISC. More specifically, four participants reported no sickness and five other participants reached the termination criterion (MISC = 6), on both the ice hockey and tennis video. The figure presents the correlation of the individual MISC_{max} between the two conditions, fitting the data to a first order polynomial ($y = a * x$). The optimal fit is achieved with $a = 0.746$, with an adjusted $R^2 = 0.40, p < .001$). Considering the decent R^2 and the significance of the result ($p < .001$), this implies that the individual MS occurrence (MISC_{max}) was by around 25% less in the condition with the ice hockey compared to the one with the tennis video. This results contradict with the non-significant differences between the average MS occurrence.

3.5 Cognitive Performance (RQ5 & RQ6)

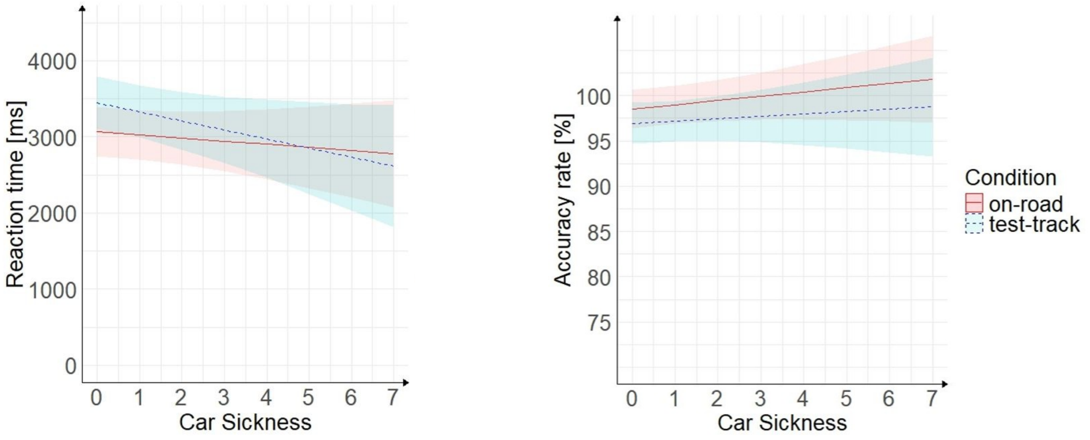
This paper explored the variation of cognitive performance before and after the experiment in relation to motion sickness (RQ5) and across testing environments (RQ6) is being explored.

For this multilevel analyses, also known as hierarchical, linear models (Hox et al., 2017), were used due to the repeated measures nested within individuals. As outcome variable, the performance measure per block for pre- and post-drive was used. The predictor of interest for RQ6 is the interaction of carsickness (MISC level) with the condition (on-road vs. test-track) to check for differences between conditions regarding the carsickness effect. The condition is also included in the model separately to see the unique effect on performance (RQ5), as was the factor carsickness. For condition, the on-road condition is labeled as the reference category. Task performance is most likely also affected by other possible confounding factors and factors that were intentionally manipulated like the difficulty level of blocks. Therefore, the time point (pre- vs. post-drive) with pre-drive being the reference category is included as predictor to control for the confounding effect of time as well as the difficulty level with the easy condition as reference category. One model for each task and dependent measure was calculated. As the significance of carsickness and its interaction with the study condition as a predictor of task performance is of interest, we choose not to report model fit tests. For further details on the used method see Metzulat et al. (2025) who applied the same analysis on the same tasks. Table 1 shows the models for each task and criteria. Each model includes $N = 376$ observations. Fig.12 and 13 show the predicted values (simple slopes) of the interaction effect of carsickness and condition with 95% CI for reaction times (a) and accuracy rates (b).

Table 1: Regression results of visual search task and reaction task

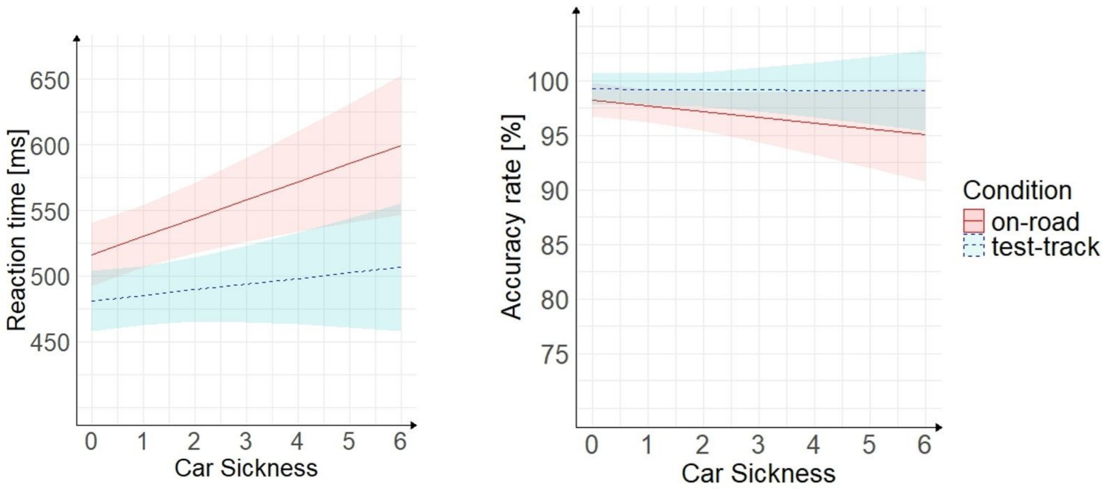
		Visual search task					***Reaction task***				
		b	***95% CI***		***β***	***p***	***b***	***95% CI***		***β***	***p***
Criterion	**Predictor**		***LL***	***UL***				***LL***	***UL***		
Reaction time [ms]	(Intercept)	3068	2741	3396	-0.54	< **.001**	516	492	540	0.3	<**.001**
	Carsickness	-42	-140	57	-0.05	0.405	14	5	23	0.25	**0.002**
	Condition	383	-74	840	0.23	0.1	-35	-67	-3	-0.56	**0.03**
	Carsickness *Condition	-78	-208	52	-0.09	0.237	-10	-20	1	-0.17	0.076
	Time Point	-213	-425	1	-0.17	0.049	-17	-34	-1	-0.21	0.039
	Difficulty	1308	1143	1473	1.03	< **.001**	19	6	32	0.23	0.004
Accuracy rate [%]	(Intercept)	98.54	96.42	100.66	0.32	< **.001**	98.26	96.72	99.8	-0.06	<**.001**
	Carsickness	0.47	-0.21	1.15	0.1	0.176	-0.53	-1.28	0.21	-0.13	0.16
	Condition	-1.58	-4.51	1.35	-0.25	0.198	0.99	-0.93	2.91	0.25	0.31
	Carsickness *Condition	-0.21	-1.11	0.69	-0.04	0.644	0.5	-0.39	1.4	0.12	0.269
	Time Point	-0.2	-1.67	1.28	-0.03	0.794	0.93	-0.51	2.37	0.15	0.207
	Difficulty	-2.79	-3.94	-1.64	-0.38	< **.001**	-1.99	-3.18	-0.8	-0.32	0.001

Notes:b represents unstandardized regression weights; 95% CI: 95% confidence interval of b; β represents the standardized regression weights; Significant effects between predictors of interest and the respective criteria are printed in bold; respectively; LL and UL indicate the lower and upper limits of a confidence interval



(a) (b)

Fig. 12: Visual search task: Simple slopes with 95 CI marked for interaction effect of carsickness and condition on reaction time (a) and on accuracy rate (b).



(a) (b)

Fig. 13: Reaction task: Simple slopes with 95% CI marked for interaction effect of carsickness and condition on reaction time (a) and on accuracy rate (b).

The difficulty level predicted the performance of all criteria significantly, showing that the manipulation was successful. The effect of difficulty and time point are not interpreted further as they were only included in the model to control for their confounding effect. When controlled for the other predictors, carsickness did not significantly predict the accuracy rate in either task. The same applies for the condition and interaction of both factors. Neither carsickness, condition nor their interaction significantly predicted reaction time in the visual search task. Thus, performance did not differ significantly between on-road and test-track in visual search. However, in the reaction task carsickness and condition predicted the reaction time significantly. The model predicts that the reaction time increases by approximately 14 ms for each point increase on the MISC scale. This means that reaction times are increasing with increasing carsickness level when reacting to sudden events. The model further predicts the reaction time to decrease by 35 ms in the test-track condition compared to on-road. The interaction of interest (carsickness*condition) is slightly not significant. So, the effect of carsickness on performance was not significantly different between the two study conditions in either task. However, looking at the simple slopes in Fig. 13a there is a tendency ($p = .076$) that reaction times in the simple reaction task are predicted to increase stronger in the on-road condition compared to the test-track condition.

4 Discussion

4.1 Carsickness (RQ1)

The experimental set-up generally led to motion sickness, measured on the MISC, comparable to other working groups (Bos et al. (2005): motion simulator, Irmak et al. (2025): vehicle on public roads). The high standard deviation of the average subjective carsickness is caused by the inclusion of susceptible and non-susceptible participants in this calculation. At the same time, the duration of the provocation had a significant effect on carsickness. This result underlines the time-dependence development of motion sickness (Irmak et al. (2022); Pham Xuan (2023)). Neither the development of the subjective carsickness (considering means, dropouts or participants without any carsickness) nor the statistical analysis showed a significant difference between the two testing environments. Additionally, the analysis of the detailed relationship between the two environments by means of fitting a linear regression between both conditions for individual maximum carsickness led to a slope of 1.028, with an adjusted $R^2 = 0.45$. This indicates that the maximum carsickness on the test track and on road were similar within individuals. The medium R^2 and the significance of the result ($p<.001$) implies that the total individual MS occurrence (MISC_{max}) was highly replicated by the method. The only motion sickness element, which did show a significant difference between the environments, were the peripheral symptoms recorded by the MSAQ. This component, in contrast to the remaining three dimensions, could have been more sensitive to environmental differences. In the test track, the peripheral surroundings in the environment were different due to the location, and more dynamically varying due to the optimized trajectory. As Harmankaya et al. (2024) pointed out, their applied approach increased the yaw motion of the vehicle, provoking excessive vehicle re-directional cornering to allow the replication of the on-road exposure in the compact test-track. Therefore, the significant different peripheral symptoms were expected given that the yaw motion is provocative for the visually induced motion sickness (Nooij et al. (2017)). Furthermore, there was a negligible tendency towards higher carsickness in the on-road ride, which could be caused by individual changes of psychological states. However, these weren't recorded. In contrast, the importance of driving dynamics on carsickness is strongly indicated, raising the question of a ranking between these different influencing factors. While Lukacova et al. (2023) investigated the influence of relevant trait characteristics (such as migraine or personality traits) using a survey, future work should shed further light on the impact factor of trait factors in comparison to the dynamics of the environment. Hence, enabling a more holistic view on influential factors of motion sickness. To sum up, these results demonstrate that the applied approach by Harmankaya et al. (2024) successfully provoked comparable levels of motion sickness in both environments: on road and on the test track.

Despite the need of this method for complex technical equipment (i.e., higher level AVs), these results mark an important milestone towards the efficient replication of motion sickness in safer and more accessible environments and the comparability of motion sickness studies in different environments. The great importance of these results is also evident when exploring the efficiency of other methods to replicate on-road exposure to MS, e.g. driving simulators. As simulators are a beneficial environment when it comes to replicability of driving conditions, few studies have also examined simulators' absolute and relative validity in terms of motion sickness. When applying a realistic driving behaviour in the simulator significant differences were found between the two conditions (on-road and driving simulator) with regards to motion sickness levels (Talsma et al. (2023), illustrating simulator disadvantages about their absolute validity. Even different driving simulators, ranging from simplified to more advanced setups, had significant difference on motion sickness exposure (Himmels et al., 2024).

Simulator and on-road studies differ in several factors such as weather, passenger and driver distraction or other road users which affect the overall user experience. Meanwhile, simulators are also capable of provoking symptoms of simulator sickness due to the visual environments and the artificial motions. The gap between artificial and real-world conditions is further narrowed to factors such as traffic and road conditions in the comparison between on-road and test track environment. This is an additional step towards absolute validity in testing environments.

4.2 Psychological factors (RQ2)

The comfort derived through the ARCA questionnaire (Figure 9) was rather positive (i.e., item mean values were higher than four which is the scale median) regardless of the testing environment (except for *Feeling of control*, *Predictability* and *Fatigue*). Interestingly, the question regarding *Sense of Safety* (Item 3) from the ARCA questionnaire ("While being driven, I felt [unsafe ... safe]) evaluated the on-road condition as more positively while the question within Part 4 for *Perceived Safety* ("I was feeling safe most of time") led to no significant differences between two driving conditions. This discrepancy could be caused by the higher resolution of the ARCA scale (7-steps vs. 5-steps), which leads to a higher sensitivity of the questionnaire. Alternatively, the different phrasing of the questions could have led to an answer reflecting either the whole ride (perceived safety) or rather just an especially memorable and emotional situation (ARCA). In addition to the significant differences in the *Sense of Safety* and *Interference with NDRT* (ARCA), *Trust* (Part 3) was also evaluated significantly more positively in the on-road condition. This suggests that both real-world driving scenarios may foster greater confidence in the system (automated system or driver as explained in the questionnaires), possibly due to perceived realism or familiarity. This, in turn, allowed to fully immerge into the NDRT. A study evaluating trust in automated driving found that trust did not increase with repeated usage in an on-road condition, while it did increase in a simulator setting, which is also more controlled and artificial Metz et al. (2025). However, in our on-road condition, participants did not experience an automated driving system, but instead a manually driven ride, which most likely increased trust due to higher familiarity. Peng et al. (2025) identified in a questionnaire that multiple prerequisites influence the willingness to immerge into NDRTs. Although trust is mostly mentioned there - it is only one factor. In parallel, the driving environment or the desire to remain vigilant in an (unfamiliar) AV, seems to play a role regarding NDRT engagement. It is important to further identify the influencing factors on NDRT engagement as it is one of the passengers' greatest benefits in AVs. Besides *NDRT engagement*, *G-Forces Acceleration*, *Naturalness* and *Predictability* were rated worse in the test-track condition indicating that participants sensed that the applied trajectory differed from natural rides and therefore felt artificial and unfamiliar. However, these did not eventually affect the occurrence of motion sickness.

Contrary to expectations, *Anxiety* did not significantly differ between the two driving conditions. This indicates a relatively stable emotional response across contexts, despite the differing nature of the environments. The Self-Assessment Manikin (SAM) results further support this interpretation. Participants reported high levels of *Pleasure* in both conditions, indicating a generally positive emotional experience. *Arousal* levels were generally neutral, with significantly higher in the test track condition before the ride. This may reflect heightened alertness or excitement due to the controlled yet unfamiliar setting with the automated vehicle. The *Sense of control* was moderate across both conditions, with slightly higher ratings in the on-road scenario, again pointing to the potential influence of environmental familiarity.

Acceptance measures revealed no significant differences in perceived *Usefulness* or *Satisfaction* between the two conditions: Both were rated positively, suggesting a generally favorable reception - even of the unfamiliar automated driving system. Nevertheless the lower trust ratings recorded for the test track condition align with the findings by Czaban and Himmels (2025), who found higher stress levels in the more artificial environment (simulators) when comparing them with on-road rides. These reduced trust ratings may indicate that artificial or staged environments evoke more caution or skepticism among users. When developing and testing new autonomous driving functions, the possibly lowered trust in artificial environments need to be considered, as trust is an important factor in the acceptance of AVs (Kenesei et al. (2025)).

4.3 Psychological factors & Carsickness (RQ3)

Besides focusing on the general effect of differences between the two study conditions, the present study aimed to investigate the relationship between subjective experience measures and carsickness. Again, this was done by comparing user perceptions between on-road and test-track conditions. The findings offer several insights into how emotional response, acceptance, trust and perceived safety relate to motion sickness and overall user experience. To the knowledge of the authors, the items of *Acceptance*, *Perceived Safety*, and *Trust* have not been investigated regarding motion sickness in a vehicle with SAE automation Level 3 or higher.

The analysis of the emotional responses regarding *Arousal*, *The feeling of being in control* and *Pleasure* from the SAM questionnaire showed a significant negative relationship between *Pleasure* and motion sickness. Greater levels of subjective motion sickness are connected with a stronger reduction in the reported *Pleasure* compared to the pre-ride condition. This aligns with the findings by Choukèr et al. (2010); Stelling et al. (2021), who found a significant connection between motion sickness and stress in parabolic flights. In terms of user workload, this finding underlines the necessity of working on motion sickness reliefs in environments which are more likely to provoke motion sickness, especially for passengers who need to work or otherwise have cognitive workload. In addition, Kaufeld et al. (2022) found a connection between stress and visually induced motion sickness. However, a discrepancy to that paper can be found regarding *Arousal*, which seems to be unaffected by the reported motion sickness. This difference could be caused by the additional influence of the environment. While Kaufeld et al. (2022) stayed in the same environment, but actively elicited certain emotions, the change of the external environment along with the used automation level to drive the participants could have additionally influenced the arousal and thereby confounded the relationship between arousal and motion sickness in our experiment. This thought also applies to *The feeling of being in control*, which has a significant relationship with motion sickness on the test-track. While being driven in an L3-automation level in a separated area is an uncommon experience to most people, this unfamiliarity connected with the feeling of being unwell due to motion sickness could have confounded the relationship between the two latter.

A key finding was the significant negative correlation between maximum carsickness ratings and participants' *Trust* (on test-track) and *Perceived Safety* (on both, test track and road). This suggests that higher levels of carsickness may undermine users' confidence in the driving system, potentially influencing their overall endorsement of both, an automated driving system as well as a human driver, especially on a test track. A negative relationship between motion sickness and *Trust* did not become significant in the Road-condition. This may reflects differences in expectations with regard to *Trust*.

Importantly, under the test-track condition the item *Interference with NDRT* was strongly negatively correlated with maximium motion sickness ratings. This is consistent with the findings of Irmak et al. (2021a), who found a significant negative correlation between motion sickness and subjective workload. Following this finding Irmak et al. (2021a) hypothesize that motion sickness may lead to task avoidance. This assumption is further corroborated in the test track condition based on the presented study. This result reinforces the notion that physical discomfort can significantly detract from the overall user experience, even when other aspects of the system are positively evaluated.

4.4 Optical flow of non-driving related tasks (RQ4)

On an individual level a difference between the influence of the two tasks could be found, with a slight tendency towards higher subjective motion sickness in the tennis condition. However, the analysis of significant differences did not reveal significant differences in the averaged MISC between the ice hockey and tennis videos. This indicates that, when generalized to a broader population, the higher optical flow in the ice hockey scenes does not lead to an increase in motion sickness. Although the difference in optical flow between the two video types was statistically significant (see 2.5), it may not have been substantial enough to meaningfully affect symptom development. The observation that participants in both conditions still experienced similar levels of motion sickness could be attributed to the fact that individual susceptibility and driving dynamics likely exerted a stronger influence on the outcome than the differences in visual motion. However, several studies indicate that the dynamic visual input does have an effect on motion sickness. Thereby, it can be distinguished between the passengers' NDRTs and the optical flow of the environment. Regarding NDRTs, Metzulat et al. (2024) reported that engaging in a dynamic visual task provokes more motion sickness than performing a static task, supporting the idea that dynamic visual stimulation of NDRTs can exacerbate motion sickness symptoms. On the other hand, studies indicate that an alignment of optical flow with the driving dynamics can, however, have a positive influence on motion sickness. In vehicles, having no external vision (Irmak et al. (2021b)) as well having a restricted view due to lower viewing angles into the vehicle interior (Brietzke et al. (2021)) increases carsickness in comparison to having external vision. Finally, Tamura et al. (2023) have investigated the relationship between visual and vestibular inputs and concluded that this relationship can be considered when mitigating motion sickness. Generally, it is hypothesized that while unsynchronized visual motion can increase symptoms, an aligned optical flow of visual input with the sensed movement can alleviate motion sickness symptoms.

Another factor that could have an influence on the interaction of the optical flow and the provoked carsickness is the eye movement. Brietzke (2023) found that different tasks in a vehicle lead to different eye movements while several papers showed that eye movement is connected to visually induced motion sickness (VIMS): Wibirama and Hamamoto (2014) found that an unstable depth gaze was related to experienced VIMS, while Diels et al. (2007) found an interaction between optic flow and the gaze angle on VIMS. In the present study design, the counting task may have influenced participants' eye movements. Although the search target always appeared in the center of attention, their eye movements due to the search likely differed from natural viewing behavior when watching a film (Castelhano and Rayner (2023)). As the visual input plays an important role in the genesis of motion sickness (Flanagan et al. (2004)), the task might have eventually triggered similar carsickness developments even though the optical flow of the videos were significantly different.

4.5 Cognitive Performance (RQ5+RQ6)

The three performance metrics (accuracy rate of visual search, reaction times in visual search, accuracy rate of simple reaction task), illustrated no difference across testing environments. However, a significant difference in reaction times was found across testing environments in the simple reaction task, with faster times in the test-track condition. This could be due to differences in the testing environment; for example, there is a higher potential for distraction due to noise, other road users, etc. on the road, or, as reported above, the higher level of arousal on the test track could have had an activating effect on the subjects. However, this was a between-subjects comparison of on-road versus test-track. Hence, this difference is confounded with individual differences in performance, leading potentially to this result. Therefore, this effect might not be robust.

For both tasks, the effect of carsickness on performance did not differ significantly between the testing environment. This result shows that the effect of carsickness on performance is comparable between environments, regardless of potential differences due to variations in the test environment. However, a non-significant difference was identified in the reaction times in the simple reaction task, which show a tendency to increase more substantially in the on-road condition (larger slope at the curve, Figure 13) compared to the increase observed in the test-track condition. This could also be due to the above mentioned differences of environment (noise, arousal, etc.). Nevertheless, as the effects are in the same direction and the interaction was not significant, the effects of carsickness on performance do not differ across testing environments.

Carsickness had no significant effect on accuracy rates in either the visual or the reaction task. Reaction times in the visual search task were not affected by carsickness, whereas in the reaction task the reaction times increased significantly with increasing carsickness. Although previous studies have reported a negative impact on visual performance (Bos et al. (2008); Golding and Kerguelen (1992); Kaplan et al. (2017)), the results of the present study are more consistent with those of Smyth et al. (2019b), in which no adverse effects of simulator sickness on visual acuity were observed. The finding of prolonged reaction times with increasing carsickness in simple reaction tasks is consistent with studies on cybersickness (Nesbitt et al., 2017; Nalivaiko et al., 2015) but only partly with the study regarding carsickness (Kantusch (2023)).

4.6 Limitations

In order to interpret the findings of this study correctly, certain limitations need to be noted. Methodologically, it should be considered, that the recruitment of the sample was primarily through a university network, which resulted in a young age range (M = 29.30 years, SD = 13.01). Literature indicates that susceptibility to motion sickness decreases with increasing age Bos et al. (2007). Therefore, the predominantly young sample has potentially increased susceptibility compared to a possible sample with a broader or older age range, but probably did not bias the qualitative findings. Meanwhile, the dataset was not gender-balanced, which might also raise certain limitations since females are more susceptible to motion sickness Flanagan et al. (2005); Lentz and Collins (1977). With regard to the psychological factors, one aspect should be noted: The scale to assess the subjective acceptance (Van-der-Laan questionnaire) was adapted from an original 5-point scale to a 7-point scale. This change may have affected the validity of the results, as participants could have interpreted the more granular scale differently. Additionally, the comparability of the results to other publications using the 5-point scale is restricted. All rides were done while participants watched a sports videos, representing a non-driving related task. Although using NDRTs to increase the risk of motion sickness in such studies is a often used method, it limits all of our findings to passengers engaged in a NDRT. Extending the results towards a ride without NDRT can only be done with reservations, due to the lack of rides without NDRTs in this study. Finally, the on-road condition was always driven manually, while the test track condition was driven using autonomous driving functions. Even if the vehicle dynamics were highly replicable according to our extended analysis, this might have affected the psychological factors across the two conditions.

5 Conclusion & outlook

The presented manuscript summarizes a human subject study on carsickness in two highly dissimilar environments. The work advances research in the area of carsickness by an improvement in testing methodologies. In detail, insights on the following key aspects were addressed: the testing environment, psychological factors, optical flow, and cognitive performance. The investigation of these factors was administered by a within-study with 47 participants. In the study, we realized the highest possible reproduction of a dynamic stimulus in on-road driving by implementing a trajectory controlled self-driving vehicle. During the rides, we collected the subjective motion sickness experience, in combination with a post drive symptom assessment. Participants were watching different sports videos, varying in visual optical flow. Cognitive performance, as measured by a simple reaction task and a visual search task, and various psychological factors were recorded before and after the rides.

Our findings

- The dynamics stimulus in combination with the “eyes off the road” task of video watching led to the expected development of motion sickness. No significant differences in motion sickness symptom reaction occurred, despite the different testing environment and the variations in dynamic stimulation due to traffic in the on-road environment. This is a first step towards making future studies, conducted either in an on-road or a test track environment, comparable.
- The consideration of psychological factors leads to the conclusion of comparability across the two testing environments for most aspects such as general ride comfort, anxiety, acceptance and satisfaction. However, it became also apparent that participants have reduced trust on the test track. This is especially important when considering that multiple psychological factors, i.e., arousal, pleasure, trust and perceived safety correlate with carsickness.
- The observed correlations between sickness with pleasure, safety perceptions were correlated across testing environments. The repeatability of these results proves that methodologically the interconnected nature of physical and psychological responses can be captured across/in testing environments and allows the assessment of perceived safety. However, this was not captured for the others psychological factors, such as acceptance and trusts.
- The two different optical flow stimuli, represented by means of a tennis and ice hockey video did not alter the motion sickness symptom reaction in the tested scenario. As passenger activities have been shown to lead to different provocations of motion sickness, it is essential to further investigate which cognitive processes contribute to motion sickness.
- Cognitive performance differed by one metric out of four between the test environments. Participants reacted faster on the test track in the simple reaction task, which could have been influenced by differences in arousal and distraction between conditions. However, in terms of carsickness, its effect on cognitive performance was comparable in both test environments. This suggests that both settings can be equally used for research on this topic.

Further work on influencing factors of motion sickness is needed to test and to develop effective countermeasures and metrics to assess it. In order to be successful, different research needs to be able to derive and transfer results from different testing environments and stimuli. The presented work is further contribution towards establishing a methodology to achieve such standards.

Acknowledgments

The authors would like to thank all partners within the Hi-Drive project for their cooperation and valuable contribution. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101006664. The sole responsibility of this publication lies with the authors. Neither the European Commission nor CINEA – in its capacity of Granting Authority – can be made responsible for any use that may be made of the information this document contains.

Contributions

R.P.X. Conceptualization, Methodology, Investigation, Resources, Data Curation, Formal Analysis, Writing - Original draft, Writing - Review & Editing, Project administration, Funding acquisition; A.B. Conceptualization, Methodology, Software, Investigation, Resources, Data Curation, Formal Analysis, Writing - Original draft, Writing - Review & Editing, Project administration; H.K. Conceptualization, Methodology, Formal Analysis, Writing - Review & Editing; M.M. Conceptualization, Methodology, Formal Analysis, Writing - original draft, Writing - Review & Editing; A.E.: Conceptualization, Methodology, Formal Analysis, Writing - original draft, Writing - Review & Editing; R.H. Resources, Writing - Review & Editing, Funding acquisition; G.P. Conceptualization, Methodology, Software, Investigation, Resources, Data Curation, Formal Analysis, Writing - Original draft, Writing - Review & Editing, Project administration;

Conflict of interest

RPX., A.B., H.K. are employed by Volkswagen AG. A.E. is employed by Audi AG. The remaining authors declare that they have no conflict of interest

References

Bansal P, Kockelman KM, Singh A (2016) Assessing public opinions of and interest in new vehicle technologies: An austin perspective. Transportation Research Part C: Emerging Technologies 67:1–14

Bengler K, Omozik K, Müller AI (2019) The renaissance of wizard of oz (wooz)–using the wooz methodology to prototype automated vehicles. Proceedings of the Human Factors and Ergonomics Society Europe pp 63–72

Bos J, Damala D, Lewis C, Ganguly A, Turan O (2007) Susceptibility to seasickness. Ergonomics 50(6):890–901

Bos J, Hogervorst M, Munnoch K, Perrault D (2008) Human performance at sea assessed by dynamic visual acuity. In: ABCD Symposium'Human Performance in the Maritime Environment', 31 January 2008 Pacific 2008 International Maritime Conference, Sydney, Australia

Bos JE (2004) How motions make people sick such that they perform less: a model based approach. In: RTO AVT Symposium on" Habitability of Combat and Transport Vehicles: Noise, Vibration and Motion, vol 110, pp 1–11

Bos JE (2015) Less sickness with more motion and/or mental distraction. Journal of Vestibular Research 25(1):23–33

Bos JE, MacKinnon SN, Patterson A (2005) Motion sickness symptoms in a ship motion simulator: effects of inside, outside, and no view. Aviation, space, and environmental medicine 76(12):1111–1118

Bos JE, Diels C, Souman JL (2022) Beyond seasickness: a motivated call for a new motion sickness standard across motion environments. Vibration 5(4):755–769

Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry 25(1):49–59

Brietzke A (2023) Kinetose Als Merkmal Der Mensch-Fahrzeug-Interaktion. Springer

Brietzke A, Pham Xuan R, Dettmann A, Bullinger AC (2021) Influence of dynamic stimulation, visual perception and individual susceptibility to car sickness during controlled stop-and-go driving. Forschung im Ingenieurwesen 85(2):517–526

Castelhano MS, Rayner K (2023) Eye movements during reading, visual search, and scene perception: An overview. Cognitive and Cultural Influences on Eye Movements p 1

Choukèr A, Kaufmann I, Kreth S, Hauer D, Feuerecker M, Thieme D, Vogeser M, Thiel M, Schelling G (2010) Motion sickness, stress and the endocannabinoid system. PloS one 5(5):e10752

Czaban M, Himmels C (2025) Investigating simulator validity by using physiological and cognitive stress indicators. Available at SSRN 5153560

Dam A, Sanford C, Jeon M (2024) Verifying motion sickness induction in automated vehicles using motion-based driving simulators. International Journal of Human–Computer Interaction pp 1–19

Diels C, Bos JE (2016) Self-driving carsickness. Applied ergonomics 53:374–382

Diels C, Ukai K, Howarth PA (2007) Visually induced motion sickness with radial displays: effects of gaze angle and fixation. Aviation, space, and environmental medicine 78(7):659–665

Donders FC (1969) On the speed of mental processes. Acta psychologica 30:412–431

Elbanhawi M, Simic M, Jazar R (2015) In the passenger seat: investigating ride comfort measures in autonomous cars. IEEE Intelligent transportation systems magazine 7(3):4–17

Emond W, Zare M (2024) Motion sickness mitigation in road vehicles with a triad of focus on safety and sustainability. In: Technological Systems, Sustainability and Safety

Emond W, Bohrmann D, Zare M (2024) Will visual cues help alleviating motion sickness in automated cars? a review article. Ergonomics 67(6):772–800

Fagnant DJ, Kockelman K (2015) Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice 77:167–181

Farnebäck G (2003) Two-frame motion estimation based on polynomial expansion. In: Scandinavian conference on Image analysis, Springer, pp 363–370

Flanagan MB, May JG, Dobie TG (2004) The role of vection, eye movements and postural instability in the etiology of motion sickness. Journal of Vestibular Research 14(4):335–346

Flanagan MB, May JG, Dobie TG (2005) Sex differences in tolerance to visually-induced motion sickness. Aviation, space, and environmental medicine 76(7):642–646

Gianaros PJ, Muth ER, Mordkoff JT, Levine ME, Stern RM (2001) A questionnaire for the assessment of the multiple dimensions of motion sickness. Aviation, space, and environmental medicine 72(2):115

Golding JF, Kerguelen M (1992) A comparison of the nauseogenic potential of low-frequency vertical versus horizontal linear oscillation. Aviation, space, and environmental medicine 63(6):491–497

Harmankaya H, Brietzke A, Pham Xuan R, Shyrokau B, Happee R, Papaioannou G (2024) Efficient motion sickness assessment: Recreation of on-road driving on a compact test track. arXiv preprint arXiv:241214982

He X, Stapel J, Wang M, Happee R (2022) Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles. Transportation research part F: traffic psychology and behaviour 86:178–195

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 101:250–266

Hox J, Moerbeek M, Van de Schoot R (2017) Multilevel analysis: Techniques and applications. Routledge

Irmak T, De Winkel K, Pattanayak A, Happee R (2021a) Motion sickness, motivation, workload and task performance in automate d vehicles. In: Comfort congress

Irmak T, Pool DM, Happee R (2021b) Objective and subjective responses to motion sickness: the group and the individual. Experimental Brain Research 239(2):515–531

Irmak T, Kotian V, Happee R, de Winkel KN, Pool DM (2022) Amplitude and temporal dynamics of motion sickness. Frontiers in systems neuroscience 16:866503

Irmak T, de Winkel KN, Happee R (2025) Visually induced motion sickness correlates with on-road car sickness while performing a visual task. Experimental Brain Research 243(4):81

Isu N, Hasegawa T, Takeuchi I, Morimoto A (2014) Quantitative analysis of time-course development of motion sickness caused by in-vehicle video watching. Displays 35(2):90–97

Jain V, Kumar SS, Papaioannou G, Happee R, Shyrokau B (2023) Optimal trajectory planning for mitigated motion sickness: Simulator study assessment. IEEE Transactions on Intelligent Transportation Systems 24(10):10653–10664

Jones ML, Le VC, Ebert SM, Sienko KH, Reed MP, Sayer JR (2019) Motion sickness in passenger vehicles during test track operations. Ergonomics 62(10):1357–1371

Kantusch T (2023) Auswirkungen einer Kinetose auf die menschliche Leistung und Ermittlung von Gegenmaßnahmen in Bezug auf das automatisierte Fahren. Technische Universitaet Berlin (Germany)

Kaplan J, Ventura J, Bakshi A, Pierobon A, Lackner JR, DiZio P (2017) The influence of sleep deprivation and oscillating motion on sleepiness, motion sickness, and cognitive and motor performance. Autonomic neuroscience 202:86–96

Karjanto J, Yusof NM, Hassan MZ, Terken J, Delbressine F, Rauterberg M (2021) An on-road study in mitigating motion sickness when reading in automated driving. Journal of Hunan University Natural Sciences 48(3)

Kaufeld M, Bourdeinik J, Prinz LM, Mundt M, Hecht H (2022) Emotions are associated with the genesis of visually induced motion sickness in virtual reality. Experimental Brain Research 240(10):2757–2771

Kenesei Z, Kökény L, Ásványi K, Jászberényi M (2025) The central role of trust and perceived risk in the acceptance of autonomous vehicles in an integrated utaut model. European Transport Research Review 17(1):8

Khusro YR, Zheng Y, Grottoli M, Shyrokau B (2020) Mpc-based motion-cueing algorithm for a 6-dof driving simulator with actuator constraints. Vehicles 2(4):625–647

Kremer C, Tomzig M, Merkel N, Neukum A (2022) Using active seat belt retractions to mitigate motion sickness in automated driving. Vehicles 4(3):825–842

Kuiper OX, Bos J, Diels C (2018) Looking forward: In-vehicle auxiliary display positioning affects carsickness. Applied Ergonomics 68:169–175

Lentz JM, Collins WE (1977) Motion sickness susceptibility and related behavioral characteristics in men and women. Aviation, space, and environmental medicine 48(4):316–322

Lukacova I, Keshavarz B, Golding JF (2023) Measuring the susceptibility to visually induced motion sickness and its relationship with vertigo, dizziness, migraine, syncope and personality traits. Experimental Brain Research 241(5):1381–1391

Madigan R, Lee YM, Merat N, Goodridge C, Lehtonen E, Wolter S, Wilbrink M, Oehl M, Dozza M, Edelmann A, Happee R, Hennes N, Horn S, Maggi D, Merlhiot G, Metz B, Metzulat M, Nordhoff S, Peng C, Portouli E, Reher A, Schmidt V, Schrank A, Tango F, Wörle J (2023) Deliverable d4.4: User evaluation methods (version 1.0 draft). Tech. rep., Hi-Drive Project, URL `https://www.hi-drive.eu/app/uploads/2023/08/Hi-Drive-SP4-D4.4-User-Evaluation-Methods-v1.0_DRAFT_for_website.pdf`

Marberger C, Otto H, Schulz M, Alt P, Carlowitz S (2022) Questionnaire for ”automated ride comfort assessment (arca)” DOI 10.13140/RG.2.2.34833.71520

Mattes S, Hallén A (2009) Surrogate distraction measurement techniques: The lane change test

Metz B, Wörle J, Engel S (2025) Behavioural adaptation to prototype level 3 automated driving systems on public roads. Transportation Research Part F: Traffic Psychology and Behaviour 109:1507–1522

Metzulat M, Metz B, Landau A, Neukum A, Kunde W (2024) Does the visual input matter? influence of non-driving related tasks on car sickness in an open road setting. Transportation Research Part F: Traffic Psychology and Behaviour 104:234–248

Metzulat M, Metz B, Landau A, Neukum A, Kunde W (2025) Too sick to take over?- impact of car sickness on cognitive performance related to driving in the context of automated driving. Transportation Research Part F: Traffic Psychology and Behaviour 109:480–500

Morimoto A, Isu N, Ioku D, Asano H, Kawai A, Masui F (2008) Effects of reading books and watching movies on inducement of car sickness. In: Proceedings of the FISITA 2008 World Automotive Congress, Munich, Germany, pp 14–19

Mühlbacher D, Tomzig M, Reinmüller K, Rittger L (2020) Methodological considerations concerning motion sickness investigations during automated driving. Information 11(5):265

Nalivaiko E, Davis SL, Blackmore KL, Vakulin A, Nesbitt KV (2015) Cybersickness provoked by head-mounted display affects cutaneous vascular tone, heart rate and reaction time. Physiology & behavior 151:583–590

Nesbitt K, Davis S, Blackmore K, Nalivaiko E (2017) Correlating reaction time and nausea measures with traditional measures of cybersickness. Displays 48:1–8

Nooij SA, Pretto P, Oberfeld D, Hecht H, Bülthoff HH (2017) Vection is the main contributor to motion sickness induced by visual yaw rotation: Implications for conflict and eye movement theories. PloS one 12(4):e0175305

Nordhoff S, Stapel J, He X, Gentner A, Happee R (2021) Perceived safety and trust in sae level 2 partially automated cars: Results from an online questionnaire. Plos one 16(12):e0260953

Organization IS (1997) Mechanical vibration and shock—evaluation of human exposure to whole body vibration. part 1: General requirements. International Standard ISO 2631–1

Paddeu D, Parkhurst G, Shergold I (2020) Passenger comfort and trust on first-time use of a shared autonomous shuttle vehicle. Transportation Research Part C: Emerging Technologies 115:102604

Papaioannou G, Shen C, Rothhämel M, Happee R (2025) Occupants' comfort: what about human body dynamics in road and rail vehicles? Vehicle System Dynamics 63(7):1241–1299

Peng C, Carlowitz S, Madigan R, Marberger C, Lee JD, Krems J, Beggiato M, Romano R, Wei C, Wooldridge E, et al. (2024a) Conceptualising user comfort in automated driving: Findings from an expert group workshop. Transportation research interdisciplinary perspectives 24:101070

Peng C, Merat N, Romano R, Hajiseyedjavadi F, Paschalidis E, Wei C, Radhakrishnan V, Solernou A, Forster D, Boer E (2024b) Drivers' evaluation of different automated driving styles: Is it both comfortable and natural? Human factors p 00187208221113448

Peng C, Öztürk , Madigan R, Nordhoff S, Hoogendoorn-Lanser S, Hagenzieker M, Merat N (2025) How do passengers of automated vehicles experience comfort on road? an interview study. An Interview Study

Pham Xuan R (2023) Evaluation multimodaler physiologischer merkmale zur objektiven detektion von kinetose im pkw. PhD thesis

Smyth J, Birrell S, Mouzakitis A, Jennings P (2019a) Motion sickness and human performance–exploring the impact of driving simulator user trials. In: Advances in Human Aspects of Transportation: Proceedings of the AHFE 2018 International Conference on Human Factors in Transportation, July 21-25, 2018, Loews Sapphire Falls Resort at Universal Studios, Orlando, Florida, USA 9, Springer, pp 445–457

Smyth J, Birrell S, Mouzakitis A, Jennings P (2019b) Motion Sickness and Human Performance – Exploring the Impact of Driving Simulator User Trials. In: Advances in Intelligent Systems and Computing, Springer Verlag, vol 786, pp 445–457, DOI 10.1007/978-3-319-93885-1{_}40, URL `https://doi.org/10.1007/978-3-319-93885-1_40`

Stelling D, Hermes M, Huelmann G, Mittelstädt J, Niedermeier D, Schudlik K, Duda H (2021) Individual differences in the temporal progression of motion sickness and anxiety: the role of passengers' trait anxiety and motion sickness history. Ergonomics 64(8):1062–1071

Tajdari F, Messiou C, Happee R, Papaioannou G (2025) Exploring the impact of seat suspension with negative stiffness elements on occupants' motion comfort. Available at SSRN 5183381

Talsma TM, Hassanain O, Happee R, de Winkel KN (2023) Validation of a moving base driving simulator for motion sickness research. Applied Ergonomics 106:103897, DOI https://doi.org/10.1016/j.apergo.2022.103897, URL `https://www.sciencedirect.com/science/article/pii/S0003687022002204`

Tamura Y, Wada T, Liu H (2023) Generating visual information for motion sickness reduction using a computational model based on svc theory. In: 2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC), IEEE, pp 5066–5071

Tomzig M, Schoemig N, Wehner T, Marberger C, Otto H, Schulz M, Kenar E, Schultz A (2023) How to make reading in fully automated vehicles a better experience? effects of active seat belt retractions and a 2-step driving profile on subjective motion sickness, ride comfort and acceptance. In: Proceedings of the 15th international conference on automotive user interfaces and interactive vehicular applications, pp 11–21

Van Der Laan JD, Heino A, De Waard D (1997) A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies 5(1):1–10

Wertheim A, Bos J, Krul A (2001) Predicting motion induced vomitting from subjective misery (misc) rating obtained in 12 experimental studies

Wibirama S, Hamamoto K (2014) Investigation of visually induced motion sickness in dynamic 3d contents based on subjective judgment, heart rate variability, and depth gaze behavior. In: 2014 36th annual international conference of the IEEE engineering in medicine and biology society, IEEE, pp 4803–4806

A Appendix

A.1 Vehicle dynamics

Table 2: Driving Dynamics Metrics for Road and Track Conditions

	On-road			**Test-track**		
Variable	n	mean	sd	n	mean	sd
Velocity_Mean	47	25.26	2.06	47	8.33	0.62
Velocity_Max	47	79.85	7.17	47	38.37	2.14
Acceleration_Max	47	2.64	0.32	47	1.63	0.07
Deceleration_Mean	47	-0.43	0.04	47	-0.36	0.03
Deceleration_Min	47	-3.03	0.63	47	-1.94	0.30
Lat_Accel_Left_Mean	47	0.05	0.01	47	0.07	0.01
Lat_Accel_Left_Max	47	0.38	0.03	47	0.44	0.04
Lat_Accel_Right_Mean	47	-0.04	0.00	47	-0.03	0.00
Lat_Accel_Right_Max	47	-0.36	0.03	47	-0.34	0.02
Yaw_counter_Mean	47	3.09	0.17	47	6.79	0.13
Yaw_counter_Max	47	30.85	1.83	47	34.42	1.54
Yaw_counterclock_Mean	47	-1.96	0.18	47	-4.44	0.50
Yaw_counterclock_Max	47	-30.71	3.28	47	-40.21	2.23
MSDV_X_Max	47	27.53	2.14	47	26.26	3.31
MSDV_Y_Max	47	0.31	0.03	47	1.74	0.17

A.2 ARCA

Kendall-Tau-Correlation between the single *ARCA* items and subjective maximum Motion Sickness (MISC).

	Item	Estimate_{Road}	p.value_{Road}	$\text{Estimate}_{Test-Track}$	$\text{p.value}_{Test-Track}$
1	Safety	-0.18	0.14	-0.29	0.02
2	VehicleNaturalAppearance	-0.14	0.24	-0.21	0.12
3	SenseOfControl	-0.15	0.22	-0.37	0.01
4	EfficientTravel	-0.19	0.09	-0.35	0.00
5	Stress	-0.29	0.01	-0.31	0.01
6	PredictVehicleBehaviour	0.06	0.60	-0.10	0.38
7	TrustSystem	-0.22	0.08	-0.26	0.04
8	NDRT	-0.13	0.25	-0.29	0.01
9	GForcesBrakingAppropriate	-0.24	0.04	-0.21	0.08
10	GForcesAccelAppropriate	-0.21	0.07	-0.16	0.20
11	GForcesTurnAppropriate	-0.26	0.02	-0.30	0.01
12	BodyRecovery	-0.52	0.00	-0.44	0.00

A.3 Trust

Kendall-Tau-Correlation between the single *Trust* items and subjective maximum Motion Sickness (MISC).

	Item	Estimate_{Road}	p.value_{Road}	$\text{Estimate}_{TestTrack}$	$\text{p.value}_{TestTrack}$
1	LaneKeeping	-0.10	0.50	-0.47	0.00
2	Hesistant2UseSystem	0.12	0.43	-0.24	0.11
3	MaintainSpeedAndDistance	-0.09	0.54	-0.42	0.00
4	NotComfortableUsing	-0.25	0.08	-0.40	0.01
5	WouldTrust	-0.01	0.97	-0.27	0.07

A.4 Perceived Safety

Kendall-Tau-Correlation between the single *Perceived Safety* items and subjective maximum Motion Sickness (MISC).

	Item	Estimate_{Road}	p.value_{Road}	$\text{Estimate}_{TestTrack}$	$\text{p.value}_{TestTrack}$
1	FeelingSafe	-0.24	0.05	-0.25	0.05
2	FeelingComfortable	-0.42	0.00	-0.44	0.00
3	FeelingAnxious	-0.17	0.17	-0.25	0.05
4	FeelingAtRisk	-0.19	0.14	-0.12	0.37
5	FeelingInDanger	-	-	-0.09	0.49
6	FeelingRideSaferThanExpected	-0.00	0.98	-0.18	0.15
7	RecommendBecauseOfSafety	-0.08	0.60	-0.34	0.01

Note: After removal of outliers, all values of the variable *FeelingInDanger* in the condition Road equaled to 5 with no variance. Therefore, the correlation is undefined.