Driver's perceived risk in relation to automated vehicle behaviour

Evaluation and mitigation of perceived risk through simulator studies, computational models, and user interface design

He, X.

DOI
10.4233/uuid:14f15a1f-7f81-43b4-9d4c-8a28c4e683cf

Publication date
2024

Document Version
Final published version

Citation (APA)
He, X. (2024). *Driver's perceived risk in relation to automated vehicle behaviour: Evaluation and mitigation of perceived risk through simulator studies, computational models, and user interface design*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:14f15a1f-7f81-43b4-9d4c-8a28c4e683cf

DRIVER'S PERCEIVED RISK IN RELATION TO AUTOMATED VEHICLE BEHAVIOUR

EVALUATION AND MITIGATION OF PERCEIVED RISK THROUGH SIMULATOR STUDIES, COMPUTATIONAL MODELS AND USER INTERFACE DESIGN

Xiaolin He

DRIVER'S PERCEIVED RISK IN RELATION TO AUTOMATED VEHICLE BEHAVIOUR

Evaluation and Mitigation of Perceived Risk through Simulator Studies, Computational Models and User Interface Design

DRIVER'S PERCEIVED RISK IN RELATION TO AUTOMATED VEHICLE BEHAVIOUR

EVALUATION AND MITIGATION OF PERCEIVED RISK THROUGH SIMULATOR STUDIES, COMPUTATIONAL MODELS AND USER INTERFACE DESIGN

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on Wednesday 27 November 2024 at 12:30 pm

by

Xiaolin HE

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. R. Happee,	Technische Universiteit Delft, promotor
Prof. dr. M. Wang,	Technische Universität Dresden, promotor
Independent members:	
Prof. dr. G. Markkula	University of Leeds, UK
Prof. dr. J. Bärgman	Chalmers University of Technology, Sweden
Prof. dr. ir. J.C.F. de Winter	Technische Universiteit Delft
Prof. dr. ir. J.W.C. van Lint	Technische Universiteit Delft
Dr. ir. H. Farah	Technische Universiteit Delft

The work presented in this thesis was made possible by the SHAPE-IT project funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410. Additionally support for the simulator study and perceived risk model validation was provided by Toyota Motor Europe.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410

Keywords: Perceived risk, trust, automated driving, simulator studies, computational models, user interface design

Cover designed by: Weijian Xu

ISBN 978-94-6384-684-4

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

The real problem is
not whether machines think but whether men do.

CONTENTS

SUMMARY

Automated vehicles (AVs) represent a significant leap forward in transportation, aiming to enhance road safety, increase comfort, and improve traffic efficiency. As technology progresses to SAE Level 3 and higher, drivers are increasingly able to engage in non-driving-related activities. However, this (r)evolution raises challenges concerning driver's perceived risk and trust in AVs, which are crucial factors influencing the acceptance of AVs. This dissertation aims to enhance perceived safety and trust in AVs through experimental studies, computational modelling and user interfaces (UIs) design.

The initial phase of this dissertation focused on how drivers' perceived risk and trust when using AVs evolve in close encounters with other road users. We developed regression-based perceived risk and trust models based on a simulator study with 25 participants involving merging and hard braking scenarios on motorways. The proposed models reveal that perceived risk is dynamically influenced by driving conditions and sensitive to individual factors such as driving experience and gender with experienced and male drivers generally perceiving lower risk. Notably, a decrease in trust after high-risk encounters was observed, indicating a close relationship between perceived risk and trust in AVs. Additionally, physiological responses were observed as potential indicators of perceived risk in critical driving scenarios.

To develop a tool for gaining insights on perceived risk, we put forward a novel computational model called potential collision avoidance difficulty (PCAD) model. Drawing inspiration from Fuller's Risk Allostasis Theory and the looming phenomenon, PCAD evaluates the difficulty of avoiding potential collisions by calculating minimal control effort through braking or/and steering needed to navigate safely. By integrating visual looming, factors in the uncertain behaviour of surrounding vehicles, control inaccuracies of the subject vehicle, and potential collision severity, PCAD provided an accurate population-level fitting of perceived risk in our own dataset on highway merging and a published dataset on obstacle avoidance. The findings highlight the need to account for both the longitudinal and lateral dimensions of driving condition, and uncertain behaviours of surrounding vehicles when interpreting perceived risk.

Further exploration of perceived risk was achieved through the creation of a large-scale dataset of perceived risk using an online survey. This new dataset provided time-continuous perceived risk in dynamic driving conditions. A total of 105 events was created including merging, hard braking and lane changes on motorways, while systematically varying multiple control parameters (such as relative speed and distance) to achieve different levels of event criticalities. Deep neural networks (DNNs) were then trained on this dataset to fit perceived risk, and SHapley Additive exPlanations (SHAP) was used to identify the key contributors to perceived risk in the continuous time domain. Aligned with the PCAD model developed previously, the results highlighted the importance of the relative motion information, particularly the distance to other road users and the uncertainty of surrounding vehicle behaviour in shaping perceived risk.

This approach not only discerns the dynamics of perceived risk by systematically analysing interactions with other road users but also provides a guide for future modelling of perceived risk. The development of this extensive dataset fills the gap by providing the lacking continuous perceived risk data, thereby supporting further research on perceived risk.

The last contribution of this dissertation was on enhancing perceived safety and trust through optimised design of UIs. A simulator experiment demonstrated that multimodal UIs incorporating both visual and auditory modalities enhanced perceived safety and trust the most. Manoeuvre information delivered through the auditory modality was particularly effective in enhancing trust and acceptance. The findings indicate the benefits of the UIs in enhancing perceived safety and trust but also showed the limitations of using UIs alone during highly critical events. This part of the work suggests that the design of UIs for partially automated vehicles shall include automation information via visual and auditory modalities to enhance perceived safety and trust.

This dissertation makes several contributions to the field of perceived risk research in AVs. First, it provides foundational insights into perceived risk, demonstrating the significant influences of driving conditions, manoeuvre uncertainties and individual personal characteristics. The computational perceived risk models demonstrate strong predictive power in perceived risk and offer a deep understanding of how perceived risk is shaped in dynamic driving conditions. Additionally, the rich dataset obtained in this dissertation, which includes event-based discrete data and time-continuous data on perceived risk, serves as a new and open resource for future perceived risk research. Lastly, the practical evaluation of the design of UI provided actionable recommendations in enhancing trust and perceived safety, particularly through manoeuvre information delivered using auditory modality. These contributions advance the understanding, modelling, and practical application of perceived risk in automated driving environments, supporting the broader acceptance and integration of AVs.

The dissertation presents various opportunities for the advancement of AV technology and its integration with human factors. Building on the comprehensive datasets, computational models and insights gained in this dissertation, future studies should focus on further refining computational models to capture perceived risk in general scenarios. Expanding data collection efforts to include on-road tests, and more diverse participants will also enhance the generalisability of the findings. Additionally, the design of adaptive UIs that fit individual preferences remains a promising direction for future research.

SUMMARY IN CHINESE

概述

自动驾驶车辆（AVs）的问世代表了交通领域的一次重要进步，其目标是改善道路安全、增加舒适性和提升交通效率。随着技术逐步发展至 SAE 3 级及以上，驾驶员能够越来越多地参与非驾驶相关任务，这改变了驾驶员的传统角色。然而，这种变化也引发了有关驾驶员在自动驾驶车辆的感知风险和信任度方面的挑战，而这两个因素对自动驾驶车辆的接受度具有重要影响。本文旨在通过实验研究、计算建模以及用户界面（UIs）设计来提升对自动驾驶车辆的感知安全性和信任度。

本文在初始阶段探究了当使用自动驾驶车辆时，驾驶员在不同驾驶场景中对自动驾驶车辆行为的感知风险和信任度的变化。基于一个 25 名被试者参与的高速公路并线和急刹场景的驾驶模拟器实验，我们建立了感知风险和信任的回归模型。此模型表明感知风险会受到驾驶环境条件的动态影响，并对驾驶经验和性别等个体因素高度敏感，其中更有经验的驾驶员和男性驾驶员感知到的风险普遍较低。值得注意的是，驾驶员的信任水平在遭遇高风险事件之后发生了显著的下降，表明感知风险与对自动驾驶车辆的信任之间存在密切关系。此外，一些生理信号能够潜在地反映感知风险，但是仅限在高危险场景中。

本文构建了一种新的计算模型——潜在碰撞规避难度（PCAD）模型，以深入理解感知风险。该模型从 Fuller 的风险稳态理论和视觉扩张出发，通过计算使用制动和/或转向来安全地规避潜在碰撞所需的最小控制投入，从而评估规避潜在碰撞的难度。通过整合视觉扩张，模型考虑了周围车辆的不确定行为、主车的控制误差以及潜在碰撞的严重程度。在我们创建的高速公路并线和制动感知风险数据集以及公开的避障感知风险数据集中，PCAD 能够在群体水平上准确拟合感知风险。研究结果强调了在解释感知风险时，纵向和横向驾驶条件以及周围车辆的不确定行为是非常重要的。

本文继而通过一个线上研究创建的大规模感知风险数据集进一步探究了感知风险。该全新感知风险数据集提供了动态驾驶条件下的连续感知风险数据。在此线上研究中，我们一共创建了在高速公路场景下包括并线，强制动和变道情况在内的 105 个事件，同时系统地改变多个控制参数（如相对速度和距离），以实现不同的危险程度。随后，在该数据集上训练了深度神经网络（DNNs）来拟合感知风险，并采用 SHapley Additive exPlanations（SHAP）来识别在连续时间域内感知风险的关键影响因素。与之前开发的 PCAD 模型一致，分析结果强调了相对运动信息，尤其是自车与其他道路使用者的距离以及周围车辆行为的不确定性对于感知风险拟合的重要性。这一方法不仅通过系统分析与其他道路使用者的互动来揭示感知风险的动态变化，还为未来感知风险的建模提供了指导。本文构建的这一大型数据集填补了时间连续感知风险数据的空白，支持了对感知风险的进一步研究。

本文的最后一个重要贡献是通过优化用户界面的设计来提升驾驶员的感知安全和信任度。另一个驾驶模拟器实验表明，结合视觉和听觉的多模态用户界面在增强感知安全和信任度方面效果最佳。通过听觉模态传递的自车短期操控信息在提升信任度和接受度方面尤为有效。研究结果显示，用户界面在提升驾驶员感知安全性和信任度方面具有显著优势，但是在高风险事件中，仅依赖用户界面也存在一定的局限性。本文这一部分表明，有条件自动驾驶车辆的用户界面设计应包含通过视觉和听觉模态传递的自动化信息，以增强驾驶员的感知安全和信任度。

本文对自动驾驶车辆中感知风险的研究领域作出了多项贡献。首先，本文提供了感知风险的基础性见解，证明了驾驶条件、操控不确定性和个体因素对感知风险的显著影响。感知风险的计算模型展现了强大的拟合能力，并深入揭示了感知风险在动态驾驶条件下的形成机制。此外，本文获得的大量感知风险数据，包括基于事件的离散数据和基于动态驾驶环境的时间连续数据，为未来的感知风险研究提供了数据支持。最后，用户界面设计的研究为如何提升驾驶员的感知安全和信任度提供了可行的建议，特别是在通过听觉模态传递自车短期操控信息方面。这些贡献推动了对自动驾驶环境中感知风险的理解、建模和实际应用，进而支持了自动驾驶车辆被更广泛地接受。

本文为自动驾驶技术的发展及其与人因工程的融合提供了多种可能性。基于本文中获得的感知风险数据集、计算模型和分析结果，未来的研究应着重进一步优化计算模型，以在更广泛的场景中预测感知风险。此外，扩展数据收集工作，包括增加实际道路测试和涉及更加多样化的参与者，将有助于提升研究结果的普适性。最后，设计适应个体偏好的自适应用户界面也是未来研究中的一个重要方向。

1

INTRODUCTION

1.1. BACKGROUND

THE advent of automated vehicles (AVs) marks a significant shift in transportation, where the principles of road safety, efficiency, and passenger comfort are being redefined [1], [2]. For the general public, the concept of AVs brings up images of cars that possess the ability to drive themselves, not only promising to deliver passengers safely to their destination, but also affording them the opportunity to engage in other activities during the journey. This optimistic vision is frequently reinforced by the media celebrating the successes of AV technology [3]–[5]. However, this enthusiasm is tempered by safety concerns arising from reports of accidents involving automated vehicles, creating a division in public perception [6]–[8]. Such incidents not only highlight the technological obstacles that AVs face, but also emphasise the critical human factors involved – particularly the reactions of drivers and passengers to AV's presence on the road [9], [10]. This subjective assessment, known as perceived risk, affects individual choices and reflects a major challenge in the advancement of AVs [11]–[14]. While ISO 26262 defines risk as a combination of the likelihood and severity of harm [15], perceived risk is a subjective measure influenced by individual encounters, system performance, and the surrounding environmental context, capturing how users personally perceive risk in automated vehicle scenarios, beyond just the objective metrics of severity and likelihood. This subjective feeling of risk is critical to shaping the willingness of users to engage with AVs, consequently influencing whether AVs can fulfil their promise of significantly reducing traffic accidents and congestion [9], [12], [13], [16]–[20].

In discussions around perceived risk of AVs, trust emerges as a critical concept to public acceptance of automated vehicles as evidenced by many studies [9], [13], [21]–[24]. Among various definitions [25]–[29], the most widely referenced in AV trust research frames trust as 'the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the trustor, regardless of the ability to monitor or control that other party' [30]. This perspective emphasises that trust is largely constructed from perceptions, thoughts, and

emotions [28]. Understanding the complexities of trust is essential for the development and broader acceptance of automated vehicles, which can guide strategies to enhance user confidence and facilitate a smoother integration of AV technology into society.

Studies typically explore perceived risk and trust through various dimensions, including their assessment, modelling, and strategies to reduce perceived risk and improve trust. In the ensuring, we first review the three aspects to position the thesis in the state of the art.

1.1.1. ASSESSING PERCEIVED RISK AND TRUST IN AVS

Numerous studies aimed to evaluate perceived risk and trust in AVs primarily using questionnaires to assess the general public opinion, leading to research that focuses on how general personal characteristics such as age, gender, and driving experience influence perceived risk and trust [16], [20], [21], [31]–[33]. However, this approach does not capture the variability of perceived risk and trust in dynamic driving scenarios. Some studies surveyed participants regarding perceived risk and trust after actual rides in specific driving scenarios [9], [14], [34], [35]. While they can indicate the changes after experiencing the AV, their analyses of perceived risk generally do not connect changes in perceived risk directly to dynamic driving situations.

To address these limitations, there have been attempts to capture real-time responses during driving or by referencing specific video frames [36]–[38]. However, these efforts tend to focus on isolated incidents rather than providing a continuous and dynamic understanding of perceived risk.

1.1.2. MODELLING PERCEIVED RISK

Modelling perceived risk in AVs involves exploring models that illustrate the connection between human feelings and AVs, in particular regarding their interaction with other road users. Modelling efforts have greatly improved our understanding of how people evaluate and respond to the risks linked to AVs. The application of Surrogate Measures of Safety (SMoS), such as Minimum Time to Collision (TTC), illustrates the complex interplay between operational and perceived risk [39], [40]. These measures can, to some extent, reflect perceived risk during driving, although the relationship between SMoS and perceived risk is not yet well established.

Advancements in data-driven approaches have further enriched computational perceived risk modelling. The driving risk field (DRF) model [41] is an example of data-driven models based on the foundational concept of the "field of safe travel" [42]. This model has made progress in capturing perceived risk continuously across spatial and temporal dimensions. Similarly, the study by Tzouras *et al.* [43] takes a focused approach by developing a regression model of perceived risk for tram drivers. This model includes factors such as the presence and type of pedestrian crossings, the existence of tram stops, and the density of vulnerable road users in the environment. Extending beyond these specific models, there has been a significant increase in the use of data-driven methods in this field, including probabilistic methods, regression analyses, and advanced machine learning strategies, particularly using computer vision. These methods are applied to predict perceived risk in specific driving contexts [37], [38], [44], [45].

Although data-driven models provide quantification of perceived risk in certain sce-

narios, their validity across diverse situations is uncertain and often lacks complete explainability. In contrast, physics-based models, which rely on the fundamental laws of physics, are more explainable in computing perceived risk, as they derive from physical principles. Such models are based on physical variables that make the model predictions transferable and interpretable. Additionally, such models can obtain new insights of perceived risk because of their transferability. Despite these strengths, physics-based models for perceived risk specifically remain largely unexplored. This gap highlights the need for developing physics-based computational perceived risk models to better understand perceived risk in automated driving.

1.1.3. REDUCING PERCEIVED RISK AND ENHANCING TRUST

Reducing perceived risk and enhancing trust in AVs have emerged as new focuses of research. Design strategies for AV motion planning and control have been developed to align AVs operations closely with human expectations and comfort levels. For instance, Tusseyeva *et al.* [46] investigated how different motion planning algorithms affect perceived safety in human-robot interaction, offering insights into how AV behaviours can be optimised for human safety perceptions. Kolekar *et al.* [41] demonstrated that a driver model based on perceived risk shows human-like behaviours when driving. Similarly, Sheng *et al.* [47] introduced a novel approach to route planning in AVs for long trips by integrating human trust as a critical factor.

Parallel to AV behavioural adjustments, human machine interface (HMI) design is crucial to improve perceived safety, trust, and acceptance of AVs. Basantis *et al.* [48] compared four different interfaces in the rear seat of AVs and found that interfaces featuring auditory notifications of manoeuvres significantly enhanced perceived safety compared to those providing only visual information about vehicle paths. Despite these findings, there remains a gap in research concerning the systematic evaluation of combined modalities and information types. Additionally, augmented reality displays and auditory notifications about vehicle manoeuvres have been used to comfort users about the AV's capabilities and intentions [49], [50]. Although these studies stress the importance of HMI in improving perceived safety and trust, they simply examined the impact of these information types in isolation or did not systematically evaluate the combined effects of different modalities (visual and auditory) on perceived safety and trust.

1.2. RESEARCH GAPS

The aforementioned literature overview identified several research gaps that underline the challenges in fully understanding and addressing perceived risk in AVs:

(1) Specificity and dynamics in perceived risk and trust evaluation: Existing research often relies on generalised opinions through static methodologies, such as surveys, to evaluate perceived risk and trust, which does not adequately link these assessments to specific AV behaviours. Particularly in the case of perceived risk, this method fails to consider the dynamic aspect of how perceived risk changes in reaction to real-time AV behaviours. There is a crucial need for methodologies that not only connect perceived risk directly with specific AV actions but also capture the dynamic factors influencing perceived risk in real-time, providing a more accurate understanding of user

perceptions.

(2) Comprehensive Models for Perceived Risk: The field lacks precise and explainable models that are applicable across diverse AV scenarios. Exploring the potential of merging the predictive power of data-driven approaches with the clarity of physics-based models remains an open question and an important area for further research. Such a model would serve as a vital tool for both predicting perceived risk and enhancing the transparency necessary for wider acceptance and understanding of AV technologies.

(3) HMI designs for enhancing perceived safety and trust: Research on HMI design typically examines the impact of individual modalities independently. There exists a significant gap in exploring how different modalities can be systematically combined to improve perceived safety and trust. Systematic studies are required to assess the integrated effects of visual and auditory information delivered through HMIs.

1.3. OBJECTIVES

To address the research gaps, this dissertation is dedicated to investigating perceived risk in relation to the behaviours of AVs. The **main objective** of this dissertation is to develop quantitative methods to measure and understand perceived risk in automated vehicles, with an emphasis on the influence of AV behaviours and dynamic driving conditions.

To achieve this main objective, the following sub-objectives have been identified:

- **Objective 1: To collect data and gain insights on perceived risk and trust based on specific AV behaviours (Chapter 2, Chapter 4)**
  Quantitative data and comprehensive insights on perceived risk and trust, as well as the continuous dynamics of perceived risk over time, in various driving scenarios. This objective establishes the foundational data necessary for subsequent analytical and modelling efforts, focusing on how perceived risk and trust are influenced by specific AV behaviours.

- **Objective 2: To develop computational models for interpreting perceived risk and trust (Chapter 2, 3, and 4)**
  Development and refinement of computational models of perceived risk and trust. To achieve this, regression models, physics-based models, and neural networks for perceived risk and trust will be developed. This objective focuses on enhancing the accuracy and explainability of perceived risk and trust predictions across diverse AV scenarios.

- **Objective 3: To assess the impact of HMI modalities and information types on perceived risk and trust (Chapter 5)**
  Systematic assessment of different HMI modalities and the information types they convey, evaluating their effects on perceived safety and trust in AVs. This research aims to identify optimal combinations of HMI modalities that enhance user comfort and confidence.

While the primary focus is on perceived risk, the examination of trust is also included as it naturally intersects with perceived risk during initial evaluations and in studies involving HMI. This inclusion acknowledges the interconnection of trust and perceived risk, essential for understanding user acceptance and perceived risk of AVs.

1.4. CONTRIBUTIONS

The research described in this dissertation contributes to multiple aspects in the field of perceived risk and trust in AVs, which are categorised into scientific and practical contributions.

1.4.1. SCIENTIFIC CONTRIBUTIONS

The scientific contributions of the thesis include:

- **Large-scale datasets of perceived risk and trust:** We created two significant datasets: one dataset containing 500 perceived risk ratings and 500 trust ratings from 25 participants in merging and hard braking on motorway during automated driving through a simulator study; One comprehensive dataset with over 140,000 perceived risk ratings from more than 2,100 participants, recording various driving scenarios on motorways. These datasets form the foundation for the analysis in this dissertation and are essential for future empirical studies on perceived risk and trust in AVs.

- **New insights on perceived risk and trust in relation to AV behaviours:** We quantified the influence of factors such as neighbouring road users' behaviours (relative motion), driving experience, and gender on perceived risk and trust. Such quantitative analysis reveals the dynamic nature of perceived risk in AVs, highlighting varying contributions of different factors to perceived risk over time and emphasising the importance of manoeuvre uncertainties. These insights provide a deeper understanding of perceived risk and trust, which are crucial for future computational modelling efforts.

- **Three computational models of perceived risk and trust:** We developed two computational models for perceived risk and trust. The first one is a one-dimensional regression model to predict perceived risk and trust in merging and hard braking events. The second one is a novel two-dimensional computational perceived risk model based on potential collision avoidance difficulty (PCAD), which outperforms three well-established models across two different datasets (Dataset Merging and Dataset Obstacle Avoidance). We further designed and trained a deep neural network-based model capable of predicting perceived risk in various motorway driving scenarios, advancing theoretical understanding of perceived risk modelling.

- **Impact of UI modalities and information type on perceived safety, trust and acceptance:** We systematically assessed the effects of different UI modalities and the types of information they convey on perceived safety, trust and acceptance of AVs. We identified optimal combinations of UI modalities that enhance user comfort and confidence, providing guidelines for future UI and interaction design in AVs.

1.4.2. PRACTICAL CONTRIBUTIONS

- **Application of regression models in AV design:** The insight from the regression models provides practical guidance for automotive designers and engineers, help-

ing to design AV system behaviour that aligns with user expectations and enhances safety and trust.

- **Application of PCAD model in AV design:** The PCAD model offers a practical tool for controller, path planner or decision-making module design to align with user expectations and enhance perceived safety.
- **User interface design for enhanced perceived safety, trust and acceptance:** The study on different UI modalities and information types has practical implications for designing user interfaces in AVs. The findings help optimise UI designs to enhance drivers' perceived safety, trust and acceptance of partially automated vehicles, informing industry best practices.

1.5. THESIS STRUCTURE

This thesis journeys through the evolving understanding and modelling of perceived risk and trust in automated vehicles (AVs), crucial for their acceptance and successful integration as shown in Figure 1.1.

The exploration begins with an assessment of how occupants' perceived risk and trust change in response to the behaviour of AVs in Chapter 2. A simulator study with 25 participants was conducted to experience simulation scenarios of merging and hard braking on motorways, observing changes in perceived risk and trust through both verbal assessments and physiological measures. Chapter 2 develops a one-dimensional perceived risk model, conceptualised as a regression model, which reveals that perceived risk dynamically varies with the criticality of driving situations in the longitudinal direction. The study also identifies that personal characteristics, such as driving experience and gender, influence perceived risk and trust. Notably, a decrease in trust after high-risk encounters was observed, indicating a close relationship between perceived risk and trust in AVs. Furthermore, Chapter 2 highlights physiological responses as potential effective indicators of perceived risk in critical driving scenarios.

Chapter 3 presents a novel computational model for assessing perceived risk in AVs, named the Potential Collision Avoidance Difficulty (PCAD) model. Drawing inspiration from Fuller's Risk Allostasis Theory and the looming phenomenon, PCAD quantifies the minimal control effort needed through braking or/and steering needed to avoid potential collisions, factoring in the visual looming, the unpredictable behaviour of surrounding vehicles, control inaccuracies of the subject vehicle, and potential collision severity. The process of developing the model is clearly explained, including its theoretical foundations and the rigorous validation method used. PCAD's validation uses two distinct datasets: Dataset Merging (collected in Chapter 2) and Dataset Obstacle Avoidance sourced from existing literature. The comparative analysis reveals PCAD's performance over three state-of-the-art models, indicating its enhanced precision in predicting human drivers' perceived risk in merging and obstacle avoidance scenarios.

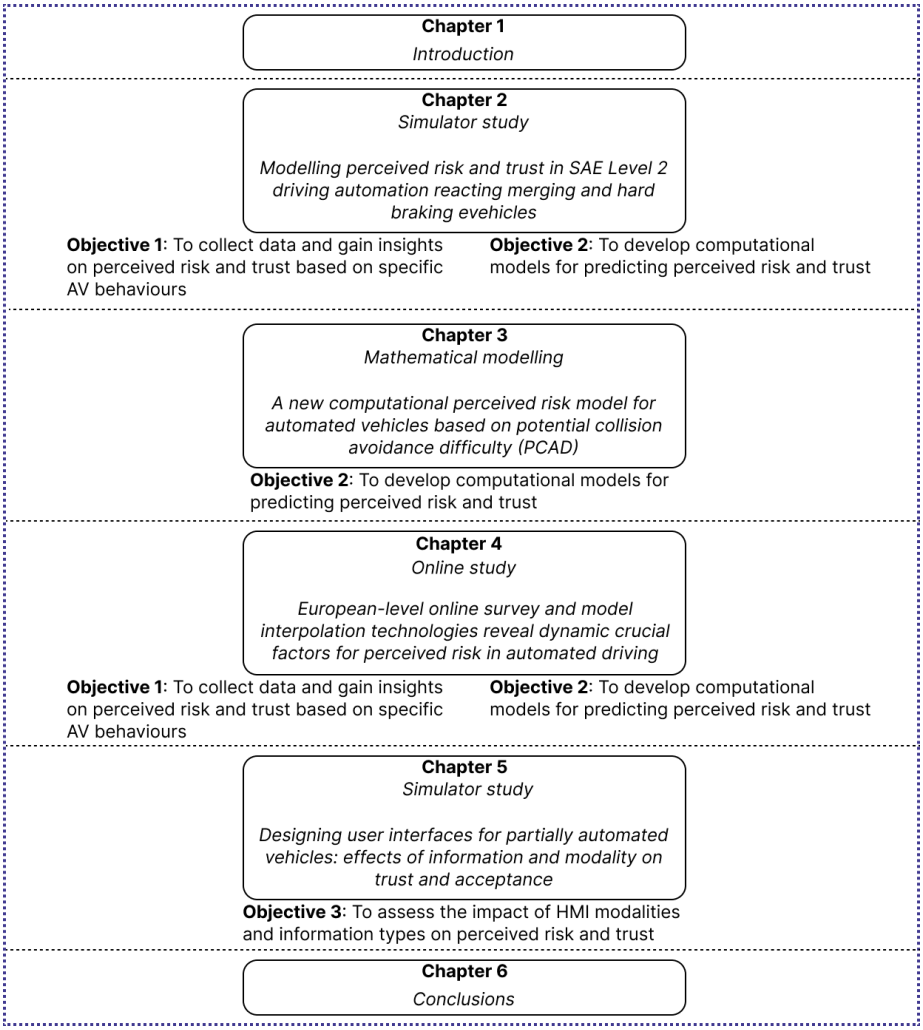


Figure 1.1: Structure of the thesis

A novel computational model of perceived risk was built in Chapter 3. However, the lack of comprehensive data for validation across various driving scenarios and in a continuous time domain leads to the motivation behind Chapter 4, which aims to construct an extensive perceived risk dataset that is continuous in the time domain and illustrates the dynamic nature of perceived risk in diverse automated driving conditions using explainable AI technologies. Through an extensive online survey, over 180,000 perceived risk ratings were collected from 2,164 participants across four common motorway scenarios. This effort involved adjusting behaviours of neighbouring and subject vehicles, with each scenario documented via video recordings and kinematic data to track perceived risk over time. This chapter expanded upon this dataset by calibrating two models, the PCAD from Chapter 3 and the DRF from existing literature, and training deep

neural networks for accurate perceived risk prediction, employing SHapley Additive exPlanations (SHAP) for input analysis. The findings reveal the dynamic aspects of perceived risk, particularly the impact of vehicle proximity and manoeuvre uncertainties, setting a foundation for future risk modelling and enhancing automated vehicle integration by understanding human perceived risk dynamics.

Building on previous work on perceived risk and trust in automated vehicles, Chapter 5 explores the use of user interfaces (UIs) to enhance perceived safety and trust in partially automated vehicles (PAVs). This chapter details a simulator study building upon Chapter 2 that tests various UI complexities to determine their effect on drivers' trust and perceived risk. By varying the type of automation information (surrounding versus manoeuvre) and the modality of that information (visual versus visual-auditory), the study aims to identify UI configurations that enhance perceived safety and trust among drivers. The results highlight the importance of multimodal information delivery in increasing trust and acceptance of PAVs, particularly through UIs that combine visual and auditory feedback with comprehensive automation details. While pointing out the effectiveness of such UIs in improving driver-automation interaction, the chapter also notes the challenges in completely reassuring drivers in critical scenarios, emphasising the need for continuous UI optimisation.

2

MODELLING PERCEIVED RISK AND TRUST IN DRIVING AUTOMATION REACTING TO MERGING AND BRAKING VEHICLES

This chapter investigates predictors of perceived risk and trust in driving automation, using a driving simulator experiment with 25 participants. These participants monitored SAE Level 2 driving automation (ACC+LC) and experienced various merging and hard braking events on a motorway. Their perceived risk and trust were measured verbally and through physiological indicators like pupil diameter and ECG signals. The findings reveal that relative motion with other road users is a primary factor influencing perceived risk and trust. Interestingly, the type of event (hard braking with or without merging) did not significantly alter perceptions. Repeated exposure to events increased trust in automation. Additionally, experienced drivers showed less sensitivity to risk, and females perceived more risk than males. There is a strong correlation between perceived risk and trust, with both sharing similar influencing factors. The study also finds that continuous perceived risk aligns with verbal risk assessments post-event, and braking behaviour is a key indicator of high perceived risk and low trust. Pupil diameter was particularly correlated with perceived risk in the most critical events, whereas heart rate increased during events but did not correlate with their criticality. These insights into the dynamics of perceived risk and trust can inform the design of human-centred automated driving systems, aiming to reduce perceived risk and enhance trust.

The content of this chapter has been published in
He, X., Stapel, J., Wang, M., & Happee, R. (2022).“Modelling perceived risk and trust in driving automation reacting to merging and braking vehicle,” *Transportation Research Part F: Traffic Psychology and Behaviour, 86, 178–195.*

2.1. INTRODUCTION

AUTOMATED vehicles (AV) have the potential to improve safety and comfort and reduce congestion [1]. Current production vehicles support SAE Level 1 or Level 2 driving automation systems with Adaptive Cruise Control (ACC) and Lane Centring (LC). Such systems still require continuous supervision by drivers to guarantee safety. Higher automation levels (SAE Level 3+) gradually allow drivers to shift attention away from dynamic driving tasks but requires them to be fallback-ready in case of automation failure [51]. Driving automation changes the driver's perceived risk and trust, influencing acceptance of driving automation [11], [20]. If the public does not widely accept driving automation, road safety and traffic benefits will not accrue [16], [52]. Therefore, understanding perceived risk and trust in driving automation is of great importance. Below, we review existing perceived risk and trust studies to synthesise definitions, influencing factors, and measurement methods for the current paper.

2.1.1. THEORIES OF PERCEIVED RISK AND TRUST

Perceived risk, or perceived safety, captures the level of risk experienced by users of driving automation. It can differ from operational (or actual) risk [36], [53], which is defined as the combination of accident probability and severity [15]. A low perceived risk leads to feeling relaxed, safe, and comfortable [9], [31], while a high risk perception results in cautious behaviour [53]. Perceived risk is highly individualised and is influenced by personal experience, personality, and attitudes [33], [38]. The driving environment, such as urban or rural roads, influences both operational and perceived risk [54].

Trust has been studied in psychology, sociology, and human factors. A survey by Kaplan *et al.* [55] listed 18 commonly-cited but distinct definitions of trust. Although definitions vary, there are three essential factors in trust: risk of losing or not gaining; uncertain outcomes and interdependence between trustor and trustee. The most cited trust definition is 'the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other party will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party' [30], which is also applicable to driving automation. Subjective thinking, feeling, and emotions account for most trusting behaviour [28]. The process of trust calibration is dynamic with a high degree of volatility [28]. Kraus *et al.* [56] proposed a dynamic trust calibration model based on the theory of Lee and See [28], and Hoff and Bashir [29], where initial learned trust is formed before the actual interaction with a system and is related to personality, provided information and driving experience. Dynamic learned trust evolves through interacting with the system and is influenced by experienced performance in similar situations and can be moderated by expectations and presentation (e.g. interface design). Trust factors can be divided into three broad categories: human-related, system-related and environment-related factors [29], [57]–[59]. Surveys and interviews demonstrated that trust is affected by personal characteristics (e.g., culture, age, and gender), external situations (e.g., task difficulty, traffic), internal situations (e.g., perceived risk, mood), driver's experience and system's performance. Kraus *et al.* [60] developed a personality model for trust in automation, indicating that a priori acceptability of automated driving is positively related to trust. Increased trust with experience is reported by Gold *et al.* [61] for driving automation and Kaplan *et al.* [59] for artificial intelligence system operators.

Exploring the influencers of perceived risk and trust in driving automation is challenging since most people do not have experience in driving automation, making it difficult to reveal user preferences based on empirics. Recent field experiments showed that real driving automation experience led to improvements in trust Xu *et al.* [9] and Walker *et al.* [62]. However, further work is needed to study the influence of driving scenarios and personal characteristics on perceived risk and trust.

Several studies reported highly negative correlations between perceived risk and trust in driving automation. Some studies regard perceived risk as the antecedent of trust [11], [35]: where perceived risk is firstly influenced by traffic condition, age, gender, etc., upon which trust is then built. In another school of thought, perceived risk is treated as the descendant of trust [9], [20], [21] where trust is firstly formed and then influences perceived risk. The relation between perceived risk and trust is also considered a mutual interaction [63], [64]. Existing studies include conceptual models, surveys and post-experiment questionnaires. However, the relation between perceived risk and trust still needs further investigation using actual driving automation.

MEASUREMENTS OF PERCEIVED RISK AND TRUST

Questionnaires are widely used to measure perceived risk, trust and other psychological constructs [9], [58], [65], [66]. Post-test questionnaires mainly reflect recent experience, and responses may be biased towards social norms. Besides, questionnaires increase participants' mental workload and are not suitable for real-time measurement. To overcome these drawbacks, continuous measurement devices are considered. Researchers have used handset controls [67], [68], sliders [69], rotary bars [70] and angle sensors within the steering wheel [36] for the continuous rating of perceived risk, trust and other subjective items.

Behaviour indicators (e.g., braking) have the potential to reflect automation reliance and compliance. Tenhundfeld *et al.* [34] used intervention by braking as an indicator of distrust in automated parking in a Tesla Model X. Naturalistic braking profiles have also been used to cluster near-miss events to different risk levels [71]. Physiological indicators (heart activity, skin response, etc.) are widely studied for non-intrusive continuous state assessment. Taylor [72] measured skin conductance (galvanic skin response; GSR) as an index of perceived risk in various traffic situations. GSR rates were 50 times higher during driving than during quiet sitting but were not correlated with actual traffic conditions (e.g. day-time off-peak, night off-peak). Morris *et al.* [73] compared driver's trust to GSR in different automation driving modes, finding lower skin conductance in safe driving modes. Ajenaghughrure *et al.* [74] identify brain activity (EEG) and gaze as the most robust indicators of trust in driving automation.

Pupil diameter is a practical tool to investigate perceived risk and trust in AV. For perceived risk, Tang *et al.* [75] showed that the change rate of pupil diameter is significantly higher in severe crashes than in minor crashes. For trust, Perello-March *et al.* [76] proposed to use increased and decreased pupil size as potential indicators to classify users' distrust and appropriate trust in driving automation. Therefore, we expect pupil diameter to reflect perceived risk and trust in driving automation.

MODELLING PERCEIVED RISK AND TRUST

Following the principle of human-centred design, modelling perceived risk in AV has gained attention. Varotto *et al.* [77] investigated perceived risk in full range ACC. They proposed a decision model where the driver would choose to deactivate ACC if the perceived risk becomes unacceptable based on the risk allostasis theory [78]. Kolekar *et al.* [36] presented the driving risk field capturing perceived risk continuously in the spatial and time domain based on driver's verbal ratings and steer response to obstacles in manual simulator driving. Surrogate Measures of Safety (SMoS) evaluate operational risk in terms of event criticality and can provide a basis to estimate perceived risk. Minimum time to collision (TTC) can show the driver's acceptance threshold of perceived risk when they take actions (e.g. braking) [39]. The inverse of TTC represents the relative visual expansion of the obstacle, which is referred to as looming [79]. Hence, SMoS reflect event criticality and the related perceived risk, but this relationship needs experimental support.

Trust in AV has been captured with conceptual models with various structures [21], [28], [30]. Marsh and Dibben [80] identified three layers of trust in information science: dispositional, situational and learned trust. Hoff and Bashir [29] extended this to trust in AV and considered different personal characteristics (e.g. culture, personality traits and mood) in the three layers. Empirical models have also been developed. Kraus *et al.* [56] investigated the dynamic process of trust calibration in partial driving automation and high driving automation, demonstrating that trust increased along with knowledge accumulation. [34] reported that drivers trust automation more and intervene less frequently when using driving automation more. Hu and Wang [81] proposed a prescribed-performance control barrier function with a dynamic model of trust in ACC, where the human will hand over the control to ACC if the system's performance reaches a certain threshold.

OBJECTIVES

Based on the discussions above, we identified the following research gaps. Firstly, a quantitative model between event criticality and perceived risk and trust is still lacking. Secondly, personal characteristics (e.g. age, driving experience) influence perceived risk and trust, but their impact has not been fully quantified in dynamic driving. Thirdly, the relations between perceived risk and trust are primarily derived from surveys and post-experiment questionnaires but hardly investigated in dynamic driving. Lastly, trust in automation has been studied mainly through surveys regarding higher automation levels (Level 3 and higher), whereas perceived risk has been mainly studied for manual driving. Hence it is unclear whether perceived risk and trust operate differently in Level 2 driving automation.

This chapter contributes to two main objectives: **Objective 1**: to model perceived risk and trust in SAE Level 2 automation. **Objective 2**: to quantify the impact of personal characteristics by verifying their contributions in the perceived risk and trust models.

We conduct a simulator study with partial automation (SAE Level 2) motorway driving with drivers continuously monitoring the automation. This allows continuous measurement of perceived risk during dynamic interactions with other road users. We focus on aggressive merging (cut in) and hard braking as safety-critical events [82]. Ques-

tionnaires, continuous measures of perceived risk (e.g. sensors for hands), physiological measures (e.g. GSR, ECG), behaviour indicators (e.g. braking behaviour) and eye behaviour (e.g. pupil diameter) are jointly evaluated to assess their ability to quantify perceived risk and trust in automation. We develop perceived risk and trust models that include the factors mentioned above to explain and predict perceived risk and trust.

2.2. METHODS

2.2.1. PARTICIPANTS

Twenty-five participants with at least 3 years of driving experience were recruited. A recruitment advertisement was distributed via email to university employees and students, and advertised on the neighbourhood app NEXTDOOR to citizens living in Delft. 25 participants (6 females and 19 males) joined the experiment. The age ranged from 24 to 76 years, with a mean of 40.6 years (SD = 16.3). Years with a driving license varied from 3 to 55 years (Mean = 19.2, SD = 15.0). 16 of the 25 participants reported no experience in adaptive cruise control (ACC) or lane centring systems (LC).

2.2.2. APPARATUS

DRIVING SIMULATOR

The experiment was conducted at Delft University of Technology on a driving simulator named DAVSi with Yaris cockpit (Figure 2.1). In this experiment, the motion platform was not actuated. The environment was shown on the cylindrical 180-degree screen using three high-quality projectors [83]. CarMaker 8.0.1 was used to create the motorway traffic environment. A model of an Auris 4 was used to simulate the subject vehicle dynamics. Subject vehicle dynamics and traffic were controlled using Simulink on a real-time simulation system (dSPACE SCALEXIO). Motion data of the subject vehicle and other vehicles were logged at 10 Hz.

Automated lateral control of the subject vehicle was performed by the IPG driver model provided by CarMaker. A non-linear full-range ACC algorithm was used with the following key parameters: $t_d = 1.2\,\mathrm{s}$ (desired time gap to the vehicle in front); $s_0 = 6\,\mathrm{m}$ (minimum space gap at a standstill); $v_0 = 27.78\,\mathrm{m/s}$ (100 km/h, desired velocity when there is no vehicle detected in front) [84].

An indicator on the dashboard was used as a basic HMI displaying the automation's working status with two colours. Green indicated that the system was activated and worked well. Yellow indicated that the driver had to take over control, but this never happened during this experiment.

QUESTIONNAIRES

A pre-questionnaire collected personal characteristics such as gender, age, years licensed, and prior automation experience (see Appendix A.1). Initial learned trust and dynamic learned trust were assessed before and after the simulator drive using the questionnaire in Table 2.1, including related questions on the willingness to hand over control, the need to monitor automation and the willingness to do other activities.

Table 2.1: Trust related questionnaire used before and after the simulator drive (scaled between 1-10)

Item	Results before the drive (Mean+Std)	Results after the drive (Mean+Std)	t	p
To what extent do you trust the described driving automation system? (adapted from Meyer-Waarden and Cloarec [85])	6.84 (1.57)	7.92 (1.50)	-2.49	0.016
To what extent are you willing to hand over control to the described automation system? (Self-developed)	6.60 (1.98)	6.56 (2.48)	0.06	0.950
To what extent do you think it is necessary to monitor the described automation system? (adapted from Nordhoff *et al.* [11])	2.96 (2.62)	4.56 (2.41)	-2.24	0.029
To what extent are you willing to do other activities (e.g., eating, drinking, checking the phone) while using the described automation system? (adapted from Xu *et al.* [9] and Gold *et al.* [61])	4.40 (2.67)	4.60 (2.57)	-0.27	0.789

Figure 2.1: Experimental setup. Left side: Driving simulator (DAVSi) at Delft University of Technology. Middle: Participant with all measurement devices. Right side: (A) Pressure sensor for reporting continuous perceived risk. (B) LED bar- visual feedback of reported continuous perceived risk. (C) ECG device TMSi to measure cardiovascular activity. (D) Eye tracker Tobii pro 2 to measure pupil dilation

PHYSIOLOGICAL MEASURES

Three physiological signals were measured to assess their predictive value in monitoring trust and perceived risk: Cardiovascular activity (ECG), galvanic skin response (GSR) and pupil dilation.

ECG was measured on Lead II (between the left inner ankle and right inner wrist, with the ground on the right inner ankle) and recorded using a TMSi amplifier at 1024 Hz (Figure 2.1C). Heartbeats were identified using BioSigKit [86]. The MTEO_QRST algorithm was found to produce the most reliable detection [87]. Peak detections were inspected manually for mislabelling and ectopic beats. The resulting detections were then converted to the rate and variability metrics Heart rate (Beats per minute, BPM), Inter-beat interval (IBI), Root mean square of successive inter-beat interval differences (RMSSD), and power in the High-frequency band (HF; 0.15 – 0.40 Hz) using the heart rate analysis toolkit heartpy 1.2.6. For calculating HF, IBI was re-sampled using 3rd order univariate spline interpolation. Metrics were calculated for "ultra-short-term" windows of 30 s, in which RMSSD and HF variability metrics are acceptable surrogates for 5-minute recordings, according to Baek *et al.* [88]. Samples were deemed too short to inspect the Low-frequency band (LF; 0.04 – 0.15 Hz). An increase in RMSSD and HF may indicate increased activity of the parasympathetic nervous system (and hence a state of ease), while a reduction could indicate increased anxiety, but HF is also influenced by breathing [89].

GSR was measured on the right palm with a Groove GSR sensor at 60 Hz and deconvolved into phasic and tonic components using Ledalab-349 (Benedek and Kaernbach, 2010).

Pupil dilation (diameter) was measured at 50 Hz using a Tobii head-mounted eye tracker (Figure 2.1D) measuring the left eye and post-processed with a 4 Hz low-pass filter [90].

VERBAL RATINGS OF PERCEIVED RISK AND TRUST AND CONTINUOUS RATINGS OF PERCEIVED RISK

The experimenter asked two questions during the simulator drive after each event (see the detailed experiment design below). The two questions were "*How dangerous do you think was the previous event?*" and "*To what extent do you trust the driving automation*

according to the previous performance of the system?" Meanwhile, participants continuously rated their perceived risk with a pressure sensor fixed on the steering wheel (Figure 2.1A), obtaining visual feedback through a LED bar (Figure 2.1A). The participants were tasked to press the sensor harder whenever they felt unsafe, where no force (zero active LED) indicated no risk and the maximum (ten active LEDs) meant very high risk. The sensor's scope and sensitivity were calibrated based on the data from Astin [91] to have a better experience on the ratings and visual feedback of the LED bar. The continuous rating was recorded at 60 Hz.

2.2.3. EXPERIMENTAL DESIGN AND SCENARIOS

Level 2 automation is mainly developed for motorway driving [92]. Car following (braking) and lane change (merging) account for most driving scenarios on motorways [93] and are used as safety-critical events in a simulator study [94]. Therefore, merging by an adjacent vehicle and hard braking by a lead vehicle were selected for our research. The driving automation maintained longitudinal velocity or kept the predesigned distance to the lead vehicle. The reference velocity of the subject vehicle and the traffic vehicles was set to 100 km/h [95].

The participants monitored the automation driving at the right lane (see Figure 2.2 and Figure 2.3). A merging vehicle entered the motorway from an on-ramp, passed the subject vehicle, and merged between the subject and lead vehicles. Detection of this merging manoeuvre was implemented as the moment when the centre of the merging vehicle crossed the line. This somewhat late detection was seen as representative of current systems. After this detection, the subject vehicle automation followed the merging vehicle instead of the original lead vehicle. At this exact moment, the original leading vehicle braked strongly to 60 km/h, followed by acceleration to 100 km/h. The merging vehicle braked and accelerated accordingly keeping a safe distance. The initial merging distance and braking intensity were both varied threefold, creating 9 merging with hard braking (MB) events with different criticalities (see Table 2.2). In addition, a hard braking (HB) event without merging was designed to investigate whether perceived risk and trust differ between hard braking after merging and normal hard braking (Table 2.2). All 10 events in Table 2.2 were repeated twice, resulting in 20 events. Hereafter, the event names with 'a' or 'b' mean the first and second exposure to the event. All events occurred in a single drive through a series of ramps (20 out of 23 ramps) along the road, as shown in Figure 2.3. The order of events was randomized with intervals around 1-min between merging locations.

Participants could overrule automation using brake or gas pedals whenever they felt that this was necessary. However, manual steering was not allowed during the simulator drive. No accidents or automation failures were designed.

2.2.4. PROCEDURE

The participants were asked to read an information letter about the experiment and sign an informed consent form. The experimenter introduced the driving automation system and the procedure. Then the participants were asked to fill the trust questionnaire (Table 2.1). Electrodes were attached to measure GSR and ECG and the eye tracker was installed and calibrated. Participants were seated in the simulator, and sensors were connected to

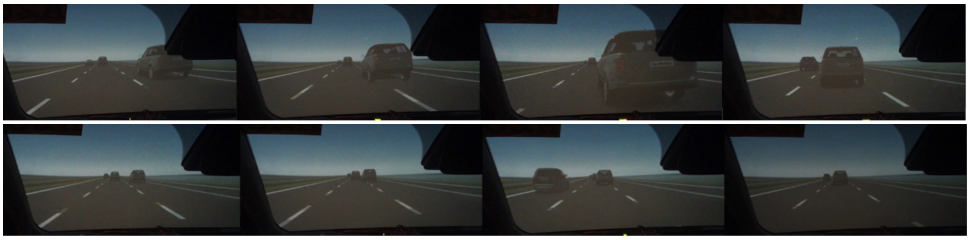


Figure 2.2: Video stills of merging with hard braking events. The first row shows the sequence of merging manoeuvre in the most critical MB3 (initial gap 5 m, braking intensity $-8\,m/s^2$), and the second row shows the sequence of the least critical MB7 (initial gap 25 m, braking intensity $-2\,m/s^2$)

Table 2.2: Events with different criticalities

Scenario	Merging gap (initial gap) (m)	Braking intensity (m/s^2)	Event name	Minimum THW (s) averaged for the two exposures (measured value)
Merging with hard braking	5	-2	MB1	0.22
		-5	MB2	0.22
		-8	MB3	0.14
	15	-2	MB4	0.48
		-5	MB5	0.48
		-8	MB6	0.48
	25	-2	MB7	0.84
		-5	MB8	0.83
		-8	MB9	0.80
Only hard braking	25	-8	HB1	0.85

their recording devices. The experimenter introduced the scenarios and the operation of the simulator.

The experimenter trained the participants to use the pressure button to rate perceived risk. Participants used the pressure button by giving a number from 0 to 10 and hold each number for at least 3 seconds. Subsequently, they had to follow a random number between 1 to 10 provided by the experimenter. In the ensuing practice drive, the participants experienced several merging events. They were asked to continuously indicate perceived risk using the pressure button and answer the experimenter's questions mentioned in section 2.2.2 *Verbal ratings of perceived risk and trust and continuous ratings of perceived risk* verbally after each event. The training lasted until the participants could handle all tasks well.

The formal drive (after practice) followed the same procedure, now including 20 events presented in a randomised order per participant. Another questionnaire with the items in Table 2.1 was filled after the simulator drive to measure changes in trust.

2.3. RESULTS

All 25 participants completed the simulator drive, and no motion sickness was reported. 7 out of 25 participants intervened by braking in at least one event. The variations induced by participants' braking and simulated control noise inside the subject vehicle

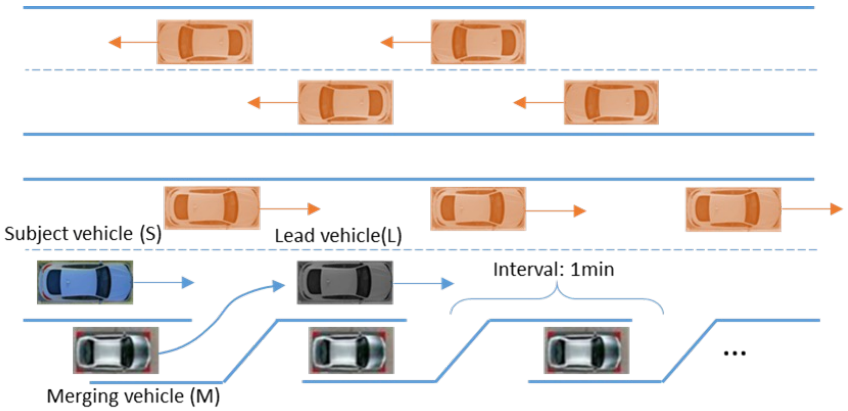


Figure 2.3: Merging scenario. The Subject vehicle (S) was driving behind the Lead vehicle (L) when the Merging vehicle (M) merged. L then braked while M and S still had a reduced headway. The scenario's objective risk was varied through merging distance (S-M) and braking intensity of L.

led to 3 events with a close-to-zero distance in the most critical event MB3, one of which was regarded as a collision by the participant. Therefore, these 3 participants were not considered during regression modelling. Besides the three collision participants, 12 outlier ratings (see Figure 2.4 for criterion) were removed, leaving 428 out of 500 events for the regression analysis. Pupil diameter and ECG signal were successfully recorded for 22 participants. The recorded GSR signal was not of good quality since skin conductance was outside the device's sensitive range for all but two participants and was excluded from the analysis. We also repeated the experiment without the verbal rating task with 5 extra participants (2 of them were new) to evaluate the influence of speaking on pupil dilation and ECG signals.

2.3.1. PERCEIVED RISK AND TRUST AS FUNCTIONS OF SCENARIO AND PERSONAL CHARACTERISTICS

The verbal risk and trust ratings for the events of different criticalities are presented in Figure 2.4. Perceived risk varied highly between conditions for all participants. Trust was lowest after the first occurrence of the most critical event (MB3a) and varied less than perceived risk. 12 participants consistently rated trust as 7 or higher in all events except in MB3.

CORRELATION ANALYSIS OF POTENTIAL INFLUENCERS FOR PERCEIVED RISK AND TRUST

Perceived risk and trust are influenced by many factors. Before the regression analysis, we selected four clusters of potential factors (see Table 2.3). Cluster 1 includes the participant's responses to the events, including verbal ratings of perceived risk (PR) and trust (TRU) themselves and the maximum braking by participants (Max_B) because they can reflect driver's perceived risk and trust [9], [34], [70], [96]. Cluster 2 captures the criticality of the event with factors related to relative motion, including the initial merging gap (IMG), minimum gap (min_gap), minimum time to collision (min_TTC), minimum

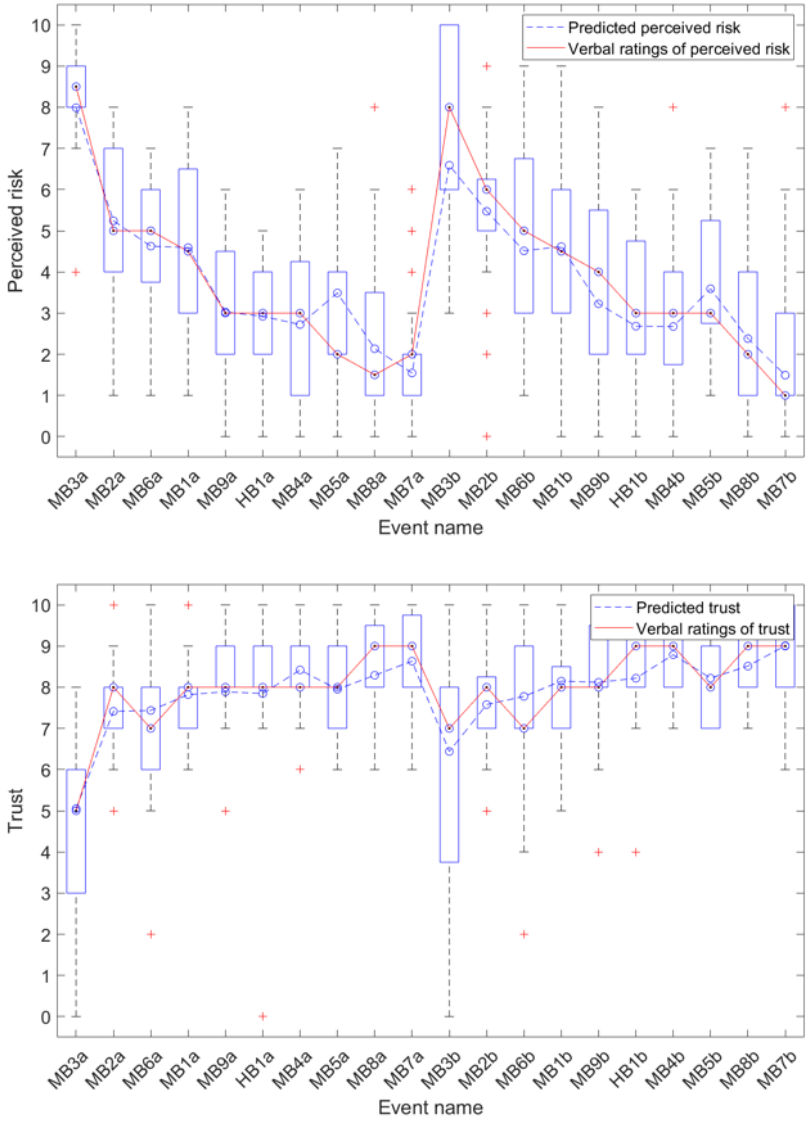


Figure 2.4: Upper: Verbal perceived risk ratings for different events. Lower: Verbal trust ratings for different events. Bars present verbal ratings (5, 25, 50, 75, 95 percentile). The dashed blue lines represent the model output in Equation (2.1) and Equation (2.2). . The red '+' represent outliers beyond 75 percentile+1.5IQR or below 25 percentile-1.5IQR (IQR is the first quartile subtracted from the third quartile).

time headway (min_THW) and braking intensity (BI) of the merging vehicle during the events. Here min_gap, min_TTC and min_THW express the smallest gap in different manners and are established SMoS to assess the criticality of vehicle interactions [97]. These factors (e.g., minimum gap, minimum TTC, minimum THW, etc.) have been used to reflect perceived risk in prior studies [39], [98]. We will also verify their substitution in the regression. Additionally, we will investigate the overlap among influencers of perceived risk and trust. Cluster 3 includes personal characteristics, age (AGE), gender (GEN), years with a driving license (YDL), and automation experience (AE). These factors play essential roles in individual modelling of perceived risk and trust [33], [38], [59], [61], [99], [100]. Cluster 4 includes repetition (REP) of the event and event type (ET) being merging or hard braking because trust may change with automation experience according to the trust formation from Lee and See [28], Hoff and Bashir [29], and Kraus *et al.* [56]. Also, participants perceived less risk in the second exposure to the same event in our experiment (Figure 2.4).

Before the stepwise regression, we checked for multicollinearity among potential influencers.

In Cluster 2, non-linear transformations on the original metrics were explored (Appendix A.2). We found that the logarithm of min_gap, min_TTC and min_THW have strong linear relationships with the participant-averaged perceived risk and the reciprocal of the min_gap, min_TTC and min_THW have strong linear relationships with trust (see Figure A.1 in Appendix A.2). Therefore, the logarithm and the reciprocal of min_gap, min_TTC and min_THW were used as potential predictors for perceived risk and trust in the regression.

We only found strong ($|r| > 0.7$, shaded in Orange in Table 2.3) correlations within Cluster 2 and Cluster 3. In Cluster 2, strong correlations exist between min_gap, min_TTC, min THW and IMG (correlation group 1). In Cluster 3, AGE highly correlates with YDL (correlation group 2). Therefore, in the following regression procedure, we will use at most one variable of each of the two correlation groups as potential predictors to resolve multicollinearity.

Table 2.3: Correlation analysis (high correlations in orange and insignificant correlations in grey)

		Cluster 1			Cluster 2								Cluster 3				Cluster 4	
		PR	TRU	max_B	Log (min_gap)	Log (min_TTC)	Log (min_THW)	1/ min_gap	1/ min_TTC	1/ min_THW	IMG	BI	AGE	YDL	GEN	AE	ET	REP
Cluster 1	PR	1.000																
	TRU	-0.607***	1.000															
	max_B	0.278***	-0.449***	1.000														
Cluster 2	Log(min_gap)	-0.632***	0.477***	-0.302***	1.000													
	Log(min_TTC)	-0.611***	0.487***	-0.373***	0.918***	1.000												
	Log(min_THW)	-0.579***	0.416***	-0.248***	0.972***	0.919***	1.000											
	1/min_gap	0.592***	-0.513***	0.337***	-0.914***	-0.818***	-0.857***	1.000										
	1/min_TTC	0.566***	-0.599***	0.690***	-0.781***	-0.829***	-0.736***	0.840***	1.000									
	1/min_THW	0.581***	-0.468***	0.275***	-0.951***	-0.871***	-0.949***	0.962***	0.805***	1.000								
	IMG	-0.519***	0.337***	-0.176***	0.907***	0.881***	0.965***	-0.705***	-0.614***	-0.840***	1.000							
	BI	-0.280***	0.307***	-0.157**	0.144**	0.172***	-0.013	-0.218***	-0.243***	-0.071	-0.090*	1.000						
Cluster 3	AGE	-0.179***	0.112*	0.063	-0.041	0.050	-0.001	0.012	-0.021	-0.013	0.005	-0.017	1.000					
	YDL	-0.186***	0.134**	0.095	-0.035	0.033	0.000	0.002	-0.002	-0.019	0.002	-0.018	0.957***	1.000				
	GEN	-0.051	-0.065	0.020	-0.004	-0.017	-0.004	-0.018	-0.010	-0.014	-0.016	-0.011	0.059	0.154***	1.000***			
	AE	0.063	-0.078	0.115*	-0.028	-0.016	-0.019	0.023	0.054***	0.019	-0.015	0.040	-0.333***	-0.240***	0.091	1.000		
Cluster 4	ET	-0.128*	0.000	-0.064	0.283***	0.282***	0.307***	-0.192***	-0.170***	-0.233***	0.326***	-0.275***	0.016	0.005	-0.049	-0.046	1.000	
	REP	0.033	0.121**	-0.037	0.028	-0.073	0.022	-0.083*	-0.026	-0.062	-0.005	0.022	0.002	0.002	-0.001	0.003	-0.015	1.000

$^{*}p < 0.05; ^{**}p < 0.01; ^{***}p < 0.001$

STEPWISE REGRESSION ANALYSIS OF PERCEIVED RISK AND TRUST

Stepwise regression is an efficient way to select suitable predictors after the elimination of multicollinearity. We used stepwise multiple regression to model the influence of event criticality and personal characteristics on perceived risk and trust. The consecutive steps of both regression models are shown as the first 4 models in Table 2.4 and Table 2.5 for perceived risk and trust respectively. Hence, the two models 4 represent the final models of perceived risk (PR) and trust (TRU) as shown in Equation (2.1) and Equation (2.2).

$$PR = 9.384 - 2.473 \cdot \ln min\_gap - 0.038 \cdot YDL - 0.201 \cdot BI + 0.470 \cdot GEN \tag{2.1}$$

$$TRU = 8.787 - 6.265 \cdot (1/min\_TTC) + 0.125 \cdot BI + 0.016 \cdot YDL + 0.372 \cdot REP \tag{2.2}$$

According to the results, perceived risk and trust mainly vary with min_gap, min_TTC and BI. A smaller minimum gap and more intense braking lead to higher perceived risk; a larger minimum TTC and more gentle braking cause higher trust. Participants with more driving experience trust the automation more and are less sensitive to risk. Female participants are more susceptible to risk. Participants trust the automation more in the second exposure to the events.

The participant-averaged results were used to validate the two models expressed by Equation (2.1) and Equation (2.2). A very good fit was obtained (Figure 2.4) with $R^2 = 0.9379, F(2,17) = 88.2940 (p = 0.000)$ for perceived risk and adjusted $R^2 = 0.8643, F(3,16) = 27.6157 (p = 0.000)$ for trust. The root mean squared error (RMSE) is 0.4044 for perceived risk and 0.3164 for trust while the Pearson correlation is 0.9812 ($p = 0.000$) for perceived risk and 0.9554 ($p = 0.000$) for trust, indicating that the models well predict the participant-averaged perceived risk and trust ratings within events. The group based regression models also describe individual events within individual participants quite well,

as shown in Figure A.2 and Figure A.3 in Appendix A.2, resulting in $RMSE = 1.7810$, Pearson coefficient = 0.7252 ($p = 0.000$) for perceived risk and $RMSE = 1.4835$, Pearson coefficient = 0.5824 ($p = 0.000$). Note that the output of these two models can be out of the feasible range [0, 10] because the linear regression models do not constrain the output, but individual data were all within the feasible range. The perceived risk model also well predicted the extra 5 participants' data with $R^2 = 0.3641, F(3,86) = 17.0563 (p = 0.000)$ in $RMSE = 2.3727$, Pearson coefficient = 0.6562 ($p = 0.000$) (see Figure A.4 in Appendix A.2).

INDIVIDUAL CALIBRATION

We calibrated the two models in Equation (2.1) and (2.2) for each participant (in Appendix A.3), where both models were statistically significant for all participants except the trust model for participant 18. Female participants were only sensitive to minimum gap but not braking intensity (the models without BI for the female participants have an average $RMSE = 1.4794$); participants with more driving automation experience perceived the risk to depend only on the minimum gap.

SUBSTITUTION OF POTENTIAL PREDICTORS

As motivated in Section 2.3.1 *Correlation analysis of potential influencers for perceived risk and trust* and *Stepwise regression analysis of perceived risk*, only the best predictor in each of the two correlation groups was adopted in the regression. Below we verify the substitution of alternative predictors.

In correlation group 1, we replaced min_gap and min_TTC in Equation (2.1) and (2.2) with the other three relevant factors. For example, in Equation (2.1), we replaced ln(min_gap) with ln(min_TTC). This generated models 5-7 (see Table A.1 and Table A.2 in Appendix A.2). The new models remain significant and the R-square decreases slightly. Hence, these common safety metrics have a similar capability to predict perceived risk and trust, indicating that they can be replaced. However, some predictors become insignificant, such as GEN in perceived risk (see model 9 in Table A.1) and REP in trust (see model 9 in Table A.2).

In correlation group 2, we replaced YDL with AGE (see model 8 in Table A.1 and Table A.2 in Appendix A.2). The new model 8 is still significant with only a slight decrease in R-square, but GEN is no longer significant. Hence, AGE and YDL have a similar performance in predicting perceived risk and trust and are thereby replaceable.

For other predictors outside the two correlation groups, REP only appears in the trust model (Equation (2.2)), meaning that participants trust the automation more after the first exposure of the events, which implies that trust will accumulate over time provided that no crash or automation failure occurs. To further validate the effects of automation exposure, we compared the trust levels of the questionnaires before and after the simulator drive for all 25 participants (Table 2.1). We found only participants 4 and 6 trusted the automation less after the simulator drive, and especially participant 6 reported a 'crash' during the experiment, where the minimum distance between the automated vehicle and the lead vehicle was close to zero. The average trust level increased from 6.84 ($SD = 1.54$) before the simulator drive to 7.92 ($SD = 1.47$) ($p = 0.016$) after the drive. Hence, both the inclusion of REP in the trust model and the questionnaire indicate that trust generally increased over time.

Table 2.4: Multiple regression results of perceived risk. Models follow the stepwise regression (Model 4* is the final perceived risk model). See Table B. 1 in Appendix B for alternative models.

Model	Variables in model	R^2	Adjusted R^2	F	p	B	95 confidence interval for B		β	t	p
							Lower bound	Upper bound			
1	Constant	0.400	0.398	283.782	0.000	10.183	9.451	10.916		27.326	0.000
	ln(min_gap)					-2.556	-2.854	-2.258	-0.632	-16.846	0.000
2	Constant	0.443	0.440	169.033	0.000	10.981	10.223	11.738		28.499	0.000
	ln(min_gap)					-2.585	-2.873	-2.297	-0.640	-17.657	0.000
	YDL					-0.035	-0.047	-0.023	-0.208	-5.743	0.000
3	Constant	0.481	0.477	130.777	0.000	9.673	8.806	10.540		21.929	0.000
	ln(min_gap)					-2.472	-2.753	-2.191	-0.612	-17.281	0.000
	YDL					-0.035	-0.047	-0.024	-0.211	-6.012	0.000
	BI					-0.202	-0.274	-0.130	-0.196	-5.538	0.000
4*	Constant	0.487	0.482	100.427	0.000	9.384	8.487	10.280		20.565	0.000
	ln(min_gap)					-2.473	-2.752	-2.193	-0.612	-17.377	0.000
	YDL					-0.038	-0.049	-0.026	-0.223	-6.327	0.000
	BI					-0.201	-0.273	-0.130	-0.195	-5.546	0.000
	GEN					0.470	0.071	0.869	0.082	2.313	0.021

Table 2.5: Multiple regression results of trust. Models 1-4 follow the stepwise regression (Model 4* is the final trust model). See Table B. 2 in Appendix B for alternative models.

Model	Variables in model	R^2	Adjusted R^2	F	p	B	95 confidence interval for B		β	t	p
							Lower bound	Upper bound			
1	Constant	0.359	0.358	238.883	0.000	8.715	8.536	8.894		95.771	0.000
	1/min_TTC					-6.770	-7.631	-5.909	-0.599	-15.456	0.000
2	Constant	0.387	0.384	134.126	0.000	9.299	8.984	9.615		57.969	0.000
	1/min_TTC					-6.300	-7.169	-5.431	-0.558	-14.247	0.000
	BI					0.125	0.069	0.180	0.171	4.379	0.000
3	Constant	0.405	0.401	96.126	0.000	8.974	8.616	9.331		49.332	0.000
	1/min_TTC					-6.290	-7.147	-5.433	-0.557	-14.425	0.000
	BI					0.126	0.071	0.182	0.174	4.509	0.000
	YDL					0.016	0.007	0.025	0.136	3.627	0.000
4*	Constant	0.416	0.410	75.290	0.000	8.780	8.399	9.161		45.326	0.000
	1/min_TTC					-6.265	-7.115	-5.414	-0.555	-14.476	0.000
	BI					0.125	0.071	0.180	0.172	4.500	
	YDL					0.016	0.007	0.025	0.136	3.651	0.000
	REP					0.372	0.107	0.638	0.102	2.754	0.006

2.3.2. CORRELATION BETWEEN PERCEIVED RISK AND TRUST

Our regression models for perceived risk and trust show a substantial similarity in agreement with the literature [9], [20], [101]. This similarity is also supported by Figure 2.5. A strong linear relationship was found with a Pearson correlation $r = -0.919 (p < 0.01)$ between participant-averaged perceived risk and trust levels for different events. Regarding all 500 events, the repeated measures Pearson correlation coefficient is $-0.649 (p = 0.000)$ [102]. This means that people trust the system more after events where they perceive a lower risk. The individual correlations between perceived risk and trust are significant for 13 out of 25 participants but are not significant for participants with a low standard deviation of trust (Table A.6 in Appendix A.2).

To evaluate how well the predictors discriminate between trust and perceived risk, we cross-validated the models of perceived risk and trust. Specifically, we used the predictors of perceived risk to model trust and the predictors of trust to model perceived risk (models 9 in Table A.1 and Table A.2). The two models are still significant with a slightly lower R-Square but REP and GEN are no longer significant in the new models.

We conclude that perceived risk and trust negatively correlate and can be modelled using the same predictors of min_gap, min_TTC, YDL, and BI. However, REP only significantly affects trust, and GEN only significantly affects perceived risk.

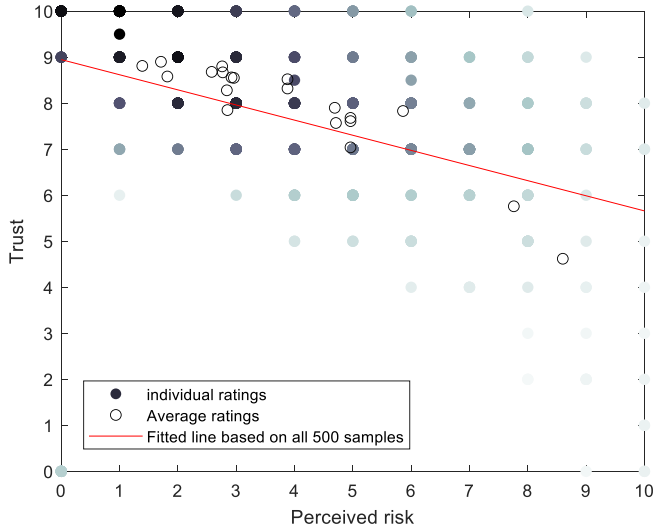


Figure 2.5: Relation between perceived risk and trust for individual events (filled dots) and averaged over participants (circles). Darker dots indicate more overlapping data points.

2.3.3. EFFECTIVE INDICATORS OF PERCEIVED RISK AND TRUST

Participants' braking signal, pupil diameter, and ECG were recorded during the experiment, along with the continuous rating of perceived risk (Figure 2.7). This section investigated whether these signals reflect perceived risk and trust effectively.

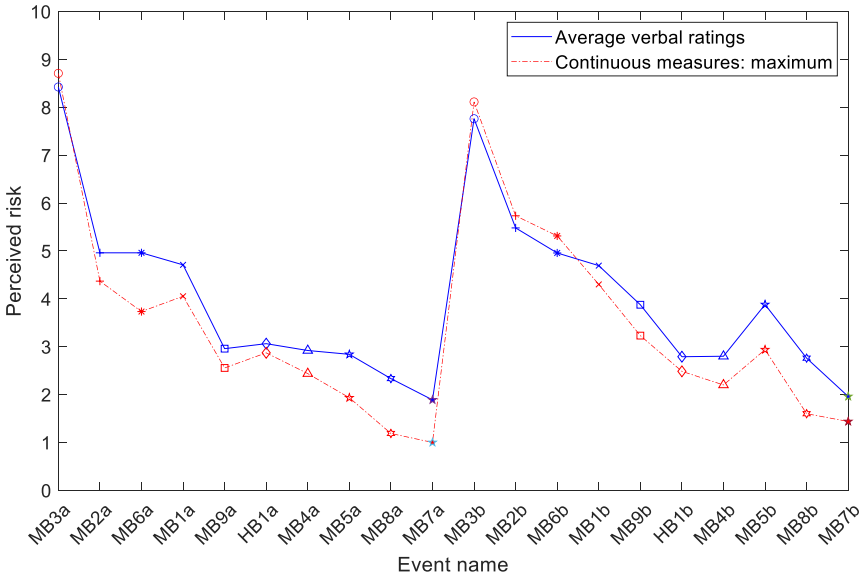


Figure 2.6: Participant-averaged perceived risk ratings within events for the two rating methods.

CONTINUOUS RATINGS OF PERCEIVED RISK

To examine the consistency between continuous and verbal ratings, we compared the participant-averaged peak continuous risk to the corresponding verbal ratings (Figure 2.6). The two measures have a strong linear relationship, as indicated by their correlation ($r = 0.983, p < 0.001$). Hence, we conclude that the continuous ratings accurately reflect participants' perceived risk.

BRAKING BEHAVIOUR AND PERCEIVED RISK AND TRUST

Braking is a signal potentially reflecting higher perceived risk and lower trust levels. This was confirmed by adding the maximum braking pedal position (Max_B) to the models in Equation (2.1) and (2.2), resulting in model 10 (see Table A.1 and Table A.2) for perceived risk and trust. Maximum braking was a significant predictor of perceived risk ($p = 0.012$), but the relation with trust was not significant ($p = 0.070$) within single events. 7 of 25 participants braked in at least one event. Therefore, the participants were divided into non-braking and braking groups (see Appendix A.4 Braking behaviour). The braking group reported a higher perceived risk ($p = 0.021$) and lower trust ($p = 0.000$) in most of the event types supported by t-test.

PUPIL DILATION AND PERCEIVED RISK

We evaluated if pupil dilation can indicate perceived risk. Pupil dilation was expressed as the difference of the maximum and minimum pupil diameter from 20 s before till 10 s after reaching a minimum gap to the lead vehicle, which spans the first moment when the on-ramp became visible until the time the participant was asked to provide subjective ratings (see Figure 2.7).

Firstly, we explore whether the pupil significantly dilated in different event conditions. Kruskal-Wallis tests showed significant variations of participants' pupil diameter within most of the event conditions (p = 0.000), indicating pupil dilation significantly changed within an event (see Appendix A.4 Pupil dilation).

To evaluate whether event criticality had a within-subjects effect on pupil dilation, we performed repeated measures ANOVA among all 20 event types. The difference between pupil dilation in different event types is statistically significant in the second exposure of the events ($F(5.881, 123.502) = 2.783, p = 0.015$), but insignificant in the first encounters ($F(4.470, 93.865) = 1.014, p = 0.409$). Repeated measures correlation analysis in 397 events shows no significant correlation between within-subject pupil dilation and the maximum continuous risk (Correlation $r = 0.03, p = 0.530$). A strong correlation is found between the participant-averaged pupil diameter signal and the participant-averaged continuous perceived risk signal across all participants (22 participants available in 18 out of the 20 event types, especially in the most critical event MB3 (see Table A.8 in Appendix A.4 Pupil dilation).

Therefore, we conclude based on the participant-averaged signals that pupil dilation highly correlates with perceived risk if the events are sufficiently risky. The merging and braking events affected pupil dilation, but the correlation between the maximum continuous perceived risk and pupil dilation across all events was not statistically significant.

ECG AND PERCEIVED RISK

To evaluate whether event criticality had a within-subjects effect on heart measures, we performed repeated measures ANOVA as well as a repeated measures correlation among all event types. Heart rate and variability metrics were calculated over the same time period as the pupil dilation in Section 2.3.3 *Pupil dilation and perceived risk*. The ANOVA results show that a difference between event types was only statistically observed for IBI with a marginal significance of $p = 0.053$ among first events and $p = 0.002$ among second events. In the most critical event MB3, IBI tended to be smaller compared to less critical events (see Table A.9 in Appendix A.4 ECG).

Repeated measures correlations were performed between the three heart metrics and two safety metrics: the minimum time headway and maximum continuous risk. A strong correlation between the safety metrics was observed but they do not correlate with the within-subject heart metrics (see Table A.10 in Appendix A.4 ECG).

In terms of the participant-averaged BPM, RMSSD and HF, heart rate increases in the period following an event, and this pattern is consistent across events and within repetitions of the same event. No such pattern was observed for RMSSD and HF in the participant-averaged signals. The participant-Averaged BPM increase did not correlate significantly with participant-averaged perceived risk ($r = 0.174, n = 20, p = 0.464$) or minimum time headway ($r = -0.378, n = 20, p = 0.100$), which indicates no consistent relation with event criticality (see Figure A.7 in Appendix A.4).

Therefore, we conclude that the merging and brake events can increase heart rate, but no significant relation was found between these heart measures and perceived or objective risk. Heart rate variability metrics were not significantly affected.

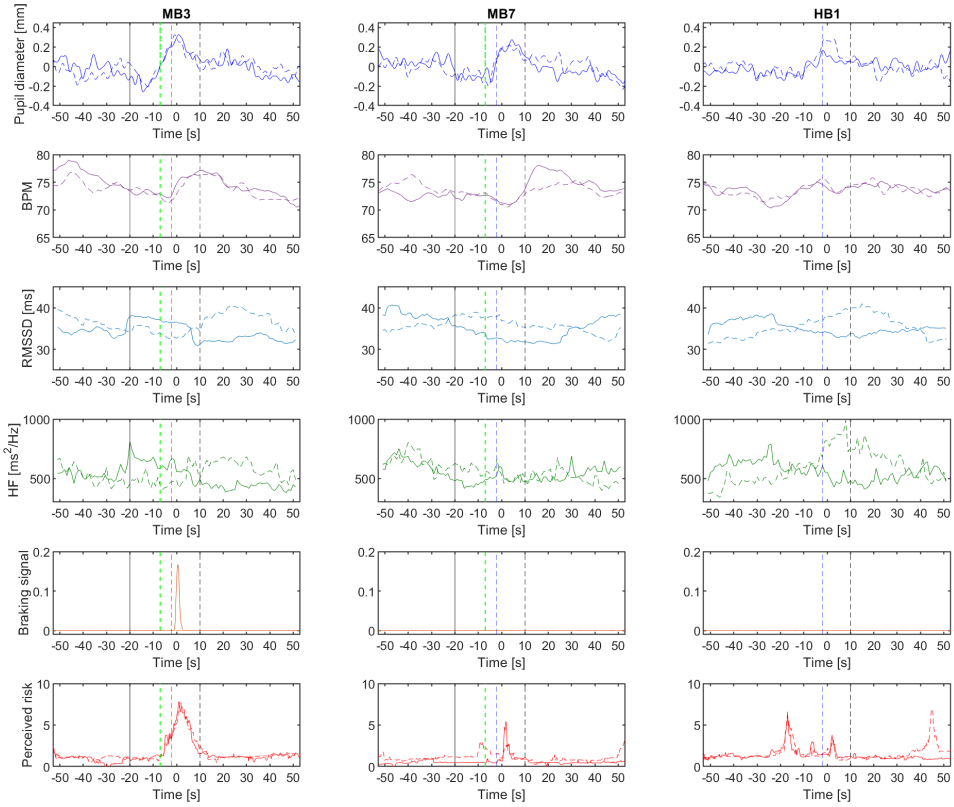


Figure 2.7: Participant-averaged signals in the most critical merge MB3 (left), the least critical merge MB7 (mid), and in hard braking (right). The time scale is $t = -55s$ to $t = 55s$, where $t = 0$ means the smallest gap to the lead vehicle. Pupil diameter is relative to the participant's overall pupil average (all events combined). BPM indicates heat rate; RMSSD represents Root mean square of successive inter-beat interval differences; HF indicates the power in the High-frequency band (HF; 0.15 – 0.40 Hz); The solid curve and the dashed curve represent the first and second exposure. Time is zero for the timing of minimum gap to the merging or leading vehicle. Black solid line: timing when the on-ramp became visible. Green dashed line: timing when the merging vehicles became visible. Blue dashed line: timing when the merging vehicle started to brake. Black dashed line: timing when the participants gave verbal ratings. Results before t = 0 include recovery from preceding events with some variation due to randomization.

2.4. DISCUSSION

This study was conducted to model perceived risk and trust in SAE Level 2 driving automation (ACC+LC) reacting to merging and hard braking vehicles based on a driving simulator experiment.

2.4.1. FACTORS INFLUENCING PERCEIVED RISK AND TRUST

The regression models of perceived risk and trust (Equation (2.1) and (2.2)) demonstrate that smaller minimum gap, minimum TTC, and stronger braking intensity lead to higher perceived risk and lower trust. Other classic surrogate metrics of safety (SMoS), including the initial merging gap and THW have similar performance predicting perceived risk and trust (Table A.1 and Table A.2 in Appendix A.2).

Our study shows that these well-known SMoS are good predictors for perceived risk. The results are in line with Ma *et al.* [98], who used TTC and THW in a regression model of driver's subjective risk in critical cut-in scenarios in manual naturalistic driving. Kondoh *et al.* [103] captured human risk perception during car-following as the summation of the time headway inverse and the time to collision. Lu *et al.* [104] also found a strong relationship between TTC, THW and perceived risk in car-following. It should be noted that non-linear transformation of the SMoS was required to obtain accurate models predicting perceived risk and trust. These transformations may express human risk perception related to the visual perception of relative motion [79]. Braking intensity of the leading vehicle affects both perceived risk and trust as an independent factor which is only mildly correlated to other factors related to vehicle motion ($|r| < 0.275$ for Cluster 2 in Table 2.3). We are not aware of other studies indicating the relevance of braking intensity in perceived risk and trust, and recommend further exploration of this factor in future studies.

For trust, our study shows that the well-known SMoS are also predictive. Specifically, participants have a lower post-event trust if the minimum TTC is smaller and the braking is stronger in the previous event. This aligns with existing studies on trust modelling and trust calibration. According to Hoff and Bashir [29] and Lee and See [28], trust is established in a dynamic process using new information (e.g., event criticality, system performance, etc.). Kraus *et al.* [56] demonstrate that drivers perceive and interpret the system behaviour and then update the dynamic learned trust based on the initial dynamic trust. Kaplan *et al.* [105] established a trust model where the environmental or contextual factors are directly used to calculate trust. In our models, the safety metrics and kinematic vehicle state represent the criticality of the previous event, which updates trust based on the initial trust.

The impacts of personal characteristics on perceived risk and trust were quantified by the regression models in Table 2.4 and Table 2.5. We found increased trust in automation with more driving experience similar to Jin *et al.* [33], Gold *et al.* [61], and Kaplan *et al.* [105]. For perceived risk, we found that driving-experienced participants perceived lower risk when using automation compared to inexperienced drivers. Apart from a higher trust in the system, this may also be a consequence of trust in their own ability to intervene, as illustrated by He and Donmez [106], where experienced drivers showed less attention relaxation with SAE Level 2 automation while performing a secondary-task. Similarly, Ping *et al.* [38] and Borowsky and Oron-Gilad [107] show that experienced

drivers have stronger hazard awareness when watching movies of real-world driving situations.

Gender effects were found in several studies, where males tend to trust the automation system more than females [59], [108], and females perceive more risk [109]. We found females to experience a higher risk but found no significant effect on trust, possibly due to a limited sample size and imperfect gender balance.

2.4.2. THE RELATION BETWEEN PERCEIVED RISK AND TRUST

Our results support that when people perceive lower risk, they trust automation more. Models 4 in Table 2.4 and Table 2.5 show considerable overlap between trust and perceived risk determinants, confirmed through cross-validation with models 9 in Table A.1 and Table A.2 in Appendix A.2. The two constructs shared a strong negative correlation, which agrees with internet surveys (e.g. Zoellick *et al.* [20] and Choi and Ji [21]). Consequently, we observed that predicting variables of either construct are interchangeable between regression models. However, there are grounds to believe that the two constructs are partially independent (see the discussion in Section 2.4.4).

2.4.3. MEASURES OF PERCEIVED RISK AND TRUST

The regression models are based on post-event verbal ratings of perceived risk and trust. Continuous perceived risk was highly correlated to the post-event verbal rating, indicating the effectiveness of the continuous measurement of perceived risk. However, the continuous measurement increases the drivers' workload, and participants occasionally forget to press the sensor. Taking into account these limitations we do consider our continuous risk measure to be a valuable reference to study and model perceived risk as a function of time. However, in a recent on-road experiment such a measurement was less effective with many missed events [110].

Significant effects were obtained for physiological measurements of pupil diameter and ECG metrics. Effects of events on pupil dilation were significant at the group level. Pupil dilation varied with perceived risk, in particular for the most critical events. ECG metrics showed significant effects of events on IBI, but showed no significant correlations with event criticality. Hence, pupil dilation may indicate the amount of perceived risk while IBI is at best indicative of the presence of perceived risk. However, they lack accuracy in quantifying perceived risk which will be an even larger drawback in on-road studies.

We found that driver intervention (braking) relates to perceived risk and trust at the event level, where the 7 participants braking in at least one event reported higher risk and lower trust averaged over all events. Such interventions demonstrate active monitoring by drivers and the somewhat lower trust levels in braking drivers can reflect well-calibrated trust levels. Our findings regarding braking behaviour are in line with Tenhundfeld *et al.* [34] and Lee *et al.* [111], who found that if participants brake more frequently, they perceive more risk and trust the system less in a Tesla automated parking test and an intersection crossing experiment on a driving simulator. Hence, braking behaviour is a relevant indicator of the presence of perceived risk and distrust but only 7 out of 25 participants braked. Braking can be easily used for research since it will not influence behaviour and requires no additional instrumentation.

2.4.4. LIMITATIONS AND FUTURE WORK

Our regression models of perceived risk and trust are based on Level 2 driving automation (ACC+LC) in limited samples (428 events). The models use surrogate metrics of safety related to longitudinal interaction (1-D) to predict perceived risk and trust for both merging with hard braking (MB) and hard braking without merging (HB). However, in the real world, perceived risk comes not only from longitudinal but also from lateral conflicts. More advanced surrogate metrics of safety already consider lateral motion and multiple risk sources [36], [40], [112]. Further experiments should consider lateral interaction to extend the current models to 2-D. The models predict perceived risk and trust per event, and need a transformation for prediction as continuous function of time suitable for real time control.

The regression models of perceived risk and trust combine existing knowledge with human response data. The selection of potential influencers of perceived risk and trust, and the clusters in Table 2.3 are knowledge-based, while multicollinearity checks and regression are data-driven. Full data-driven methods, like machine learning, can dig out more valuable information from data beyond human experience, and may be promising in future modelling of perceived risk and trust in particular when large datasets become available from multiple experiments and on-road observations.

A close correlation was observed between perceived risk and trust. However, we still cannot answer the causality question 'which one is the determinant of the other'. The interaction between perceived risk and trust can be studied further using longitudinal data monitoring the process of trust calibration in relation to event criticality and performance of the driving automation. We found that participants trust the automation more in the second exposure to events, which is shown as REP in trust regression model (Equation (2.2)). This is related to trust calibration and is also supported by Kraus *et al.* [56], who found that trust increases over the course of system interaction if an automated system works without malfunctions. During our experiment, no malfunctions or accidents were simulated. Therefore, we can only conclude that participants trust the system more with more automated system interaction, but our results cannot support accident-related trust calibration. The current results only support short-term learned trust calibration. Self-reported trust can still change in several weeks, even in several months, according to Walker *et al.* [62], where driver's trust toward SAE level 2 cars still changed two weeks after the automation experience. Note that the trust here usually refers to dynamic learned trust calibration, which is steadily updated in a dynamic calibration feedback loop according to Lee and See [28], Hoff and Bashir [29], and Kraus *et al.* [56].

It shall also be pointed out that the applied verbal rating procedure captured perceived risk and trust with only two questions for each event. This simplification was needed to capture ratings after each event but can be complemented by more complex ratings in future studies.

The models of perceived risk and trust can be used to calibrate SMoS so that human factors can be included in risk prediction and assessment. These subjectively calibrated SMoS can be subsequently used as cost function, constraint or reference in AV path planning, decision making and controller design [81]. This has the potential to make users feel safe and trust AV, enhancing user acceptance of driving automation.

2.5. CONCLUSIONS

This chapter investigated perceived risk and trust in Level 2 driving automation (ACC+LC) in motorway driving with a simulator experiment. We developed regression models that accurately predict perceived risk and trust in specific events and the models reveal that neighbouring road users' behaviours (relative motion) significantly influence occupants' perceived risk and trust. No difference was found in perceived risk and trust between merging with hard braking and hard braking without merging. Our models show that experienced drivers are less sensitive to risk and trust the automation more, while female participants perceive more risk than males. The findings confirm that perceived risk and trust are highly correlated. The proposed models indicate that trust and perceived risk shared the predictors of minimum gap, minimum TTC, years of driving, and braking indicator but differ in using event repetition in the trust model and gender in the perceived risk model. Additionally, the results show that people who perceive lower risk trust the automation more. Regarding the indicators of perceived risk and trust, continuous ratings of perceived risk and braking behaviour can effectively indicate perceived risk or trust. Pupil dilation can reflect perceived risk if the event is sufficiently risky. The merging and braking events increased heart rate, but there was no quantified relation between heart rate increase (variability) and perceived risk.

Future research will focus on extending the perceived risk and trust models towards more complex interactions and applying the models in designing control strategies and human-machine interfaces leading to desirable levels of perceived risk and trust.

3

A NEW COMPUTATIONAL PERCEIVED RISK MODEL FOR AUTOMATED VEHICLES BASED ON POTENTIAL COLLISION AVOIDANCE DIFFICULTY (PCAD)

Perceived risk is crucial in designing trustworthy and acceptable vehicle automation systems. However, our understanding of perceived risk dynamics remains limited, and corresponding computational models are scarce. This chapter formulates a new computational perceived risk model based on potential collision avoidance difficulty (PCAD) for drivers of SAE Level 2 automated vehicles. PCAD quantifies task difficulty using the gap between the current velocity and the safe velocity region in 2D, and accounts for the minimal control effort (braking and/or steering) needed to avoid a potential collision, based on visual looming, behavioural uncertainties of neighbouring vehicles, imprecise control of the subject vehicle, and collision severity. The PCAD model predicts both continuous-time perceived risk and peak perceived risk per event. We analyse model properties both theoretically and empirically with two unique datasets: Datasets Merging and Obstacle Avoidance. The PCAD model generally outperforms three state-of-the-art models in terms of model error, detection rate, and the ability to accurately capture the tendencies of human drivers' perceived risk, albeit at the cost of longer computation time. Our findings reveal that perceived risk varies with the position, velocity, and acceleration of the subject and neighbouring vehicles, and is influenced by uncertainties in their velocities.

The content of this chapter has been published in
He, X., Happee, R., & Wang, M. (2024). "A new computational perceived risk model for automated vehicles based on potential collision avoidance difficulty (PCAD)," *Transportation Research Part C: Emerging Technologies, 166, 104751.*

LIST OF NOTATIONS

Categories	Variable*	Description
Kinematic and Geometric Variables	$\boldsymbol{a}_i$	Acceleration of vehicle i
	$d_{s,n}$	Distance between the subject (s) and neighbouring (n) vehicles
	$\dot{d}_{s,n}$	Distance changing rate between the subject (s) and neighbouring (n) vehicles
	$\boldsymbol{p}_i$	Position of vehicle i
	$\boldsymbol{p}_{il}, \boldsymbol{p}_{ir}$	Left and right reference points on vehicle i (in this chapter at the front for the subject and at the rear for the neighbouring vehicle)
	$\boldsymbol{v}_i$	Velocity of vehicle i
	$\Delta\boldsymbol{v}_{i,a}$	Acceleration-based velocity of vehicle i based on the known acceleration
	$\Delta\boldsymbol{v}_{i,u}$	Uncertain velocity of vehicle i based on manoeuvre uncertainties
	$\boldsymbol{v}_i'$	Perceived velocity of vehicle i taking into account $\boldsymbol{v}_i$, $\Delta\boldsymbol{v}_{i,u}$ and $\Delta\boldsymbol{v}_{i,a}$
	$\boldsymbol{\theta}_{sj_1,nj_2}$ **	Bearing between reference points on subject (s) and neighbouring (n) vehicles
	$\dot{\boldsymbol{\theta}}_{sj_1,nj_2}$	Bearing rate between reference points on subject (s) and neighbouring (n) vehicles
	φ	Heading angle
	$\mathbf{X}_i$	State vector of vehicle i
	X, Y	Longitudinal and lateral directions of the coordinate system
	L, W	Length and width of a vehicle
Avoidance Difficulty Computation	$\boldsymbol{V}$	Safe velocity set
	$\boldsymbol{V}'$	Safe velocity set considering uncertainties and known acceleration
	$\boldsymbol{v}_g$	Velocity gap for collision avoidance difficulty derived as distance between $\boldsymbol{v}_s'$ and $\boldsymbol{V}'$
	$\boldsymbol{v}_{s,\boldsymbol{V}}$ $\boldsymbol{v}_{s,\boldsymbol{V}'}$	Subject velocity in the safe velocity set $\boldsymbol{V}$ and $\boldsymbol{V}'$
	r_i	Direction of the uncertain velocity of vehicle i
	$\mathscr{D}$	Probability density function of a truncated Gaussian distribution
	fb, bb, lb, rb	Forward, backward, left and right bounds of the uncertainty velocity in the probability density function of a truncated Gaussian distribution
	$\boldsymbol{N}$	Probability density function of a normal Gaussian distribution
	$\mathscr{N}$	Cumulative distribution function of a normal Gaussian distribution
	l	Length of an uncertain velocity vector
	$\mathscr{P}$	Conditional probability of an uncertain velocity vector
PCAD Model Components	$R_{PCAD}(t)$	Perceived risk function in PCAD model
	$\mathscr{V}_i$	Perceived velocity function for vehicle i
	$\mathscr{A}$	Avoidance difficulty function
	$\mathscr{W}$	Weighting function
PCAD Model Parameters	$\sigma_{i,X}, \sigma_{i,Y}$	Standard deviations for uncertain velocity distributions of vehicle i
	$t_{i,a}$	Anticipated time for acceleration-based velocity of vehicle i
	α	The exponent of the Weighting function.
	β	The mass ratio $\beta = \frac{M_n}{M_s+M_n}$. M_s and M_n represent the mass of the subject (s) and neighbouring (n) vehicles.
	v_{ref}	A reference velocity in the weighting function $\mathscr{W}$, which can be a velocity limit under specific conditions.

* $i \in \{s, n\}$ with s representing the subject vehicle and n representing the neighbouring vehicle.
** j_1, j_2 represent the numbering of different reference points, which can be 1, 2, 3, ...

3.1. INTRODUCTION

Road crashes are a leading cause of injury and death worldwide, resulting in approximately 1.35 million deaths and 20-50 million non-fatal injuries each year [113]. Most traffic accidents arise from human misjudgements [114]. Specifically, distorted perception of driving risk by human drivers is one of the important causes of road accidents [115].

Perceived risk captures the level of risk experienced by drivers, which can differ from operational (or actual) risk [41], [53]. A low perceived risk leads to feeling safe, relaxed, and comfortable, while a high-risk perception results in cautious behaviour [53]. The advent of active safety and driving automation systems has reduced actual risk, but changes in drivers' risk perception have been observed. Human drivers will inversely perceive a high level of risk if the driving automation shows inappropriate driving behaviours, causing decreased trust, low acceptance, and even refusal of vehicle automation [9]. In manual driving, maintaining perceived risk below a specific threshold motivates drivers' actions, such as steering and braking [116]. Consequently, misperception of risk during automated driving may cause drivers' to distrust and intervene unnecessarily while in other cases drivers may fail to recognise dangerous situations that require drivers' intervention. Therefore, it is essential to comprehend and quantify drivers' perceived risk in driving automation and in turn, use it to design driving automation which is not only technically safe, but is also perceived as safe.

Computational models for estimating perceived risk have been developed, falling into two categories: empirical models reliant on data, and mechanistic models grounded in first principles. In the first category, Kolekar *et al.* [41] established a driving risk field (DRF) model considering the probability of an event occurring and the event consequence based on drivers' subjective risk ratings and steering responses. Ping *et al.* [38] used deep learning methods to model perceived risk in urban scenarios with factors related to the subject vehicle and the driving environment. Chapter 2 [117] built a regression-based perceived risk model to explain and compute event-based perceived risk in highway merging and braking scenarios. Among other factors, the model captures the influence of relative motion with respect to other road users on drivers' subjective perceived risk ratings.

Mechanistic perceived risk models typically rely on surrogate measures of safety (SMoS). The minimum time to collision (TTC) can show the drivers' threshold of perceived risk when they take last-moment braking actions [39]. The inverse TTC represents drivers' relative visual expansion of an obstacle, which can indicate drivers' risk perception [79]. Additionally, Kondoh *et al.* [103] and Kondoh *et al.* [118] further analysed the relationship between drivers' risk perception regarding the leading vehicle and inverse TTC and time headway (THW) in car-following situations. Models using TTC and THW only capture one-dimensional (1D) interaction and are mainly validated for car following. Attempts have been made to model risk for two-dimensional (2D) motion based on the driving risk field theory [119] and develop collision warning algorithms [120]. This research line is advanced by the probabilistic driving risk field model (PDRF) [40] by considering motion probability distributions of other road users and the collision severity to estimate the collision risk. Although the above-mentioned models estimate the actual collision risk rather than the perceived risk, they are promising to predict human drivers' risk per-

ception thanks to the strong connection between the actual risk and the perceived risk.

The empirical models reviewed above accurately quantify perceived risk in certain scenarios, but lack validation across diverse situations and are not fully explainable. Mechanistic models, while explainable, can compute the actual risk. However, the mapping between the actual collision risk and perceived risk remains ambiguous and the thresholds of the SMoSs lack empirical support. Hence, an explainable and validated computational perceived risk model is still lacking.

This chapter has two primary objectives: **Objective 1** is to formulate an explainable computational perceived risk model grounded in the human drivers' risk perception mechanism applicable to general 2D movements. **Objective 2** is to analyse and compare our new model against existing models both theoretically and empirically. The model uses the velocity gap to the safe velocity region as the potential collision avoidance difficulty to quantify perceived risk. The safe velocity region accounts for vehicles' kinematics with uncertainty, as well as collision severity. The model describes perceived risk in continuous time and per event, and is validated using event-based reported perceived risk. The model is developed towards the general driver population instead of personalised modelling but can capture individual differences by tuning model parameters.

The remainder of this paper is structured as follows. We first revisit three computational perceived risk models from literature in Section 3.2, and then present the formulation of the new model in Section 3.3. Perceived risk data, model calibration approach and model performance indices are introduced in Section 3.5. The model evaluation results are represented in Section 3.6 followed by a discussion in Section 3.7, and conclusions in Section 3.8.

3.2. RELATED PERCEIVED RISK MODELS

This section introduces the preliminaries for perceived risk modelling and three baseline models for comparison and performance evaluation, while referring to Appendix B.1 for details.

3.2.1. COORDINATE SYSTEM, REFERENCE POINTS DEFINITION AND VEHICLE MODEL

All models in this chapter employ the same coordinate system. The road space is modelled as a flat Euclidian plane. The X-axis aligns with the direction of the road, while the Y-axis is perpendicular to it, oriented counter-clockwise, as illustrated in Figure 3.1. Given our focus on perceived risk based on relative motion, rather than vehicle dynamics, we employ a simple point mass model incorporating vehicle dimension. According to the point mass model, the positions, velocities and accelerations of the geometric centre for both the subject vehicle s and a neighbouring vehicle or an obstacle n are $\boldsymbol{p}_s = [x_s, y_s]^T$, $\boldsymbol{p}_n = [x_n, y_n]^T$, $\boldsymbol{v}_s = [v_{s,X}, v_{s,Y}]^T$, $\boldsymbol{v}_n = [v_{n,X}, v_{n,Y}]^T$, $\boldsymbol{a}_s = [a_{s,X}, a_{s,Y}]^T$, $\boldsymbol{a}_n = [a_{n,X}, a_{n,Y}]^T$ respectively. The heading angle ψ follows from the vehicle velocity direction for the point mass model.

Vehicle dimensions are incorporated into the perceived risk models. Figure 3.1 illustrates that the leftmost and the rightmost points in the front side of the subject vehicle and the rear side of the neighbouring vehicle that is closest to the subject vehicle are the

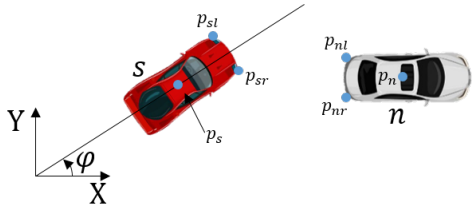


Figure 3.1: The coordinate system in the definition

reference points in this case. Given the vehicle's length L and width W in straight drive, the positions of the reference points are $\boldsymbol{p}_{sl} = \boldsymbol{p}_s + [L/2 \;\; W/2]^T$ and $\boldsymbol{p}_{sr} = \boldsymbol{p}_s + [L/2 \;\; -W/2]^T$ for the subject vehicle, $\boldsymbol{p}_{nl} = \boldsymbol{p}_s + [-L/2 \;\; W/2]^T$ and $\boldsymbol{p}_{nr} = \boldsymbol{p}_s + [-L/2 \;\; -W/2]^T$ for the neighbouring vehicle.

3.2.2. EXISTING PERCEIVED RISK MODELS

In this section, we briefly revisit three perceived risk models that are fundamental to understanding driver's perceived risk in different driving scenarios. These models, including the Regression Perceived Risk Model (RPR) [117], the Perceived Probabilistic Driving Risk Field Model (PPDRF) [40], and the Driving Risk Field Model (DRF)[41], offer diverse approaches to quantifying and analysing perceived risk. For a comprehensive overview of these models, including their key assumptions and mathematical definitions, please refer to Table 3.1 with more details in Appendix B.1 and corresponding literature.

Table 3.2, summarises model features and factors used in risk calculation. RPR and DRF are validated but do not take into account all factors known to be relevant in risk perception. PPDRF takes into account all listed factors, but its parameters are not based on empirical data and it has not been validated. Hence, this paper presents the new Potential Collision Avoidance Difficulty (PCAD) model which is inspired by the three existing models and validates the resulting four models with the two available perceived risk datasets used to develop and validate the RPR and the DRF model.

3.3. POTENTIAL COLLISION AVOIDANCE DIFFICULTY MODEL (PCAD)

Our proposed model is grounded in Fuller's Risk Allostasis Theory which proposes that a feeling of risk can be indicated by the driving task difficulty [78] and drivers' primary driving task is to perform avoidance actions to moderate the perceived risk to a preferred range [121]. Consequently, we develop a dynamic perceived risk model by quantifying the driving task difficulty, which computes real-time perceived risk and explain its underlying mechanism. We quantify the task difficulty considering the 2D velocity change to avoid a potential collision. Additionally, the model accounts for the behaviour uncertainties of other road users reflected in their velocities and imprecision in longitudinal and lateral control as motion uncertainties of the subject vehicle. In this section, we introduce the primary assumptions and the general structure of the model followed by a detailed explanation of each component, including the potential collision judgement method and the perceived velocity of a neighbouring vehicle and the subject vehicle, and a weighting function that considers the collision severity.

Table 3.1: Introduction of existing perceived risk models

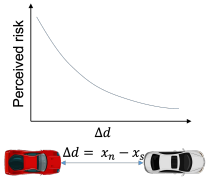
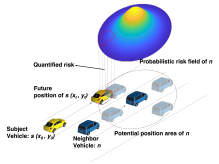
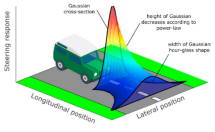
Model	**Regression Perceived Risk Model (RPR)[117]**	**Perceived Probabilistic Driving Risk Field Model (PPDRF)[40]**	**Driving Risk Field Model (DRF)[41]**
Introduction	Event-based model derived from a simulator experiment involving 18 merging and braking event types on a 2-lane highway.	Perceived Probabilistic Driving Risk Field Model accounting for behaviour uncertainties of surrounding vehicles in 2D based on predicted collision probability and collision severity.	Represents human drivers' risk perception as a 2-dimensional field combining probability and consequence of an event.
Assumptions	• Perceived risk stems from vehicles directly in front after entering the lane. • Drivers can accurately estimate motion information.	• Uncertainties of neighbouring vehicles are represented by independent Gaussian distributions of 2D accelerations. • The subject vehicle maintains the current acceleration over the prediction horizon.	• Perceived risk is the product of the probability of a hazardous event and its severity. • The risk field widens with distance and decays with lateral and longitudinal distance. • The height of the perceived risk field decays as the lateral and longitudinal distance from the vehicle increases.
Definitions	The perceived risk is calculated using the longitudinal position of the neighbouring vehicle and the subject vehicle, along with the current acceleration of the neighbouring vehicle. The perceived risk is given as Equation (B.2).	• Total perceived risk is calculated as a sum of kinetic risk and potential risk as Equation (B.3). • Kinetic risk is given by Equation (B.4), involving the subject mass, mass ratio, relative velocity, and estimated collision probability. • Potential risk is modelled as Equation (B.6), involving the subject mass, relative velocity and the distance to the obstacle.	• The overall perceived risk is quantified as the product of the probability and the severity of events at different positions relative to the vehicle as shown in Equation (B.7). • The probability field is modelled as a Gaussian distribution in the lateral direction in Equation (B.8) • The height of the probability field is a function of the longitudinal distance as shown in Equation (B.9), and the width of the probability field increases linearly with the distance to the obstacle, reflecting widening of the risk field as shown in Equation (B.10).
Risk diagram			

Table 3.2: Model features and factors used in risk calculation

	RPR	PPDRF	DRF	PCAD
Dimension	1-D	2-D	2-D	2-D
Distance	●	●	●	●
Using relative velocity	-	●	-	●
Using acceleration	●	●	-	●
Using subject velocity	-	●	●	●
Considering crash consequence	-	●	●	●
Considering manoeuvre uncertainties	-	●	-	●
Usable on curved lanes	-	●	●	●

●indicate "yes" and – indicate "no".

3.3.1. ASSUMPTIONS

To operationalise the model, we adopt several simplifying assumptions:

- **Assumption 1**: Human drivers perceive risk based on an estimation of the difficulty in avoiding a potential collision according to their visual perception of the relative motion of the subject vehicle and neighbouring vehicles [78], [121], [122]. They judge whether a vehicle is on a collision course based on looming [123].
- **Assumption 2**: The known acceleration and manoeuvre uncertainties of neighbouring and subject vehicles cause extra perceived risk. The latter is presented as an uncertain acceleration following a specific probability distribution. In this chapter, we assume a Gaussian distribution with zero means. This is grounded in existing literature. In stable highway driving, the longitudinal and lateral accelerations of a neighbour follow a Gaussian distribution [124]. Specifically, Ko *et al.* [125] observed that in a vehicle field test with GPS, the modelling results of acceleration as a response variable indicated that it followed a Gaussian distribution. Additionally, Jansson [126] argued that a constant acceleration model is sufficiently accurate for tracking vehicle motion. Both the known acceleration and this uncertain acceleration will remain constant in a short period of time.
  - **2a**: The known accelerations of the subject and neighbouring vehicles influence perceived risk [117], [127], [128].
  - **2b**: The uncertain acceleration of neighbouring vehicles comes from a potential manoeuvre change (e.g., a sudden brake or steer) [129], [130].
  - **2c**: The uncertain acceleration of the subject vehicle is caused by imprecise control in steering and throttle/braking, which is relevant to human drivers' control ability or driving automation's performance [131].
- **Assumption 3**: Human drivers perceive higher perceived risk with higher subject vehicle velocity [132].

3.3.2. GENERAL STRUCTURE OF PCAD

Let $\mathbf{X}_s = (\boldsymbol{p}_s, \boldsymbol{v}_s, \boldsymbol{a}_s)^T$ and $\mathbf{X}_n = (\boldsymbol{p}_n, \boldsymbol{v}_n, \boldsymbol{a}_n)^T$ denote the state of the subject vehicle s and the neighbouring vehicle n respectively, with $\boldsymbol{p}_s$ and $\boldsymbol{p}_n$, $\boldsymbol{v}_s$ and $\boldsymbol{v}_n$, $\boldsymbol{a}_s$ and $\boldsymbol{a}_n$ being the

position, velocity and acceleration vectors, and T the transpose of a vector. The PCAD is formulated as Equation (3.1)

$$R_{PCAD}(t) = \mathscr{A}(\boldsymbol{p}_s, \boldsymbol{p}_n, \mathscr{V}_s(\boldsymbol{x}_s, \boldsymbol{x}_n), \mathscr{V}_n(\boldsymbol{x}_s, \boldsymbol{x}_n)) \cdot \mathscr{W}(\boldsymbol{v}_s) \tag{3.1}$$

Here, $\mathscr{A}$ represents the avoidance difficulty function. This function quantifies the required 2D velocity change to bring the subject vehicle to the safe velocity region in the velocity domain to avoid a potential collision with the neighbouring vehicle, considering factors such as their relative positions, velocities and accelerations. $\mathscr{V}_i$ denotes the 2D perceived velocity for vehicle $i \in \{s, n\}$, thereby capturing absolute and relative motion of the interacting vehicles. Finally, $\mathscr{W}$ is the weighting function, being a Power function with v_s, which accounts for the influence of the subject vehicle's speed on perceived risk. Higher speeds generally increase the perceived risk, as the consequence of a potential collision is more severe.

The perceived velocity function $\mathscr{V}$ can be represented as

$$\boldsymbol{v}_i' = \mathscr{V}_i(\mathbf{X}_s, \mathbf{X}_n) = \boldsymbol{v}_i + \Delta\boldsymbol{v}_{i,a} + \Delta\boldsymbol{v}_{i,u} \tag{3.2}$$

where $\mathscr{V}_i$ is the functional operator to compute the perceived velocity $\boldsymbol{v}_i'$ of the vehicle $i \in \{s, n\}$ by human drivers for perceived risk computation. The perceived velocity combines three components: the velocity $\boldsymbol{v}_i$ ($i \in \{s, n\}$), an acceleration-based velocity $\Delta\boldsymbol{v}_{i,a}(i \in \{s, n\})$ that accounts for the influence of the known acceleration (**Assumption 2a**) and an uncertain velocity $\Delta\boldsymbol{v}_{i,u}$ that accounts for uncertainties in vehicle motion (**Assumption 2b** and **Assumption 2c**). For example, consider a driver who notices that a car ahead is braking rapidly. The driver might perceive the car's velocity to be lower than it actually is because the driver anticipates its future motion based on the acceleration. The uncertain velocity component captures uncertainties in vehicle motion, such as a neighbouring vehicle suddenly swerving or the subject vehicle's imprecise control.

3.3.3. COLLISION AVOIDANCE DIFFICULTY FUNCTION $\mathscr{A}$ IN DETERMINISTIC CONDITIONS

In this section, the collision avoidance difficulty is formulated to capture part of human drivers' perceived risk under constant speed and deterministic motion conditions. The perceived velocity (Equation (3.2)) relaxes to the actual velocity $\boldsymbol{v}$ under such conditions. Uncertainties and acceleration are incorporated in the next section.

POTENTIAL COLLISION JUDGEMENT OF HUMAN DRIVERS —LOOMING DETECTION

A precedent step for collision avoidance is to detect a potential collision based on the current environment information. One observation from aircraft pilots is that two aircraft are on a crossing course if they remain in the same position in their field of view. Similarly, in road traffic, one vehicle lies on a crossing course at a specific moment, if its relative bearing to you does not change [123]. Additionally, if the vehicle is simultaneously approaching, a phenomenon known as *looming* is occurring. This situation indicates a risk of collision (see two vehicles in interaction in Figure 3.2). To identify this phenomenon and anticipate a potential collision is referred to as *looming detection*.

Our method of looming detection combines the two criteria introduced above, requiring that: **(i)** the bearing $\boldsymbol{\theta}$ of a neighbouring vehicle remains constant (see Figure

3.2b), and **(ii)** the distance between the two vehicles is decreasing. Figure 3.2b illustrates the bearing $\boldsymbol{\theta}$ considering each vehicle as a single point, but to detect looming we must also consider vehicle size. Here we approximate vehicle shape by rectangles and use corners as reference points. In the side impact example in Figure 3.2c the relevant reference points are the front left and front right of the subject vehicle and the front left and rear left of the neighbouring vehicle. We consider four interactions for looming detection. Figure 3.2c illustrates the four relevant interactions between corners $\boldsymbol{p}_{sl}/\boldsymbol{p}_{nl}$, $\boldsymbol{p}_{sl}/\boldsymbol{p}_{nr}$, $\boldsymbol{p}_{sr}/\boldsymbol{p}_{nl}$, and $\boldsymbol{p}_{sr}/\boldsymbol{p}_{nr}$ respectively. From the perspective of the subject vehicle, the left reference point on the neighbouring vehicle moves to the left (anticlockwise), but the right one moves to the right (clockwise). At an intermediate point the heading rate will be zero, representing a collision. Meanwhile, the distance between the two vehicles is decreasing and hence this is a looming case. Alternatively if both points would move to the left the subject vehicle would pass at the right, and if both points would move to the right the subject vehicle would pass at the left.

In this chapter, the reference points are chosen at the front left and right on the subject vehicle and the rear left and right on the neighbouring vehicle (Figure 3.3). This simplification is justified since the datasets used contain only obstacle avoidance events and merging events in the front of the subject vehicle.

In our method, the relative bearing rate $\dot{\boldsymbol{\theta}}_{sj_1,nj_2}$ of four pairs of reference points on the subject vehicle and the neighbouring vehicle is calculated using Equation (3.3)[1] (See Figure 3.2c for more details).

$$\dot{\boldsymbol{\theta}}_{sj_1,nj_2} = \frac{\left(\boldsymbol{p}_{sj_1} - \boldsymbol{p}_{nj_2}\right) \times \left(\boldsymbol{v}_{sj_1} - \boldsymbol{v}_{nj_2}\right)}{\left\|\boldsymbol{p}_{sj_1} - \boldsymbol{p}_{nj_2}\right\|^2}, \quad j_1, j_2 \in \{l, r\} \tag{3.3}$$

Looming is indicated when the product of the minimum and maximum values of $\dot{\boldsymbol{\theta}}_{sj_1,nj_2}$ is negative (one is positive and one is negative), as shown in Equation (3.4) and Figure 3.2c.

$$\min \dot{\boldsymbol{\theta}}_{sj_1,nj_2} \cdot \max \dot{\boldsymbol{\theta}}_{sj_1,nj_2} < 0, \quad j_1, j_2 \in \{l, r\}, \tag{3.4}$$

The second criterion for *looming* is that the neighbouring vehicle is approaching the subject vehicle. That is, a neighbouring vehicle may only collide with the subject vehicle if it is getting closer. This is assessed by examining the distance changing rate between the two vehicles (centre), defined by Equation (3.5) and its derivative in Equation (3.6). A negative rate indicates that the neighbouring vehicle is approaching.

$$d_{s,n} = \sqrt{(\boldsymbol{p}_s - \boldsymbol{p}_n)^T (\boldsymbol{p}_s - \boldsymbol{p}_n)} \tag{3.5}$$

$$\dot{d}_{s,n} = \frac{1}{d_{s,n}} (\boldsymbol{p}_s - \boldsymbol{p}_n)^T (\boldsymbol{v}_s - \boldsymbol{v}_n) < 0 \tag{3.6}$$

Considering the two criteria, if Equations (3.4) and (3.6) are satisfied at the same time, the neighbouring vehicle is looming (Figure 3.2). Conversely, if Equations (3.4) and (3.6)

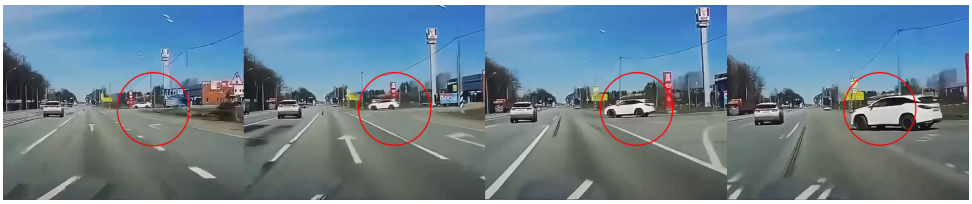
[1] In straight driving, the velocity of reference points $\boldsymbol{v}_{i,j}$ ($i \in \{s, n\}, j \in \{l, r\}$) can be simplified as the vehicle's linear velocity $\boldsymbol{v}_i$ ($i \in \{s, n\}$) without considering vehicle's yaw rate.

are not met simultaneously, the neighbouring vehicle is classified as non-looming (Figure 3.3), namely

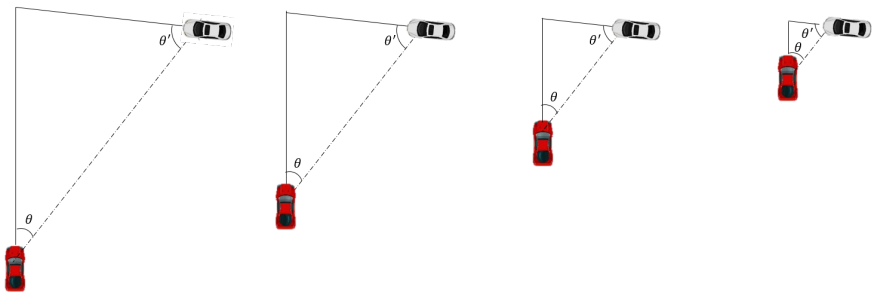
$$\min \dot{\boldsymbol{\theta}}_{sj_1,nj_2} \cdot \max \dot{\boldsymbol{\theta}}_{sj_1,nj_2} \geqslant 0, \quad j_1, j_2 \in \{l, r\}, \tag{3.7}$$

or

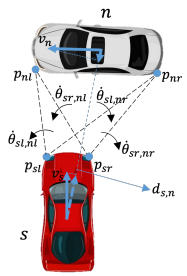
$$\dot{d}_{s,n} = \frac{1}{d_{s,n}} (\boldsymbol{p}_s - \boldsymbol{p}_n)^T (\boldsymbol{v}_s - \boldsymbol{v}_n) \geqslant 0, \tag{3.8}$$



(a) Video stream for a potential collision. The neighbouring vehicle (white) stays at the same bearing (the red circle) and becomes larger when on a crossing course with the subject vehicle [133]. Here the bearing is the orientation in the field of view in which another object is observed.

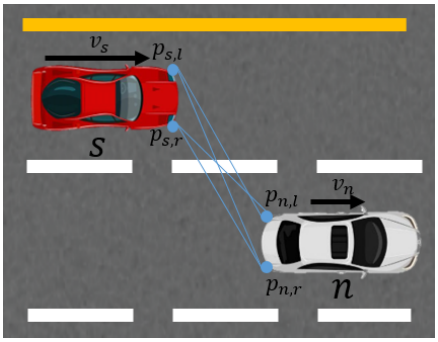


(b) Bird eye view for the potential collision above. The bearings θ and θ' of subject vehicle (red) and neighbouring vehicle (white) remain the same when they are on a crossing course.

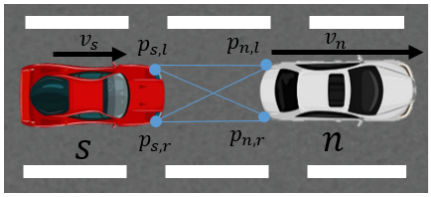


(c) Details for the potential collision above. From the perspective of the subject vehicle, the left reference point on the neighbouring vehicle moves to the left (anticlockwise), but the right one moves to the right (clockwise). Meanwhile, the distance between the two vehicles is decreasing. This case meets Equation (3.4) and Equation (3.6) simultaneously, indicating that it is a looming case.

Figure 3.2: An example of looming. The subject vehicle and a neighbouring vehicle are on a crossing course.



(a) The subject vehicle (Red) is overtaking a neighbouring vehicle (white). Although the neighbour vehicle (White) approaches the subject vehicle (red), the orientation of it in the subject vehicle's view field is not staying constant (all bearing rate regarding all pairs of reference points $\dot{\theta}$ is negative, meaning that the neighbouring vehicle is rotating around the subject vehicle (clockwise)). The situation meets Equation (3.6) but does not meet Equation (3.4), indicating that it is a non-looming case.



(b) The subject vehicle (red) is following a leading vehicle (white) with a lower speed. Although the orientation of the neighbouring vehicle stays the same in the subject vehicle's view field ($\dot{\theta}_{sl,nr}$ is positive but $\dot{\theta}_{sr,nl}$ is negative), it is not approaching the subject vehicle. The situation meets Equation (3.4) but does not meet Equation (3.6), indicating that it is a non-looming case.

Figure 3.3: Two examples of non-looming.

Note that we examine four pairs of reference points in Equation (3.3) and Equation (3.4) for a simpler computation. In general circumstances, reference points for collision detection are ideally positioned at the four corner points of each vehicle. If the computation capability permits, we can examine all 16 pairs of reference points for Equation (3.3) and Equation (3.4). Additionally, *Looming detection* is directly valid when the subject vehicle only has translational motion with constant acceleration and thereby follows a straight path. When the subject vehicle has a yaw rate ($\dot{\varphi}$), and follows a curved path the theory still stands based on a conformal mapping.

COLLISION AVOIDANCE DIFFICULTY

We define a safe velocity set $\boldsymbol{V}$, which comprises all non-looming subject velocity vectors that meet Equation (3.7) and/or (3.8) based on the position of the two vehicles $\boldsymbol{p}_s$, $\boldsymbol{p}_s$ and the velocity of the neighbouring vehicle $\boldsymbol{v}_n$ at the current moment. The safe velocity set $\boldsymbol{V}$ is defined in Equation (3.9).

$$\forall \boldsymbol{v}_s \in \boldsymbol{V} \Rightarrow \min \dot{\boldsymbol{\theta}}_{sj_1,nj_2} \cdot \max \dot{\boldsymbol{\theta}}_{sj_1,nj_2} \geqslant 0 \,(j_1, j_2 \in \{l, r\},) \text{ or } \dot{d}_{s,n} \geqslant 0, \tag{3.9}$$

where the equality holds when $\boldsymbol{v}_s$ is at the boundary of velocity set $\boldsymbol{V}$.

The collision avoidance difficulty $||\boldsymbol{v}_g||$ is defined as the 2D distance from the current subject velocity $\boldsymbol{v}_s$ to the nearest point on the boundary of the safe velocity set $\boldsymbol{V}$, which is the end point of the vector denoted as $\boldsymbol{v}_{s,V}$ (Equation (3.10)) (See Figure 3.4 for an illustration). Hence, the collision avoidance function $\mathscr{A}$ is defined as

$$||\boldsymbol{v}_g|| = \mathscr{A} = ||\boldsymbol{v}_{s,V} - \boldsymbol{v}_s|| \tag{3.10}$$

where $\boldsymbol{v}_{s,V}$ is the vector in the safe velocity set $\boldsymbol{V}$, the end point of which is closest to the subject velocity vector $\boldsymbol{v}_s$, satisfying

$$\boldsymbol{v}_{s,V} = \arg\min_{\boldsymbol{v}\in\boldsymbol{V}} ||\boldsymbol{v} - \boldsymbol{v}_s|| \tag{3.11}$$

$\boldsymbol{v}_g$ represents the vector pointing from the current subject velocity $\boldsymbol{v_s}$ towards $\boldsymbol{v}_{s,V}$, indicating the direction and magnitude of the adjustment needed to reach the safe velocity set $\boldsymbol{V}$ from the current velocity $\boldsymbol{v_s}$. If the current subject velocity $\boldsymbol{v}_s$ already lies within the safe velocity set $\boldsymbol{V}$, then the velocity gap $||\boldsymbol{v}_g||$ is zero, implying no collision avoidance difficulty.

In this chapter, the technique of grid search is employed to identify $\boldsymbol{v}_g$, ensuring compliance with both Equation (3.10) and Equation (3.11).

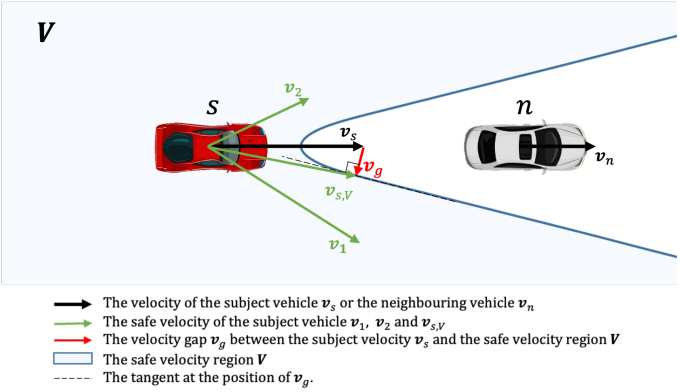


Figure 3.4: An example to show the collision avoidance difficulty. In this case, the subject vehicle (red) is following a leading vehicle (white, 50 m ahead, $\boldsymbol{v}_n = 8.33$ m/s) with a higher velocity ($\boldsymbol{v}_s = 16.67$ m/s). Equation (3.7) and (3.8) define the safe velocity set $\boldsymbol{V}$ as the blue area, e.g., if the current subject velocity is any one of the elements in $\boldsymbol{V}$ (e.g., $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$), the neighbouring vehicle (white) is not looming, and the collision avoidance difficulty is zero. In this example, since the subject vehicle is driving faster than the leading vehicle, the current subject velocity $\boldsymbol{v_s} \notin \boldsymbol{V}$, indicating that the neighbouring is looming regarding the subject vehicle. The distance from the subject velocity $\boldsymbol{v_s}$ to the safe velocity set $\boldsymbol{V}$ (the safety boundary) is $\boldsymbol{v}_g$ (the red arrow), the length of which is the defined collision avoidance difficulty.

3.3.4. PERCEIVED VELOCITY FUNCTION $\mathcal{V}_i$ OF A NEIGHBOURING VEHICLE AND THE SUBJECT VEHICLE CONSIDERING KNOWN ACCELERATION AND MANOEUVRE UNCERTAINTIES

The collision avoidance difficulty calculated using the actual (deterministic) motion information (i.e., $\boldsymbol{p}_i$ and $\boldsymbol{v}_i$) (Equation 3.10) presented in the previous section can already account for most of the perceived risk, which is shown in Section 3.6 (Figure 3.10(e) and Figure 3.11(e)). However, these calculations overlook how human drivers process environmental information considering known accelerations and uncertainties. In this section, we define a perceived velocity function denoted as $\mathcal{V}$ shown in Equation (3.2) for both the subject and the neighbouring vehicles. This function is based on a more comprehensive understanding, which outputs a perceived velocity $\boldsymbol{v}_i'$ consisting of three

components: the actual velocity $\boldsymbol{v}_i$, the velocity derived from the known acceleration $\boldsymbol{v}_{i,a}$ and the velocity derived from manoeuvre uncertainties $\boldsymbol{v}_{i,u}$. The perceived velocity yields an adjusted safe velocity set $\boldsymbol{V}'$ and thereby a new velocity gap $\boldsymbol{v}_g$.

THE PERCEIVED VELOCITY

The perceived velocity $\boldsymbol{v}_i'$ is the final output of perceived velocity function $\mathcal{V}$ based on the state of the subject vehicle $\mathbf{X}_s = (\boldsymbol{p}_s, \boldsymbol{v}_s, \boldsymbol{a}_s)^T$ and the state of the neighbouring vehicle $\mathbf{X}_n = (\boldsymbol{p}_n, \boldsymbol{v}_n, \boldsymbol{a}_n)^T$, which consists of the actual velocity $\boldsymbol{v}_i$, the known acceleration-based velocity $\Delta\boldsymbol{v}_{i,a}$ and the uncertain velocity $\Delta\boldsymbol{v}_{i,u}$ as shown in Equation (3.2). This integrated perceived velocity function considers the acceleration and uncertainties, thus contributing to extra perceived risk. Figure 3.5 illustrates the relationship between the actual velocity $\boldsymbol{v}_i$, the uncertain velocity $\Delta\boldsymbol{v}_{i,u}$, the acceleration-based velocity $\Delta\boldsymbol{v}_{i,a}$ and the final perceived velocity $\boldsymbol{v}_i'$. The perceived velocity $\boldsymbol{v}_i'$ is utilised for computing perceived risk. The known acceleration-based velocity $\Delta\boldsymbol{v}_{i,a}$ and the uncertain velocity $\Delta\boldsymbol{v}_{i,u}$ will be detailed in Section 3.3.4 and 3.3.4 below.

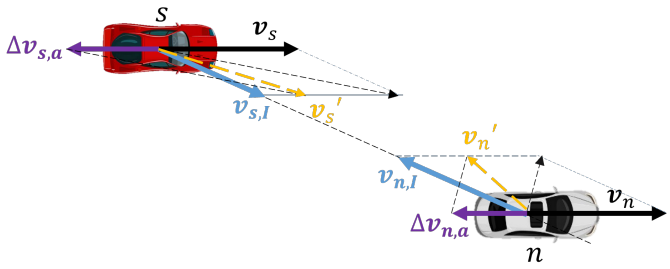


Figure 3.5: The relationship between the actual velocity $\boldsymbol{v}_i$ the uncertain velocity $\Delta\boldsymbol{v}_{i,u}$, the acceleration caused velocity change and the perceived velocity $\boldsymbol{v}_i'$ ($i \in \{s, n\}$) of the subject vehicle s and the neighbouring vehicle n. In this case, the subject vehicle (red) is passing by a neighbouring vehicle (white). Both vehicles are decelerating causing acceleration-based velocities $\Delta\boldsymbol{v}_{s,a}$ and $\Delta\boldsymbol{v}_{n,a}$ (the purple arrows). The uncertain velocities $\Delta\boldsymbol{v}_{s,u}$ and $\Delta\boldsymbol{v}_{n,u}$ are pointing to each other. The final perceived velocity $\boldsymbol{v}_s'$ and $\boldsymbol{v}_n'$ contain the contribution of the acceleration-based velocity and the uncertain velocity.

THE VELOCITY COMPONENT DERIVED FROM KNOWN ACCELERATION — THE ACCELERATION-BASED VELOCITY

Previous studies have shown that human drivers consider the acceleration of the subject and other vehicles during driving [127]. A collision avoidance behaviour model for drivers achieves 20% more accuracy when the acceleration of other vehicles is considered [128]. For example, the action of braking by a leading vehicle can initially cause perceived risk to the following subject vehicle, even if the distance between the vehicles does not close rapidly at the initial stage. The perceived risk may decrease once the subject vehicle also brakes, even as the gap between the vehicles continues to decrease.

To account for the influence of known acceleration (**Assumption 2a**), we introduce a component to the perceived velocity (Equation (3.2), which reflects human drivers' anticipation of velocity based on the current known acceleration. We name this component as acceleration-based velocity represented by Equation (3.12).

$$\Delta \boldsymbol{v}_{i,a} = \boldsymbol{a}_i \cdot t_{i,a}, \quad i \in \{s, n\} \tag{3.12}$$

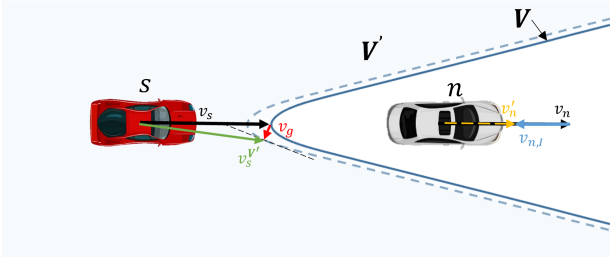
where $\Delta \boldsymbol{v}_{i,a}$ represents the component of perceived velocity caused by known acceleration; $\boldsymbol{a}_i$ is the current acceleration, and $t_{i,a}$ is an anticipated time for computation that varies for the subject vehicle and the neighbouring vehicle. $t_{i,a}$ is determined by calibration. The impact of $t_{i,a}$ duration on the model behaviour depends on the known acceleration direction. If the known acceleration tends to decrease the gap between two vehicles, a longer anticipated time results in a higher perceived risk output by the model, and vice versa.

THE VELOCITY COMPONENT DERIVED FROM MANOEUVRE UNCERTAINTIES — THE UNCERTAIN VELOCITY

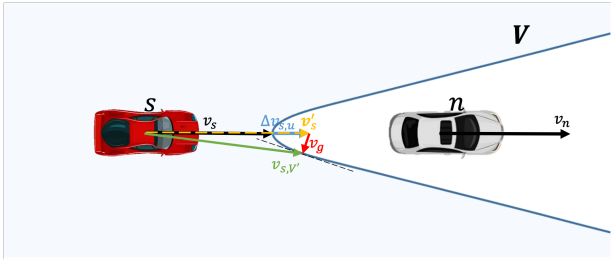
Assumption 2b and **Assumption 2c** specify that manoeuvre uncertainties cause additional perceived risk. For instance, when we pass by a car in the adjacent lane, we unconsciously shift to the other side of the lane to keep away from the car for safety because the velocity of the other car can suddenly change [130]. Accordingly, we define an uncertain velocity perceived by human drivers based on the manoeuvre uncertainties as a component of the perceived velocity (Equation (3.2)).

The uncertain velocity in human driver's mind caused by the uncertainties of each vehicle in interaction makes the situation being perceived as more dangerous. Figure 3.6 shows an example of the uncertain velocity and its influence on the velocity set $\boldsymbol{V}$.

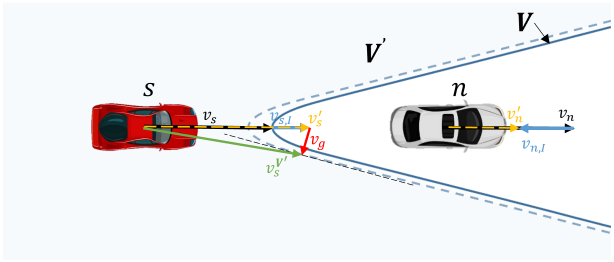
The uncertain velocity exists in all directions on both subject and neighbouring vehicles, but its impact for different directions on perceived risk varies. We assume that acceleration with a direction reducing the distance between vehicles most strongly increases perceived risk. Hence we only consider this direction in the perceived risk model, which also reduces computational complexity. This direction is illustrated in Figure 3.5, with a detailed explanation in B.4.



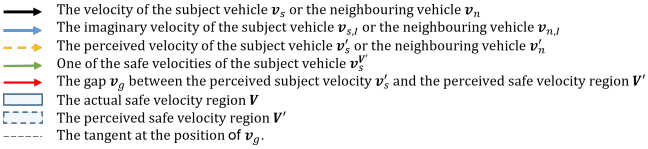
(a) The uncertain velocity $\Delta \boldsymbol{v}_{n,u}$, the final perceived velocity of the neighbouring vehicle $\boldsymbol{v}_n'$ and their influence on safe velocity set $\boldsymbol{V}$. Note that $\Delta \boldsymbol{v}_{s,a}$, $\Delta \boldsymbol{v}_{s,u}$ and $\Delta \boldsymbol{v}_{n,a}$ are not taken into account to show only the influence of the uncertain velocity $\Delta \boldsymbol{v}_{n,u}$ on the safe velocity set $\boldsymbol{V}$ and finally the velocity gap $\boldsymbol{v}_g$ for collision avoidance difficulty. $\boldsymbol{v}_s \in \boldsymbol{V}$ indicates collision avoidance difficulty is originally zero. However, the human driver inside the subject vehicle perceives an uncertain velocity $\Delta \boldsymbol{v}_{n,u}$ (the blue arrow on the neighbouring vehicle) of the leading vehicle considering the uncertainties, changing the neighbouring vehicle's velocity from $\boldsymbol{v}_n$ to $\boldsymbol{v}_n'$ (the yellow dashed arrow on the neighbouring vehicle, which is the perceived velocity of the neighbouring vehicle). Correspondingly, the velocity set $\boldsymbol{V}$ becomes $\boldsymbol{V}'$ that is smaller than $\boldsymbol{V}$ based on the perceived velocity $\boldsymbol{v}_n'$, which leads to $\boldsymbol{v}_s \notin \boldsymbol{V}'$ causing extra perceived risk.



(b) The uncertain velocity $\Delta \boldsymbol{v}_{s,u}$ and the final perceived velocity $\Delta \boldsymbol{v}_{n,u}$ of the subject vehicle. Note that $\Delta \boldsymbol{v}_{s,a}$, $\Delta \boldsymbol{v}_{n,u}$ and $\Delta \boldsymbol{v}_{n,a}$ are not taken into account to show only the influence of the uncertain velocity $\Delta \boldsymbol{v}_{s,u}$ on the velocity gap $\boldsymbol{v}_g$ for collision avoidance difficulty. $\boldsymbol{v}_s \in \boldsymbol{V}$ indicates collision avoidance difficulty is originally zero. However, the human driver inside the subject vehicle perceives an uncertain velocity $\Delta \boldsymbol{v}_{s,u}$ (the blue arrow on the subject vehicle) due to human drivers' inaccurate control or imperfect control of driving automation, making the subject velocity change from $\boldsymbol{v}_s$ to $\boldsymbol{v}'_s$ (the yellow dashed arrow on the subject vehicle, which is the perceived velocity of the subject vehicle). The perceived velocity $\boldsymbol{v}'_s \notin \boldsymbol{V}$, causing extra perceived risk.



(c) The perceived velocity set $\boldsymbol{V}'$ based on the perceived velocity. Note that $\Delta \boldsymbol{v}_{s,a}$ and $\Delta \boldsymbol{v}_{n,a}$ are not taken into account to show only the combined influence of uncertain velocities $\Delta \boldsymbol{v}_{s,u}$ and $\Delta \boldsymbol{v}_{n,u}$ on the safe velocity $\boldsymbol{V}$ and the velocity gap $\boldsymbol{v}_g$ for collision avoidance difficulty. With the larger perceived subject velocity $\boldsymbol{v}'_s$ and the smaller perceived velocity set $\boldsymbol{V}'$ simultaneously, $\boldsymbol{v}'_s \notin \boldsymbol{V}'$. $\boldsymbol{v}_g$ is the distance from the perceived subject velocity to the boundary of the perceived safe velocity set $\boldsymbol{V}'$, which is the perceived risk in this chapter.



(d) The legend of subfigures (a), (b) and (c)

Figure 3.6: An example of the uncertain velocity and its influence on the perceived velocity and the velocity set $\boldsymbol{V}$. The subject vehicle is following a leading vehicle (50 m ahead) with the same velocity $\boldsymbol{v}_s = \boldsymbol{v}_n = 16.67\,\mathrm{m/s}$ and we have the velocity set $\boldsymbol{V}$ according to Equation (3.9). In all cases, $\boldsymbol{v}_s \in \boldsymbol{V}$ indicating that the collision avoidance difficulty is originally zero. The velocity $\Delta \boldsymbol{v}_{i,a}$ is not considered due to zero acceleration. This scenario illustrates that uncertainties can cause drivers to perceive vehicle velocities differently from their actual values, which can increase the perceived difficulty of collision avoidance, and thus, perceived risk.

Based on the discussion above, we have

$$\Delta \boldsymbol{v}_{i,u} = l \cdot \frac{\boldsymbol{p}_s - \boldsymbol{p}_n}{||\boldsymbol{p}_s - \boldsymbol{p}_n||} \cdot r_i, \quad i \in \{s, n\}, \quad r_n = 1 \quad \text{and} \quad r_s = -1. \tag{3.13}$$

where $\Delta \boldsymbol{v}_{i,u}$ ($i \in \{s,n\}$) is the uncertain velocity; l is the length of the uncertain velocity vector which is derived below; $\frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}$ is a unit vector pointing from the neighbouring vehicle to the subject vehicle; r determines the direction of the uncertain velocity where $r_n = 1$ is for the neighbouring vehicle representing the direction from the neighbouring vehicle to the subject vehicle, and $r_s = -1$ is for the subject vehicle representing the opposite direction.

According to **Assumption 2b** and **2c**, the manoeuvre uncertainties are presented as an uncertain acceleration, which is assumed to follow Gaussian distributions as motivated under Assumption 2, and this acceleration will remain constant over a short period of time. Hence, given a specific duration, the uncertain velocity $\Delta \boldsymbol{v}_{i,u}$ ($i \in \{s,n\}$) also follows Gaussian distributions. With the consideration of physical restrictions of the vehicle velocity, the Gaussian is

$$\begin{aligned} v_{i,u,X} &\sim \mathscr{D}(v_{i,u,X}|0,\sigma_{i,X},fb,bb) \\ v_{i,u,Y} &\sim \mathscr{D}(v_{i,u,Y}|0,\sigma_{i,Y},lb,rb) \end{aligned} \tag{3.14}$$

where $v_{i,u,X}$ and $v_{i,u,Y}$ are the uncertain velocity in X and Y directions; $\mathscr{D}$ is the probability density function of the uncertain velocity in each direction. fb, bb, lb, rb are the forward, backward, left and right bounds for the uncertain velocity in the density function , which are set to 50 m/s, −14 m/s, 8.5 m/s and −8.5 m/s respectively in this chapter [134]. The truncated distribution $\mathscr{D}$ becomes

$$\begin{aligned} \mathscr{D}(v_{i,u,X}|0,\sigma_{i,X},fb,bb) &= \begin{cases} \frac{\frac{1}{\sigma_{i,X}}N(\frac{v_{i,u,X}}{\sigma_{i,X}})}{\mathscr{N}(\frac{fb-v_{i,X}}{\sigma_{i,X}})-\mathscr{N}(\frac{bb-v_{i,X}}{\sigma_{i,X}})}, & bb \leqslant v_{i,u,X} \leqslant fb, i \in \{s,n\} \\ 0, \text{ otherwise} \end{cases} \\ \mathscr{D}(v_{i,u,Y}|0,\sigma_{Y},lb,rb) &= \begin{cases} \frac{\frac{1}{\sigma_{i,Y}}N(\frac{v_{i,u,Y}}{\sigma_{i,Y}})}{\mathscr{N}(\frac{lb-v_{i,Y}}{\sigma_{i,Y}})-\mathscr{N}(\frac{rb-v_{i,Y}}{\sigma_{i,Y}})}, & rb \leqslant v_{i,u,Y} \leqslant lb, i \in \{s,n\} \\ 0, \text{ otherwise} \end{cases} \end{aligned} \tag{3.15}$$

where N is the probability density function of the Gaussian distribution and $\mathscr{N}$ is its cumulative distribution function.

To obtain the final uncertain velocity, its length and direction should be considered simultaneously. Hence, we use the mathematical expectation of Equation (3.13) as the length of the uncertain velocity, which can be calculated as follows

$$\begin{aligned} E(||\Delta \boldsymbol{v}_{i,u}||) &= \int_0^{+\infty} \mathscr{P}\left(\Delta \boldsymbol{v}_{i,u} \middle| \frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}, l\right) \cdot l \, dl \\ &= \int_0^{+\infty} \mathscr{D}(v_{i,u,X}|0,\sigma_{i,X},fb,bb) \cdot \mathscr{D}(v_{i,u,Y}|0,\sigma_{i,Y},lb,rb) \cdot \frac{1}{\mathscr{P}\left(\frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}\right)} \cdot l \, dl \end{aligned} \tag{3.16}$$

This conditional probability is denoted by $\mathscr{P}\left(\Delta \boldsymbol{v}_{i,u} \middle| \frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}, l\right)$, representing that an uncertain velocity $\boldsymbol{v}_{i,u}$ with length l is on the line connecting the subject vehicle and the neighbouring vehicle. To ensure that this direction-specific probability is considered,

we divide the product of the two probability density functions, $\mathscr{D}(v_{i,u,X}|0,\sigma_{i,X},fb,bb)$ and $\mathscr{D}(v_{i,u,Y}|0,\sigma_{i,Y},lb,rb)$, by the aforementioned conditional probability $\mathscr{P}\left(\frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}\right)$. This division normalises the probability densities and allows for the proper calculation of the mathematical expectation of the length of the uncertain velocity $E(||\Delta\boldsymbol{v}_{i,u}||)$.

Accordingly, the uncertain velocity is

$$\Delta\boldsymbol{v}_{i,u} = E(||\Delta\boldsymbol{v}_{i,u}||)\cdot\frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}\cdot r_i,\quad i\in\{s,n\},\quad r_n=1\quad\text{and}\quad r_s=-1. \tag{3.17}$$

Note that this uncertain velocity is not the most probable one but it is the probabilistic average in the most dangerous direction. Although the integral is expressed in an analytical format, *integral* function is used in MATLAB for numerical evaluation.

3.3.5. WEIGHTING FUNCTION $\mathscr{W}$

The subject velocity influences perceived risk, as it affects the accident rate and the consequence of a crash. The relationship between velocity and crash outcome is related to the kinetic energy ($E_k=\frac{1}{2}mv^2$) released during a collision but the relationship is not a simple linear mapping. A scaling function ranging on [0,1] is needed to show the relationship between the subject velocity and perceived risk. Previous studies tried to examine the relationship between the subject velocity and the crash outcome based on real-world crash data and found that a power function best fits the relationship [135]. We employ a power function proposed by Finch *et al.* [136] to describe such a relationship:

$$\mathscr{W} = \beta\left(\frac{||\boldsymbol{v}_s||}{v_{\text{ref}}}\right)^{\alpha} \tag{3.18}$$

where $||\boldsymbol{v}_s||$ is the subject velocity; v_{ref} is a reference velocity and it can be set as the velocity limit in specific conditions; α is the power coefficient. This equation also takes into account the mass ratio between neighbouring and subject vehicle. A heavier and larger vehicle will induce a higher perceived risk as it will yield higher subject vehicle accelerations in case of impact. Building upon the PPDRF [40] we introduce the additional scaling $\beta=\frac{M_n}{M_s+M_n}$ where; M_s and M_n are the mass of the subject vehicle and the neighbouring vehicle. Given a speed limit in a specific scenario and a specific vehicle type, the $\mathscr{W}\propto v_s^{\alpha}$ ranging on $[0,1]$ if $||\boldsymbol{v}_s||$ stays below v_{ref}, which can be used as a weight for the final perceived risk based on functions $\mathscr{A}$ and $\mathscr{V}$ as shown in Equation (3.1).

3.3.6. PCAD MODEL PARAMETERS

Table 3.3 summarises the parameters of the PCAD model to be calibrated. Details regarding the avoidance difficulty function $\mathscr{A}$, the perceived velocity function $\mathscr{V}$, and the weighting function $\mathscr{W}$ can be found in sections 3.3.3, 3.3.4, and 3.3.5, respectively.

Table 3.3: The key parameters of PCAD model

Parameters	Explanations
$\sigma_{i,X}, \sigma_{i,Y}$	Standard deviations for uncertain velocity distributions of the subject ($i = s$) and neighbouring ($i = n$) vehicles
$t_{i,a}$	An anticipated time for acceleration-based velocity of the subject ($i = s$) and neighbouring ($i = n$) vehicles
α	The exponent of the velocity weighting function

3.4. ANALYTICAL MODEL PROPERTIES

This section offers an analysis of the PCAD model and the three baseline models. Table 3.2 in Section 3.2.2 summarises model properties, covering aspects such as dimension, the usage of distance, relative motion, acceleration, subject speed, manoeuvre uncertainties, crash consequences, and usability on curved lanes. In summary, PCAD is a comprehensive model based on Risk Allostasis Theory which considers all aspects listed in Table 3.2. It is a 2-D model capturing both longitudinal and lateral perceived risk, and can also be used on curved lanes.

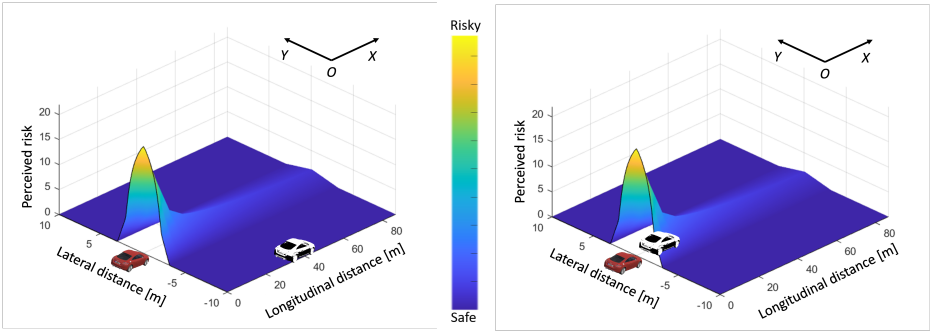
For an intuitive understanding, we visualise the perceived risk variations of the four models in a 2-D coordinate system describing the relative position of a neighbouring vehicle. As demonstrated in Figure 3.7, perceived risk varies with different relative velocities (Figure 3.7b), different decelerations (Figure 3.7c), and different subject velocities (Figure 3.7d). Figure 3.7a provides the legend for these diagrams.

The PCAD model indicates that perceived risk amplifies as an object or neighbouring vehicle nears the subject vehicle, demonstrating a sharp rise both longitudinally and laterally. The non-linear relationship caused by non-linear *looming detection* in PCAD prevails in the other three typical models but is described by different functions such as Gaussian (i.g., the lateral risk in PPDRF and DRF), Exponential (i.g., the potential risk in PPDRF), logarithmic (i.g., the risk in RPR) and Quadratic functions (i.g., the longitudinal risk in DRF). Note that RPR cannot capture perceived risk in the lateral direction since it is only defined in the same traffic lane.

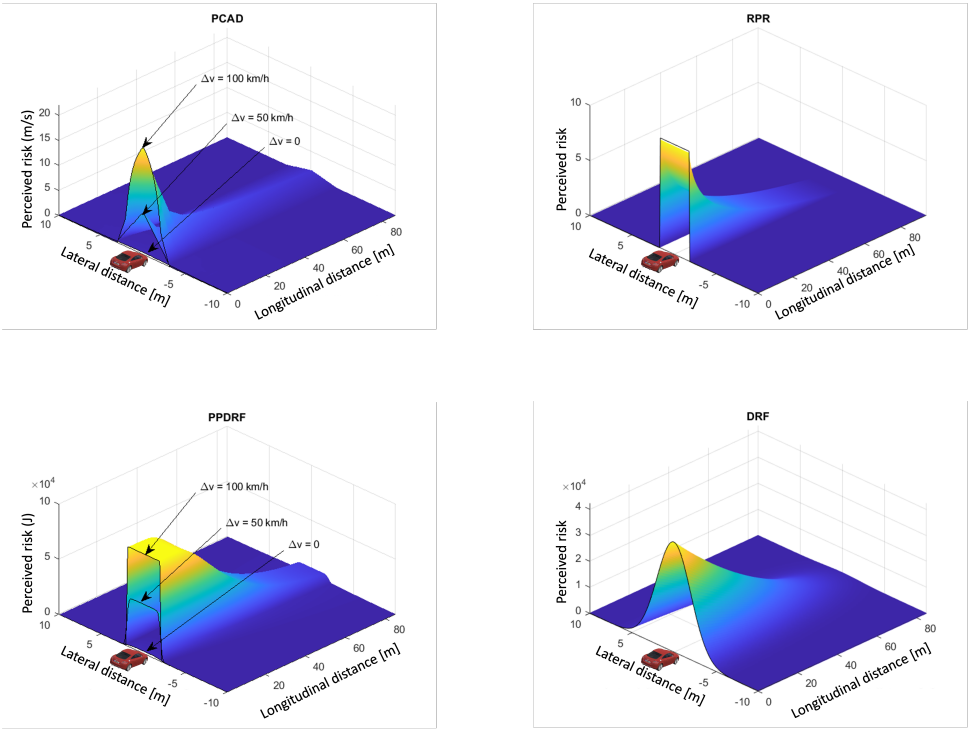
PCAD shows that human drivers perceive more risk when approaching an object faster. Compared to the other three models, PCAD and PPDRF can output different perceived risk values facing different relative velocities (Figure 3.7b). RPR and DRF do not include velocity information of the neighbouring vehicles or objects.

Reacting to neighbouring vehicles' velocity changes (e.g., braking) is a common task in daily driving. PCAD can clearly describe effects on perceived risk (Figure 3.7c) where $a_n = -8\,\mathrm{m/s^2}$ leads to the highest perceived risk and $a_n = 0$ causes the lowest perceived risk. RPR and PPDRF also capture effects of acceleration but DRF cannot indicate the change of perceived risk caused by a neighbouring vehicle's deceleration due to a lack of acceleration information in the model.

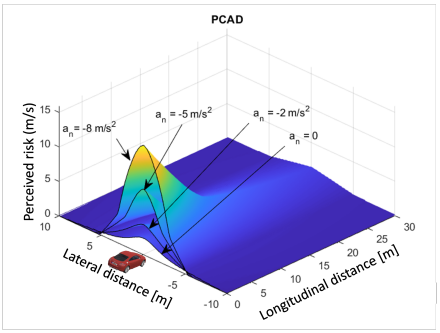
The subject velocity significantly influences perceived risk. In Figure 3.7d, PCAD demonstrates that, given the same following gap, human drivers perceive more risk with a higher subject velocity, which is similar to PPDRF and DRF. However, RPR does not contain subject speed information in the model and cannot capture the perceived risk variance in this condition.

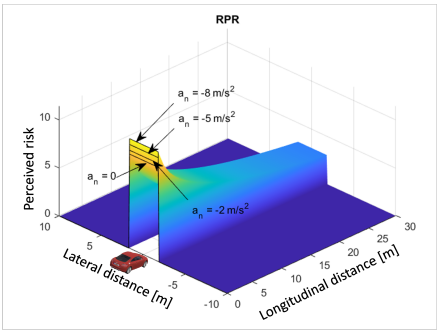


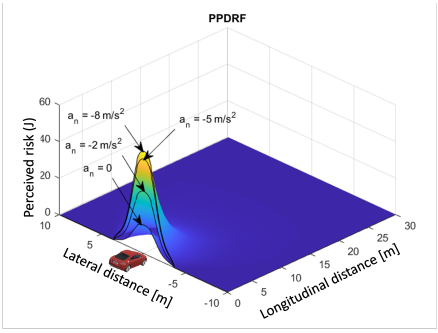
(a) Legend and explanation of figures (b), (c), (d). Surfaces represent the computed perceived risk value of different models as function of the relative 2D position of the subject vehicle (red) and neighbouring vehicle (white). The subject vehicle is at the origin (0,0), moving with velocity $\boldsymbol{v}_{s,X} = 100\,\text{km/h}$ along the X-axis. The neighbouring vehicle's velocity is $\boldsymbol{v}_{n,X} = 50\,\text{km/h}$, also along the X-axis. In the left figure, the neighbouring vehicle at $(40,-10)$ implies a front-right position relative to the subject vehicle, indicating low perceived risk. In the right figure, the neighbouring vehicle at $(10,0)$ shows the subject vehicle approaching the leader with a relative velocity $\Delta\boldsymbol{v} = \boldsymbol{v}_{s,X} - \boldsymbol{v}_{n,X} = 50\,\text{km/h}$ and a longitudinal gap of 10 m, resulting in high perceived risk.

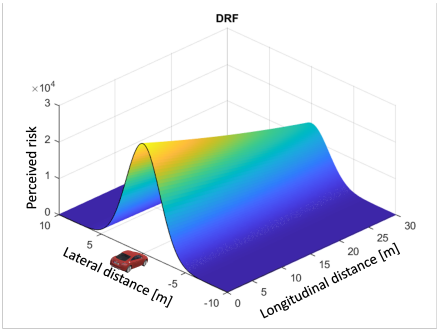


(b) The effect of relative velocity on human driver's perceived risk. The constant subject vehicle velocity is $v_{s,X} = 100$ km/h but the neighbouring vehicle has different velocities $v_{n,X}$; no acceleration of the subject vehicle and the neighbouring vehicle $a_{s,X} = a_{n,X} = 0$. Note that PPDRF's kinetic risk instead of potential risk is used here for a more understandable visualisation.

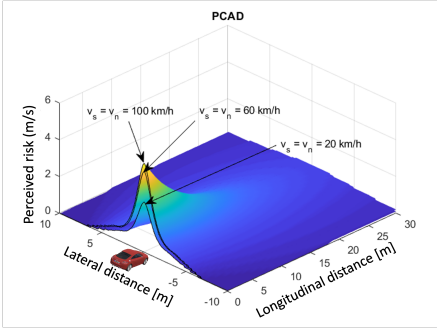


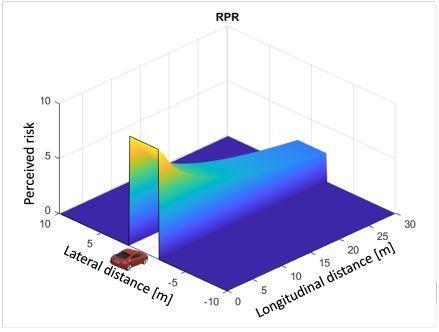


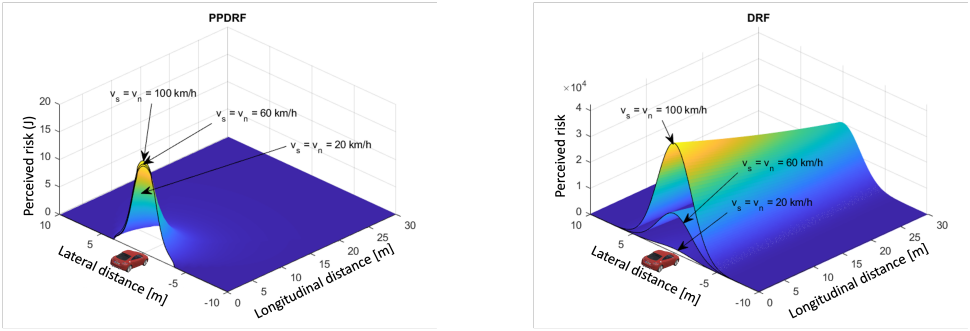




(c) The effect of neighbouring vehicle's known acceleration (braking) on human driver's perceived risk. The velocity of the subject vehicle and the neighbouring vehicle are equal $v_{s,X} = v_{n,X} = 100$ km/h; the subject vehicle has no acceleration $a_{s,X} = 0$ but the neighbouring vehicle's known acceleration ($a_{n,X}$) varies.







(d) The effect of subject velocity. Human driver's perceived risk. The subject vehicle and the neighbouring vehicle have equal velocity ($v_{s,X} = v_{n,X}$) varying from 20 km/h-100 km/h; no acceleration of the subject vehicle and the neighbouring vehicle $a_{s,X} = a_{n,X} = 0$.

Figure 3.7: The effect of relative velocity (b), the acceleration of the neighbouring vehicle (c) and the subject velocity (d) with legend in (a)

3.5. MODEL EVALUATION METHOD

To conduct a comparative evaluation of the proposed model and the baseline models, model calibration with empirical data is indispensable. This section details the experimental datasets, calibration method, and performance indices for the models.

3.5.1. DATASET INTRODUCTION

We employ two datasets for model calibration and evaluation. The first dataset (**Dataset Merging**) was collected in our previous simulator experiment where the subject automated vehicle reacts to merging and hard-braking vehicles. The experiment simulated 18 merging event types with different merging distances and braking intensities on a 2-lane highway [117]. Figure 3.8 shows an example of the simulated events during the experiment. The participants were asked to monitor the scenario as fall-back ready drivers for an SAE Level 2 automated vehicle. They used a pressure sensor on the steering wheel to provide perceived risk ratings from 0-10 continuously in the time domain (see the lower row in Figure 3.8), which are the continuous perceived risk data. After each event, the participants were also asked to give a verbal perceived risk rating from 0-10 regarding the previous event, which is the discrete event-based perceived risk data. The corresponding kinematic data (e.g. position, speed and acceleration of the subject vehicle and neighbouring vehicles) were collected simultaneously.

The second dataset (**Dataset Obstacle Avoidance**) includes drivers' verbal perceived risk ratings (i.e., unlimited numbers) and steering angle signals when the participants face static obstacles suddenly appearing in front the subject vehicle driving at 25 m/s in manual driving mode [36]. Figure 3.9 shows the distribution of the obstacles. The corresponding vehicle kinematic data and the positions of the obstacles were recorded at the same time.

The following reference data is utilised for model calibration:

- Dataset Merging: the **event-based perceived risk rating** and the **peak of the continuous perceived risk** in specific events

- Dataset Obstacle Avoidance: the **event-based perceived risk**, and the **peak of steering wheel angle** in specific events

Figures B.2 and B.3 in Appendix B.2 illustrate the kinematic data from the two datasets, along with the continuous risk predicted by PCAD.

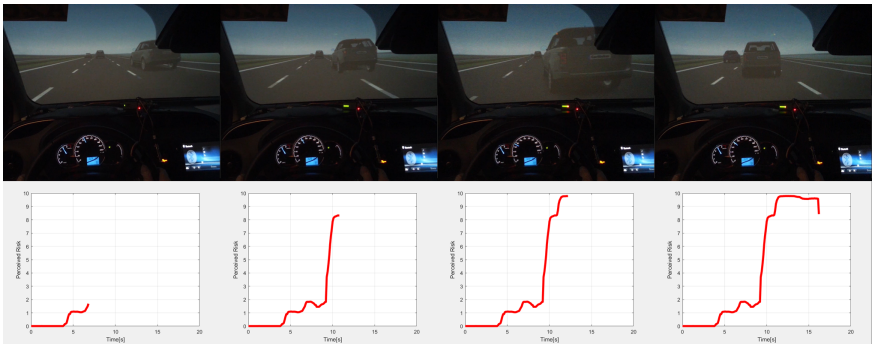


Figure 3.8: The experiment where Dataset Merging was collected. Upper row: Video stream of a merging with hard braking event simulated in the experiment. Lower row: Corresponding perceived risk values indicated by a participant with the pressure sensor.

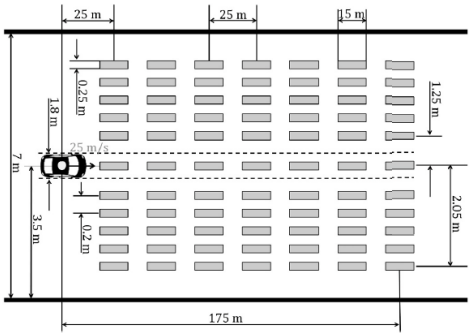


Figure 3.9: Dataset Obstacle Avoidance, with stationary obstacle positions from Kolekar *et al.* [36]

3.5.2. MODEL CALIBRATION

While our aim is to develop general models considering the average characteristics of all participants, we cannot ignore the influence of group features and scenarios. To optimise performance, we perform a dataset-level calibration of parameters for all models. We have $RMSE_i$ defined as

$$RMSE_q = \sqrt{\frac{\sum_{k=1}^{K} \left(\hat{y}_k - y_k\right)^2}{K}} \tag{3.19}$$

Here, $RMSE_q$ denotes the root mean square error between the collected perceived risk data and the model output. For Dataset Merging, $q = event$ and $q = peak$ represent the

$RMSE$ for event-based perceived risk and the peak of continuous perceived risk respectively; for Dataset Obstacle Avoidance, $q = event$ and $q = peak$ denote the $RMSE$ for event-based perceived risk and the maximum steering wheel angle separately. In Equation (3.19), $\hat{y}_k$ represents the model output, while y_k refers to the perceived risk rating. The variable k, which falls within the set of $1, 2, 3, ..., K$, represents the event number in the specific dataset. K signifies the number of available events in different datasets, with $K = 414$ for Dataset Merging and $K = 2496$ for Dataset Obstacle Avoidance. Note that the first sample point of the kinematic data when the obstacle suddenly appears in Dataset Obstacle Avoidance is used for the calibration since the participants were asked to give a verbal perceived risk rating as soon as the obstacle appeared.

The calibration aims to minimise $\sum RMSE_q$ for all models by tuning the key model parameters based on perceived risk and corresponding kinematic data. Given the variability in perceived risk data ranges across the two datasets and the differing output ranges of the four models, min-max feature scaling is employed to normalise both perceived risk data and model outputs to a uniform range of [0,10]. This normalisation facilitates direct comparison and calibration, as encapsulated in Equation (3.20).

$$\hat{z}_k = \frac{z_k - z_{\min}}{z_{\max} - z_{\min}} \times 10 \tag{3.20}$$

where $\hat{z}_k$ represents the scaled value, either model output or perceived risk data. For model outputs, $z_{\max}$ and $z_{\min}$ are the global maximum and minimum values across all outputs for a specific model per dataset. In contrast, for perceived risk ratings within the Dataset Obstacle Avoidance, the scaling is conducted individually for each participant, reflecting the participant-specific range of ratings. This distinction is crucial, as it allows for the individual scaling of perceived risk ratings in the Dataset Obstacle Avoidance due to its unrestricted numerical range and participant-specific variation, while maintaining a universal scaling framework for model outputs and the already bounded ratings with the range [0, 10] in Dataset Merging.

This scaling approach ensures that both participant-specific variations in perceived risk assessment and the diverse output ranges of different models are appropriately normalised for accurate calibration and comparison.

3.5.3. PERFORMANCE INDICATORS

We use five indicators to evaluate the model performance: Correlation, Model error, Detection rate, Computation cost, and Linear Time Scaling Factor.

- **Correlation** The predicted perceived risk has to be correlated with the event-based perceived risk. We use R-Square to quantify how well model outputs fit the real perceived risk. Since the perceived risk output by the models is linearly rescaled to 0-10 (Equation (3.20)), different ranges of perceived risk data or model outputs have no influence on R-Square in our case.
- **Model error** We use Root Mean Squared Error (RMSE) to quantify a model's overall Model error, which is the same as the model calibration criterion (Equation (3.19)). This indicator reflects the model's ability to compute the overall perceived risk in a certain dataset. A model with a smaller $RMSE$ can more accurately predict the overall perceived risk for a given scenario.

- **Detection rate** The Detection rate represents the model's ability to detect an event's risk that is also perceived by human drivers. We defined Detection rate as in Equation (3.21)

$$R_{det} = \frac{K_{detected}}{K_{event}} \times 100\% \tag{3.21}$$

where $K_{detected}$ represents the number of events where the model manages to detect the risk with non-zero output; K_{event} is the total number of the events where human drivers gave perceived risk ratings in a certain dataset with $K_{event} = 414$ for Dataset Merging and $K_{event} = 2496$ for Dataset Obstacle Avoidance. In this chapter, every event carries a 'risk' due to experimental settings, which simulate scenarios where some level of risk is always present. Consequently, the detection of an event, reflected by a non-zero output, implies the recognition of this risk. Therefore, a higher detection rate correlates with better model performance, as it indicates the model's consistent ability to recognise the presence of risk in every event.

- **Computation cost** It is essential that all models possess real-time risk computation capability, so the computation cost is critical. More complex models may offer a better performance in other aspects such as model errors but tend to take longer to compute. We define the computation cost as the model's computation time per computation step. If the time consumption per computation step exceeds the on-board computation capability, it means that the computation of perceived risk cannot be completed in real-time.

The above metrics validate the event-based perceived risk. We also compared the continuous perceived risk measured for Dataset Merging. However, we observed that participants pressed the button following a fixed pattern regardless of the actual real-time risk level. This suggests that the timing of their responses was more likely influenced by the given instructions and their interpretation, rather than reflecting a valid measure of continuous perceived risk over time. Consequently, these responses, although appearing as a 'continuous perceived risk', do not offer reliable time-domain information. Due to this lack of time-domain validation, we have chosen not to report on the validation of continuous perceived risk.

3.6. MODEL EVALUATION RESULTS

In this section, we illustrate the applicability of the four models and evaluate their performance with the performance indicators introduced previously regarding the two datasets.

3.6.1. MODEL CALIBRATION RESULTS

The calibration is performed separately for the two datasets. According to the model structure and dataset features, the calibrated parameters are listed in Table 3.4.

3.6.2. PERFORMANCE EVALUATION RESULTS

We test the four models using both datasets including the perceived risk data and the corresponding kinematic data with the calibrated parameters shown in Table 3.4. The following sections present different aspects of performance.

CORRELATION

The correlation between predicted and measured event-based perceived risk data plays a crucial role in assessing the performance of risk assessment models. This is particularly important given the uncertainty in defining the unit of perceived risk. Figure 3.10 and Figure 3.11 display the correlation between the predicted perceived risk and event-based perceived risk for Dataset Merging and Dataset Obstacle Avoidance respectively. The adjusted R-Square is calculated based on the averaged event-based perceived risk across the same event type (●in Figure 3.10 and Figure 3.11).

In both datasets, the PCAD model demonstrates a stronger correlation with event-based perceived risk data compared to other models. Furthermore, the regression models RPR and DRF exhibit strong performance in Dataset Merging and Dataset Obstacle Avoidance, respectively, for which they were originally developed.

MODEL ERROR

As discussed in Section 3.5.3, the Root Mean Square Error ($RMSE$) is an indicator of the overall Model error. Table 3.5 presents the RMSE values for all four models across the two datasets.

The $RMSE$ values (both event and peak) reveal that the PCAD model achieves a comparable performance level to the regression models (e.g., RPR in Dataset Merging and DRF in Dataset Obstacle Avoidance), albeit with a slightly larger model error. In Dataset Merging, the lower $RMSE$ values for PCAD and RPR suggest better performance, as these models directly incorporate the neighbouring vehicle's acceleration. In Dataset Obstacle Avoidance, the lower $RMSE$ values for PCAD and DRF indicate better performance, as these models also consider lateral perceived risk, resulting in reduced model error when applied to a 2-D dataset (Table 3.5). The PPDRF model was originally designed to evaluate actual collision risk in traffic, and was not previously calibrated or validated. We now performed such a calibration and demonstrate moderate performance in both datasets.

DETECTION RATE

As per Equation (3.21), the detection rates for the four models across both datasets are presented in Table 3.5. In Dataset Merging, the merging vehicle primarily poses longitudinal risk in the same lane. Consequently, all models are capable of detecting dangerous events, regardless of whether they are 1-D or 2-D models. However, in Dataset Obstacle Avoidance, the obstacles are dispersed across a 2-D plane. As a result, only models that account for lateral risk can effectively identify dangerous vehicles outside the forward path. This leads to a lower detection rate for the RPR model, while the other three models are able to recognise all dangerous events that human drivers also perceive as risky. Note that Figure 3.11(e) displays outputs with marginal values that appear to be zero but are, in fact, detected by PPDRF.

COMPUTATION COST

Table 3.5 presents the computation cost for different models, tested on a workstation with an Intel Core i7-8665U 1.9Ghz processor and 8GB RAM. Generally, models that take more factors into account require longer computation times. In both datasets, RPR is the fastest model, as it only involves logarithmic calculations.

In Dataset Merging, PCAD is the most time-consuming model since it relies on a grid search to identify the optimal velocity gap vector to the safe velocity region. PPDRF requires spatial overlap computations and multiple integrals over variations of acceleration probability density function in the overlap area, making it a time-intensive process. Although DRF involves discretising a 2D area of an object or a vehicle with a grid and summing the risk values over each grid cell to obtain the final perceived risk, its overlap computations are simpler than those of PPDRF, as the risk field and severity field are static and no motion prediction of neighbouring vehicles is needed.

In Dataset Obstacle Avoidance, PPDRF takes less time than DRF and PCAD, as it only computes potential risk, which is a simpler process compared to the kinetic risk computation in Dataset Merging.

SUMMARY OF MODEL PERFORMANCE EVALUATION

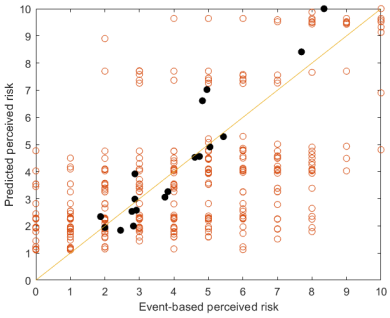
Based on the results discussed above, we utilise radar charts to illustrate the performance of each model across various aspects, as shown in Figure 3.12. Generally, PCAD demonstrates strong performance in terms of overall model error, R-square, and detection rate. However, the primary drawback of PCAD is its high computation cost, which results from its complexity. The regression models (i.e., RPR in Dataset Merging and DRF in Obstacle Avoidance) exhibit the best performance in their respective datasets. We remark that the advantage of PPDRF in capturing the manoeuvre uncertainties of the surrounding vehicles vanishes in the second dataset due to the specific experimental setting. As a result, the PPDRF models used in the two datasets are two different models. This largely explains the poor performance of PPDRF, albeit it clearly showed advantages in the analytical model properties.

It is worth mentioning that PCAD demonstrates consistent performance across both datasets. We also conducted cross-validation between the two datasets, in which the four models and their parameters were calibrated using one dataset and then used to predict perceived risk in the other dataset. PCAD performs quite well even without recalibration where as seen in Figure B.4, Figure B.5, and Figure B.6 in Appendix B.3, PCAD maintains its strong performance in cross-validation. Additionally, this suggests that the calibration process has a low risk of overfitting, further highlighting the robustness of the PCAD model.

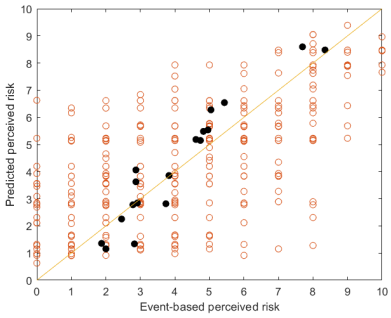
Table 3.4: Calibrated parameters for all models

Model	Parameters	Explanation	Values for Dataset Merging	Values for Dataset Obstacle Avoidance
PCAD[1]	$\sigma_{n,X}$	The standard deviation in X of the velocity Gaussian of a neighbouring vehicle (m/s)	4.28	/
	$\sigma_{n,Y}$	The standard deviation in Y of the velocity Gaussian of a neighbouring vehicle (m/s)	3.86	/
	$\sigma_{s,X}$	The standard deviation in X of the velocity Gaussian of the subject vehicle (m/s)	0.80	6.58
	$\sigma_{s,Y}$	The standard deviation in Y of the velocity Gaussian of the subject vehicle (m/s)	1.70	1.20
	$t_{s,a}$	The anticipated time for the acceleration-based velocity of the subject vehicle (s)	0.13	/
	$t_{n,a}$	The anticipated time for the acceleration-based velocity of a neighbouring vehicle (s)	0.01	/
	α	The exponent of the power function in velocity weighting function	0.52[2]	/
RPR	C_0	The intercept in the regression model	12.10	20.70
	C_1	The coefficient of gap to the leading vehicle	-3.70	-3.68
	C_2	The coefficient of leading vehicle's braking intensity	-0.36	/
PPDRF	$\tilde{\sigma}_x$	The standard deviation of longitudinal acceleration distribution of neighbouring vehicle (m/s^2)	2.01	/
	$\tilde{\sigma}_y$	The standard deviation of lateral acceleration distribution of neighbouring vehicle (m/s^2)	0.02	/
	D	The steepness of descent of the potential field	/	0.14
DRF[3]	s	The steepness of the height parabola of the risk field	0.15	0.005
	t_{la}	Human driver's preview time (s)	1.20	8.12
	m	The rate of the risk field width expanding	3.98×10^{-8}	3.66×10^{-4}
	c	The initial width of the DRF (m)	0.45	1.10

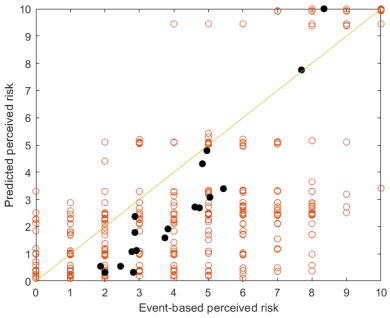
[1] The experimental design of two datasets, featuring only one category of obstacle or other road users, means that the mass ratio β remains constant. Consequently, in the calibration phase, this ratio is set to 1.
[2] This is the calibrated value regarding the specific dataset. Due to the lack of subject velocity change, α has limited influence on Dataset Merging. α ranging on [0, 2.5] leads to an R-square ranging on [0.80, 0.90]. For Dataset Obstacle Avoidance, α was set to 0 since it almost has no influence. Additionally, the v_{ref} in the weighting function $\mathcal{W}$ was set to 27.78 m/s for Dataset Merging and 25 m/s for Dataset Obstacle Avoidance.
[3] The best performance for DRF was obtained when the subject velocity $\boldsymbol{v}_{s,X}$ in Equation (B.9) was fixed to its initial value 27.78 m/s when the vehicle decelerated for Dataset Merging. For Dataset Obstacle Avoidance, the subject velocity $\boldsymbol{v}_{s,X}$ was a constant 25 m/s.



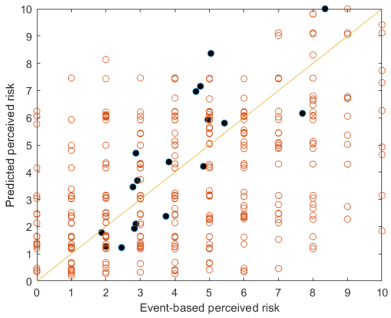
(a) PCAD (Adjusted R-Square = 0.90)



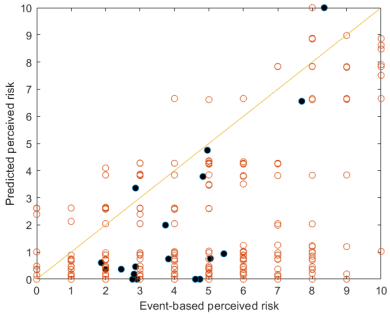
(b) RPR (Adjusted R-Square = 0.90)



(c) PPDRF (Adjusted R-Square = 0.90)

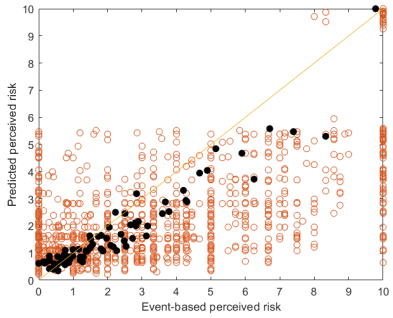


(d) DRF (Adjusted R-Square = 0.67)

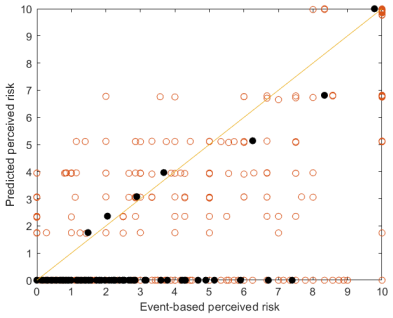


(e) PCAD with only deterministic motion information (Adjusted R-Square = 0.58)

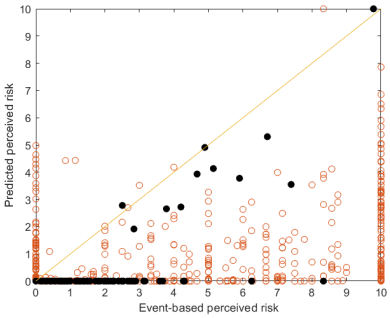
Figure 3.10: Predicted and measured event-based perceived risk in Dataset Merging. "○" indicates individual event-based perceived risk and "●" indicates the averaged event-based perceived risk across the same event type, for which the Adjusted R-Square is also given.



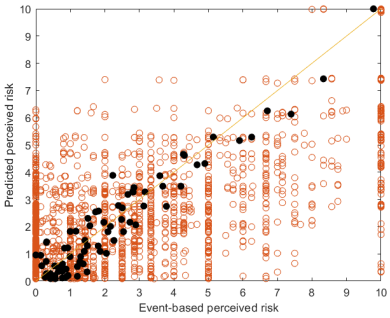
(a) PCAD (Adjusted R-Square = 0.90)



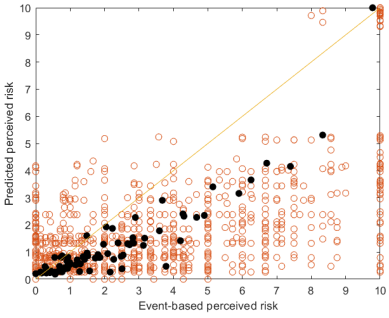
(b) RPR (Adjusted R-Square = 0.38)



(c) PPDRF (Adjusted R-Square = 0.50)



(d) DRF (Adjusted R-Square = 0.90)



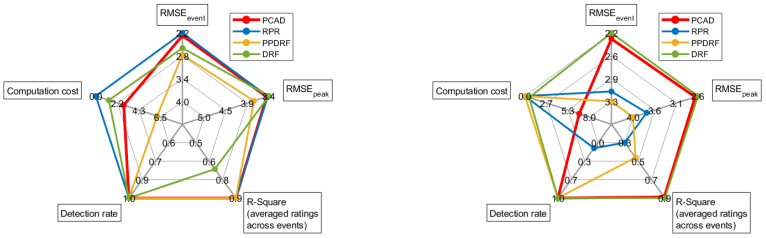
(e) PCAD with only deterministic motion information (Adjusted R-Square = 0.81)

Figure 3.11: Predicted and measured event-based perceived risk in Dataset Obstacle Avoidance. "○" indicates individual event-based perceived risk and "●" indicates the averaged event-based perceived risk across the same event type. Note that in (c), there are many dots with small values but non-zero, indicating that they are actually detected by PPDRF.

Table 3.5: Model performance represented by the performance indicators

Dataset	Performance indicators	PCAD	RPR	PPDRF	DRF
Dataset Merging	$RMSE_{event}$	2.25	2.18	2.76	2.58
	$RMSE_{peak}$	3.41	3.39	3.73	3.35
	Adjusted R-Square	0.90	0.90	0.90	0.67
	Detection rate	1.00	1.00	1.00	1.00
	Computation cost (ms)[1]	2.79	1.77×10^{-4}	6.14[5]	1.30
Dataset Obstacle Avoidance	$RMSE_{event}$	2.27	3.20	3.34	2.17
	$RMSE_{peak}$	2.71	3.84	4.02	2.61
	Adjusted R-Square	0.90	0.38	0.50	0.90
	Detection rate	1.00	0.09[4]	1.00	1.00
	Computation cost (ms) [2]	6.70[3]	2.01×10^{-4}	1.08×10^{-2}	1.22

[1] The average value of computing 124614 steps.
[2] The average value of computing 349440 steps.
[3] PCAD consumed more time in Dataset Obstacle Avoidance because the searching algorithm worked in a larger searching area to find the velocity gap $\boldsymbol{v}_g$.
[4] Only the vehicles directly in front of the subject vehicle can be detected by RPR, which leads to a low detection rate. See [36] for more experiment details.
[5] PPDRF consumed much more time in Dataset Merging because the model contains numerical integration when facing moving vehicles.



(a) Performances in Dataset Merging

(b) Performances in Dataset Obstacle Avoidance

Figure 3.12: Radar charts of model performance indicators in two datasets

3.7. DISCUSSION

In this paper, we present a computational perceived risk model based on the Risk Allostasis Theory [78], [121], [137], capturing task difficulty using the gap between the current velocity and the safe velocity region in 2D. Our model quantifies event-based perceived risk and the peak of continuous perceived risk in both longitudinal and lateral directions. We validated the model on two datasets of human drivers' perceived risk and compared its performance with three baseline perceived risk models. Our work contributes to addressing the challenge of perceived risk computation for SAE Level 2 driving automation, while also illustrating the mechanisms underlying human drivers' risk perception.

Traditional perceived risk models consider collision probability and the collision consequence [138], such as DRF and PPDRF, while our PCAD model is developed based on the concept of potential collision judgement using looming, which originates from aerospace and maritime experience [123]. PCAD demonstrates improved performance

in estimating perceived risk in various driving conditions, aligning with the argument of McKenna [139] and Rundmo and Nordfjærn [140] that drivers are incapable of monitoring infrequent event probabilities, thus supporting the underlying theory of PCAD.

The demonstrated better performance of our PCAD model unveils new insights into perceived risk. Firstly, PCAD considers all motion information in Table 3.2, highlighting the importance of position, velocity, and acceleration for risk perception. Secondly, the models that can capture lateral risk lead to a higher detection rate in Dataset Obstacle Avoidance, indicating that perceived risk is 2-D and human drivers perceive the risk from all directions in a 2-D plane. Thirdly, manoeuvre uncertainties of the subject vehicle and other road users cause extra perceived risk, which is supported by Kolekar *et al.* [36] and Ding *et al.* [130]. Lastly, perceived risk is a dynamic concept and varies with the changing traffic conditions as illustrated in Figure 3.7, which presents the perceived risk variations in three distinct driving conditions (i.g., different relative velocities, subject velocities and accelerations). This observation motivates the need for models, such as the proposed PCAD model, which can adjust to varying driving scenarios even without recalibration.

It is plausible that drivers associate trucks with higher threats compared to cars, and tend to maintain greater distances accordingly. Our model incorporates this by adjusting perceived risk levels to the mass ratio. Furthermore, the distance between reference points on neighbouring vehicles also affects the model's output. A larger vehicle like a heavy truck will have reference points that are farther apart, resulting in a smaller safe velocity set and consequently, a higher perceived risk. Our model can thus explain the cautious behaviour when driving around trucks, which can lead to more realistic assessment of perceived risk.

We note that our model has limitations. There is only one traffic object considered in this chapter. If multi-road users or even infrastructure is added, PCAD still has the potential to estimate perceived risk. In that case, we need to compute the potential collision avoidance difficulty for multiple neighbouring vehicles considering them simultaneously. Based on this we calculate the comprehensive safe velocity set and derive the total perceived risk.

The PCAD model's identification of the minimal velocity change $\boldsymbol{v}_g$ does not account for drivers' tendency to assign different weights to braking and steering based on the driving experience and situational complexity. This anisotropic weighting in driver's decision-making is not captured by the current model. Exploring this difference presents a potential opportunity to improve the PCAD model to more accurately simulate real-world driver behaviour and improve its applicability in safety-critical scenarios.

It is important to acknowledge that integrating advanced motion prediction techniques into models assessing collision avoidance and perceived risk enhances their predictive accuracy. Our model, within its current scope, incorporates aspects of motion prediction through the inclusion of acceleration-based velocity and uncertain velocity, offering a simplified but effective approach to understanding driving dynamics. The primary objective of this research has been to establish a foundational model capable of predicting perceived risk via collision avoidance difficulty. This aim has shaped our methodological choices, balancing model complexity with practical applicability. While there are limitations associated with the simplified approach to motion prediction, these

do not diminish the contributions of our work. Instead, they highlight areas for future research and development, suggesting paths for incorporating more advanced motion prediction methodologies to further enhance the model's precision and usefulness in studies on perceived risk.

Regarding perceived risk data, Dataset Merging covers human driver's perceived risk with SAE Level 2 driving automation where drivers need to monitor the system and environment and be ready to intervene. This makes this data suitable to assess perceived safety when using automation, but the lateral risk is not systematically explored. Dataset Obstacle Avoidance explores human drivers perceived risk data in 2-D, including lateral perceived risk. However, this dataset is collected from human-automation transitions (i.e., human drivers' taking-over process in this case), which may cause bias in automated driving conditions. Additionally, the objects in the experiment are fixed and suddenly displayed during driving. The additional perceived risk caused by surprise cannot be ignored.

Dataset Merging measured perceived risk on a scale from 0 to 10 for no to very high risk, while Dataset Obstacle Avoidance captured perceived risk as a non-negative real number without predetermined upper limit. To facilitate comparison, all results from Dataset Obstacle Avoidance and all models were scaled to 0-10 to match the scale used in Dataset Merging. More specific scales of perceived risk can be developed for experimental studies, including factors such as accident risk and severity, and the driver's tendency or need to intervene and overrule the driving automation.

To further advance perceived risk modelling, we recommend collecting more perceived risk data in various scenarios through online surveys with videos, simulator experiments and on-road observations. Such additional data can help to assess the validity of the PCAD model also in multi-vehicle interactions, and infrastructure interactions including curve negotiation. Additional data can also serve to examine perceived risk at different driving automation levels should be examined in the future. Moreover, internal HMIs have positive effects in reducing human drivers' perceived risk and the perceived risk modelling will be further improved to capture different internal HMI conditions [141], [142]. Our PCAD model can also be used as a cost function, a constraint, or a reference of perceived risk in driving automation decision making, trajectory planning, or controller design, enhancing trust [81] and acceptance.

3.8. CONCLUSIONS

In this chapter, we have formulated, calibrated, and validated a novel computational perceived risk model, and compared its performance with three well-established models across two different datasets. Our findings reveal valuable insights into the understanding and quantification of perceived risk in various driving situations. The key conclusions drawn from our analysis are as follows: (1) Driving task difficulty serves as an effective indicator of perceived risk; (2) Perceived risk is two-dimensional, originating from both longitudinal and lateral directions, and exhibits a non-linear increase as the distance to surrounding vehicles decreases; (3) Incorporating uncertainties in the model is crucial for an accurate representation of perceived risk; (4) Perceived risk is dynamic and changes with driving conditions.

4

DECODING PERCEIVED RISK IN AUTOMATED VEHICLES THROUGH 140K RATINGS

Automated vehicles offer a potential solution to improve road safety, but the insufficient understanding of driver's perceived risk hinders the acceptance of automated driving. This chapter reveals factors determining the perceived risk dynamics during vehicle interactions using a novel and large-scale online survey. We obtained time-continuous perceived risk data from more than 140k ratings by 2,164 participants. A total of 105 events was created, including merging, hard braking, and lane changes on motorways, while systematically varying multiple control parameters (such as relative speed and distance) to achieve different levels of event criticality. Our analysis, using deep neural networks, demonstrates that manoeuvre uncertainties of surrounding road users are crucial in shaping perceived risk, in addition to proximity and relative velocity. Our results also show that the influence of these factors of perceived risk is non-stationary and this non-stationarity is well captured by deep neural networks but not by existing (physics-based) models. This comprehensive analysis provides insights into the temporal dynamics of perceived risk, guiding the future development of automated vehicles to improve user subjective safety.

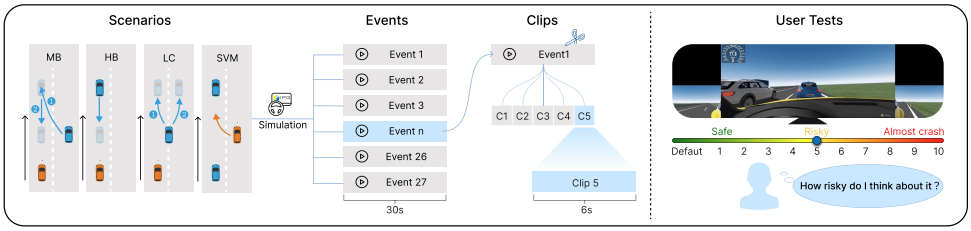
The content of this chapter is prepared for submission as
He, X., Li, Z., Wang, X., Happee, R., & Wang, M. "Reading minds on the road: Decoding perceived risk in automated vehicles through 140K ratings," in revision.

4.1. INTRODUCTION

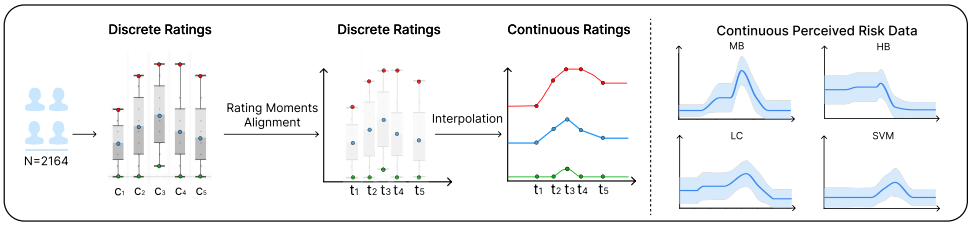
Road accidents cause around 1.19 million deaths and up to 50 million injuries globally every year [113]. Automated vehicles have attracted substantial attention due to their potential to improve road safety with advanced sensing and control systems [2], [143]. Despite their promise, AVs have yet to prove themselves as a definitive solution to road safety. One key hurdle stems from user interaction with AVs - how we, as drivers and passengers, perceive risks when using the system on the road [9], [20], [21], [144]. Such a subjective assessment of a possible negative outcome, known as perceived risk, affects individual choices and reflects a major challenge in the advancement of AVs [11]–[14]. This highlights the importance of addressing perceived risk of AV users when interacting with this technology on the road [35], [117].

Substantial efforts have been made to investigate perceived risk in AVs, particularly exploring how factors such as visibility [145], [146], weather conditions [69], [147], driving experience [33], personality [148] and driving environment [54], [149] shape perceived risk. While these studies have shed light on contributing factors and their effect on perceived risk, they typically analyse perceived risk after one or more events with (critical) vehicle interactions. The event-based perceived risk does not capture the risk evolution prior to the event and hence only impart partial understanding of risk perception [36], [110], [117], [150]. To effectively manage perceived risk, it is essential to dynamically analyse it in real time, aligning with the driving behaviours of AVs. This leads to our primary goal of this chapter: to reveal how different factors dynamically affect perceived risk during driving.

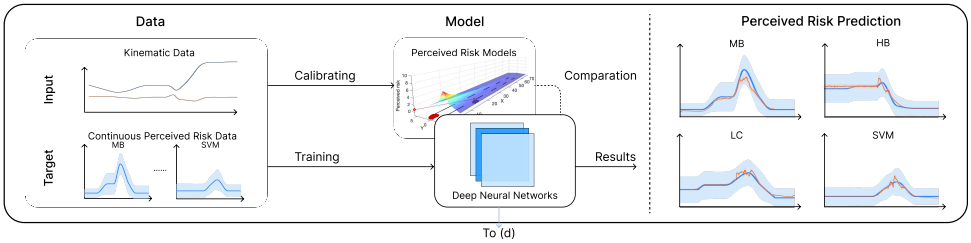
Current technology does not yield robust direct analysis of the brain to measure perceived risk [151]. Therefore, we first need subjective perceived risk evaluations in different driving conditions, specifically, time-continuous perceived risk data, to study the dynamic nature of perceived risk. Unfortunately, there is no continuous observation of perceived risk at present. As a psychological concept, perceived risk in AVs is commonly measured by questionnaires after rides or based on public knowledge [9], [11], [20], [21]. These methods offer but a static snapshot of perceived risk, falling short of capturing perceived risk's dynamic nature. A few studies have measured perceived risk in real time during driving or by referencing a specific video frame [36]–[38], [45], but these remain limited to singular moments. Innovative approaches such as sliders [69] and pressure buttons [110], [117] have been introduced for continuous measurement of perceived risk, but validation mainly refers to peak signals reflecting the highest perceived risk within an event. To bridge this continuous measurement gap, researchers have explored physiological signals such as the galvanic skin response (GSR) and heart rate changes through electrocardiogram (ECG) to indicate perceived risk in various traffic situations [110], [117], [152]–[155]. Pupil dilation, reactive behaviours such as braking and hand position on the steering wheel have also been examined as potential perceived risk indicators [34], [75], [117], [141], [156]–[160]. Although these methods suggest the possibility of continuous measurement of perceived risk, their precision remains questionable.



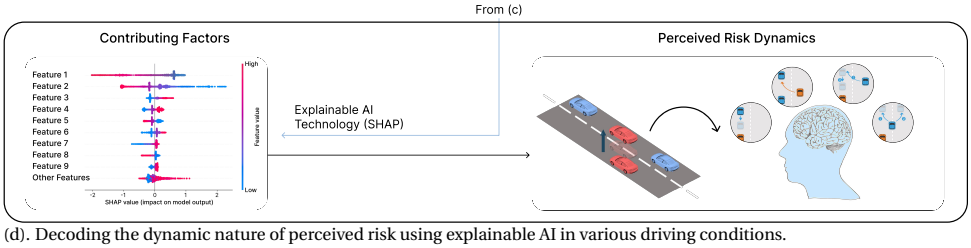
(a). Design of driving scenarios, events and the online questionnaire. Four driving scenarios were designed: The subject vehicle reacts to a merging and hard-braking vehicle (MB); the subject AV reacts to a hard-braking vehicle (HB); the subject vehicle reacts to merging vehicles with different lateral behaviour (LC); the subject AV merges onto the main road (SVM). A series of events were generated by controlling the behaviour of neighbouring vehicles and the subject AV (See Experimental methods and Table 4.1). Each video was fragmented into 5 or 6 clips of 6 s.



(b). Data collection and process. 2,164 participants were kept in this chapter. Continuous perceived risk ratings were generated based on the kinematic data and interpolation.



(c). Two state-of-the-art perceived risk model were calibrated and evaluated and compared with deep neural networks trained using perceived risk data.



(d). Decoding the dynamic nature of perceived risk using explainable AI in various driving conditions.

Figure 4.1: Overview of the method.

Our methodological framework is shown in Figure 4.1. We used a novel method that balanced accuracy, efficiency, difficulty, and cost associated with collecting large-scale

perceived risk data. We defined four common traffic event scenarios in motorway driving [93]. In each scenario, we systematically varied the behaviours of neighbouring vehicles and the subject AV to generate a wide range of events with different criticality levels. These events were captured in simulated videos using a high-fidelity simulator. Subsequently, each video was fragmented into five or six shorter clips that were arranged to be viewed one after another in their original sequential order to collect continuous perceived risk ratings within each event from participants. All videos were integrated into an online questionnaire for data collection (Figure 4.1a). We collected more than 180,000 perceived risk ratings from 2,164 participants primarily from Europe, resulting in more than 140,000 ratings used for further analysis. Based on the relationship between simulated kinematic data and the peaks of perceived risk levels identified in previous studies [110], [117], we synchronised the perceived risk ratings per video clips with participants' actual rating moments. By connecting these perceived risk ratings per video clip in one event, the continuous perceived risk ratings were derived (Figure 4.1b).

Continuing from the collected perceived risk data, the next step is to analyse these data to find out the dynamic nature of perceived risk. Unlike many studies, we did not collect and analyse human physiological responses in this chapter, such as EEG [161], [162], ECG [117], GSR [110], [117], and pupil size signals [34], [75], [117], [141], [156]–[160], although indicative of perceived risk, remain questionable in terms of their precision. Therefore, a more viable approach involves the use of an accurate model that serves as an ideal observer [163]–[166]. These models, if fed with data perceptible to humans and capable of accurately predicting current perceived risk, offer a distinct advantage over human participants in that they can be directly analysed by us. To further our understanding of the dynamic nature perceived risk, we proceeded to select suitable candidates for the role of the ideal observer. Among the candidates were several pioneer computational models based on first principles, notably the potential collision avoidance difficulty (PCAD) model [167] and the driving risk field (DRF) model [41]. We calibrated the two existing models using the perceived risk data we had collected. Simultaneously, we developed and trained deep neural networks (DNNs) to predict perceived risk. After evaluation, we discovered that DNNs outperformed PCAD and DRF in terms of accuracy. The superior performance of DNNs suggested that it was a more suitable candidate to serve as the ideal observer in this chapter (Figure 4.1c).

Following the training and validation of our DNNs as ideal observers, we used SHapley Additive exPlanations (SHAP) [168]–[170] to analyse the trained models (Figure 4.1d). Unlike many studies using neural networks to study human perception, our primary focus is not on the correspondence between the network structure and the internal structure of the human brain. Our goal is to use neural networks as an ideal observer that can substitute humans but allows for more in-depth analysis, to then compare and analyse its characteristics with those of humans. This analysis provided a ranking of contributions of influencing factors across specific scenarios. Moreover, their contributions in the time domain in every driving event were derived. The findings aligned with existing knowledge, confirming that factors such as a smaller distance to other road users [36], [117], [129], [150], a higher speed [167], and a higher relative speed [98], [167] increase perceived risk non-linearly. Notably, we uncovered dynamic insights: the contributions of influence factors to perceived risk varied over time, and manoeuvre uncertainty of

other road users played a substantial role in perceived risk, especially when the subject vehicle maintains a stable but close distance from neighbouring vehicles. This analysis helps refine our understanding of the dynamic nature perceived risk, emphasising the fluctuating importance of different risk factors.

In summary, our work reveals the dynamic nature of perceived risk based on interactions with other road users in common driving scenarios. This creates guidelines for the future modelling of perceived risk. Concurrently, a huge dataset is created on perceived risk, which is beneficial for perceived risk investigation and development facilitating acceptance and successful deployment of automated vehicles. This chapter successfully pioneers a path in the dynamic analysis of perceived risk, providing valuable guidance for future continuous data collection and analysis.

4.2. EXPERIMENTAL METHODS

4.2.1. DESIGN OF DRIVING SCENARIOS

To capture the dynamic nature of perceived risk in various automated driving conditions, we designed four driving scenarios (Figure 4.2): The subject AV reacting to merging and hard-braking vehicles (MB); The subject AV reacting to hard-braking vehicles in front (HB); The subject AV reacting to lane-changing vehicles with various lateral behaviours (LC); The subject AV merging onto the main road with dense traffic (SVM).

In MB scenario (Figure 4.2a), a merging vehicle entered the motorway from an on-ramp, passed the subject AV, and merged between the subject and lead vehicle. Detection of this merging manoeuvre by the AV was implemented as the moment when the centre of the merging vehicle crossed the line. After this, the subject AV followed the merging vehicle instead of the original lead vehicle. Meanwhile, the merging vehicle braked to keep a safe distance from the original lead vehicle until the velocity of the merging vehicle decreased to 60 km/h. Then the merging vehicle accelerated again to the desired cruising speed. The initial merging distance to subject AV, desired cruising speed, and braking intensity were all varied with three levels, leading to 27 events with different levels of criticality (Table 4.1 row MB).

In HB scenario (Figure 4.2b), a lead vehicle initially drove at the desired cruising speed in front of the subject AV. Then the leading vehicle braked to a lower speed of 60 km/h due to certain safety considerations (e.g. its leading vehicle brakes) and then accelerated again to the desired cruising speed. Correspondingly, the subject AV braked to keep a safe distance and accelerated again. The initial car-following distance, the desired cruising speed, and the braking intensity were varied threefold, generating 27 events with different levels of criticality (Table 4.1 row HB).

The LC scenario (Figure 4.2c) was based on the MB scenario. The desired cruising speed, braking intensity, and the ending braking velocity were fixed to be 100 km/h, $-2\,\mathrm{m/s^2}$ and 60 km/h respectively. The initial merging distance was set to 5 and 15 m. Then four different kinds of lateral behaviours were designed: lane change with a low lateral speed (1 m/s) , lane change with a high lateral speed (3 m/s), a fragmented lane change with a low lateral speed (1 m/s)[1] and an aborted lane change with a low lateral

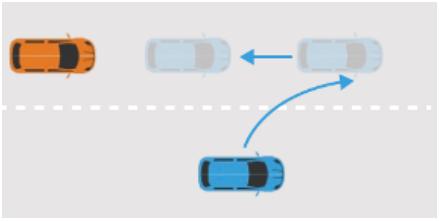
[1] The neighbouring vehicle attempts to merge into the ego lane from the adjacent lane, but pauses midway for 6 seconds before completing the lane change.

speed (1 m/s)[1]. The subject AV's behaviour was also controlled to generate cautious, mild and aggressive longitudinal behaviour. Consequently, 24 events were generated with different levels of criticality (Table 4.1 row LC).

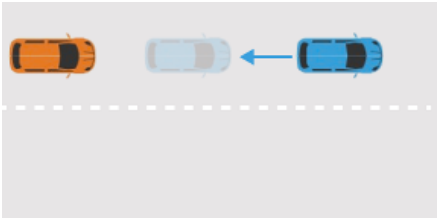
In SVM scenario (Figure 4.2d), the subject AV was arranged to merge onto a highway with dense traffic. At the exact moment when AV entered the highway, the AV started to brake in order to ensure a safe distance from the lead vehicle on the main road. The initial distance between the subject AV and the leading vehicle, the desired cruising speed, and braking intensity were varied threefold, generating 27 events with different levels of criticality (Table 4.1 row SVM).

We created videos of the predefined events using CarMaker 8.0.1. The video of each event was further fragmented into 5 or 6 clips, with each clip lasting 6 seconds (Figure C.1 in Appendix C.1 provides examples of video streams of all scenarios). There were no overlapping clips and no specific rule was applied for cutting the clips. Finally, from the 105 different events, a total of 549 clips were generated. Kinematic data from all vehicles, including the subject AV and neighbouring vehicles, were recorded at 10 Hz.

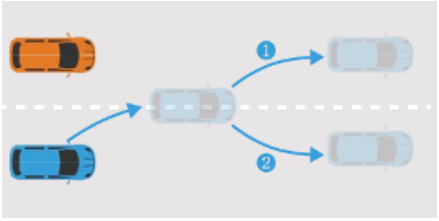
All videos were uploaded to the video sharing platform Vimeo [171] and embedded in an online questionnaire on Qualtrics [172] (See Appendix C.4.1 for the online questionnaire).



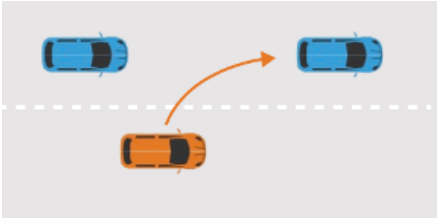
(a). The subject AV reacting to merging and hard-braking vehicles (MB)



(b). The subject AV reacting to hard-braking vehicles (HB)



(c). The subject AV reacting to lane-changing vehicles with various lateral behaviours (LC)



(d). The subject AV merging onto the main road with dense traffic (SVM)

Figure 4.2: Driving scenarios in this study.

4.2.2. PROCEDURE

The study in this chapter was approved by the Human Research Ethics Committee of Delft University of Technology under application number 1245 and the digital consent

[1] The neighbouring vehicle initiates a lane change into the ego lane from the adjacent lane, pauses midway for 6 seconds, and then returns to the original adjacent lane without completing the lane change.

Table 4.1: Designed scenarios for the perceived risk experiments.

Scenarios	Varied parameters	Number of events
Merging with hard braking (MB)	Initial merging distance (m): 5, 15, 25	27
	Desired cruising speed (km/h): 80, 100, 120	
	Braking intensity (m/s^2): -2, -5, -8	
Hard braking (HB)	Initial distance (m): 5, 15, 25	27
	Desired cruising speed (km/h): 80, 100, 120	
	Braking intensity (m/s^2): -2, -5, -8	
Reacting to lane-changing vehicle with lateral control (LC)	Initial merging distance (m): 5, 15	24
	Lateral categories: normal (1 m/s), normal (3 m/s), fragmented (1 m/s), aborted (1 m/s)	
	ACC categories: cautious, mild, aggressive	
Subject AV merging with hard braking (SVM)	Initial distance (m): 5, 15, 25	27
	Desired cruising speed (km/h): 80, 100, 120	
	Braking intensity (m/s^2): -2, -5, -8	

form was provided at the beginning of the online questionnaire to all participants. Participants were recruited and rewarded through the Prolific crowd-sourcing platform [173].

The participants began the study reading an introduction that explains the experimental procedures and objectives. This was followed by the acquisition of informed consent. Subsequently, the individuals provided demographic information. Before viewing and rating the videos, participants underwent a training module designed to rate perceived risk in a standardised manner. In this module, they learned how to answer the question, *"How risky do you perceive the clip above?"* by adjusting a digital slider along a scale from 0 to 10. This included the presentation of a video clip that shows high-risk and low-risk scenarios to establish a baseline for risk assessment. After training, the formal experimental phase was initiated. Participants encountered a series of events, each segmented into 5 to 6 video clips that are arranged to be viewed one after another in their original sequential order in the time domain. After viewing each video clip, the participants rated their perceived risk using the slider on a scale from 0 to 10. In general, each participant was asked to view 16 out of 105 events individually, with 4 events in each scenario. The order of events for each participant was randomly selected.

4.2.3. VALIDATION OF THE CONTROLLED VARIABLES OF DRIVING SCENARIOS

LINEAR MIXED MODEL (LMM) APPROACH

We used a mixed-effects analysis to examine the influence of controlled parameters on the perceived risk ratings. In each driving scenario, each video clip was characterised by

different levels of controlled variables (e.g., desired cruising speed, initial merging distance, braking intensity, etc.). The perceived risk ratings were organised in a long-format structure, where each row represented a participant's perceived risk rating for a specific video clip. Each row contained columns for the participant ID, the clip number, the controlled parameters and the perceived risk rating. The participant ID was treated as a random effect to account for individual biases in the ratings, while the event-specific parameters were considered fixed effects to evaluate their influence on the perceived risk ratings (see Table 4.1 for more details of the parameters). This analytical approach provided a nuanced understanding of the responses within subjects, while also accommodating the between-subjects variability introduced by the random assignment of events. The analysis was conducted in IBM SPSS Statistics 29 [174].

JENSEN-SHANNON (J-S) DIVERGENCE APPROACH

Although the perceived risk ratings are discrete, they are continuous in the time domain within the same event. Therefore, when analysing whether different events within the same scenario have resulted in variations in perceived risk ratings, we need to consider the continuity of time. To this end, we used the Jensen-Shannon (J-S) divergence, a method that quantitatively assesses the similarity between two probability distributions [175]. This approach is particularly suited for comparing the distribution of perceived risk ratings between different events within the same driving scenario (i.e., MB, HB, LC and SVM), revealing the difference in various events over time.

Rating transfer matrix computation The foundational step in our analysis involves the construction of Rating Transfer Matrices for each event in one scenario. These matrices represent the transition probabilities of moving from one perceived risk rating to another between successive video clips. The process is as follows:

1. For every pair of successive video clips, we collect the transitions of perceived risk ratings made by all participants. This collection captures how frequently participants transition from one risk rating to another across clips.

2. These transitions are then normalised to probabilities, forming the matrix M, where each element m_{ij} indicates the probability of transitioning from rating i at time t to rating j at time $t+1$.

Defining probability distributions P and Q For our analysis, we define the probability distributions P and Q as follows:

- P: The probability distribution of perceived risk ratings for a given video clip, represented by a row in the Rating Transfer Matrix corresponding to time t.

- Q: The probability distribution of perceived risk ratings for the subsequent video clip, represented by the next row in the matrix corresponding to time $t+1$.

These distributions are essential for computing the Kullback-Leibler (K-L) divergence, which in turn is used to calculate the Jensen-Shannon (J-S) divergence.

Computing the Jensen-Shannon (J-S) divergence The J-S divergence is calculated by first determining the midpoint distribution $M = \frac{1}{2}(P+Q)$, and then computing the average of the K-L divergence from P to M and from Q to M:

$$JS(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M) \tag{4.1}$$

where the K-L divergence $D_{KL}(P||Q)$ is defined as:

$$D_{KL}(P||Q) = \sum_i P(i)\log\frac{P(i)}{Q(i)} \tag{4.2}$$

This symmetrical and bounded measure allows us to comprehensively compare the evolution of perceived risk perceptions across video clips within a scenario. Our MATLAB implementation systematically applies this method to all pairs of consecutive video clips, enabling a detailed investigation into the changing landscape of risk perception as participants engage with different driving scenarios.

4.2.4. GENERATION OF CONTINUOUS PERCEIVED RISK RATINGS

To convert discrete perceived risk ratings into a continuous timeline, we use a novel systematic approach rooted in empirical observations. The process involves pinpointing the moment of rating within the video clips and then applying interpolation methods to create a seamless perceived risk profile over time.

RATING MOMENT ALIGNMENT

According to our instructions, participants were asked to provide their perceived risk ratings based on what they considered the most dangerous moment in each video clip. The alignment of rating moments with these dynamic events is guided by the following assumptions based on empirical findings [110], [117]:

1. Perceived risk changes in response to stimuli, which in this chapter include various actions of neighbouring vehicles, such as merging, hard braking, and lane changing [176], [177].

2. The perceived risk remains constant when the relative position of neighbouring vehicles remains stable.

The alignment of rating moments is based on the statistics of perceived risk signals in motorway driving scenarios. According to statistics in our previous studies [110], [117], for instance, in merging scenarios, the peak perceived risk appears 2.95 s after the onset of a neighbouring vehicle's merging manoeuvre, whereas in braking scenarios, the peak is observed at about 1.15 s after the minimum gap between vehicles is reached. Following the peak, perceived risk returns to baseline levels within 3.93 s. Algorithm 1 in Appendix C.1 details the steps taken to align these moments.

Based on the algorithm above, we obtained the rating moment for all ratings as shown in Table C.1-C.4 in Appendix C.1. Based on the assumption that perceived risk remains constant when the relative position of neighbouring vehicles is stable, some ratings were duplicated to facilitate subsequent perceived risk interpolation. Therefore, in these tables, some ratings have two rating moments.

FROM DISCRETE PERCEIVED RISK RATINGS TO CONTINUOUS PERCEIVED RISK THROUGH INTERPOLATION

Upon identifying the moments corresponding to each discrete perceived risk rating, the subsequent phase involved interpolating these data points to generate continuous perceived risk ratings. By using Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) (see Algorithms 4 in Appendix C.1), discrete perceived risk ratings were interpolated into time-continuous perceived risk. See Appendix C.2 for more details of the selection of the interpolation method.

4.3. MATHEMATICAL METHODS

This section introduces three kinds of computational perceived risk models: potential collision avoidance difficulty (PCAD) model [167], driving risk field (DRF) model [150] and deep neural networks (DNNs).

4.3.1. POTENTIAL COLLISION AVOIDANCE DIFFICULTY (PCAD) MODEL

The PCAD model represents a computational framework designed to estimate perceived risk in driving scenarios, with a particular focus on automated vehicles. Rooted in Fuller's risk allostasis theory, the PCAD model assesses driving task difficulty by quantifying the minimal two-dimensional (2D) velocity change necessary to avoid potential collisions. This quantification takes into account the manoeuvre uncertainties of other road users as well as the control imprecision of the subject vehicle.

Let $\mathbf{X}_s = (\boldsymbol{p}_s, \boldsymbol{v}_s, \boldsymbol{a}_s)^T$ and $\mathbf{X}_n = (\boldsymbol{p}_n, \boldsymbol{v}_n, \boldsymbol{a}_n)^T$ denote the state of the subject vehicle s and the neighbouring vehicle n respectively, with $\boldsymbol{p}_s$ and $\boldsymbol{p}_n$, $\boldsymbol{v}_s$ and $\boldsymbol{v}_n$, $\boldsymbol{a}_s$ and $\boldsymbol{a}_n$ being the position, velocity and acceleration vectors, and T the transpose of a vector. The PCAD is formulated as Equation (4.3)

$$R_{PCAD}(t) = \mathscr{A}(\boldsymbol{p}_s, \boldsymbol{p}_n, \mathscr{V}_s(\mathbf{X}_s, \mathbf{X}_n), \mathscr{V}_n(\mathbf{X}_s, \mathbf{X}_n)) \cdot \mathscr{W}(\boldsymbol{v}_s) \tag{4.3}$$

Here, $\mathscr{A}$ represents the avoidance difficulty function. This function quantifies the required 2D velocity change to bring the subject vehicle to the safe velocity region in the velocity domain to avoid a potential collision with the neighbouring vehicle, considering factors such as their positions, velocities and accelerations. $\mathscr{V}_i$ denotes the 2D perceived velocity for vehicle $i \in \{s, n\}$, thereby capturing absolute and relative motion of the interacting vehicles. Finally, $\mathscr{W}$ is the weighting function, being a power function with v_s, which accounts for the influence of the subject vehicle's speed on perceived risk. Higher speeds generally increase the perceived risk, as the consequence of a potential collision is more severe.

4.3.2. DRIVING RISK FIELD (DRF) MODEL

The DRF model represents human drivers' risk perception as a 2D field, combining the probability (probability field) and consequence (severity field) of an event, the product of which provides an estimation of driver's perceived risk. The DRF model was derived from a simulator experiment involving obstacle avoidance with 77 obstacles distributed on a 2D plane in front of the subject vehicle. During each drive, one obstacle was randomly chosen and suddenly appeared, after which participants needed to steer to avoid

the obstacle and gave a non-negative number indicating required steering effort. Based on the position information of the obstacles, the maximum steering angle, and the subjective ratings, the DRF model was fitted to the data, and thereby it is essentially an empirical model.

The DRF model quantifies overall perceived risk as

$$R_{DRF}(t) = \sum p(x(t), y(t)) \cdot sev(t) \tag{4.4}$$

where $p(x(t), y(t))$ is the probability of an event happening at position $(x(t), y(t))$; $sev(t)$ is the severity field of events. Specifically, in straight drive, the probability field can be simplified as

$$p(x(t), y(t)) = h \cdot \exp\left(\frac{-y(t)^2}{2\sigma^2}\right) \tag{4.5}$$

$$h = s \cdot \left(x(t) - v_{s,X}(t) \cdot t_{la}\right)^2 \tag{4.6}$$

$$\sigma = m \cdot x(t) + c \tag{4.7}$$

where the subject vehicle is at the origin $(0,0)$ with h and σ representing the height and the width of the Gaussian at longitudinal position $x(t)$; s defines the steepness of the height parabola; t_{la} is the human driver's preview time (s); m defines the widening rate of the 2D probability field; c is the quarter width of the subject vehicle (m). $v_{s,X}(t)$ is the subject vehicle's velocity (m/s). The lateral cross-section of the 2D probability field is a Gaussian. Note that the height of the Gaussian h and the width σ are separately modelled as a parabola and linear function of longitudinal distance x in front of the subject vehicle.

The severity field of the events in this chapter can be defined as

$$sev(t) = \begin{cases} C_{sev}, & (x(t), y(t)) \in A^O, \\ 0, & (x(t), y(t)) \notin A^O. \end{cases} \tag{4.8}$$

where C_{sev} is the severity value that is set empirically and A^O represents a neighbouring vehicle's spatial area.

4.3.3. STATE-OF-THE-ART MODEL CALIBRATION

Model input PCAD and DRF models use a range of observed kinematic variables, specifically, the position, velocity, and acceleration of both the subject AV and neighbouring vehicles. The detailed inputs used by PCAD and DRF, which are essential for their operation, are listed in Table 4.4 in Chapter 4.4.3.

Model calibration Calibrating the computational models used to assess perceived risk, such as PCAD and DRF, is critical for enhancing their accuracy and reliability. The calibration process specifically focuses on minimising the Root Mean Square Error (RMSE),

a widely used measure of the difference between values predicted by a model and the values actually observed. The RMSE is calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.9}$$

where n is the number of observations. In this chapter, $n = 8,127$ for MB, HB and SVM scenarios while $n = 8,664$ for LC scenario, since the time-continuous perceived risk lasted for either 36 s for 24 events in the LC scenario or 30 s for 27 events respectively in the other three scenarios, and the sampling rate was 10 Hz; y_i are the observed values, which are the average perceived risk for each sampling moment across all participants; $\hat{y}_i$ are the predicted values that are the rescaled model output of PCAD and DRF. By minimising RMSE, we ensure that the model's predictions closely align with perceived risk data, thereby improving the model's precision.

The calibration involves using kinematic data as input and producing a continuous perceived risk signal as output. Given the diverse ranges of computational model outputs, we employ min-max feature scaling to standardise these outputs to a uniform scale of [0, 10], which simplifies comparisons and integration between different models and scenarios. The scaling is described by:

$$\hat{y}_k = \frac{y_k - y_{\min}}{y_{\max} - y_{\min}} \times 10 \tag{4.10}$$

where $\hat{y}_k$ is the scaled model output; $y_{\max}$ and $y_{\min}$ are the maximum and minimum model output values, respectively.

For the optimisation process, we used the *fmincon* function from MATLAB [178]. This method is particularly effective for dealing with the nonlinear optimisation problems often encountered during model calibration, enabling precise adjustments of model parameters under various constraints. Figure 4.5a and Figure 4.5b show examples of prediction performance of PCAD and DRF in all four scenarios. The full results are shown in Figure C.7 in Appendix C.1.

4.3.4. DEEP NEURAL NETWORKS (DNNS)

DNNs were designed to predict perceived risk. This chapter uses the superiority of DNNs in learning complex, nonlinear patterns in large datasets to establish the relationship between observed kinematic variables such as position, velocity, acceleration, etc., and perceived risk [179].

DNNs structure We trained six DNNs for four scenarios. Specifically, three DNNs were trained for MB, HB and SVM scenarios separately. For LC scenario, events were categorised into normal lane change, fragmented lane change, and aborted lane change so that three DNNs were trained for these three categories respectively. Each of the six neural networks is composed of three fully connected layers: the first layer matches the dimension of the input, the hidden layer contains 500 neurons, and the final layer is designed to output two values — mean and variance [180]. The mean provides the network's best prediction of perceived risk, while the variance quantifies the uncertainty of

this prediction, helping assess how reliable or confident the prediction is based on the input data, which will be detailed later. Table 4.3 in Chapter 4.4.3 outlines the DNNs' structure for different scenarios.

DNNs input features In this chapter, DNNs used a comprehensive set of kinematic variables as inputs, including the position, velocity, and acceleration of both the subject AV and neighbouring vehicles. To address the uncertainties in vehicular manoeuvres, we also incorporated processed variables such as "uncertainty velocity" and "deceleration rate to avoid a crash" into DNNs. These additional variables are important as they enhance the model's ability to predict perceived risk by accounting for the dynamic and potentially unpredictable behaviours of the subject AV and other vehicles on the road. A detailed list of all input variables, including these advanced metrics, is provided in Table 4.4 in Chapter 4.4.3. Further explanation of "uncertainty velocity" and "deceleration rate to avoid a crash" will be introduced below.

Uncertain velocity DNNs use a concept named "uncertainty velocity" to include the manoeuvre uncertainties of the subject vehicle and neighbouring vehicles into perceived risk computation, which is the same concept as that in Chapter 3.3.4 and Ref. [167].

The uncertain velocity of each vehicle, as perceived by the human driver, contributes to the driving situation being perceived as more dangerous. The uncertain velocity exists in all directions on both subject and neighbouring vehicles, but its impact for different directions on perceived risk varies. We assume that the uncertain velocity with a direction that reduces the distance between vehicles most strongly increase perceived risk as shown in Figure C.6 in Appendix C.1. The uncertain velocity is represented as $\Delta v_{i,u,j}$ in Table 4.4 in Chapter 4.4.3, where i can be s, n, nb representing the subject vehicle, the neighbouring vehicle and the following vehicle in SVM scenario; j can be X or Y representing the longitudinal or lateral direction. The calculation of the uncertain velocity in this chapter was based on the parameter values of PCAD for Dataset Merging in Table 3.4 in Chapter 3.

Deceleration rate to avoid a crash (DRAC) DRAC was originally defined in car-following as the squared differential speed between a following vehicle and its corresponding leading vehicle, divided by their closing gap [181]–[183]. We extend this definition to two-dimensional motion as the squared differential speed between two vehicles in one direction divided by the their gap in this direction as shown in Equation (4.11).

$$DRAC = \begin{cases} \frac{(v_s - v_n)^2}{gap}, & \text{if } \dot{gap} < 0; \\ 0, & \text{if } \dot{gap} \geqslant 0. \end{cases} \tag{4.11}$$

where v_s and v_n represent the velocity of the subject vehicle and the neighbouring vehicle in one specific direction (i.e., longitudinal or lateral direction); gap is the distance between the subject vehicle and the neighbouring vehicle in the same direction. In Table 4.4 in Chapter 4.4.3, $DRAC_R$ represents the DRAC caused by the real velocity and $DRAC_u$ represents the DRAC caused by the uncertain velocity.

DNNs training In our approach to training neural networks, we begin by standardising input variables using Z-score normalisation to ensure uniformity across different feature scales [184]. This normalisation is important for our model, as it aligns all input variables to have zero mean and unit variance, thus facilitating more stable and effective learning. The normalisation process is defined mathematically as

$$X_{\text{norm}} = \frac{X - X_{\text{mean}}}{X_{\sigma}} \tag{4.12}$$

where X_{mean} and X_{σ} denote the mean and standard deviation of the input variable X, respectively.

To optimise DNNs' performance, the training is conducted over 200 epochs with a learning rate of 0.001 [185]. This particular learning rate assists in the gradual and precise adjustment of the network's weights, helping to minimise RMSE between the average perceived risk across all participants and the DNNs' output. Additionally, a dropout rate of 0.1 is implemented in the final layer of the network for uncertainty quantification and avoiding overfitting [185]. The dataset allocation for this training consists of 80% for training and 20% for validation randomly [186]. Without leaving out some entire events from the training set, the division is applied across individual data points from all events in one scenario, ensuring that each event is represented in both training and validation sets. Methods to create more independent validation datasets are being proposed in the discussion. The whole dataset contains 8,127 samples for MB, HB and SVM scenarios and 8,664 samples for LC scenario.

UNCERTAINTY OF PERCEIVED RISK PREDICTION

It is critical to acknowledge the variability in human judgements. Our dataset, collecting ratings from hundreds of participants on the same events, shows a certain variability in the ratings, highlighting the complexity and subjective nature of perceived risk assessment. This variability indicates the necessity of quantifying the uncertainty in predictions made by models regardless PCAD, DRF or DNNs. In this regard, two fundamental types of uncertainty are recognised: aleatoric uncertainty and epistemic uncertainty [187], [188] . Aleatoric uncertainty arises from the inherent variability in the data itself, reflecting the stochastic nature of the observations, which remains irreducible regardless of the amount of data collected. On the other hand, epistemic uncertainty refers to the uncertainty in the model parameters, often caused by the lack of knowledge or insufficient data, and can be reduced with the accumulation of more data.

Aleatoric uncertainty With contributions from hundreds of participants in the same events, the dataset shows a broad range of perceived risk ratings. This variability, known as aleatoric uncertainty, is fundamental to the data itself and remains unaffected by perceived risk models.

Aleatoric uncertainty, therefore, reflects the stochastic nature of the data, attributed to the variability in participant ratings for the same event. We employ a robust statistical technique, bootstrapping, to directly quantify aleatoric uncertainty from our dataset. This non-parametric approach enables the estimation of uncertainty without assuming a specific underlying distribution, making it particularly suited for our diverse and

subjective dataset. In our analysis, multiple resampled datasets were generated. Each dataset is created by randomly selecting ratings with replacement from the original dataset, thereby simulating the natural variability present in the collected ratings. For each resampled dataset, we calculate the standard deviation of the ratings. By averaging these standard deviations across all bootstrap samples, we derive a robust estimate of aleatoric uncertainty for each event as follows [187]:

$$U_{\text{aleatoric}} = \frac{1}{B}\sum_{b=1}^{B}\sigma_b, \tag{4.13}$$

where B represents the number of bootstrap samples, and σ_b denotes the standard deviation of participant ratings in the b^{th} bootstrap sample. This equation is the way of measuring aleatoric uncertainty, giving a clear method to understand how much participants' perceived risk varies. Bootstrapping is a common technique for estimating uncertainty but note that if the dataset is large enough, the average of the bootstrap standard deviations could be very similar to the standard deviation of the whole dataset.

Epistemic uncertainty Epistemic uncertainty arises from perceived risk model's limited knowledge. Epistemic uncertainty, unlike aleatoric uncertainty, can be reduced as we improve our models or collect more data.

For DNNs, epistemic uncertainty quantification can be effectively achieved through a technique known as Monte Carlo (MC) Dropout [189]. This method leverages the randomness introduced by applying dropout during each forward pass, simulating training on different subsets of the data. Given a dataset $D = \{(\mathbf{X}_i, Y_i)\}_{i=1}^{N}$ with N samples, where $\mathbf{X}_i$ and Y_i represent the input features and corresponding labels of perceived risk, the prediction process for a new input $\mathbf{X}^*$ can be expressed as:

$$\hat{Y}, \hat{\sigma}^2 = f_{\boldsymbol{\theta}}(\mathbf{X}^*), \tag{4.14}$$

where $\hat{Y}$ and $\hat{\sigma}^2$ denote the mean and variance of the predicted output, respectively, and $f_{\boldsymbol{\theta}}$ is the neural network function parameterised by weights $\boldsymbol{\theta}$. To estimate epistemic uncertainty, MC Dropout utilises the variance of predictions obtained from multiple forward passes with dropout enabled:

$$U_{\text{epistemic}} = \text{Var}_{\text{MC}}(\hat{Y}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{Y}_i - \overline{\hat{Y}})^2, \tag{4.15}$$

where $\hat{Y}_i$ is the prediction from the i-th forward pass, $\overline{\hat{Y}}$ is the average prediction over all M forward passes, and $\text{Var}_{\text{MC}}(\hat{Y})$ represents the variance of these predictions, indicative of the model's epistemic uncertainty.

For PCAD and DRF, we use a different approach to quantify epistemic uncertainty. Recognising that these models do not support dropout-based methods, we employ bootstrapping to simulate the process of training multiple versions of each model [190]. By calibrating 200 models through bootstrapping, we can capture the variability in the models' predictions, which reflects their epistemic uncertainty. The epistemic uncertainty for the two models is calculated as follows:

$$U_{\text{epistemic}} = \frac{1}{\text{iter}}\sum(\text{value} - \text{mean}(\text{value}))^2, \tag{4.16}$$

where "value" represents the predictions from each bootstrapped model, "iter" denotes the number of bootstrap iterations, and the operation captures the variance among these predictions. This method allows us to estimate the epistemic uncertainty by measuring how much our predictions vary as we slightly alter the data or model configuration, providing insights into the confidence we should have in the models' predictions.

Through these methodologies, we are able to quantify epistemic uncertainty for each model type. This quantification not only enhances our understanding of where our models stand in terms of predictive reliability, but also guides future improvements in model development and data collection strategies.

4.3.5. ANALYSING THE DYNAMIC NATURE OF PERCEIVED RISK WITH SHAPLEY ADDITIVE EXPLANATIONS (SHAP)

SHapley Additive exPlanations (SHAP) is a method to explain the output of machine learning models by computing the contribution of each feature to the prediction for each instance. It uses Shapley values, a concept from game theory, to assign an importance value to each feature, showing how much each feature contributes to the prediction [168], [169].

In applications, based on neural network models that have been well-trained and demonstrate good performance, we further aim to analyse the contribution of various features to the prediction. Specifically, some features may have a higher impact on perceived risk (e.g., longitudinal relative distance), while others may have a lower impact (e.g., acceleration). For a sample $\mathbf{X} = \{x_i\}_{i=1}^{D_{\text{feature}}}$ fed into model f_D, resulting in prediction $\hat{Y} = f(\mathbf{X})$, where x_i represents the i^{th} features collected from the scenario, this process of analysing the feature impact can be formulated as follows:

$$\phi_i : (x_i, f_D, \hat{Y}) \rightarrow \mathbb{R}. \tag{4.17}$$

The Shapley value is often used to analyse the contribution of each input feature in predictive models, which employs Game Theory and is theoretically justified. To illustrate using the computation of the i^{th} feature as an example, the fundamental principle of the Shapley value involves traversing all permutations of feature subset coalitions S without considering the feature x_i:

$$\phi_i^{\text{Shapley}} = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! \cdot (|F| - |S| - 1)!}{|F|!} \left[f(x_{S \cup \{i\}}) - f(x_S) \right] \tag{4.18}$$

where $|F| = D_{\text{feature}}$. $x_{S \cup \{i\}}$ and x_S are the features with and without in the set $S \cup \{i\}$ and S. $f(x_{S \cup \{i\}})$ and $f(x_S)$ are neural network models trained in corresponding feature set.

The SHAP method is an unified approach that offers both global and local interpretability for inputs, which is the Shapley value of a conditional expectation function of the original model. It attributes to each feature the change in the expected model prediction when conditioning on that feature. A surprising characteristic of SHAP is the presence of a single unique solution in this class with three desirable properties: local accuracy, missingness, and consistency [170].

- **Local accuracy**. Local accuracy in the context of SHAP refers to the property that ensures the explanation model's output matches exactly with the original model's

$f(x)$ output for individual predictions. This means that for any specific instance, the sum of SHAP values assigned to each feature, along with the base value, will accurately equal the prediction of the original model. This characteristic ensures that SHAP provides precise and faithful explanations at the individual sample level. Local accuracy ensures that the sum of SHAP values and the base value exactly matches the model's output for any given input.

$$\begin{aligned} f(x) = g(x') &= \phi_0 + \sum_{i=1}^{D_{\text{feature}}} \phi_i x'_i \\ &= \text{bias} + \sum \text{contribution of each feature} \end{aligned} \tag{4.19}$$

where x is related to its original feature by $x = h_x(x')$ by mapping function $h_x(\cdot)$. And $\phi_0 = h_x(0)$ means all simplified features are toggled off (i.e. missing).

- **Missingness**. The concept of missingness in SHAP relates to how the method handles features that are absent or missing in a given data instance. SHAP accounts for missing features by allocating a Shapley value of zero to them, reflecting their non-contribution to the model's prediction for that specific instance. This approach acknowledges the absence of data and ensures that only the present features contribute to the explanation of a model's output. The property of missingness can be formulated as:

  If

  $$f_x(S \cup i) = f_x(S) \tag{4.20}$$

  for all subsets of features $S \subseteq F$, then $\phi_i(f, x) = 0$.

- **Consistency**. It refers to the principle that if a model changes so that the contribution of a feature increases or stays the same, regardless of other features, the SHAP value for that feature should not decrease. This ensures that the explanation model remains consistent with changes in the feature's impact on the prediction. In other words, SHAP values faithfully represent the proportional impact of each feature on the model's output, adhering to changes in feature importance. The property of consistency can be formulated as:

  For any two models f and f', if

  $$f'_x(S) - f'_x(S \setminus i) \geq f_x(S) - f_x(S \setminus i) \tag{4.21}$$

  for all $S \subseteq F \setminus \{i\}$, where F is the set of D_{feature} input features, then $\phi_i(f', x) \geq \phi_i(f, x)$.

According to [168], only one possible explanation satisfies all three properties:

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(D_{\text{feature}} - |z'| - 1)!}{D_{\text{feature}}!} [f_x(z') - f_x(z' \setminus i)] \tag{4.22}$$

where $|z'|$ is the number of non-zero entries in z', and $z' \subseteq x'$ represents all z' vectors where the non-zero entries are a subset of the non-zero entries in x'.

$$\begin{aligned} f(h_x(z')) &= \mathbb{E}[f(z)|z_S] && \text{SHAP explanation model simplified input mapping} \\ &= \mathbb{E}_{z_{\overline{S}}|z_S}[f(z)] && \text{expectation over} z_{\overline{S}}|z_S \\ &\approx \mathbb{E}_{z_{\overline{S}}}[f(z)] && \text{assume feature independence (as in [191]–[194])} \\ &\approx f((z_S, \mathbb{E}[z_{\overline{S}}])). && \text{assume model linearity} \end{aligned} \tag{4.23}$$

note that $z_{\overline{S}}$ is the set of features not in S.

4.4. RESULTS

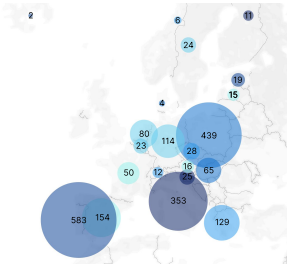
4

4.4.1. SELF-REPORTED PERCEIVED RISK DATA COLLECTION

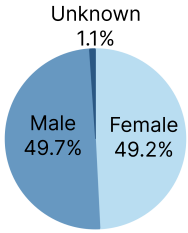
The self-reported perceived risk data were collected through the large-scale online experiment from July 11 to September 19, 2023. A diverse group of 2,341 participants from countries with right-hand traffic participated through the Prolific crowd-sourcing platform [173] (see Figure 4.3a).

In the experiment, participants evaluated a series of motorway traffic event videos, displaying various driving conditions, created using high-fidelity simulation software. Events are categorised in four typical scenarios in motorway driving: the subject vehicle reacting to hard braking vehicles (HB), merging and braking vehicles (MB), lane changing vehicles with different lateral behaviours (LC) and the subject vehicle merging onto the main road with intense traffic (SVM). Each video, capturing a distinct traffic event, was divided into 5 or 6 shorter, consecutive clips. These clips are continuous and follow sequentially in the time domain, ensuring a coherent representation of the event from start to finish. The participants were presented with videos composed of these clips, arranged to be viewed one after another in their original sequential order. After viewing each clip, participants were asked to rate the perceived risk on a scale from 0 to 10. Each participant viewed a total of 16 events, chosen randomly from a set of 105, with four events from each of four scenarios (Chapter 4.2.1; Figure 4.1a, Figure C.1 in Appendix C.1 and Appendix C.4.1).

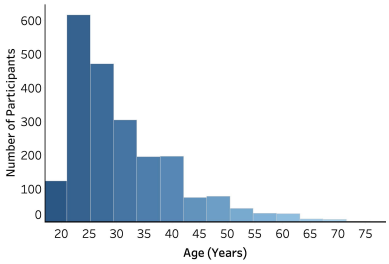
We excluded 177 participants due to incomplete responses or unusually rapid completion (under 504 seconds, less than the total duration of all videos), resulting in a final sample of 2,164 participants from Austria ($n = 16$), Belgium ($n = 23$), Czech Republic ($n = 28$), Denmark ($n = 4$), Estonia ($n = 19$), Finland ($n = 11$), France ($n = 50$), Gambia ($n = 1$), Germany ($n = 114$), Greece ($n = 129$), Grenada ($n = 1$), Hungary ($n = 65$), Iceland ($n = 2$), Italy ($n = 353$), Latvia ($n = 15$), Netherlands ($n = 80$), Norway ($n = 6$), Poland ($n = 439$), Portugal ($n = 583$), Qatar ($n = 2$), Slovenia ($n = 25$), South Korea ($n = 4$), Spain ($n = 154$), Sweden ($n = 24$), Switzerland ($n = 12$), and others (position information unavailable, $n = 3$). Within this chapter, participant ratings for each event were selected for consistency with the broader participant group. Specifically, if the correlation between a participant's series of ratings (five ratings for the MB, HB, and SVM scenarios, and six for the LC scenario) and the average ratings for the same event across all participants was below a threshold of $r < 0.3$ that participant's ratings were excluded. This exclusion criterion was applied to ensure data reliability and consistency across our sample. After



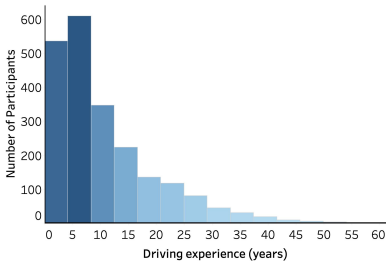
(a). Distribution of participants in Europe.



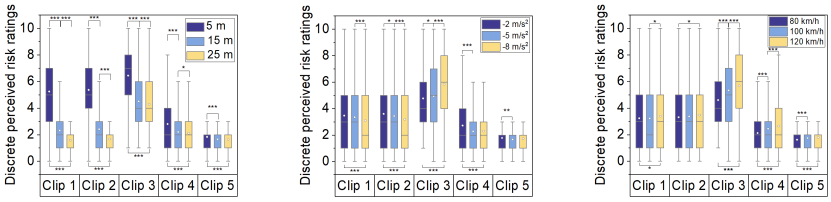
(b). Gender distribution



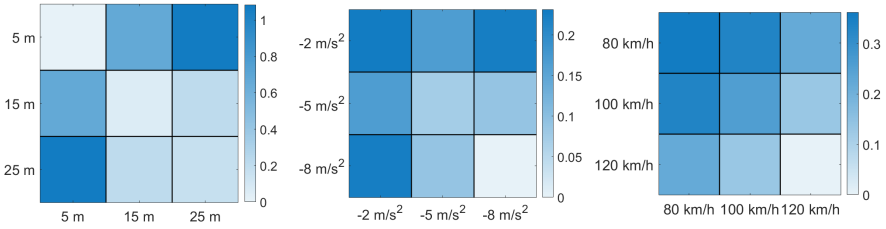
(c). Participants' age distribution in years



(d). Participants driving experience in years



(e). The influence of controlled parameters in MB scenario: initial merging gap (left), braking intensity (middle) and desired cruising speed (right) on perceived risk ratings per clip. Each bar chart shows perceived risk ratings categorised based on a specific controlled parameters with different clip numbers. For example, in the left bar chart, the dark blue bars represents all perceived risk ratings where the initial merging is 5 m, while other controlled parameters such as braking intensity and desired cruising speed are different. Similarly, the light blue and yellow bars represent the initial merging gap 15 m and 25 m respectively. This same pattern applies to the middle and right bar charts, which illustrate the impact of different levels of braking intensity and desired cruising speed on perceived risk ratings respectively.



(f). The J-S divergence of perceived risk ratings with respect to controlled parameters in MB scenario: initial merging gap (left), braking intensity (middle) and desired cruising speed (right). Each block within the matrix represents the J-S divergence between two groups categorised by the respective controlled parameter. For example, in the left heatmap (initial merging gap), the value in the third block of the first row represents the J-S divergence between perceived risk ratings for the 5 m and 25 m initial merging gap groups. Diagonal blocks (e.g., the first block in the first row) represent the J-S divergence of perceived risk within the same group, such as the 5 m group compared to itself. The colour intensity indicates the magnitude of divergence, with higher values representing greater differences between perceived risk ratings

Figure 4.3: Demographic statistics and variety of perceived risk ratings.

applying these criteria, a total of 141,628 perceived risk ratings remained valid for analysis. The breakdown of these ratings across the four scenarios is detailed in Table 4.2. The

Table 4.2: All controlled parameters significantly affect perceived risk.

Scenarios	Sample size	Controlled parameters	F	p-value	η^2
MB reacting to merging and hard braking	36,525	clip number	11475.048	< 0.001	0.571
		Initial merging distance (m)	3686.385	< 0.001	0.169
		Desired cruising speed (km/h)	68.125	< 0.001	0.004
		Braking Intensity of the merging vehicle (m/s^2)	29.243	< 0.001	0.002
HB reacting hard braking vehicles	33,355	clip number	6141.696	< 0.001	0.439
		Car following distance (m)	7753.686	< 0.001	0.321
		Desired cruising speed (km/h)	388.387	< 0.001	0.023
		Braking intensity of the leading vehicle (m/s^2)	93.563	< 0.001	0.006
LC reacting to merging vehicles with lateral control	39,018	clip number	3011.400	< 0.001	0.289
		Initial merging distance (m)	2083.866	< 0.001	0.052
		Lateral categories	63.492	< 0.001	0.005
		Driving style	133.461		0.007
SVM the subject AV merges onto the main road	32,720	clip number	6237.328	< 0.001	0.448
		Initial merging distance (m)	2539.261	< 0.001	0.136
		Desired cruising speed (km/h)	300.175	< 0.001	0.018
		Braking intensity of the leading vehicle (m/s^2)	87.656	< 0.001	0.005

participant demographics were diverse. In terms of gender distribution (as shown in Figure 4.3b), 49.7% of the participants identified as male, 49.2% as female, and 1.1% chose not to specify their gender. The age range of participants was broad, spanning from 18 to 73 years, with an average age of 31.2 years and a standard deviation of 9.5 years (refer to Figure 4.3c). All participants held valid driving licences for duration ranging from 1 to 55 years, with an average of 11 years and a standard deviation of 9.0 years (Figure 4.3d).

We analysed the effects of the controlled parameters (e.g., initial merging distance, desired driving speed, etc.) of all events with linear mixed models and the results show that all controlled parameters (e.g., initial merging distance, desired cruising speed, etc.) significantly affected perceived risk in the four scenarios ($p < 0.001$). Statistical results can be found in Table 4.2. Additionally, within individual clips, these controlled parameters, such as initial merging gap, braking intensity and desired cruising speed as illustrated in Figure 4.3e, lead to significantly different perceived risk levels. The evolution of perceived risk in time as measured after each video clip is shown in Figure 4.3e for Scenario MB and Figure C.2-C.4 in Appendix C.1 for HB, LC and SVM scenarios. These results confirmed that our designed scenarios effectively produce varying perceived risk levels among participants, validating their use for further analysis.

Considering the continuous nature of the video clips in each event, we used a rating transfer matrix created using the Jensen-Shannon (J-S) method to account for changes over time, which quantifies the similarity of perceived risk ratings among different events

(Chapter 4.2.3). Figure 4.3f shows the J-S divergence regarding different conditions in MB scenario (see Figure C.2-C.4 for HB, LC and SVM scenarios in Appendix C.1), where darker colour indicates a greater divergence and a smaller similarity, indicating a more significant influence of the controlled parameters. In Figure 4.3f, the left subfigure represents the J-S divergence between three groups controlled by different initial merging gaps 5 m, 15 m and 25 m; the middle subfigure shows the J-S divergence between three groups controlled by different braking intensity $-2\,m/s^2$, $-5\,m/s^2$ and $-8\,m/s^2$; the right subfigure represents the J-S divergence between three groups controlled by different cruising velocity 80 km/h, 100 km/h and 120 km/h.

In the left subfigure that represents the influence of initial merging gap, the J-S divergence is smaller within each group (i.e.,the diagonal), but is larger between different groups, for example, the first row shows different J-S divergence levels between three groups. This indicates that the initial merging gap affects perceived risk over time significantly. In the middle subfigure that represents the influence of braking intensity, the J-S divergence is small with the highest braking intensity ($-8\,m/s^2$). With lower braking intensity, the J-S divergence is large, which means that perceived risk is strongly influenced by other factors. However, the third row shows a significant J-S divergence difference between different groups, indicating that braking of the merging vehicle can significantly influence perceived risk over time. In the right subfigure that shows the influence of desired cruising speed, the J-S divergence is smaller only within group with the desired cruising speed 120 km/h. Within other two groups, the J-S divergence is large, indicating that when the desired cruising speed is not high enough, other factors influence perceived risk more rather than the desired cruising speed. In the third row, J-S divergence is significantly different between different groups, meaning that the desired cruising speed significantly influences perceived risk over time.

The results revealed that the controlled parameters influenced participants' perceived risk ratings significantly. These findings are consistent with results in Table 4.2 and previous studies [36], [117], [167].

4.4.2. CONTINUOUS PERCEIVED RISK RATINGS

Through rating moment alignment and interpolation, all rating moments are provided in Table C.1-Table C.4 in Appendix C.1. This transformation is critical to understanding the dynamic nature of perceived risk over time in various driving scenarios.

CONTINUOUS PERCEIVED RISK RATINGS IN VARIOUS DRIVING SCENARIOS

Examples of these continuous ratings from each of the four scenarios are shown in Figure 4.4. All for four scenarios can be found in Figure C.5 in Appendix C.1, illustrating the detailed variation in perceived risk in real time.

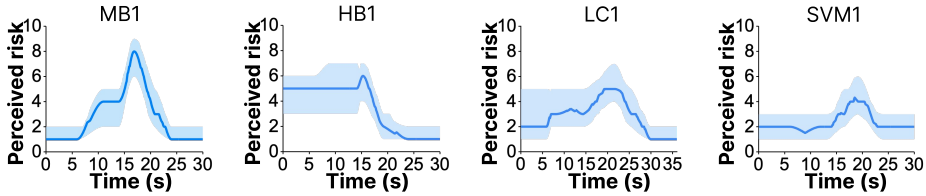
Figure 4.4a MB1 presents an example of all 27 MB events, where the subject AV reacts to a vehicle merging onto the current lane. At first, the merging vehicle is far from the subject vehicle, so the perceived risk is at a relatively low baseline. At $t = 6\,s$, the merging vehicle becomes visible on the on-ramp and approaches the subject AV and accordingly, perceived risk becomes higher. At $t = 12\,s$, the merging vehicle is in the merging area, waiting for merging onto the main road. At $t = 15\,s$, the merging vehicle starts to merge onto the current lane, increasing the perceived risk dramatically. After the merge ($t >$

18 s), the merging vehicle accelerates, and perceived risk decreases.

Figure 4.4b HB1 shows an example of all 27 HB events, where the subject AV reacts to a hard braking vehicle in front in the same lane. At the beginning, the subject AV is following a leading vehicle with a gap of only 5 m, leading to a considerable perceived risk. At $t = 13$ s, the leading vehicle abruptly brakes, causing a sharp increase in perceived risk which quickly peaks. Once the vehicle resumes speed, perceived risk decreases.

Figure 4.4c LC1 illustrates an example of all 24 LC events, where the subject AV reacts to a vehicle trying to change lane. The perceived risk is initially low, with the neighbouring vehicle behind the subject AV. As the neighbouring vehicle aligns beside the subject AV at around $t = 7$ s, perceived risk increases moderately and remains relatively stable. Perceived risk slightly fluctuates as the neighbouring vehicle begins to cut into the current lane in front at $t = 13$ s, peaking once the merging is completed at $t = 22$ s. Following the completion of the lane change, perceived risk returns to lower levels after $t = 24$ s as the vehicles stabilise in the current lane.

Figure 4.4d SVM1 displays an example of all 27 SVM events, where the subject AV merges to the main road. Initially, the subject AV is on an acceleration lane after an on-ramp with low perceived risk. The subject vehicle stays on the acceleration lane stably for a while, causing a relatively stable perceived risk. As the AV begins merging onto the main road at $t = 15$ s, perceived risk increases sharply, peaking when the merge is completed. The subject AV is stably following the leading vehicle on the main road after $t = 21$ s, and the perceived risk drops back to a low level.



(a). MB1 (b). HB1 (c). LC1 (d). SVM1

Figure 4.4: Examples of continuous perceived risk ratings in four driving scenarios. The solid blue curve illustrates the median perceived risk over time, providing a central tendency of participant responses. The light blue shading indicates the perceived risk range, spanning from the 25th to the 75th percentile, which captures the variability and dispersion of perceived risk ratings among participants. This visual representation emphasises the dynamic nature of perceived risk and its sensitivity to specific driving events and manoeuvres, showcasing the effective interpolation method and analysis of risk fluctuations in real-time driving scenarios.

4.4.3. MODEL PERFORMANCE IN PREDICTING PERCEIVED RISK

We used three computational perceived risk models as ideal observer candidates — the PCAD model, the DRF model, and DNNs — to predict continuous perceived risk ratings across the four driving scenarios. Each model offers a unique perspective on assessing and predicting perceived risk in driving scenarios, especially in automated vehicles. We presented the PCAD model and benchmarked it against other models including DRF in Ref. [167] where we found that PCAD matched the human perceived risk best followed by DRF in merging and hard braking and in obstacle avoidance scenarios. PCAD and DRF are grounded in first principles with clear and structured methodologies. If they could

accurately predict continuous perceived risk ratings, it would significantly simplify the analysis process due to their transparent and straightforward frameworks, making them particularly valuable for understanding the dynamics of perceived risk. Additionally, we used the superiority of DNNs in learning complex, nonlinear patterns in large datasets to predict continuous perceived risk ratings [179]. Each of the three models PCAD, DRF and DNNs were captured in depth in Chapter 4.3. Table 4.3 shows the structure of DNNs. All three models predict perceived risk based on the corresponding kinematic data of all events, such as the position, velocity, and acceleration of the subject AV and neighbouring vehicles (see Table 4.4 for more details).

We conducted a calibration for PCAD and DRF and a training process for DNNs using the dataset obtained previously. Our calibration aimed to minimise the models' prediction error, thereby aligning the model outputs closely with the continuous perceived risk ratings in these scenarios (Chapter 4.3.3 and 4.3.4).

Figure 4.5 illustrates the prediction performance of the three models throughout the event timelines (examples for four scenarios of the three models; See Figure C.7 in Appendix C.1 for full results).

PCAD (Figure 4.5a) shows a close alignment with peak risk periods, particularly in periods with abrupt changes. However, it is less accurate when the relative position of the subject AV and the neighbouring vehicles are relatively constant, such as the beginning stage in MB1, HB1 and LC1. Additionally, it overestimates perceived risk in SVM1.

DRF (Figure 4.5b) also closely captures the peak risk periods across all scenarios. Like PCAD, DRF's predictive performance diminishes in scenarios where the positions of the subject AV and neighbouring vehicles are relatively constant.

DNNs (Figure 4.5c) demonstrate superior adaptability across all scenarios, accurately mirroring the observed risk profiles with high fidelity. The DNN models especially outperform in complex scenarios like SVM1, when people are facing multiple other road users.

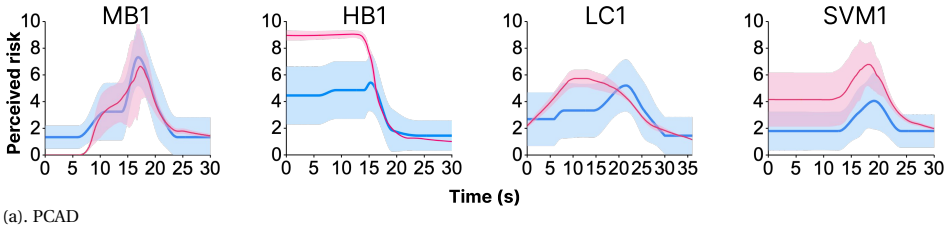
The epistemic uncertainties, represented by the light-coloured areas around the model outputs, are notably narrower for DNNs compared to PCAD and DRF (Figure 4.5). This indicates a higher confidence level in the predictions made by DNNs, showcasing their robustness in dealing with complex driving scenarios. The comparative analysis of prediction error across all models (Figure 4.5d) highlights the overall superior performance of DNNs, which consistently exhibit lower prediction errors and less variability in their predictions, confirming their robustness and reliability in perceived risk prediction. The box plots reveal that while PCAD and DRF are effective in certain conditions, their limitations are evident in the broader range of driving scenarios where DNNs provide more consistently accurate predictions.

Table 4.3: Overview of DNNs configurations for different scenarios

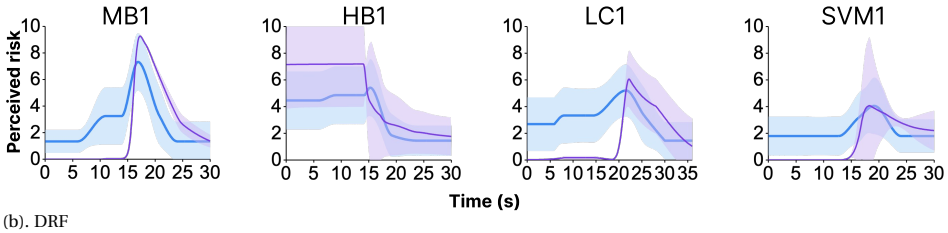
Scenarios		Input dimensions	Input normalisation	Activation	Network size
MB		21	Z-Score	ReLU	21 × 500 × 2
HB		11	Z-Score	ReLU	11 × 500 × 2
LC	Normal	20	Z-Score	ReLU	20 × 500 × 2
	Fragmented	20	Z-Score	ReLU	20 × 500 × 2
	Aborted	20	Z-Score	ReLU	20 × 500 × 2
SVM		32	Z-Score	ReLU	32 × 500 × 2

Table 4.4: Features that models use

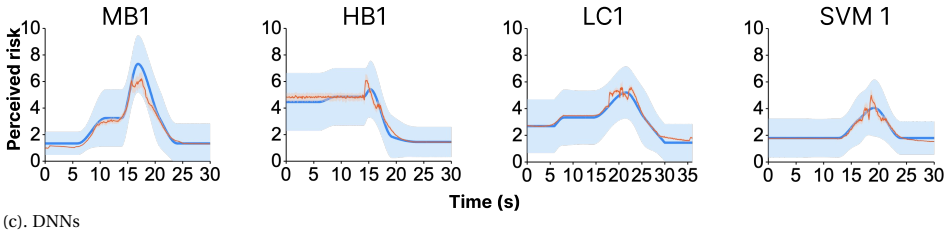
Variables	Explanations	Inputs of models
$v_{s,x}$	Longitudinal velocity subject vehicle	PCAD, DRF, and DNNs
$v_{s,y}$	Lateral velocity subject vehicle	PCAD and DNNs
$a_{s,x}$	Longitudinal acceleration subject vehicle	PCAD and DNNs
$a_{s,y}$	Lateral acceleration subject vehicle	PCAD and DNNs
$v_{n,x}$	Longitudinal velocity neighbouring vehicle	PCAD and DNNs
$v_{n,y}$	Lateral velocity neighbouring vehicle	PCAD and DNNs
$a_{n,x}$	Longitudinal acceleration neighbouring vehicle	PCAD and DNNs
$a_{n,y}$	Lateral acceleration neighbouring vehicle	PCAD and DNNs
$a_{n,y}$	Lateral acceleration neighbouring vehicle	PCAD and DNNs
$v_{nb,x}$	Longitudinal velocity neighbouring vehicle (behind)	DNNs in SVM
$v_{nb,y}$	Lateral velocity neighbouring vehicle (behind)	DNNs in SVM
$a_{nb,x}$	Longitudinal acceleration neighbouring vehicle (behind)	DNNs in SVM
$a_{nb,y}$	Lateral acceleration neighbouring vehicle (behind)	DNNs in SVM
Δx	Longitudinal distance to the neighbouring vehicle	PCAD, DRF and DNNs
Δy	Lateral distance to the neighbouring vehicle	PCAD, DRF and DNNs
Δv_x	Longitudinal relative velocity to the neighbouring vehicle	DNNs
Δv_y	Lateral relative velocity to the neighbouring vehicle	DNNs
Δa_x	Longitudinal relative acceleration to the neighbouring vehicle	DNNs
Δa_y	Lateral relative acceleration to the neighbouring vehicle	DNNs
Δx_b	Longitudinal distance to the neighbouring vehicle (behind)	DNNs in SVM
Δy_b	Lateral distance to the neighbouring vehicle (behind)	DNNs in SVM
$\Delta v_{x,b}$	Longitudinal relative velocity to the neighbouring vehicle (behind)	DNNs in SVM
$\Delta v_{y,b}$	Lateral relative velocity to the neighbouring vehicle (behind)	DNNs in SVM
$\Delta a_{x,b}$	Longitudinal relative acceleration to the neighbouring vehicle (behind)	DNNs in SVM
$\Delta a_{y,b}$	Lateral relative acceleration to the neighbouring vehicle (behind)	DNNs in SVM
$\Delta v_{s,u,x}$	Subject vehicle's longitudinal uncertain velocity to the neighbouring vehicle	PCAD and DNNs
$\Delta v_{s,u,y}$	Subject vehicle's lateral uncertain velocity to the neighbouring vehicle	PCAD and DNNs
$\Delta v_{n,u,x}$	Neighbouring vehicle's longitudinal uncertain velocity to the subject vehicle	PCAD and DNNs
$\Delta v_{n,u,y}$	Neighbouring vehicle's lateral uncertain velocity to the subject vehicle	PCAD and DNNs
$\Delta v_{nb,u,x}$	Neighbouring vehicle's longitudinal uncertain velocity to the subject vehicle (behind)	DNNs in SVM
$\Delta v_{nb,u,y}$	Neighbouring vehicle's lateral uncertain velocity to the subject vehicle (behind)	DNNs in SVM
$DRAC_{u,x}$	Longitudinal acceleration to avoid a crash caused by uncertain velocity	DNNs
$DRAC_{u,y}$	Lateral acceleration to avoid a crash caused by uncertain velocity	DNNs
$DRAC_{u,b,x}$	Longitudinal acceleration to avoid a crash caused by uncertain velocity (behind)	DNNs in SVM
$DRAC_{u,b,y}$	Lateral acceleration to avoid a crash caused by uncertain velocity (behind)	DNNs in SVM
$DRAC_{R,x}$	Longitudinal acceleration to avoid a crash caused by real velocity	DNNs
$DRAC_{R,y}$	Lateral acceleration to avoid a crash caused by real velocity	DNNs
$DRAC_{R,b,x}$	Longitudinal acceleration to avoid a crash caused by real velocity (behind)	DNNs in SVM
$DRAC_{R,b,y}$	Lateral acceleration to avoid a crash caused by real velocity (behind)	DNNs in SVM



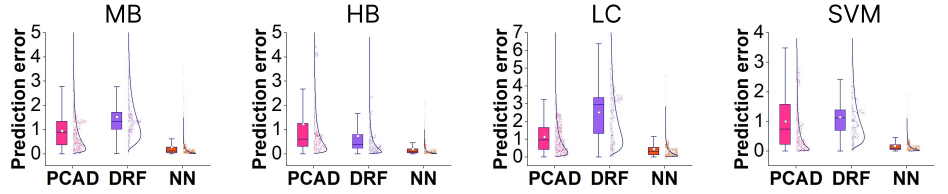
(a). PCAD



(b). DRF



(c). DNNs



(d). Overall prediction error

Figure 4.5: Prediction results (red, purple, and orange) versus measured perceived risk (blue) of the three computational perceived risk models (**a**) PCAD , (**b**) DRF and (**c**) DNNs for the specific events MB1, HB1, LC1, SVM1 as a function of time where solid curves represent model output and the light coloured areas represents the epistemic uncertainty of the corresponding model, with the solid blue curve being the averaged perceived risk data and light blue area being the aleatoric uncertainty. (**d**) shows the prediction error with box plots in for all events within the four scenarios. Boxes represent 25 ($Q1$) and 75 ($Q3$) percentile. The middle dash and the circle represent the median ($Q2$) and mean. The lower and upper whiskers represent $Q1 - 1.5IQR$ and $Q3 + 1.5IQR$ (IQR is the $Q3 - Q1$)

4.4.4. DECODING PERCEIVED RISK

SHAP was employed to analyse the contributions of various kinematic factors to perceived risk based on DNNs as they performed the best in predicting continuous perceived risk ratings [168]. This analysis provides insights into the global feature importance of different factors in four driving scenarios. The SHAP values illustrate how cer-

tain key factors impact perceived risk overall, indicating a core similarity in the dynamic nature of perceived risk across diverse driving scenarios. Additionally, a local feature importance analysis was conducted, which examined how the importance of these factors varies over time within each event in various scenarios, providing a dynamic view of perceived risk.

To ensure that the SHAP analysis based on the trained DNNs provides meaningful results and that the models are not overfitted, we compared the original DNNs with a simplified version with much fewer neurons. Both models were trained and validated under the same conditions to assess the overfitting risk and robustness against variations during training, such as neuron dropout. The results, detailed in Appendix C.3, confirm that the trained DNNs do not suffer from significant overfitting and provide a reliable basis for SHAP analysis.

The SHAP analysis reveals a notable consistency in the main contributing factors across different scenarios (Figure 4.6, Figure 4.7a-4.7f). Key factors such as the relative longitudinal distance, subject vehicle's longitudinal velocity, and longitudinal manoeuvre uncertainty recurrently emerge as significant across all scenarios. This consistency suggests that despite the different nature of the driving events, certain fundamental dynamics — like the spatial and speed relationships between vehicles — uniformly influence perceived risk.

Longitudinal distance exhibits a critical influence on perceived risk in all scenarios except LC (aborted) (Figure 4.7e); a lower value, indicating closer proximity to the neighbouring vehicle, sharply increases perceived risk. Conversely, a higher value, where the neighbouring vehicle is farther away, generally leads to a decrease in perceived risk, maintaining a stable but low influence on the model's output as the immediate threat potential decreases. However, lateral distance did not appear as a contributing factor across all scenarios but was only observed in the SVM scenario. The pattern, however, remains consistent with that of relative longitudinal distance: the smaller the distance, the higher the perceived risk, and vice versa.

The contributions of velocity (both longitudinal and lateral) of the subject vehicle and neighbouring vehicles notably rank among the top factors influencing perceived risk across all scenarios (Figure 4.7b-4.7f). We observed notable patterns of how longitudinal and lateral velocity impacts perceived risk in MB and LC scenarios. For instance, in LC (normal)(Figure 4.7c), higher longitudinal velocities of both the subject AV and neighbouring vehicles are associated with increased perceived risk. Similarly patterns were also observed in LC (fragmented) (Figure 4.7d) and LC (aborted) (Figure 4.7e). However, in MB (Figure 4.7a), the impact of the subject vehicle's velocity on perceived risk exhibits a different trend: an increased velocity correlates with a reduction in perceived risk. To explore this inconsistency, it is essential to examine the local feature importance over time (Figure 4.7g). We found that a lower longitudinal velocity exactly appears when the merging vehicle is entering the lane and undergoing intense braking. At these moments, although the subject vehicle's velocity sharply decreases, other factors like the relative distance dominate perceived risk. This does not contradict our driving experience. Overall, a higher velocity still generally correlates with an increased perceived risk.

The relative velocities between the subject AV and neighbouring vehicles also significantly influence perceived risk. As observed in MB (Figure 4.7a) and LC (fragmented)

(Figure 4.7d) scenarios, a trend emerges where increased relative longitudinal velocities seem to increase perceived risk. However, in LC (aborted) scenario (Figure 4.7e), it is observed that increased relative longitudinal velocities decrease perceived risk. Upon examination of Figure 4.7j and Figure 4.7k, it becomes apparent that the primary difference between LC (fragmented) and LC (aborted) arises from the initial phase where the neighbouring vehicle was approaching the subject AV from the rear. In later stages, the impact of relative longitudinal velocity displays a similar trend. Despite identical settings for these two scenarios, variability exists in perceived risk ratings as illustrated in Figure C.5c, which may be a plausible reason for the differences observed in SHAP value. In summary, the overall pattern suggests that, generally, larger relative velocities are associated with higher perceived risk.

The analysis of acceleration effects on perceived risk reveals different patterns in various driving scenarios. In MB, HB and SVM scenarios (Figure 4.7a, Figure 4.7b and Figure 4.7f) a notable increase in perceived risk correlates with larger values of subject longitudinal acceleration, which means that the subject AV is primarily accelerating. This alignment with common driving experiences, where rapid deceleration often signals an immediate hazard, while a start of acceleration indicates a lower risk.

However, the influence of lateral acceleration is less consistent across the scenarios. Relative lateral acceleration shows a significant impact on perceived risk only in HB scenario where there is no intentional lateral motion (Figure 4.7b). Interestingly, the local feature importance analysis (Figure 4.7h) highlights that this impact becomes pronounced during the transition from braking to acceleration, coinciding with a lateral movement observed at the end of a hard braking phase due to vehicle dynamic constraints (see videos in Appendix C.4.1. This movement occurs as the vehicle ends its braking and begins to accelerate, a phase typically associated with lower perceived risk. It is evident that this lateral movement, captured by the DNNs, might be a contributory factor to the observed perceived risk pattern. Despite these specifics, the general trend suggests that when the subject acceleration is closer to forward (e.g., when the lead vehicle stops braking and starts to accelerate), indicating reduced necessity for braking, perceived risk tends to decrease. In contrast, no clear pattern emerges for lateral acceleration across the scenarios, suggesting that its impact on perceived risk is less pronounced or context-dependent.

Longitudinal manoeuvre uncertainty indicated as the longitudinal acceleration needed by the subject vehicle (represented as Longitudinal collision-avoidance acceleration (uncertainty) in Figure 4.7 with forward as positive and backward as negative; Chapter 4.3.4) and the concept of “uncertain velocity” (represented as subject or neighbour’s longitudinal manoeuvre uncertainty in Figure 4.7 with forward and left as positive; Chapter 4.3.4) are associated with the potential manoeuvre of the neighbouring vehicle and the imprecise control of the subject AV. In general, as uncertainty increases, so does perceived risk, reflecting concerns about the effectiveness of possible evasive manoeuvres under limited conditions. Specifically, in MB, HB and LC scenarios (Figure 4.7a-4.7e), when

longitudinal uncertainty is shown in red, it indicates higher values, suggesting that the subject AV may need to accelerate to avoid a collision, although computationally indicated, maintaining speed would be adequate. Conversely, when values are shown in blue, indicating lower values, it suggests that the subject AV needs to decelerate to avoid a collision, thereby indicating a higher perceived risk. An interesting exception occurs in Figure 4.7e, representing an aborted lane change scenario, where the opposite trend is observed: perceived risk increases when acceleration is necessary. This observation does not contradict the general trend, but highlights the unique circumstances of this scenario. In LC (aborted), the neighbouring vehicle does not enter the current lane as designed, leading to different risk dynamics. Typically, as the neighbouring vehicle approaches from behind and aligns parallel to the subject AV, perceived risk peaks. In this scenario, in particular, acceleration is required to maintain a safe distance from the neighbouring vehicle, thus aligning with the increased perceived risk. This detailed dynamic indicates the complex interaction of manoeuvre uncertainty and perceived risk in dynamic driving scenarios.

In terms of lateral manoeuvre uncertainty, regardless of whether it is indicated by uncertain velocity or collision avoidance acceleration in the lateral direction, a clear pattern emerges: greater lateral uncertainty generally correlates with increased perceived risk, particularly when the neighbouring vehicle is in close proximity and maintaining a constant relative position, such as during side-by-side driving stages (Figure 4.7i). This trend is consistently observed across most scenarios; however, exceptions are noted in MB (Figure 4.7a) and LC (aborted) (Figure 4.7e) scenarios. In these cases, the computation of lateral uncertain velocity yields lower values during stages after merging or when the neighbouring vehicle paused a lane change but stayed in front of the subject AV. During these stages, despite the lower calculated uncertainty, perceived risk remains high due to the potential longitudinal hazard, which was captured by DNNs demonstrating the model's sensitivity to driving contexts.

Comparative analysis reveals distinct patterns in how kinematic factors contribute to perceived risk. In MB and HB scenarios, longitudinal collision-avoidance acceleration, particularly caused by manoeuvre uncertainty, and the subject's longitudinal velocity prominently influence perceived risk, highlighting the critical role of speed and response urgency in these high-stakes situations. The SVM scenario shows a strong impact from both longitudinal and lateral distances and both leading and following vehicles, indicating that spatial relationships are pivotal when merging into traffic. Conversely, the LC scenarios depict a nuanced contribution landscape where lateral manoeuvre uncertainty becomes more pronounced, reflecting the complex dynamics of lane-changing manoeuvres. These variations emphasise how different driving contexts prioritise various aspects of vehicular movement and proximity in perceived risk.

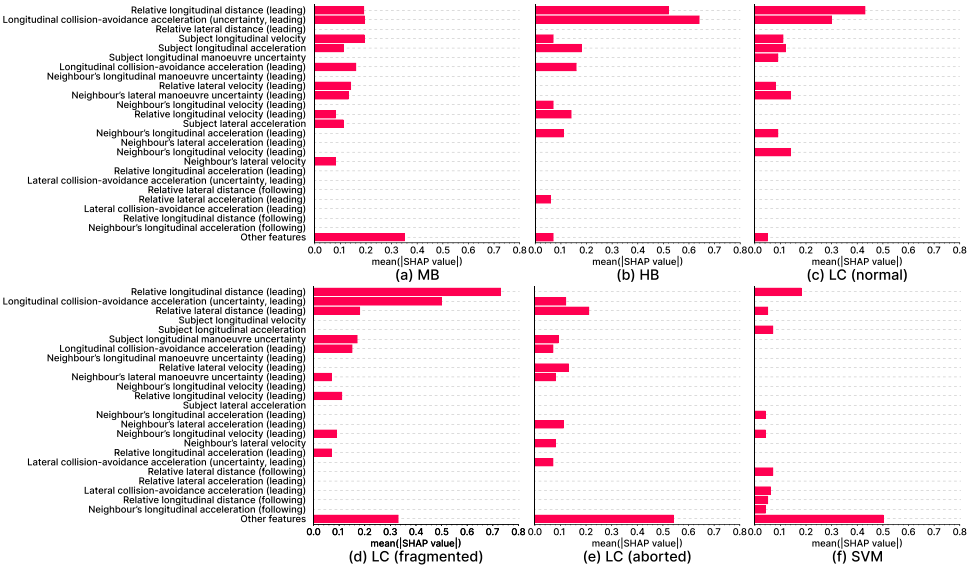
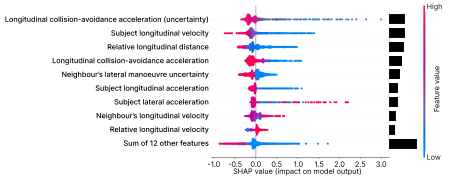
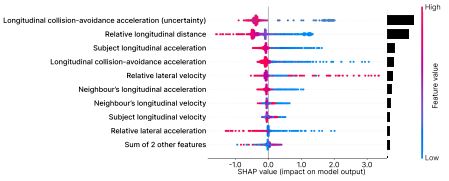


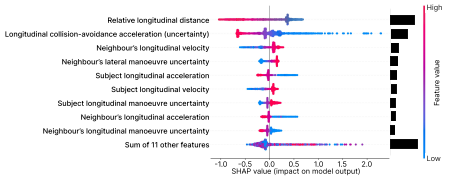
Figure 4.6: Factor SHAP value ranking in four scenarios. This ranking is based on the mean SHAP value in corresponding scenarios.



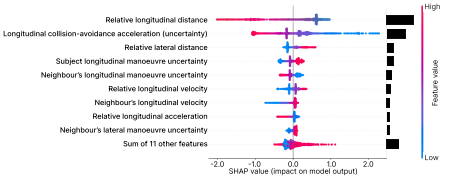
(a). Global feature contributions and ranking to perceived risk in MB



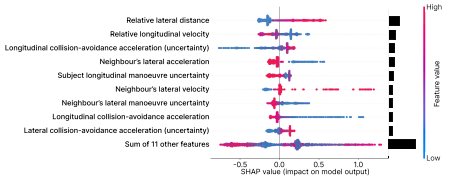
(b). Global feature contributions and ranking to perceived risk in HB



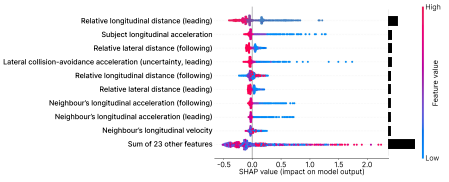
(c). Global feature contributions and ranking to perceived risk in LC (normal)



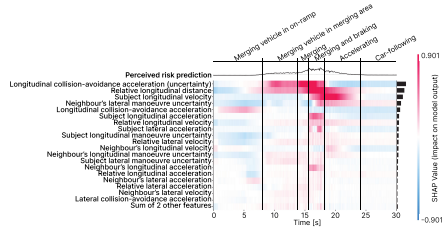
(d). Global feature contributions and ranking to perceived risk in LC (fragmented)



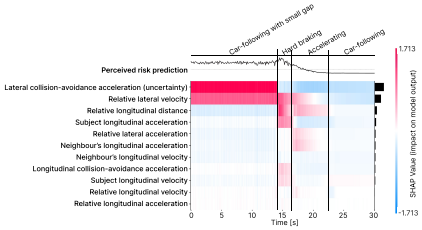
(e). Global feature contributions and ranking to perceived risk in LC (aborted)



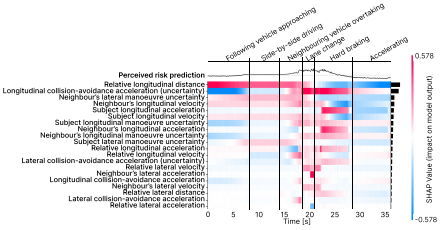
(f). Global feature contributions and ranking to perceived risk in SVM



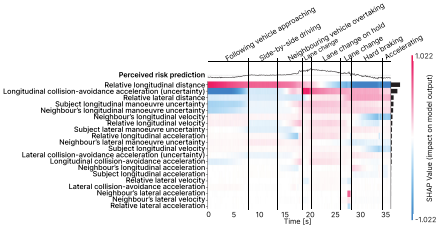
(g) Feature contributions to perceived risk over time in MB1 event



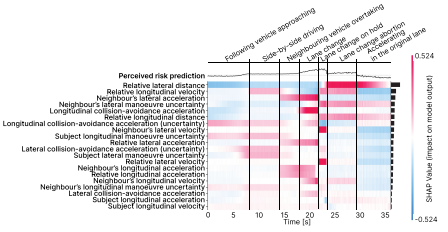
(h) Feature contributions to perceived risk over time in HB1 event



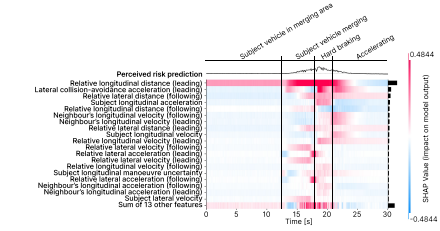
(i) Feature contributions to perceived risk over time in LC1 (normal) event



(j) Feature contributions to perceived risk over time in LC13 (fragmented) event



(k) Feature contributions to perceived risk over time in LC19 (aborted) event



(l) Feature contributions to perceived risk over time in SVM1 event

Figure 4.7: Global feature contributions and ranking to perceived risk in **(a)** MB scenario, **(b)** HB scenario, **(c)** LC (normal), in **(d)** LC (fragmented), **(e)** LC (aborted), **(f)** SVM scenario; Local feature contributions to perceived risk over time in **(g)** MB1, **(h)** HB1, **(i)** LC1 (normal), **(j)** LC13 (fragmented), **(k)** LC19 (aborted), **(l)** SVM1. See Appendix C.4.2 for the local feature contributions in all 105 events. For the global feature contributions, red indicates higher feature values while blue indicates lower feature values. Take the longitudinal relative distance as an example, red indicates a larger distance corresponding to a lower perceived risk with a negative SHAP value on model output. However, for local feature contributions, red and blue indicate the feature contribution to model output not the feature values themselves.

4.5. DISCUSSION

This chapter undertook a comprehensive examination of perceived risk within automated vehicle environments, with a focus on dynamically assessing how different factors influence perceived risk during driving. Using sequential video clips, we developed a novel method to continuously collect perceived risk data, resulting in 141,628 ratings after selection from participants mainly in Europe. Our approach surpasses traditional methods by providing a real-time, continuous assessment rather than static snapshots often limited to post-experience surveys. The core of our research involved altering various driving behaviours and conditions to understand their impact on perceived risk, revealing critical insights into both global and local feature importance in the time do-

main. These insights not only align with existing theories on perceived risk but also highlight new dynamics in how people perceive risk in automated driving. The findings pave the way for refining automated vehicle technology and improving safety protocols by focusing on human-centred design principles in automated vehicle interactions.

A significant advancement was represented by our method for collecting perceived risk data, which moved beyond traditional approaches that typically depend on static post-experience surveys [13], [14], [109]. By employing high-fidelity simulation software to generate dynamic driving scenarios, we were able to collect real-time, continuous perceived risk ratings from participants. This continuous data collection mirrors the fluid dynamics of actual driving more closely than traditional methods at population level [11], [20], [21], [31]–[33] or event level [9], [14], [35], [37], [45], [141], capturing transient continuous time changes in driver perception that would otherwise be overlooked. Additionally, this approach offers a cost-effective alternative to real-world driving tests, which are logistically complex and financially demanding [195], [196]. Simulations not only reduce these burdens but also enhance the safety and repeatability of the experiments, making this method an efficient and effective tool for large-scale research without compromising the depth and quality of data [197].

Our findings demonstrate a clear alignment with existing studies in terms of how perceived risk is influenced by specific driving conditions. Particularly, our data reaffirmed the well-established notion that closer relative distances and higher (relative) speeds correlate strongly with increased perceived risk [36], [69], [110], [117]. These results are consistent with the broader literature on driving safety and risk perception, which has long recognised that reduced following distances and elevated speeds elevate drivers' sense of danger due to decreased reaction times and increased potential for severe outcomes in the event of a collision [127], [128].

Insights into how specific driving conditions and manoeuvre uncertainties dynamically impact perceived risk were unveiled through our analysis of local feature importance (Figure 4.7h-Figure 4.7l). We discovered that manoeuvre uncertainties, variations in the actions of both the subject vehicle and nearby traffic, significantly impact perceived risk, particularly during complex manoeuvres such as merging or lane changing. These findings highlight the importance of advanced predictive algorithms in automated driving systems that can anticipate and mitigate perceived risk associated with unpredictable driving behaviours [129], [187]. In Chapter 3, a model of "uncertain velocity" within the PCAD model was derived with the calibrated parameters fitting perceived risk data from a simulator study. Now we use the same "uncertain velocity" model and parameters as inputs of DNNs where they strongly contributed to perceived risk. Here, it should be noted that this uncertain velocity is a construct explaining perceived risk, but is not based on actual uncertainty or direct subjective evaluation of uncertainty. However, it is also important to note that more detailed factors representing the motion prediction of all road users could be integrated into future analyses to provide a richer understanding of uncertainty in these dynamic driving conditions. Such an approach could gain deeper insights into manoeuvre uncertainties, enhancing predictive models for automated driving systems.

Moreover, this chapter emphasises the temporal variability of perceived risk, showing that the influence of certain conditions can intensify as drivers approach critical

points within a driving scenario. This underscores the necessity for automated systems to not only react to immediate risks but also adapt to potential future risks [198], [199], enhancing their predictive capabilities to improve safety and driver trust in automated technologies.

The PCAD and DRF models have been selected as candidates of ideal observers to predict continuous perceived risk ratings in dynamic driving scenarios. However, their predictive capabilities are less effective than DNNs in stable conditions, such as when vehicles maintain close yet constant distances either in car following or side-by-side driving situations (Figure 4.7h, Figure 4.7c-Figure 4.7e). This limitation may stem from the foundational principles upon which these models are built. Both PCAD and DRF are grounded in collision avoidance theory and using datasets focusing on scenarios where a change in the traffic environment might lead to a collision [150], [167]. These models are able to predict perceived risk well in situations where rapid responses to avoid immediate threats are necessary, such as sudden stops or evasive manoeuvres. In stable driving conditions, where the relative positions of vehicles do not change significantly, these models struggle because they are designed to react to changing traffic conditions. This issue highlights an important aspect of automated driving systems and computational perceived risk models: the need for an expanded framework that can effectively interpret risk even when driving conditions are stable but potentially risky. Although stable conditions may not trigger immediate alerts from collision avoidance algorithms, they can still present considerable risks, especially when vehicles are travelling at high speeds while being close to each other [200], [201]. Enhancing these models to better understand and predict perceived risk in these scenarios would improve their utility and reliability, ensuring that automated systems can maintain safety across a broader range of driving conditions.

We chose DNNs for their superior ability to model complex, non-linear interactions within large datasets, which traditional models like PCAD and DRF struggle with, especially in scenarios requiring time-continuous prediction of perceived risk [37], [38]. Unlike these traditional models, which predict event-based discrete perceived risk well, DNNs excel in handling dynamic and unpredictable scenarios, making them ideal for studying perceived risk continuously and in depth. Our primary goal with DNNs is not just to predict perceived risk accurately but to use these predictions to understand underlying patterns, essentially treating the model as an ideal observer [164]–[166]. By using the advanced learning capabilities of DNNs, we bridged the gap between human perceptions and AVs behaviours, enhancing both the safety and the acceptance of automated vehicle technologies.

In this chapter, we intentionally excluded direct positional and velocity variables of the subject AV and neighbouring vehicles as inputs for DNNs due to the nature of our experimental design. These variables exhibit specific, predictable patterns over time within the controlled driving scenarios we used. If included, the DNNs might learn to associate these time-dependent patterns with perceived risk, rather than learning from more substantive risk factors that generalise across different driving contexts [202], [203]. This decision aims to prevent the DNNs from overfitting to scenario-specific conditions that do not necessarily translate to real-world situations. By focusing on broader, more general factors of perceived risk, our DNNs aim to capture essential elements of risk that

apply across various driving environments. This approach enhances the model's robustness and guarantees that its predictive abilities are based on truly significant driving dynamics, instead of being influenced by the specific characteristics of our experimental arrangement.

An unexpected finding was the absence of lateral relative distance to neighbouring vehicles as a significant contributing factor, particularly during the side-by-side driving stages where this variable would apparently play an important role in perceived risk [204], [205]. Interestingly, what did emerge as significant was lateral uncertainty — the variability in the lateral movements of neighbouring vehicles represented by the concept "uncertain velocity", suggesting that the unpredictability of these movements contributes more to perceived risk than mere distance. This indicates that drivers may be more sensitive to the potential for sudden changes in the positions of adjacent vehicles than to their steady state lateral distance. Perceived risk, in this case, seems to stem more from the uncertainty or the possibility of sudden lateral movements (such as an unexpected lane change). This finding can significantly inform the development of safety mechanisms in automated driving systems and computational perceived risk modelling, where detecting and responding quickly to potential lateral movements could reduce perceived risks more effectively than simply maintaining a constant lateral distance.

The uncertain velocity that indicates longitudinal manoeuvre uncertainty was a significant factor to perceived risk only in scenario LC and did not emerge at all in the other three scenarios. This suggests that while uncertain velocity does reflect some aspects of perceived risk, it does not comprehensively capture the risk perception mechanism on its own. This may be because the complexity of perceived risk usually involves several different factors working together. Important factors such as the relative distance between vehicles, their relative speeds, and environmental conditions often come together to shape a driver's perceived risk [206], [207]. Uncertain velocity alone may not sufficiently account for how drivers assess the safety and risk of their immediate driving environment.

While this chapter provides significant insights into perceived risk in AVs, there are several limitations that indicate directions for further investigation. In this study, we trained six different DNNs for specific scenarios, which makes the models quite scenario-specific and raises the question of transferability despite our additional evaluations confirming no significant overfitting. To overcome this, training a more general neural network capable of fitting all scenarios could offer deeper and more universal insights, increasing both the generalisability and robustness of the model's predictions. However, this approach would require addressing several challenging issues, such as the variability in input parameters across different driving scenarios. Developing a unified model would need careful consideration to generalise these inputs. Additionally, the DNN predicts perceived risk based on the variables in Table 4.4 while the SHAP analysis identifies the factors that contribute the most to the variance within the current datasets, see Figure 4.6. A well-founded reduction of variables used may enhance model robustness against overfitting. Furthermore, a more independent validation of the model is required to further assess its generalisability. This can involve training the model on data from three different scenarios and then evaluating its prediction performance on the fourth scenario, or even testing it on entirely new data collected from different experimental

settings. Such validation would mitigate the risks of overfitting to specific scenarios.

The simulated scenarios and rating method used in this chapter, although valuable for controlled analysis, do not fully capture the variability and complexity of real traffic. Future studies might consider incorporating on-road testing, or collection of on-road videos to validate and enrich the findings obtained from simulated environments. Additionally, our research primarily focused on kinematic factors such as speed and relative distance, which represent only a part of the broader range of factors influencing perceived risk. We did not investigate how demographic characteristics such as age and driving experience influence perceived risk. These factors are known to affect perceived risk, and can be systematically and efficiently explored in future online studies. Future research can focus more on personalised approaches to predict perceived risk in AVs. Moreover, the driving scenarios used were limited to motorway driving and car to car interactions. To advance our understanding of perceived risk across diverse driving environments, future research should also explore interactions with various road users such as trucks, motorcycles, pedestrians, and cyclists across different road types. This approach will allow us to develop models that reflect the complexities of real-world traffic dynamics, enhancing the predictive accuracy and reliability of computational perceived risk models in varied scenarios.

5

DESIGNING USER INTERFACES FOR PARTIALLY AUTOMATED VEHICLES: EFFECTS OF INFORMATION AND MODALITY ON TRUST AND ACCEPTANCE

Trust and perceived safety are pivotal in the acceptance of automated vehicles and can be enhanced by providing users with automation information on the (safe) operation of the vehicle. This chapter aims to identify how user interfaces (UI) can enhance drivers' trust and acceptance and reduce perceived risk in partially automated vehicles. Four interfaces were designed with different levels of complexity. These levels were achieved by combining automation information (surrounding information vs surrounding and manoeuvre information) and modality (visual vs visual and auditory). These interfaces were evaluated in a driving simulator in which a partially automated vehicle reacted to an event of a merging and braking vehicle in its front. The criticality of the events was manipulated by the factors merging gap (in meters) and deceleration (m/s^2) *of the vehicle in front. The reaction of the automation was either to brake or to change lanes. The results show that an optimal combination of automation information and modality enhances drivers' trust, communication with automation, perceived ease of use, and perceived usefulness. More specifically, the most complex UI, which provided surrounding and manoeuvre information via the visual and auditory modalities, was associated with the highest trust and acceptance ranking and the lowest perceived risk. Manoeuvre information delivered*

The content of this chapter has been published in
Kim, S.*, He, X.*, van Egmond, R., & Happee, R. Designing user interfaces for partially automated vehicles: effects of information and modality on trust and acceptance *Transportation Research Part F: Traffic Psychology and Behaviour, 103, 404-419.* * Co-first author

through the auditory modality was particularly effective in enhancing trust and acceptance. The benefits of the UIs were consistent over events. However, in the most critical events, drivers did not feel entirely safe and did not trust the automation completely. This chapter suggests that the design of UIs for partially automated vehicles shall include automation information via visual and auditory modalities.

5.1. INTRODUCTION

Automated vehicle technology is rapidly developing, promising increased safety and comfort to drivers [208]. As technology continues to progress, it is expected to bring disruptive changes to transportation systems and people's lifestyles [209]. Automated vehicles may enable drivers to engage in non-driving activities, such as working, reading, or resting [210]. However, the successful diffusion of automated vehicles depends on the acceptance of the new technology [211]. Trust is an essential prerequisite for using automation, as it is a key predictor of acceptance and a positive user experience [21], [29], [212]–[215]. Trust in automation refers to the attitude that the system will help users achieve their goals in a situation characterised by uncertainty and vulnerability [28]. Perceived risk captures the level of risk experienced in driving [53]. Li *et al.* [35] considered perceived safety as an antecedent of trust, while Nordhoff *et al.* [11] found that perceived safety emerges from trust. We consider trust and perceived safety (or risk) to be interacting perceptions which are essential in the interaction between drivers and automated vehicles. Trust and perceived safety primarily derive from the automation performance as perceived by the user and the driving conditions, including the (dangerous) behaviour of other road users. User interfaces can thereby help to calibrate trust and perceived risk as they can inform users of the (safe) operation of the automated vehicle and its capability to deal with other road users [35]. The potential of user interfaces to enhance trust and perceived safety and to foster acceptance of automated driving was demonstrated in our recent survey [216]. However, previous research primarily focused on the overall effect of user interfaces and provided limited insights into the effects of different information types and modalities on driver's trust, perceived risk, and acceptance. In this chapter, therefore, we design and evaluate user interfaces conveying different types of information in various modalities to investigate their effects on trust, perceived risk, and acceptance in partially automated vehicles.

5.1.1. TRUST IN AUTOMATED VEHICLES

Trust is crucial for the acceptance of vehicle automation [21], [29], [212], [213], [215], [217]. It is important to adjust users' trust to an appropriate level depending on the systems' performance [218]. To leverage advanced technologies, driver's trust needs to be maintained at an appropriate level [219] to avoid both under-trust (or distrust) and over-trust [28]. Over-trust can lead to misuse and unintended use, which can result in various, even fatal accidents [220]. Conversely, many (potential) users distrust vehicle automation, which may lead to disuse [221]. Transparency is crucial to evoke trust [222]. Trust issues may result from a lack of information on the behaviour of a complex system, e.g., a car [223]. Transparency, as defined by Endsley *et al.* [224], encompasses the clarity and predictability of systems. It enables users to grasp the system's operations, rationale, and anticipated actions [225]. In automated vehicles, a deficiency in transparency, such as the absence of information regarding future actions, may cause inherent distrust [48]. Well-designed user interfaces can reduce unnecessary interventions by enhancing the driver's understanding of the vehicle's intentions and capabilities [226]. Automation system transparency has been shown to enhance trust calibration [28], [29], [227]–[230]. Nevertheless, the existing studies examine the importance of transparency, with less emphasis on how transparency in user interfaces influences driver's trust. Therefore, we

design user interfaces to enhance system transparency and investigate their effects on trust in this chapter.

5.1.2. SURROUNDING AND MANOEUVRE INFORMATION

To foster trust in and acceptance of automated vehicles, it is important to design transparent automated vehicle behaviour supported by a user interface explaining the operation of the automated vehicle. Previous studies have emphasised the necessity of system transparency by providing automation information, which consists of surrounding and manoeuvre information [49], [50], [231]–[234]. Surrounding information includes other road users detected by the vehicle, and manoeuvre information relates to the decisions made by the automated vehicle. Both information types enable users to anticipate and understand upcoming vehicle behaviour.

Wilson *et al.* [215] observed on-road driver behaviour in partially automated vehicles. They confirmed that one obstacle to trusting automated vehicles is a lack of information provided to the driver regarding what the automation "perceives" of the driving environment and how the automation will behave afterwards. When the vehicle detected other vehicles and presented this on the visual interface, drivers were reassured that the vehicle would respond adequately and continued to use the automation. Providing surrounding and manoeuvre information increases trust and convinces drivers to use automation [232]. Oliveira *et al.* [49] and Sawitzky *et al.* [50] have shown that augmented reality displays can increase trust by providing different visual aids for displaying driving routes as manoeuvre information. Koo *et al.* [233] and Ma *et al.* [234] confirmed that information provided using a single modality, auditory and visual, respectively, increased trust, but the impact of different levels of automation information on drivers varied between studies. Koo *et al.* [233] compared four different transparency levels of information, with and without surrounding information (the reasons for action) and manoeuvre information (how the car will act), via auditory modality and found that surrounding information increased trust, but the effect of manoeuvre information was not significant. Ma *et al.* [234] investigated three transparency levels of information (1. none; 2. surrounding information; 3. surrounding and manoeuvre information) via visual modality and showed that a combination of surrounding and manoeuvre information increased trust more than surrounding-only information. Basantis *et al.* [48] compared four different interfaces (1. No feedback; 2. Vehicle path on the visual display; 3. Manoeuvre notification sound 4. Mix of 2 and 3) in the rear seat. The results show enhanced trust and perceived safety with the auditory manoeuvre notification compared to only visual automation information. Although these studies highlight the benefits of providing surrounding and manoeuvre information, they typically examined the impact of these information types in isolation or did not systematically evaluate the combined effects of different modalities (visual and auditory) on trust and perceived safety.

While Mackay *et al.* [235] and Chang *et al.* [231] suggested that more information does not always lead to increased trust, the nuances of how different levels of information interact with modality to influence trust and acceptance in partially automated vehicles have not been fully explored. Examination of how auditory information, when synchronised with visual cues, can maintain driver attention without causing distraction or irritation is still needed, as highlighted by Liu [236] and Edworthy [237].

Thus, further research is essential to address the current gap in understanding the interaction of modality and information and how the combination of information types and modality affects trust, perceived safety and acceptance in the specific context of partially automated vehicles. Existing studies have not fully explored the systematic effects of auditory and visual information, indicating a pressing need for further research to systematically examine the impact, guiding the design of effective user interfaces that provide the necessary information to establish an appropriate trust level in partially automated vehicles.

5.1.3. THE CURRENT STUDY

The study in this chapter systematically investigates how different information types and modalities of user interfaces in driving automation information affect drivers' trust, perceived safety, and acceptance during partially automated driving. We hypothesised that user interfaces providing surrounding information, manoeuvre information, or both enhance drivers' trust, perceived safety and acceptance in driving automation. We expected that user interfaces that provide more information enhance trust and perceived safety. As a result we also expected a reduced frequency of drivers' interventions (e.g., braking) during driving automation. We evaluated visual and auditory UI to compare their effectiveness and user acceptance. For the challenge of maintaining the driver's attention, we expect that visual displays impact the driver's eye gaze distribution, which is significantly correlated with the driver's trust and perceived risk levels.

To validate our hypothesis, we designed four user interfaces using four combinations of information (surrounding information vs surrounding and manoeuvre information) and modalities (visual modality vs visual and auditory modality). The interfaces were intended to support drivers in understanding the reactions of automated vehicles to other vehicles merging in front, where the automated vehicles could react by either braking or changing lanes.

5.2. METHODS

We designed four user interfaces (UI) providing automation information via visual and auditory modalities and evaluated the interfaces in a driving simulator, adding No UI as a baseline condition in a partially automated vehicle (Table 5.1). Effects of UI were assessed objectively through brake behaviour and eye-gaze behaviour, as well as subjectively through perceived risk, trust and acceptance. In a preliminary experiment (see Appendix D.1), participants evaluated one type of UI among these five UI conditions (between-subject experiment design). The results showed significant benefits for all four UIs compared to the No UI condition, but differences between the four UIs were not significant, presumably due to large individual differences. To further investigate the effects of UI information type and modality, the main experiment was performed using a within-subject design, which is less sensitive to individual differences.

5.2.1. PARTICIPANTS

Twenty-two drivers participated in the experiment. All had driving licenses for more than a year. The average age of the participants was 28.3 years (SD = 13.1). Thirteen were

male, and nine were female. Eleven had experience with adaptive cruise control (ACC), seven with lane-keeping assist (LKA), and four with both ACC and LKA. Eight drove a few times per year, ten drove a few times per month, and four drove a few times per week.

5.2.2. APPARATUS

Participants experienced the scenarios in the DAVSi driving simulator with a Toyota Yaris cockpit (Figure 5.1) at Delft University of Technology. It used three high-quality projectors to display the environment on a cylindrical 180-degree screen. Two 7-inch tablets were used as side mirrors. The automation UI presented visual information on a 10.1-inch tablet at the centre console, while an in-vehicle embedded speaker presented the auditory information. A 5.8-inch tablet was placed on the left side of the steering for a questionnaire. The instrument panel showed vehicle speed and engine revolutions per minute. A fixed four-camera Smart Eye Pro tracked the participant's eye gaze and was used to classify the region of interest.

Figure 5.1: Exterior and interior of the DAVSi simulator, with visual automation UI right of the steering wheel and tablet for the questionnaire left of the steering wheel

5.2.3. EXPERIMENTAL CONDITIONS

The experiment evaluated the effects of two information types (i.e., surrounding and manoeuvre) and two modalities (i.e., visual and auditory). The visual modality was used to provide continuous information, and event-based information was presented using visual or auditory cues. We always included surrounding information in four UIs which was shown beneficial in a range of studies as outlined above and explored the benefits of adding manoeuvre information. Table 5.1 shows the five UI conditions: 1) Baseline (No UI), with no additional (automation) information provided, 2) Surrounding information via the visual modality (S-V UI), 3) Surrounding and manoeuvre information via the visual modality (SM-V UI), 4) Surrounding information via the visual and auditory modality (S-VA UI), 5) Surrounding and manoeuvre information via the visual and audi-

tory modality (SM-VA UI). Each participant executed all five UI conditions in randomised order.

Table 5.1: User Interface Conditions (Note. baseline is No UI)

Modality	**Information**	
	Surrounding	**Surrounding & Manoeuvre**
Visual	S-V UI	SM-V UI
Visual & Auditory	S-VA UI	SM-VA UI

5.2.4. SCENARIO DESIGN

This experiment selected highway driving scenarios with other vehicles merging into the driving lane of the ego-vehicle. Participants drove partially automated vehicles where they monitored the driving environment and kept their hands on the steering wheel. An adjacent vehicle (the yellow car in Figure 5.2) entered the highway and approached to merge into the right lane where the ego-vehicle (the blue car in Figure 2) was driving. Participants could see the adjacent vehicle approaching their lane. After detecting the merging vehicle, the ego-vehicle braked (Figure 5.2 left) or steered to the left lane (Figure 5.2 right). The velocity of the ego-vehicle and the traffic vehicles was set to 100 km/h. In the braking manoeuvres, the braking lasted until the merging vehicle velocity decreased to 60 km/h, followed by acceleration back to 100 km/h. In the lane change manoeuvres, the ego-vehicle steered into the left lane, overtook the merging vehicle and returned to the original lane. No accidents or automation failures were designed.

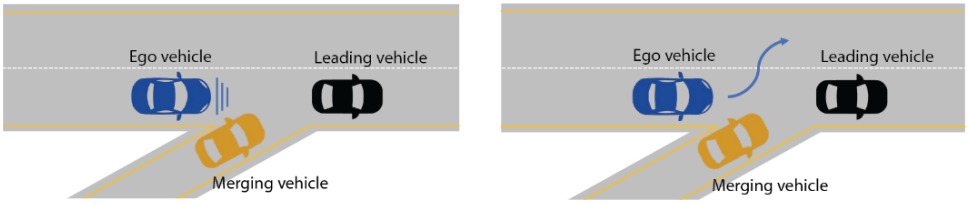


Figure 5.2: Merging scenario (Blue: ego-vehicle, Black: leading vehicle, Yellow: merging vehicle), slowing down manoeuvre (Left) and lane change manoeuvre (Right)

Six merging events were studied (Table 5.3), varying in terms of merging gap (5 m and 25 m) and automation action ($-2\,m/s^2$ or $-8\,m/s^2$ deceleration, or lane change). The slowing-down manoeuvre and lane change manoeuvre contained four and two merging events with different criticalities (automation action) respectively. To integrate these into a single drive, a series of onramps were designed along the road with one-minute intervals between merging locations. In total, merging events occurred in eight of the ten ramps. The order of events was randomised for every drive.

Table 5.3: Event types

Manoeuvre	Merging gap (m)	Automation action
Slowing down	5	$-2\,m/s^2$ deceleration
	25	$-8\,m/s^2$ deceleration
Lane change	5*	Lane change
	25*	Lane change

5.2.5. UI DESIGN

VISUAL INTERFACE

A bird-eye view with pop-up messages provided visual surrounding information (Figure 5.3 left). The bird-eye view was visible the entire time while driving. It showed the driving environment 60 m forward and 10 m backwards, including two adjacent lanes. The colour of the ego-vehicle was red to ensure that participants easily recognised their car, while the colour of other vehicles was grey. A pop-up message displayed safety-related surrounding information (i.e., merging vehicle detected). In SM-V UI and SM-VA UI conditions, manoeuvre information was also provided as a pop-up message with text and an icon after the surrounding information pop-up message in the same location (Figure 5.3 right) and was presented just before the ego-vehicle performed a manoeuvre (i.e., when the ego-vehicle is slowing down or changing lane).

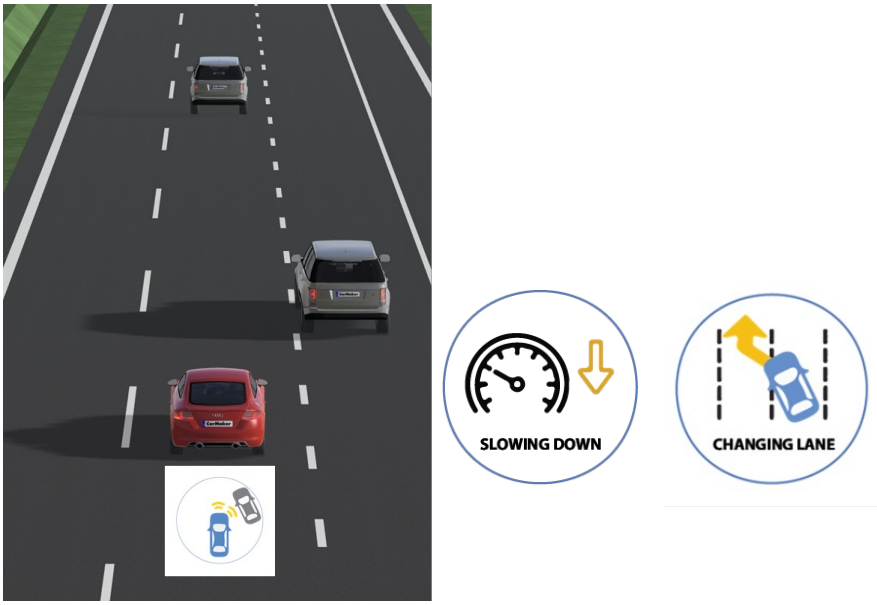


Figure 5.3: Visual UI with surrounding and manoeuvre information. Bird-eye view with pop-up message of surrounding information (Left) and pop-up messages for manoeuvre information (Right)

AUDITORY INTERFACE

A combination of abstract sounds and language-based explanations was used. An abstract sound of a low alarm level was provided to draw attention to prevent participants from being surprised by the language-based explanations. A wood and xylophone sound with a fundamental frequency of 625 Hz was two times repeated and lasted a total of 1.24 s. It was chosen because it can provide a feeling of simplicity [238]. Comprehension-level perception information was provided [239] as surrounding information. Explanations were generated using a female voice to be more likeable and comfortable [240] on the Google text-to-speech engine. The surrounding information was provided as: 'merging vehicle detected' of 1.40 s. Manoeuvre information used a first-person pronoun (i.e., 'we') as an anthropomorphism to increase trust [241]. Manoeuvre information was provided as: 'we are slowing down' of of 1.00 s seconds or 'we are changing lane' of of 1.10 s seconds. These sounds can be found in Appendix D.2.

TIMING

The provision of automation information prior to the action of the automated vehicle has been found to improve trust, as demonstrated in previous studies [219], [242]. Therefore, the information was provided before the vehicle took action in the experiment. To determine the optimal timing for information provision, we conducted a pilot test using an online survey to compare on-time and early-timing conditions. The on-time condition provided surrounding and manoeuvre information as soon as the merging vehicle changed direction to the ego-vehicle lane. The early-timing condition provided the information four seconds before the on-time condition. Twenty-four participants watched videos with different information provision timing and answered the trust comparison question. The manoeuvre was when a merging vehicle approached with a 5 m merging gap with $-5\,\mathrm{m/s^2}$ deceleration. As a result, fifteen participants answered that they trusted automation more with early timing. One participant trusted more on time, and eight had no preference. Therefore, we decided to provide the information four seconds before the merging started. Hereby, we assumed the AV to timely detect the merging intention from the directional indicator or V2V communication.

5.2.6. MEASUREMENT

During the simulation, brake pedal signals, eye gaze behaviours, trust, and perceived risk were collected. The brake pedal signal was recorded by the driving simulator automatically as braking is an effective indicator of distrust and perceived risk during automated driving [34], [117], [243]. We deactivated the option for participants to take over control by steering to ensure a controlled study environment because different traffic conditions in the left lane influence the driver's steering behaviour, which would introduce additional factors into the analysis. Eye gaze behaviour, an indicator of the driver's attention and situation awareness, is impacted by user interfaces as they can change the driver's eye gaze distributions [244]. The redistribution of eye gaze is particularly important as it is indicative of the driver's engagement with the driving environment and the automated system. Eye gaze behaviour was recorded at 60Hz using a smart-eye system with four infrared cameras mounted in the vehicle cockpit. It was measured to evaluate the eye gaze fixation time ratio on the display and the road and the eye gaze transition num-

bers between the display and the road. Participants were requested to report the level of trust and perceived risk after each merging event on the tablet on the left side of the steering wheel [117]. After each merging event, the experimenter verbally asked two 10-point Likert scale questions: "how much do you trust the driving automation according to the previous performance of the system?" and "how risky do you perceive the previous event". After each UI condition, participants answered three questions related to communication and acceptance on a 7-point Likert scale. Communication with automation measured whether drivers understood the system operation through the interface. We measured perceived usefulness and perceived ease of use to evaluate acceptance based on the Technology Acceptance Model (TAM) [245]. Perceived usefulness measures the degree to which the technology is useful and enhances driving performance. Perceived ease of use measures the degree to which the technology is easy to use and understandable. After participants experienced five UI conditions, they were asked to rank the five UIs on communication with automation, perceived usefulness, and perceived ease of use. In addition, the preferred modality (visual vs auditory vs both) of given the type of information was evaluated using a 7-point Likert scale.

5.2.7. PROCEDURE

Participants were welcomed and introduced to the experiment. They were asked to read the experiment information and sign an informed consent form before they filled out a demographic questionnaire, including age, gender, and vehicle automation experience. After finishing the questionnaire, they moved into the driving simulator. Participants adjusted the sitting position according to their individual preferences, and an experimenter calibrated the eye-tracking system. Participants were informed that they would be driving a partially automated vehicle, with the vehicle performing lateral and longitudinal motion control while they monitored the driving automation and kept their hands on the steering wheel. They were instructed that they could intervene in the automation by braking whenever they felt it was necessary, and partially automated driving would automatically reactivate right after their intervention. In the training session, participants drove partially automated driving on the highway to familiarise themselves with the simulator and learn how to answer the trust and perceived risk questions in the tablet when they were asked. This training lasted until participants could handle all tasks well. Then, the simulator experiment started. For each UI condition, participants experienced eight merging events in randomised order. Participants were informed they could stop if they felt uncomfortable or experienced motion sickness. During driving, participants rated their level of trust and perceived risk using a 10-point Likert scale questionnaire on the tablet located on the left side of the steering wheel after each event. Each UI condition took around ten minutes. After each UI condition, participants answered the questionnaire about communication with automation and acceptance. This was repeated five times to experience five UI conditions. The order of five UI conditions was randomised. Participants had a break between the third and the fourth UI conditions. After five UI conditions, they answered the ranking questionnaire on preferred information and modalities. The entire procedure took around two hours.

5.2.8. DATA ANALYSIS

Statistical analysis was conducted using IBM SPSS ver.27. A two-way repeated-measure ANOVA was used to analyse the effects of UI and Event type on trust, perceived risk and eye gaze behaviour. The data were analysed using a separate repeated-measures analysis for each dependent variable (trust, perceived risk and eye gaze behaviour) with independent factors UI (5 levels) and Event type (6 levels) as within-subject variables. To analyse the effects of UI on communication with automation, perceived ease of use, and perceived usefulness, a one-way ANOVA was used. Effects were declared statistically significant if $p < 0.05$. Post-hoc analysis was conducted with a Bonferroni test where the α value was adjusted by dividing it by the number of comparisons. Therefore, 0.005 and 0.003 were used as the adjusted α levels for post-hoc analysis on the effects of UI and event type, respectively.

5.3. RESULTS

All twenty-two participants completed the experiment, and no motion sickness was reported. Eye gaze signals were successfully collected in 106 simulations (22 participants × 5 UIs with four UI conditions missing eye gaze data). 880 answers (22 participants × 5 UIs × 8 events) about trust and perceived risk and 110 answers (22 participants × 5 UIs) about communication with automation and acceptance, and 22 answers (22 participants) about information and modality preference were collected from questionnaires and analysed.

5.3.1. TRUST AND PERCEIVED RISK

Figure 5.4 shows the mean score for trust (Left) and perceived risk (Right) over all events for each user interface. The SM-VA UI received the highest trust. The effect of user interface on trust was significant ($F(4,84) = 5.30, p < 0.001, \eta^2 = 0.20$). The SM-VA UI received significantly higher trust compared to the S-V UI ($p = 0.029$) and the S-VA UI ($p = 0.025$) (Figure 5.4 left). As expected, perceived risk showed an opposite trend as trust, where the lowest risk was perceived with the SM-VA UI (Figure 5.4 right). However, the effect of UI conditions on perceived risk was marginally non-significant ($F(4,84) = 2.48, p = 0.050, \eta^2 = 0.11$).

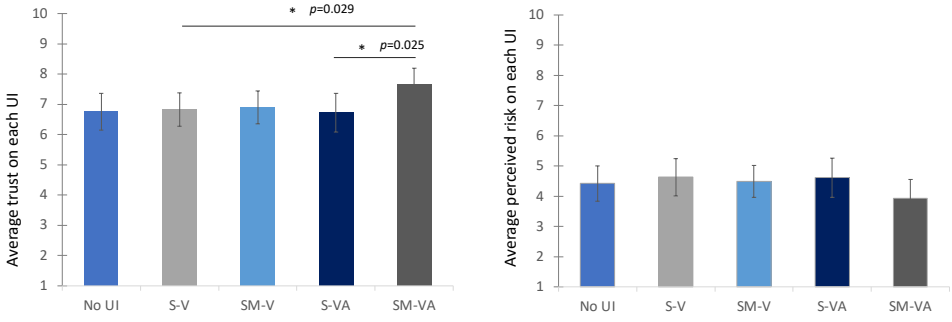


Figure 5.4: Drivers' Trust (Left) and Perceived risk (Right) on each user interface over all events, means and standard error over 22 participants (* $p < 0.05$)

Figure 5.5 shows the mean score for trust (Left) and perceived risk (Right) over all user interfaces for each event type. Here, no significant effects of the order were found between these events so lane change results were averaged over the two equivalent events tested. The most critical event (Slowing down with a 5 m merging gap and $-8\,m/s^2$ deceleration) received the lowest trust and the highest perceived risk. The least critical event (Slowing down with a 25 m merging gap and $-2\,m/s^2$ deceleration) received the highest trust and the lowest perceived risk. As shown in Table 5.3, the effect of each event element (merging gap and automation action) on trust and perceived risk was analysed. The merging gap included 5 m and 25 m, and the automation action included $-2\,m/s^2$ and $-8\,m/s^2$deceleration and lane change. The merging gap significantly affected trust ($F(1,21) = 89.48, p < 0.001, \eta^2 = 0.81$) and perceived risk ($F(1,21) = 179.09, p < 0.001, \eta^2 = 0.90$). The post-hoc analysis indicated that 25 m merging gap events received higher trust and lower perceived risk than 5 m merging gap events ($p < 0.001$). The automation action also significantly affected trust ($F(1.28,26.88) = 55.03, p < 0.001, \eta^2 = 0.72$) and perceived risk ($F(1.64,34.52) = 76.97, p < 0.001, \eta^2 = 0.79$) with a Greenhouse-Geisser adjustment. The post-hoc analysis indicated that $-2\,m/s^2$ deceleration events received the highest trust and the lowest perceived risk ($p < 0.001$), and $-28\,m/s^2$ deceleration events received the lowest trust and the highest perceived risk ($p < 0.001$). There was an interaction effect between the merging gap and automation action on trust ($F(1.44,30.28) = 45.88, p < 0.001, \eta^2 = 0.69$) and perceived risk ($F(1.40,29.34) = 58.84, p < .001, \eta^2 = 0.74$) with a Greenhouse-Geisser adjustment. The trust and perceived risk difference between $-2\,m/s^2$ deceleration or lane change and $-8\,m/s^2$ deceleration were much greater when the merging gap was 5 m than 25 m. There was no interaction effect between UIs and elements of events on trust and perceived risk.

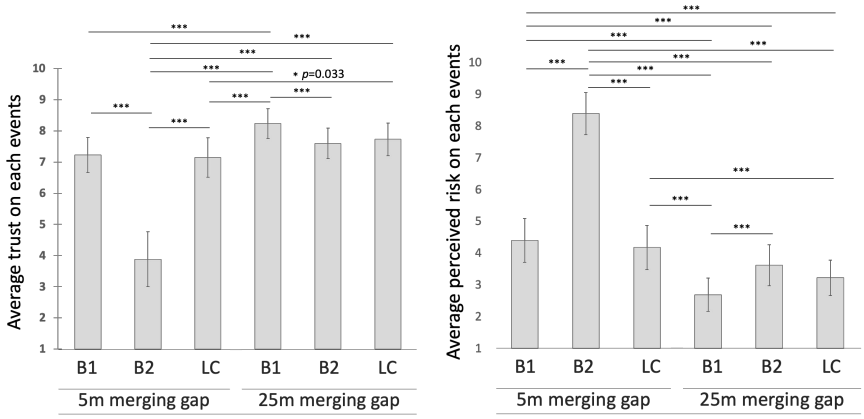


Figure 5.5: Drivers' trust (Left) and perceived risk (Right) as a function of automation action and merging gap (* $p < 0.05$, *** p<0.001)

5.3.2. BRAKING BEHAVIOUR

he use of the brake pedal was detected in 58 out of 110 UI conditions (22 participants × 5 UIs). All instances of braking occurred in the most critical events (slowing down with

a 5 m merging gap and $-8\,m/s^2$ deceleration). Eight participants braked in all five UI conditions in at least one event, two participants used the brake pedal in four UI conditions, three participants in three, and one participant in one UI condition, whereas eight participants did not use the brake pedal at all. As shown in Figure 5.6, there is almost no brake pedal behaviour difference between the five UI conditions. The eight participants who braked in all UI conditions had lower trust levels ($F(1,14) = 4.96, p = 0.04$) and higher perceived risk levels ($F(1,14) = 4.56, p = 0.05$) than the eight participants who never braked. There was no effect of experiment order on braking behaviour.

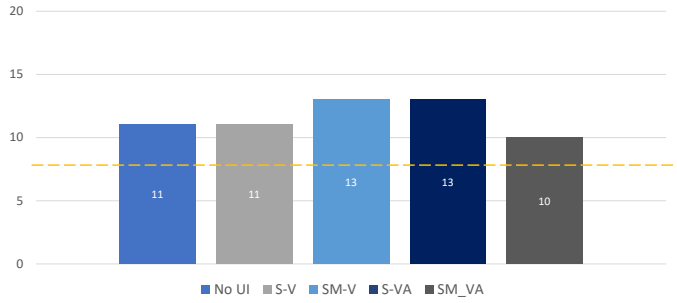


Figure 5.6: Number of participants who used the brake pedal in at least one event with each user interface. The yellow dashed line represents the eight participants who braked in all five UI conditions.

5.3.3. EYE GAZE BEHAVIOUR

As shown in Figure 5.7, eye gaze behaviour (i.e., the eye gaze fixation time on the display and the road and the eye gaze transition number between the road and the display) differs over all four UIs (S-V, SM-V, S-VA, and SM-VA UI), compared to No UI, primarily due to the visual display on the centre console. No significant difference was found between the four UIs. UI presence significantly impacted the eye gaze fixation time ratio on the display ($F(4,72) = 8.56, p < 0.001, \eta^2 = 0.32$), the eye gaze fixation time ratio on the road ($F(4,72) = 7.69, p < 0.001, \eta^2 = 0.30$), and the transition number between the road and the display ($F(4,72) = 10.38, p < 0.001, \eta^2 = 0.37$). The Bonferroni test reveals that the eye gaze fixation time ratio on display and the eye gaze transition number between the road and the display are significantly higher with the four UIs than with No UI. Significant differences were also found with the four UIs and with No UI except SM-V UI regarding the eye gaze fixation time ratio on the road. No significant effect of different events on eye gaze behaviour (i.e., the fixation duration ratio on the road and the display and the transition number between the road and the display) was found, as shown in Figure 5.8 and Table 5.4. A marginally significant ($p = 0.045$) interaction was observed between the merging gap and automation action on eye gaze fixation time ratio on the display.

The results showed insignificant effects of user interface on the eye gaze distribution (except UI versus No UI). However, there was a notable individual difference in the average eye gaze distribution on the display. Cronbach's analysis showed the high reliability

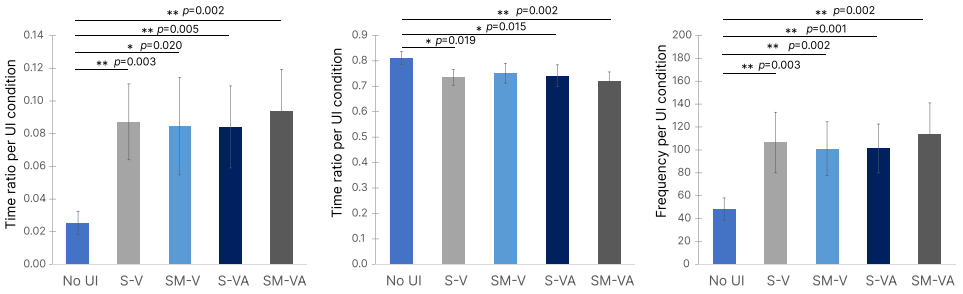


Figure 5.7: Eye gaze fixation time ratio on the display per UI (Left); Eye gaze fixation time ratio on the road per UI (Middle); Transition numbers between the road and the display per UI (Right) (* $p < 0.05$, ** $p < 0.01$)

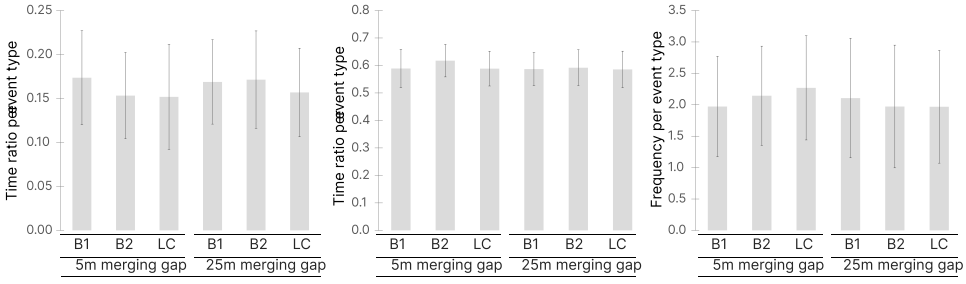


Figure 5.8: Eye gaze fixation duration ratio on the display per event (Left); Eye gaze fixation duration ratio on the road per event (Middle); Transition numbers between the road and the display per event (Right)

Table 5.4: Statistics of the event's effect of merging gap and automation action on eye gaze behaviours

	Event elements	F	Sig.	Effect size (η^2)
Eye gaze fixation time ratio on the display	Merging gap	$F(1.00, 15.00) = 2.04$	0.112	0.15
	Automation action	$F(1.05, 15.73) = 0.08$	0.860	0.01
Eye gaze fixation time ratio on the road	Merging gap	$F(1.00, 16.00) = 3.33$	0.087	0.17
	Automation action	$F(1.26, 20.20) = 0.62$	0.478	0.04
Eye gaze transition numbers between the road and the display	Merging gap	$F(1.00, 16.00) = 1.52$	0.236	0.09
	Automation action	$F(2.00, 32.00) = 0.10$	0.908	0.01

between each participant's fixation duration ratio on the display of four interfaces, excluding the No UI (Cronbach's $\alpha = 0.86$). There was no correlation between the fixation duration ratio on the display of each participant and their trust and perceived risk. The eye gaze behaviour results indicate that participants indeed checked the visual display during driving but kept the same eye gaze behaviour regardless of different event types and UI.

5.3.4. COMMUNICATION WITH AUTOMATION, PERCEIVED EASE OF USE AND PERCEIVED USEFULNESS

As shown in Figure 5.9, the SM-VA UI received the highest score on all attributes. The main effects of UI were significant for communication with automation ($F(4,84) = 5.08, p < 0.001, \eta^2 = 0.26$), perceived ease of use ($F(4,84) = 4.54, p < 0.001, \eta^2 = 0.62$) and perceived usefulness ($F(4,84) = 3.99, p < 0.001, \eta^2 = 0.42$). The post-hoc analysis indicated that participants preferred the SM-VA UI to the No UI, and S-V UI on all attributes ($p < 0.001$).

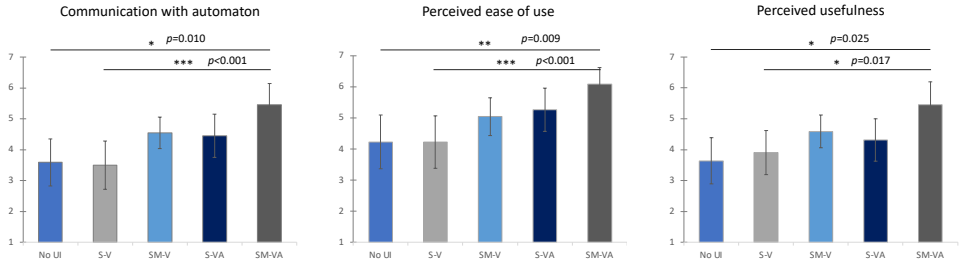


Figure 5.9: Drivers' communication with automation (Left), perceived ease of use (Middle) and perceived usefulness (Right) scores on each User interface (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$)

Concerning the UI ranking, (ordinal) data were analysed using categorical principal component analysis (CATPCA). Answers from two participants were shown to contain outliers, according to CATPCA, and therefore, we excluded all answers from these two participants.

The Friedman test examined the differences in the ranking among UI conditions. Participants ranked the five UI conditions significantly different on communication with automation ($\chi^2(4,20) = 69.8, p < 0.001$), perceived ease of use ($\chi^2(4,20) = 70.77, p < 0.001$), and perceived usefulness ($\chi^2(4,20) = 69.92, p < 0.001$). As the results of CATPCA, biplots of all attributes (communication with automation, perceived ease of use and perceived usefulness) on the five UIs are shown in Figure 5.10. The eigenvalues and percentage of total variance are presented in Table 5.5. The results of the analysis explained 100% of the total variance. Dimension 1 accounted for around 90% of the variance in the ranking of communication with automation, perceived ease of use and perceived usefulness. SM-VA UI is ranked highest over the three attributes (communication with automation, perceived ease of use and perceived usefulness), followed by S-VA UI, SM-V UI, S-V UI, and No UI. The No UI and S-V UI were least preferred in perceived ease of use. When examining the x-coordinate pertaining to Dimension 1 across all three

attributes in Figure 5.10, participants are consistently positioned on the right side of the graph. This is because participants tended to evaluate the UI with the ranking SM-VA UI, S-VA UI, SM-V UI, S-V UI, and No UI, which were displayed from left to right in the graph. Dimension 2 accounted for around 10% of the total variance. It corresponds to the difference in the preferred interface between S-VA UI and SM-V UI. The results showed differences in preference for S-VA UI and SM-V UI as the second highest-rank interface, depending on the individual. Considering the y-coordinate reflecting Dimension 2 in Figure 5.10, the S-VA UI and SM-V UI are positioned on opposite sides, while the remaining UIs congregate around zero. This discrepancy arises from varying participant preferences, particularly regarding the SM-V UI ranking. Participants clustered around zero expressed a preference for SM-V UI as their third choice. Conversely, participants positioned near the SM-V UI reported a lower preference for it compared to other participants. Notably, participants in proximity to the S-VA UI indicated a heightened preference for SM-V UI compared to their counterparts, resulting in a relatively lower ranking for S-VA UI.

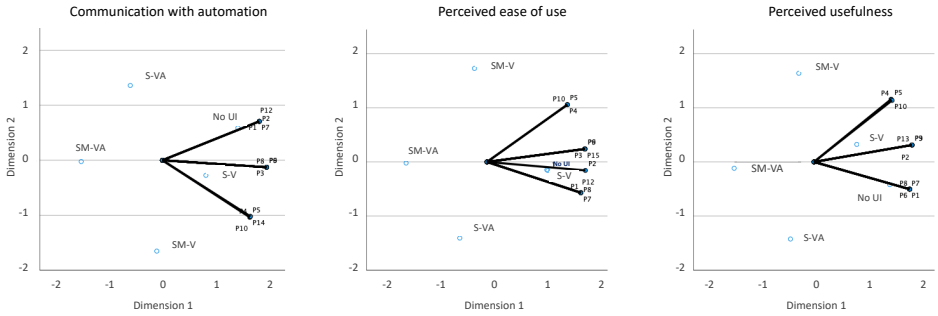


Figure 5.10: UI ranking, CATPCA results of communication with automation (Left), perceived ease of use (Middle) and perceived usefulness (Right)

Table 5.5: UI ranking, CATPCA eigenvalues and the percentage total variance explained

	Communication with automation	**Perceived ease of use**	**Perceived usefulness**
	Total (Eigenvalue)	Percentage of variance	Cronbach's Alpha
Dimension 1	17.80	89.00	0.99
Dimension 2	2.20	11.00	0.57
Total	20.00	100.00*	1.00*

5.3.5. INFORMATION TYPE AND MODALITY PREFERENCE

Participants highly appreciated both UI information types, where surrounding information received 6.18 (SD = 1.2) points and manoeuvre information received 6.50 (SD =

0.96) points on a 7-point Likert scale. Figure 5.11 shows the preferred modality for surrounding and manoeuvre information. Participants preferred receiving the surrounding information via both visual and audio modalities. Among twenty-two participants, fourteen participants (64%) preferred surrounding information in both visual and audio, four participants (18%) chose only audio, and four chose only visual (18%). The right figure indicates the modality preference for manoeuvre information. Compared to the surrounding information modality preference, more participants preferred to receive the manoeuvre information through audio only. Ten participants (45%) preferred both visual and audio manoeuvre information. Nine participants (41%) preferred only audio, and three participants (14%) preferred only visual.

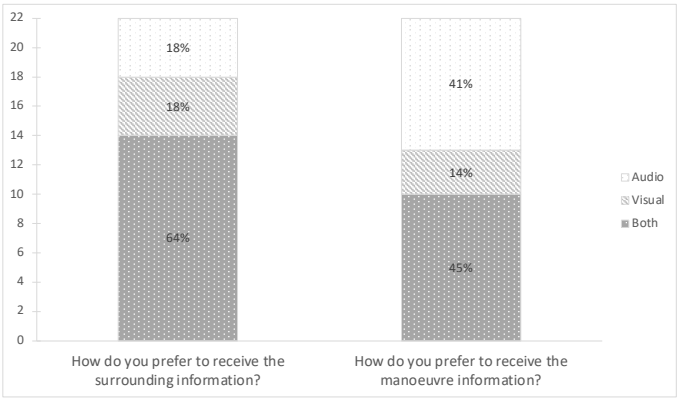


Figure 5.11: Ratio of drivers' modality preference on surrounding information (Left) and manoeuvre information (Right)

5.4. DISCUSSION

This chapter investigated the effects of user interface (UI) on trust, perceived risk, and acceptance in partially automated highway driving with a simulator experiment. Four interfaces were designed, combining surrounding and manoeuvre information and visual and auditory modalities.

5.4.1. EFFECTS OF UIS ON TRUST, PERCEIVED RISK AND ACCEPTANCE

We systematically added information types (surrounding and manoeuvre) and modalities (visual and auditory). The most advanced UI, SM-VA UI, providing surrounding and manoeuvre information via both visual and auditory modality, received the highest trust and lowest perceived risk scores and the highest communication with automation and acceptance scores. In addition, ranking showed that participants chose SM-VA UI as the best and No UI or S-V UI as the worst ranking communication with automation, perceived ease of use and perceived usefulness. Effects are highly significant comparing the most advanced SM-VA to No UI and thereby support our hypothesis that "user interfaces providing surrounding information, manoeuvre information, or both enhance drivers' trust, perceived safety and acceptance in driving automation". We did not find

the expected "reduced frequency of drivers' interventions (e.g., braking) during driving automation" in the main study but found some reduction in the preliminary experiment. Adding only manoeuvre information has more effect than only adding surrounding information, but this trend is not significant. Regarding modality, with manoeuvre information displayed through both visual and auditory modalities, trust and acceptance increased compared to when manoeuvre information is displayed through only the visual modality, and this effect is significant for trust. Eye gaze behaviour showed that drivers check the UI at the centre console when present. However, at the same time, there was no significant difference in UI gaze time between the four UIs. Presumably, drivers check the driving situation on the road after receiving surrounding information instead of perceiving visual manoeuvre information.

Interestingly, the effects of UIs show more significance in acceptance compared to driving behaviour, trust and perceived risk. Acceptance increases when receiving more information with more modalities. Drivers want to monitor the safe operation of partially automated vehicles [246]. Therefore, regardless of the actual use of the interface, the presence of the interface can support the acceptance of automation [247].

Drivers evaluated both the surrounding and manoeuvre information positively. However, the preference for modality differed between participants and between surrounding and manoeuvre information. More than half of the participants preferred that surrounding information be delivered in both visual and audio modalities. When the surrounding information was provided in both visual and auditory modalities, there was no significant difference in the gaze time on the display compared to when it was provided only in the visual modality. Hence, surrounding information via auditory modality can be interpreted as supplementing the visual modality, not replacing it. On the other hand, the participants preferred manoeuvre information to be delivered in only an auditory modality or a combination of visual and auditory modalities. Drivers checked the centre console display when it showed the detected vehicle. After the participants perceived the merging car, their view moved to the road, with no difference in gaze time across the four UIs. Hence, the visual manoeuvre information was presumably not attended to, explaining the lack of benefits of SM-V vs. S-V and SM-VA vs. S-VA.

The preliminary between-subject design experiment presented in the Appendix D.1 already indicated that UI could increase trust, compared to no UI, but disclosed no significant differences between the four UIs. The main experiment, using a within-participant design, disclosed significant differences between the four UIs, with the best overall results for the most advanced SM-VA UI. A possible explanation is that individual differences obscured the effects of UI in the preliminary between-subject experiment. An alternative explanation is that, being exposed to multiple UIs, participants develop expectations regarding automation behaviour and UI affecting their behaviour and subjective evaluation. This could include learning and trust calibration with exposure to specific UIs affecting responses with following UIs. However, we found no significant effects of order in the main results, which indicates that such learning has no strong effects. Anyhow, we see benefits in both the within and between-participant experimental design. The within-participant design discloses significant effects with a limited cost-effective sample, whereas the between-participant design better represents real-life exposure where users will presumably use one UI only.

5.4.2. EFFECTS OF CRITICALITY OF EVENT TYPES AND INDIVIDUAL DIFFERENCES

Event criticality (Figure 5.5) had a much larger effect on trust and perceived safety as compared to UIs (Figure 5.4). We additionally compared the effects of UI on trust and perceived risk in the most critical event (slowing down with 5 m gap and $-8\,m/s^2$ deceleration) and least critical event (slowing down with 25 m gap and $-2\,m/s^2$ deceleration). In both events, the effect of UIs on perceived risk was not significant. The effect of UIs on trust was marginally non-significant in the most critical event ($F(4,84) = 2.43$, $p = 0.054$), but the effect was significant in the less critical event ($F(4,84) = 3.06$, $p = 0.021$). Apparently, the effects of UI on trust and perceived risk are insufficient to make participants feel entirely safe and trust automation in the most critical events. This may be explained by Hoff and Bashir [29], who described three layers of variability in human-automation trust: dispositional trust, situational trust, and learned trust. Situational trust depends on the context of interaction, while learned trust represents users' evaluations of systems drawn from previous experience or the current interaction. The surrounding and manoeuvre information through the interface affects the learned trust, at the same time, situational trust is affected by the driving situation, such as different events. This chapter evaluated three automation manoeuvres: strong braking ($-8\,m/s^2$), mild braking ($-2\,m/s^2$) and lane changing. The latter two manoeuvres were tested with identical behaviours of the merging vehicle and resulted in similar trust and perceived safety, where the UIs provided similar benefits with positive effects of manoeuvre information.

The relationship between UI and braking behaviour appears to be moderated by individual driver characteristics. Eight out of twenty-two participants did not brake in any of the five UI conditions, while another eight used pedals in all five UI conditions. Those who used the brake pedal less tended to have higher trust and lower perceived risk, which is consistent with findings by He *et al.* [117], where trust of the braking group is lower than that of the non-braking group. It will be interesting to investigate trust calibration and its expected effect on braking in prolonged experiments or observations. The braking behaviour was quite different in the preliminary experiment, where participants braked the most in the No UI and the least in the SM-VA UI. However, the individual differences in braking behaviour may mask UIs' effect on the braking behaviour in the preliminary experiment [248]. Regarding the eye gaze behaviour, each participant looked at the display similarly regardless of the interface condition, which supports the notion that it is challenging to evaluate drivers' understanding of information in vehicles as eye gaze behaviours, as noted by Cohen-Lazry *et al.* [249]. The result is aligned with the preliminary experiment (see Figure D.3 in Appendix D.1) and confirmed the trend with significant effects between the No UI and other UI conditions on drivers' eye gaze behaviour.

5.4.3. LIMITATIONS AND PERSPECTIVE

Several limitations must be considered when interpreting our findings. The sample size, while sufficient to identify trends, is relatively small, which could potentially lead to biased effects. The artificial nature of the experimental setting, despite its high control level, may not fully capture the complexity of real-world driving dynamics. In addition, the results of user interface experiments under controlled conditions may vary de-

pending on changes in the user interface [250] (i.e., aesthetics and layout) or changes in the environment (i.e., the urgency of the scenario) [216]. These factors could limit the ecological validity of our findings. For example, the lack of significant variation in eye-tracking measures across UI conditions prompts further investigation into how different designs may influence drivers' visual attention and performance to detect the driving environment. Nevertheless, our results show significant benefits of UIs, enhancing trust and acceptance and reducing perceived risk. We provided a visual interface in the centre console display, which is common in commercial cars. However, a head-up display (HUD) could yield even better results, allowing drivers to keep their eyes on the road. HUD cannot easily present spatial surrounding information but can present event-based information such as pop-up messages. We should also consider that drivers may not perceive the auditory UI correctly when engaged in secondary tasks or may find it annoying if presented too often [251]. For auditory UI recognition, the volume of other audio systems shall be controlled. It is also necessary to consider irritation or stress when exposed to auditory information for a longer time. Future research will focus on interfaces providing a broader range of manoeuvre information, considering various human factors such as annoyance, workload, as well as trust and acceptance. Additionally, future studies should be extended towards UI enhancing trust and perceived safety in higher automation levels, allowing users to take their eyes off the road.

5.5. CONCLUSIONS

This chapter confirms that automation UI can enhance drivers' trust and acceptance of partially automated vehicles. Significant benefits were found for both surrounding (perception) and manoeuvre (action) information. Specifically, the most advanced UI (SM-VA UI), which displayed surrounding and manoeuvre information via the visual and auditory modalities, received the highest trust and acceptance ranking and the lowest perceived risk among drivers. Manoeuvre information displayed through the auditory modality was particularly effective in enhancing drivers' trust and acceptance. Current partially automated vehicles show the image received by sensors on the display, similar to the UI in this chapter that displays surrounding information visually (S-V UI). Our study in this chapter shows that the surrounding information displayed via the visual modality draws the driver's attention to the display, but it needs additional auditory communication by the UI to enhance driver's trust and acceptance. Therefore, including manoeuvre information via the auditory modality should be considered for partially automated vehicles. This may make the UI more complex but also more understandable and acceptable. To paraphrase Donald Norman, people hate things to be complicated but like complexity, which this chapter supports. Furthermore, this chapter has shown the impact of the user interface in relation to the risk level of the driving situation. When the driving situation poses a high risk, even with UI, drivers do not feel entirely safe and do not trust the automation completely. At the same time, drivers accept driving automation more with UI, regardless of perceiving the information, which was also shown by Kim *et al.* [247]. This demonstrates both the impact and limitations of UI.

6

CONCLUSION

This chapter includes a summary of the findings and conclusions from Chapters 2-5, and recommendations for practice and future research.

6.1. FINDINGS

In Chapter 1, we established three research objectives. In this section, we reflect on the extent to which these objectives were addressed by the findings of this dissertation.

Objective 1: To collect data and gain insights on perceived risk and trust based on specific AV behaviours

To meet this objective, in Chapter 2, we conducted a driving simulator experiment with 25 participants using SAE Level 2 automation in merging and hard braking events. This experiment collected event-based ratings of perceived risk and trust, along with time-continuous perceived risk data, pupil diameter, and ECG signals. The study validated various methods for measuring perceived risk and trust. The validation shows the reliability of using the peak of the time-continuous perceived risk ratings collected through the pressure sensor to represent event-based perceived risk. Additionally, results indicate that physiological indicators such as pupil dilation and ECG have the potential to represent perceived risk. Pupil dilation captures high perceived risk only in sufficiently risky events, and merging and hard braking events increase heart rate. However, no quantified relationship is found between pupil dilation, heart rate, and perceived risk. Braking is noted as a strong behavioural indicator of perceived risk and distrust: the more frequent and intense active braking, the higher the perceived risk and the lower the trust levels. Surrogate metrics of minimum gap, minimum time to collision (TTC), and maximum braking intensity are significant in predicting event-based perceived risk and trust. The data highlighted how driving experience and gender influence perceived risk and trust, with experienced drivers typically displaying lower perceived risk and higher trust, and females generally perceiving higher risk. The correlation analysis confirms that reduced perceived risk often correlates with increased trust in automated driving.

Chapter 4 extended the experimental investigation of perceived risk through a large-scale online study involving 2,164 participants from 27 countries. This study collected

141,628 discrete perceived risk ratings from more than 233 hours of time-continuous perceived risk data. The discrete perceived risk ratings were analysed using generalised linear regression and Jensen-Shannon (J-S) divergence to assess the impact of simulation parameters on perceived risk. The results indicate significant effects of factors such as initial merging distance, desired cruising speed, and braking intensity on perceived risk. Perceived risk is significantly higher when the initial merging distance is smaller, the desired cruising speed is higher, or the braking intensity is stronger. Furthermore, Chapter 4 obtained time-continuous perceived risk ratings from the discrete perceived risk ratings using the interpolation method. The continuous perceived risk ratings reflect detailed variations in perceived risk in real-time corresponding to the driving scenarios. For instance, perceived risk increases when the subject vehicle encounters other vehicle's actions such as merging, hard braking and lane changing. Perceived risk decreases when the driving scenarios become stable again.

Objective 2: To develop computational models for interpreting perceived risk and trust

To achieve this objective, Chapter 2 formulated regression models to capture the effect of merging and braking events and the influence of personal characteristics on perceived risk and trust. These models integrate crucial kinematic measures like minimum gap, time to collision (TTC), and braking intensity, along with personal characteristics, such as driving experience and gender, showing the influence of these factors on perceived risk. The regression models demonstrate capability in predicting event-level perceived risk and trust.

Chapter 3 introduced a novel physics based computational model (PCAD) to quantify continuous perceived risk in automated driving. This model offers a new perspective on how perceived risk is determined, formulating it as the least amount of control effort (braking and steering) needed to prevent possible collisions with neighbouring vehicles. By integrating factors such as visual looming, behavioural uncertainties of neighbouring vehicles, imprecise control by the subject vehicle, and collision severity, PCAD offers a comprehensive assessment of task difficulty in a 2D plane. Test results with two unique datasets, Dataset Merging and Dataset Obstacle Avoidance, show that the PCAD model outperforms existing models in model accuracy, detection rates in predicting perceived risk but with longer computation times. Its theoretical and empirical analyses reveal that perceived risk is not static, but varies with evolving positions, velocities, and accelerations of both the subject and neighbouring vehicles.

Learning-based deep neural networks (DNNs) model were developed to predict perceived risk in Chapter 4, allowing for a nuanced understanding of how perceived risk fluctuates with real-time changes in driving conditions. Additionally, explainable AI technology SHAP (SHapley Additive exPlanations) based on the trained DNNs reveal that the factors contributing to perceived risk are dynamic, changing over time with variations in speed, distance, and driver behaviour. Consistently, distance to other road users emerged as the most significant factor affecting perceived risk; concerns about potential collisions due to manoeuvre uncertainties become particularly pronounced when vehicles were in close proximity to each other. The success of these neural networks in accurately predicting perceived risk underscores their potential in enhancing automated driving systems' adaptability and safety.

Objective 3: To assess the impact of UI modalities and information types on perceived risk and trust

Chapter 5 investigated how user interfaces (UIs) in partially automated vehicles influence drivers' trust, perceived risk, and acceptance. The study designed four UIs by combining automation information (surrounding vs. surrounding and manoeuvre) and modalities (visual vs. visual and auditory), to investigate how they affected drivers' interactions with automation in a simulator experiment, based on Chapter 2.

It was found that a UI providing both surrounding and manoeuvre information through visual and auditory modalities significantly enhanced drivers' trust and acceptance while reducing perceived risk. This multimodal and comprehensive information approach was the most favoured, highlighting the importance of effective communication between drivers and automation systems.

Drivers' preferences show that auditory delivery of manoeuvre information significantly enhanced trust and acceptance without distracting from road attention, highlighting auditory feedback's role as a complementary source to visual information. This chapter also noted that event criticality and individual driver differences impacted trust and perceived risk, with drivers feeling less safe and trusting in highly critical situations despite UI improvements.

6.2. CONCLUSIONS

The comprehensive exploration conducted in Chapters 2-5 provides essential findings on perceived risk, trust and the effect of user interfaces (UIs) in automated vehicles. This section provides overarching conclusions from the combined results of all chapters.

1. Perceived risk as a context-dependent and multi-dimensional construct

The research in this dissertation demonstrates that perceived risk in automated driving is not a static concept, but rather one that is highly context-dependent and influenced by multiple factors. This is evidenced by the findings in Chapter 2, the physics-based potential collision avoidance difficulty (PCAD) model in Chapter 3 and the interpretation of the deep neural networks in Chapter 4, where perceived risk varies dynamically along with factors such as relative distance to neighbouring vehicles, braking intensity, and behaviour uncertainties during automated driving. The findings of Chapters 2-4 also emphasise the need to account for factors from both longitudinal and lateral directions, which demonstrates that perceived risk is multi-dimensional. These insights align with existing studies on driving risk and perceived safety Ping *et al.* [38], Kolekar *et al.* [41], Chen *et al.* [44], Ma *et al.* [98], and Kondoh *et al.* [103]. By incorporating simulator studies, real-time data collection and new computational modelling approaches, this dissertation pushes beyond the traditional, event-level understanding of perceived risk and captures its continuous nature.

2. Perceived risk is highly personalised. This variability should be accounted for to enable more reasonable interpretations and accurate predictions of perceived risk at the individual level.

Perceived risk is not only context-dependent but also highly personalised. The studies in this dissertation reveal that individual characteristics, such as driving experience and gender, significantly influence how perceived risk is formed. In Chapter 2, the regression model showed that experienced drivers tend to perceive lower risk compared

to less experienced drivers, and female drivers reported higher perceived risk compared to male drivers. In Chapter 3, the PCAD model effectively captured perceived risk at the population level, but showed limited effectiveness at the individual level, highlighting the variability of personal perceived risk. In Chapter 4, the data collected from a large and diverse group of participants exhibited considerable variability, indicating that perceived risk is highly influenced by personal sensitivity and interpretation. These findings align with previous studies that demonstrated the significant influence of personal characteristics on perceived risk in automated driving [53], [54], [252], [253], but extend beyond static evaluations by incorporating a more dynamic, context-sensitive approach.

3. Modelling and interpreting perceived risk based on driving conditions is feasible

The findings in Chapter 2, the PCAD model developed in Chapter 3, and the interpretation of the DNNs in Chapter 4 all indicate that modelling perceived risk based on driving conditions, particularly focusing on relative motion such as distances, speed, and accelerations, is feasible. These driving conditions are shown to account for a significant portion of perceived risk, suggesting that understanding and quantifying the driving conditions is a feasible way to interpret perceived risk. Insights from both physics-based modelling (as presented in Chapter 3 and in Chen *et al.* [44]) and data-driven modelling approaches (as presented in Chapter 2 and 4 and in Kolekar *et al.* [36], de Winter *et al.* [37], and Ping *et al.* [38]), demonstrates the viability of such models. Although these models differ in their methodologies, the results consistently show that variables related to driving conditions, particularly relative motion are critical determinants of perceived risk, supporting the idea that these driving dynamics can be effectively transformed into measurable perceived risk. Moreover, the large-scale data collected from simulator studies and the online study confirmed the feasibility of using motion-related data for modelling perceived risk.

4. Anticipating the behaviour of other road users is crucial for interpreting perceived risk.

The interpretation of perceived risk should not only focus on the current moment but also account for future uncertainties in the driving environment. In Chapter 3, the "uncertain velocity" construct within the PCAD model was formulated using calibrated parameters that effectively fits perceived risk data from a simulator study. Similarly, in Chapter 4, this "uncertain velocity" was used as an input to DNNs, where it emerged as a significant contributor to perceived risk. Although this construct does not directly quantify the anticipation of other road users' behaviour, it does reflect the anticipation mechanisms of human drivers. Given that this anticipation plays a critical role in predicting perceived risk in both the physics-based model and the data-driven model, it suggests that anticipating the behaviour of other road users is essential for accurately interpreting perceived risk. This also aligns with the observation that drivers tend to maintain longer distances than necessary when following a vehicle with a small gap or overtaking a heavy truck [129], [130].

5. UIs can mitigate perceived risk and enhance trust, but have limitations

The findings from Chapter 5 demonstrate that user interfaces (UIs) are crucial in enhancing perceived risk, trust, and user acceptance. The research shows that multimodal interfaces, which integrate both auditory and visual elements, are effective in reducing

perceived risk and increasing trust, particularly in moderately critical driving scenarios. The findings support existing theories regarding the use of multi-channel communication to enhance system usability [246]. However, UIs cannot sufficiently make people feel safe during highly critical driving situations. This indicates that UIs are influential factors but not definitive determinants of perceived risk. Perceived risk and trust are not only influenced by information access but are more significantly shaped by the driving context.

6.3. RECOMMENDATIONS FOR PRACTICE

The practical applications derived from this dissertation are primarily focused on enhancing the design and interaction of AVs with users. The insights from the regression model, the PCAD model and deep neural networks developed in this dissertation indicate that automotive designers and engineers are encouraged to integrate specific behavioural adjustments into AVs. These adjustments should focus on maintaining safe distances from neighbouring vehicles, adjusting speeds appropriately in real time, and optimising manoeuvre responsiveness to reduce user perceived risk and enhance trust. In addition, special attention should be paid to female users, who are often more sensitive to perceived risk, ensuring that AVs are able to accommodate diverse user sensitivities and safety expectations. Furthermore, the PCAD model offers a practical and computationally efficient tool for controller, path planner, or decision-making module design to align with user expectations and enhance perceived safety. While the deep neural networks (DNNs) developed in this dissertation outperformed PCAD in fitting perceived risk, their applicability remains highly scenario-specific. Therefore, one intending to use DNNs in predicting perceived risk in AVs must ensure that these models are generalised enough to be effective across a wide range of driving conditions.

The study on different UI modalities and information types has practical implications for designing user interfaces in AVs. The findings help optimise UI designs to enhance drivers' perceived safety, trust, and acceptance of partially automated vehicles. Specifically, practitioners are encouraged to provide the UIs that display both surrounding and manoeuvre information using both visual and auditory modalities as this approach enhances perceived safety, trust and acceptance. If this is not feasible, integrating auditory information can still get considerable benefits. These recommendationscan inform industry best practices and contribute to the broader acceptance and efficiency of AV technology in everyday use.

6.4. RECOMMENDATIONS FOR FUTURE RESEARCH

Future research should continue to explore the dynamic and multidimensional aspects of perceived risk and trust, as current models including those developed in this dissertation provide valuable insights but are not yet comprehensive enough to fully capture the complexity of perceived risk in automated vehicles. Therefore, further improvements are needed to enhance their accuracy and applicability across a wider range of driving conditions. Researchers should continue work on refining these models or developing new models that consider the complex, nonlinear interactions among various factors influencing perceived risk and trust, which could offer deeper insights into perceived risk,

trust and user behaviour. Moreover, incorporating multiple road users, different road user types and infrastructure elements in risk models would provide a more accurate estimation of perceived risk, especially in complex driving conditions.

Extensive data collection and model validation remain crucial. Current models require more diverse and robust datasets for validation, and expanding perceived risk data collection through various methodologies and scenarios, including on-road experiments, simulator studies and online experiments to advance perceived risk modelling. Integrating driving data in real world will help employ these models from research to practical scenarios, while exploring personalised risk models fitting individual drivers' perceptions and behaviours could improve the accuracy and reliability of these models, contributing significantly to the development of safer AV technologies.

In addition, more research is necessary on the effects of different UI designs on driver attention and behaviour, considering the variability introduced by aesthetic design, layout changes, and environmental urgencies. Investigating the impacts of head-up displays and addressing the challenges associated with auditory UIs, including potential distractions over long time exposure, would be beneficial. Future work could also explore UIs that provide extensive manoeuvre information, considering human factors such as annoyance and workload, alongside trust and acceptance, particularly for higher levels of driving automation.

A

APPENDIX FOR CHAPTER 2

A.1. QUESTIONNAIRE FOR PERSONAL CHARACTERISTICS COLLECTION

To get started, we would like to introduce our study purpose. Vehicle automation has been developed fast, with substantial mobility and safety potentials. The interaction between the automation and users, and the user's reaction to the automation are still not considered sufficiently. For instance, we don't know to what extent is the automation reliable and will we feel safe in a driving automation. The problems need to be solved.

In this study, we want to know the mechanism of perceived safety and trust in automated vehicles. You will experience a driving automation on a driving simulator. The driving data will be collected. Before, during, and after the simulator driving, you will be asked to fill the questionnaire and tell your feelings of safety & trust regarding the driving automation and the driving scenarios.

There are 15 questions in this questionnaire and it will approximately takes 2 min to fill it. All the personal information collected (e.g., name, age, gender, driving experience, etc.) will not be shared beyond the study team.

1. Email address: _____
2. Age (in years): _____
3. Gender: _____
    - Female
    - Male
    - Don't want to respond
4. What is your educational level? (if currently enrolled, highest degree received.)
    - No schooling completed

- Primary school graduate
- Middle school graduate
- Some high school, no diploma
- High school graduate, diploma or the equivalent
- Some college credit, no degree
- Trade/technical/vocational training
- Associate degree
- Bachelor's degree
- Master's degree
- Professional degree
- Doctorate degree
- Other _____

5. Years with driving license: _____

6. Approximately how many kilometres did you drive in your whole life as a driver?

   - I didn't drive in the past 12 months
   - Less than 2000 km
   - 2,000~5,000 km
   - 5,000~10,000 km
   - 10,000~15,000 km
   - 15,000~20,000 km
   - 20,000~50,000 km
   - 50,000~100,000 km
   - 100,000~200,000 km
   - More than 200,000 km
   - I don't know

7. Approximately how many kilometres did you drive in the past 12 months as a driver?

   - I didn't drive in the past 12 months
   - Less than 2000 km
   - 2,000~5,000 km
   - 5,000~10,000 km
   - 10,000~15,000 km
   - 15,000~20,000 km
   - 20,000~50,000 km

- More than 50,000 km
- I don't know

8. How much do you know about automated driving?
   - None at all
   - A little
   - A moderate amount
   - A lot
   - A great deal
   - I am engaged in automated driving related work.
9. Experience of using Adaptive Cruise Control (ACC) in your whole life? (in years) Note: Adaptive cruise control (ACC) is a system for road vehicles that automatically adjusts the vehicle speed to maintain a safe distance from vehicles ahead or maintain a target speed.
   - I never used it
   - Less than 1 year
   - 1~3 years
   - 3~5 years
   - More than 5 years
   - I don't know
10. Experience of using Adaptive Cruise Control (ACC) in your whole life? (in kilometres)
    - I never used it
    - Less than 100 km
    - 100~500 km
    - 500 km~1,000 km
    - More than 1,000 km
    - I don't know
11. Experience of using Lane Keeping System (LKS) in your whole life? (in years) Note: Lane keeping support (LKS) is a system for road vehicles that can steer you back into the current lane if you begin to drift out of it or automatically keep the car in the lane centre.
    - I never used it
    - Less than 1 year
    - 1~3 years

- 3~5 years
- More than 5 years
- I don't know

12. Experience of using Lane Keeping System (LKS) in your whole life? (in kilometres)
    - I never used it
    - Less than 100 km
    - 100~500 km
    - 500 km~1,000 km
    - More than 1,000 km
    - I don't know
13. Experience of using Adaptive Cruise Control (ACC) and Lane Keeping System (LKS) at the same time in your whole life? (in years)
    - I never used both at the same time
    - Less than 1 year
    - 1~3 years
    - 3~5 years
    - More than 5 years
    - I don't know
14. Experience of using Adaptive Cruise Control (ACC) and Lane Keeping System (LKS) at the same time in your whole life? (in kilometres)
    - I never used both at the same time
    - Less than 100 km
    - 100~500 km
    - 500 km~1,000 km
    - More than 1,000 km
    - I don't know

A.2. EXTRA FIGURES AND TABLES FOR THE REGRESSION ANALYSIS

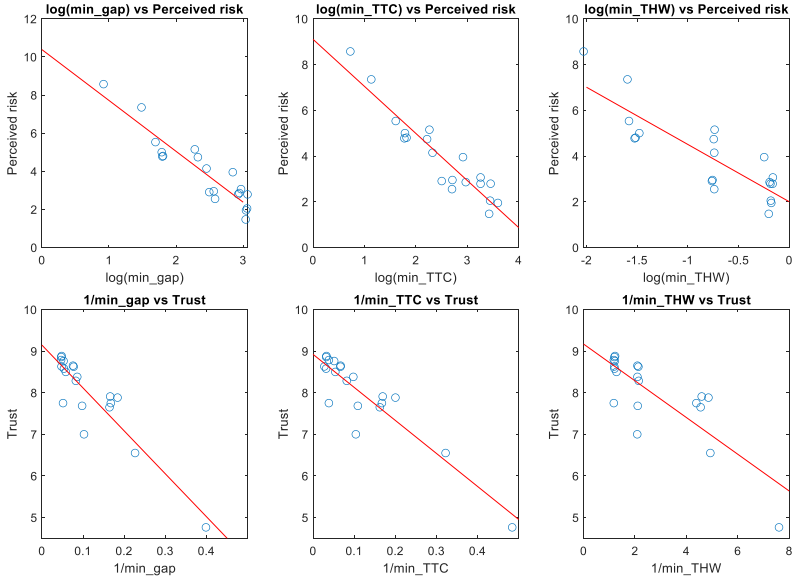


Figure A.1: Non-linear transformations. Upper row min_gap vs perceived risk, min_TTC vs perceived risk and min_THW vs perceived risk (mean value regarding different events in logarithmic scale); Lower row: 1/min_gap vs trust, 1/min_TTC vs trust and 1/min_THW vs trust

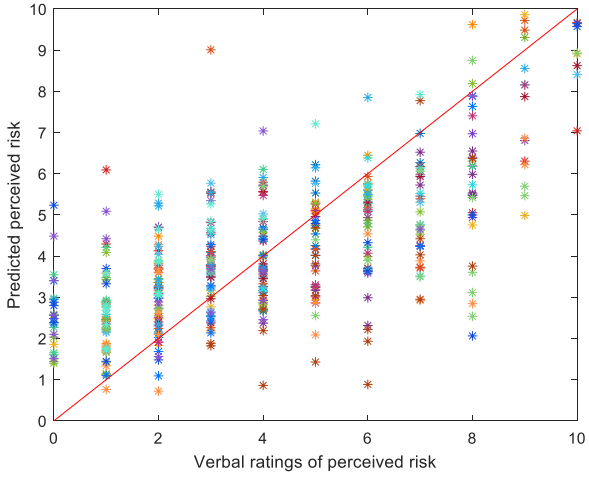


Figure A.2: Regression results of perceived risk (model 4 in Table 2.4). The dots with the same colour are from a specific participant.

A

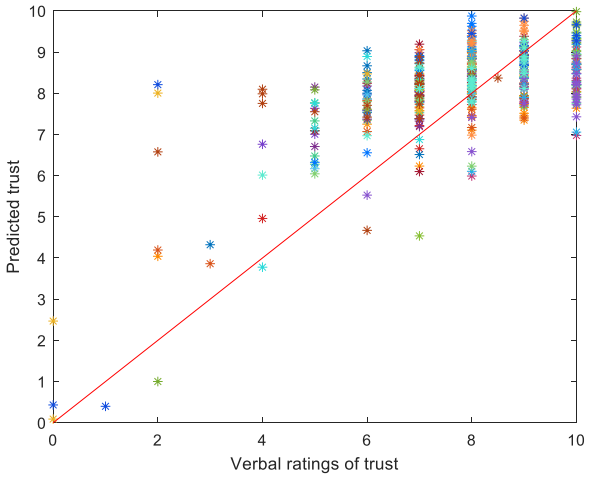


Figure A.3: Regression results of trust (model 4 in Table 2.5). The dots with the same colour are from a specific participant.

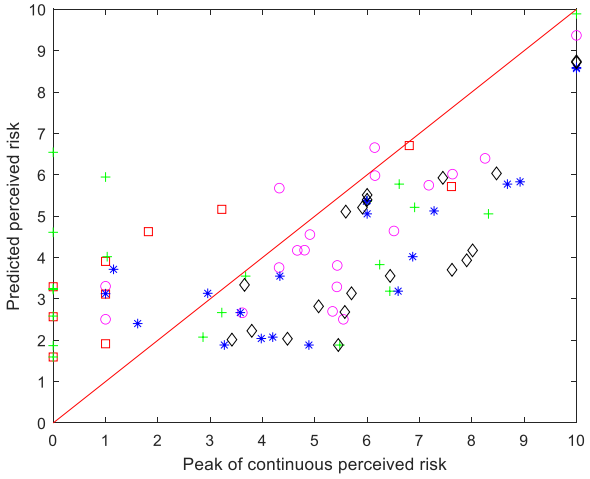


Figure A.4: Validation results of perceived risk (based on the data of 5 extra participants). One kind of marker represents a specific participant

Table A.1: Perceived risk model variants where predictor candidates from the specific groups are substituted

Model	Variables in model	R^2	Adjusted R^2	F	p	B	95% confidence interval for B		β	t	p
							Lower bound	**Upper bound**			
5	Constant	0.438	0.433	82.372	0.000	7.689	6.873	8.505	-	18.519	0.000
	ln(min_TTC)					-1.717	-1.936	-1.499	-0.572	-15.442	0.000
	YDL					-0.030	-0.043	-0.018	-0.181	-4.979	0.000
	BI					-0.190	-0.265	-0.115	-0.184	-4.888	0.000
	GEN					0.390	0.268	0.511	0.068	1.834	0.067
6	Constant	0.460	0.455	90.009	0.000	9.980	0.385	1.575	-	3.771	0.000
	ln(min_THW)					-2.453	-2.749	-2.157	-0.582	-15.918	0.000
	YDL					-0.034	-0.046	-0.022	-0.203	-5.734	0.000
	BI					-0.299	-0.372	-0.227	-0.290	-7.575	0.000
	GEN					0.444	0.034	0.854	0.077	2.142	0.033
7	Constant	0.425	0.419	78.029	0.000	8.227	7.338	9.116	-	20.129	0.000
	ln(IMG)					-2.118	-2.397	-1.839	-0.547	-14.940	0.000
	YDL					-0.034	-0.046	-0.022	-0.202	-5.403	0.000
	BI					-0.335	-0.410	-0.260	-0.332	-8.788	0.000
	GEN					0.413	0.010	0.836	0.070	1.920	0.056
8	Constant	0.483	0.478	98.652	0.000	10.052	9.071	11.034	-	20.130	0.000
	ln(min_gap)					-2.476	-2.757	-2.195	-0.553	-17.324	0.000
	AGE					-0.032	-0.043	-0.022	-0.202	-6.004	0.000
	BI					-0.201	-0.273	-0.129	-0.325	-5.510	0.000
	GEN					0.343	-0.054	0.739	0.072	1.927	0.055
9	Constant	0.380	0.374	64.678	0.000	8.484	7.239	9.729	-	13.394	0.000
	1/min_TTC					-0.161	-0.244	-0.081	-0.156	-3.951	0.000
	BI					-0.032	-0.041	-0.019	-0.187	-4.893	0.000
	YDL					-0.032	-0.129	0.648	0.050	1.312	0.190
	REP					0.260	0.129	0.648	0.050	1.312	0.190
10	Constant	0.495	0.489	82.651	0.000	9.163	8.655	10.077	-	19.845	0.000
	ln(min_gap)					-2.367	-2.257	-2.477	-0.586	-16.046	0.000
	YDL					-0.039	-0.051	-0.027	-0.231	-6.561	0.000
	BI					-0.190	-0.262	-0.119	-0.185	-5.239	0.000
	GEN					0.467	0.067	0.863	0.081	2.314	0.021

Table A.2: Trust model variants where predictor candidates from the specific group are substituted

Model	Variables in model	R^2	Adjusted R^2	F	p	B	95% confidence interval for B		β	t	p
							Lower bound	**Upper bound**			
5	Constant	0.329	0.322	51.738	0.000	9.086	8.665	9.506		42.450	0.000
	1/Min_gap					-8.106	-9.518	-6.694	-0.462	-11.281	0.000
	BI					0.150	0.092	0.209	0.207	5.165	0.000
	YDL					0.016	0.007	0.026	0.138	3.471	0.001
	REP					0.282	-0.003	0.568	0.078	1.942	0.053
6	Constant	0.319	0.313	49.615	0.000	9.504	9.053	9.955		41.439	0.000
	1/Min_THW					-0.406	-0.479	-0.333	-0.441	-10.947	0.000
	BI					0.201	0.143	0.258	0.276	6.866	0.000
	YDL					0.015	0.006	0.025	0.130	3.244	0.001
	REP					0.317	0.030	0.604	0.087	2.169	0.031
7	Constant	0.259	0.252	36.942	0.000	9.451	8.967	9.934		28.805	0.000
	1/IMG					-9.555	-11.715	-7.394	-0.365	8.767	0.000
	BI					0.240	0.180	0.300	0.330	8.094	0.000
	YDL					0.017	0.007	0.026	0.139	3.329	0.001
	REP					0.419	0.020	0.719	0.115	2.744	0.006
8	Constant	0.408	0.403	72.927	0.000	8.645	8.172	9.117		35.972	0.000
	1/Min_TTC					-6.244	-7.100	-5.387	-0.553	-14.328	0.000
	BI					0.115	0.077	0.158	0.172	4.463	0.000
	AGE					0.011	0.003	0.019	0.103	2.763	0.006
	REP					0.373	0.105	0.640	0.102	2.738	0.006
9	Constant	0.311	0.305	47.802	0.000	5.235	8.665	9.506		14.060	0.000
	ln(min_gap)					-1.272	-9.518	-6.694	0.447	10.956	0.000
	BI					0.178	0.092	0.209	0.246	6.027	0.000
	YDL					0.017	0.007	0.026	0.147	3.589	0.000
	GEN					0.191	-0.003	0.568	0.047	1.152	0.250
10	Constant	0.420	0.414	61.222	0.000	8.730	8.392	9.067		44.175	0.000
	1/Min_TTC					-5.539	-6.621	-4.396	-0.490	-9.351	0.000
	BI					0.126	0.072	0.181	0.174	4.544	0.000
	YDL					0.016	0.008	0.026	0.145	3.871	0.000
	REP					0.367	0.101	0.631	0.101	2.711	0.007
	Max_B					-1.488	-3.097	0.120	-0.094	-1.819	0.070

Table A.3: Coefficients of Pearson correlation between perceived risk and trust. Correlations with '*' are significant ($p < 0.05$).

Participant	Average perceived risk (std)	Average trust (std)	Pearson Coefficient	p-value
1	3.95 (3.08)	7.16 (1.04)	-0.043	0.858
2*	3.50 (1.06)	8.40 (1.76)	-0.779	0.000
3	5.47 (2.32)	7.05 (1.31)	-0.105	0.659
4*	2.60 (2.61)	8.10 (2.11)	-0.913	0.000
5*	4.90 (1.22)	6.40 (0.97)	-0.473	0.036
6*	3.55 (2.27)	7.05 (1.43)	-0.840	0.000
7*	2.55 (2.77)	8.70 (1.93)	-0.880	0.000
8*	4.55 (2.25)	7.55 (1.88)	-0.851	0.000
9*	3.65 (2.63)	7.80 (1.97)	-0.905	0.000
10*	4.90 (2.70)	6.65 (2.69)	-0.859	0.000
11	6.16 (1.49)	7.10 (0.97)	0.337	0.146
12*	2.80 (2.23)	9.00 (1.97)	-0.830	0.000
13	3.89 (2.12)	8.21 (1.15)	-0.129	0.588
14*	4.35 (2.24)	7.70 (1.10)	-0.708	0.000
15	5.61 (2.09)	6.83 (1.12)	0.499	0.025
16	3.52 (2.01)	7.58 (0.82)	0.008	0.974
17	3.11 (2.71)	8.56 (1.34)	-0.065	0.785
18	4.30 (1.93)	9.55 (0.59)	-0.277	0.236
19	3.80 (2.82)	8.35 (0.57)	-0.081	0.736
20	3.10 (2.17)	9.20 (0.87)	-0.249	0.290
21	3.44 (3.61)	7.61 (3.06)	-0.355	0.125
22	6.60 (2.18)	7.90 (1.26)	-0.233	0.323
23	2.40 (2.27)	8.90 (2.28)	-0.098	0.679
24*	5.45 (1.50)	6.23 (1.60)	-0.449	0.047
25*	2.30 (1.52)	8.00 (1.22)	-0.913	0.000
Mean	3.97 (1.77)	7.99 (0.97)	-0.919	0.000
Overall	4.01 (2.60)	7.83 (1.83)	-0.649	0.000

A

A.3. INDIVIDUAL CALIBRATION OF THE REGRESSION MODEL

For perceived risk, the individual calibration results are shown in Table A.4 YDL and GEN were meaningless for individual data so they were excluded from the regression. According to the calibrated model structure, participants were divided into 2 clusters (Table A.5). In Cluster 1, we find that there is only one female participant and there are only 2 participants having driving automation experience. In Cluster 2, 8 out of 9 female participants and 9 out of 11 automation-experienced participants are in this cluster. Participants in Cluster 2 are only sensitive to minimum gap during the events.

For trust, the calibration results are shown in Table A.6. We find that the variance among different participants is huge. Only participant 19 did not obtain the significant calibrated trust model because this participant always rated 8 or 9 during the simulator drive. According to the trust model structure, the participants can be divided into 5 clusters that are shown in Table A.7. However, we did not find significant features regarding different trust clusters.

For both perceived risk and trust, the individual calibration results show that the coefficients on the models are stable to some degree, where values vary slightly with consistent signs.

Table A.4: Calibration results of perceived risk based on individual data (For the data within individual participants, years with a driving license (YDL) and gender (GEN) were always constant, which were not applicable and were omitted in the table)

Participant	intercept	Coefficient ln(min_gap)	Coefficient BI	R^2	Adjusted R^2	F	p
1	7.927	-2.684	-0.498	0.551	0.494	9.802	0.002
2*	9.481	-2.446	0	0.772	0.759	60.913	0.000
3	13.160	-3.598	-0.249	0.782	0.755	28.748	0.000
4*	12.032	-3.726	0	0.617	0.596	29.000	0.000
5	8.454	-1.468	0	0.718	0.702	45.778	0.000
8*	9.843	-2.263	0	0.637	0.616	29.841	0.000
10*	13.513	-3.702	0	0.729	0.714	48.494	0.000
11	12.259	-2.430	0	0.734	0.719	46.959	0.000
12	9.791	-3.229	-0.184	0.816	0.795	37.792	0.000
13	8.418	-2.575	-0.316	0.724	0.689	20.971	0.000
14	12.010	-3.288	0	0.839	0.830	93.934	0.000
15	11.233	-2.503	0	0.529	0.499	17.953	0.001
16	7.765	-2.423	-0.258	0.782	0.755	28.622	0.000
17	11.678	-3.534	0	0.706	0.688	38.491	0.000
18	7.144	-1.842	-0.273	0.551	0.498	10.412	0.001
19	5.082	-2.069	-0.663	0.642	0.600	15.234	0.000
20	5.481	-2.133	-0.488	0.806	0.783	35.343	0.000
21	11.747	-3.627	0	0.442	0.408	12.695	0.003
22	10.575	-1.720	0	0.275	0.235	6.837	0.018
23*	3.015	-1.258	-0.437	0.405	0.335	5.784	0.012
24*	8.595	-1.359	0	0.308	0.207	8.030	0.011
25	5.244	-1.831	-0.249	0.833	0.814	42.501	0.000
Overall	9.384	-2.473	-0.201	0.487	0.482	100.427	0.000

Table A.5: Clustering results according to the perceived risk regression equation

Clusters	Participants
1	1, 3, 12, 13, 16, 18, 19, 20, 23, 25
2	2, 4, 5, 8, 10, 11, 14, 15, 17, 21, 22, 24

Table A.6: Calibration results of trust based on individual data (For the data within individual participants, years with a driving license (YDL) was always constant, which was not applicable and omitted in this table)

Participant	intercept	Coefficient 1/min_TTC	Coefficient BI	Coefficient REP	R^2	Adjusted R^2	F	p
1	9.978	-5.450	0	0	0.499	0.469	16.921	0.001
2	9.908	-9.920	0	0	0.671	0.653	36.764	0.000
3	9.752	-11.102	0.233	0	0.838	0.818	41.360	0.000
4	10.116	-14.797	0	0	0.476	0.446	16.326	0.001
5	6.414	-3.827	0	0.969	0.623	0.578	14.032	0.000
8	9.600	-7.709	0.242	0	0.887	0.873	62.971	0.000
10	7.990	-7.711	0	0	0.718	0.703	45.917	0.000
11	8.322	-8.801	0	0	0.606	0.582	26.107	0.000
12	10.264	-5.961	0.165	1.107	0.772	0.729	18.063	0.000
13	9.236	-8.341	0	0	0.464	0.434	15.552	0.001
14	8.445	-7.002	0	0	0.404	0.371	12.221	0.003
15	7.544	-2.128	0.203	1.154	0.657	0.584	8.957	0.001
16	8.579	-4.888	0.123	0.398	0.777	0.732	17.391	0.000
17	9.443	-5.191	0	0	0.403	0.365	10.782	0.005
18	9.335	-2.171	-0.892	0	0.718	0.685	21.663	0.000
19	8.760	0.239	0.092	0.101	0.161	0.003	1.020	0.410
20	8.374	-0.897	-0.023	1.597	0.854	0.826	31.135	0.000
21	9.047	-6.808	0	0	0.702	0.684	37.760	0.000
22	8.522	-5.495	0	0	0.260	0.219	6.337	0.022
23	9.389	-4.889	0	0	0.414	0.382	12.741	0.002
24	9.376	0	0.458	-1.450	0.713	0.679	21.121	0.000
25	9.705	-10.152	0.188	0	0.840	0.822	44.783	0.000
overall	8.780	-6.265	0.125	0.372	0.416	0.410	75.290	0.000

Table A.7: Clustering results according to regression model structure of trust

Clusters	Participants
1	1,2,4,10,11,13,14,17,21,22,23
2	3,8,25
3	5,18
4	12,15,16,19,20
5	24

A

A.4. BRAKING BEHAVIOUR, PUPIL DILATION AND ECG VERSUS PERCEIVED RISK AND TRUST

A.4.1. BRAKING BEHAVIOUR

We used the medians of the bars in different events, and then compare the medians between braking group and non-braking group. Braking groups reported a higher perceived risk and lower trust level supported by t-test ($p = 0.021$ for perceived risk and $p = 0.000$ for trust). (See Figure A.5)

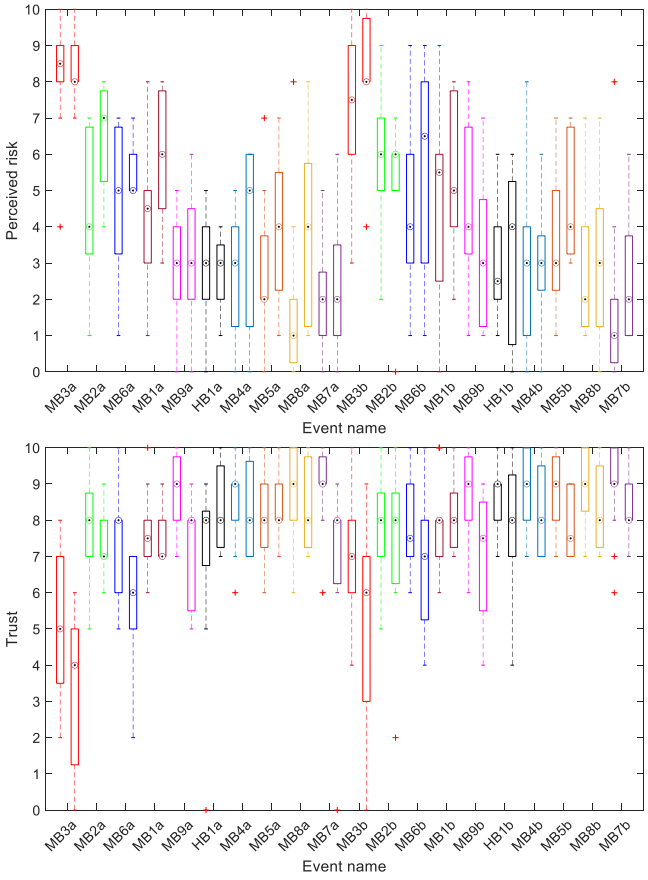


Figure A.5: Comparison between the braking and non-braking group (left bar: non-braking group; right bar: braking group). Upper chart: Perceived risk; Lower chart: Trust

A.4.2. PUPIL DILATION

The first row of Figure 2.7 shows the participant-averaged pupil diameter signal in the most critical merge MB3, the least critical merge MB7, and the hard braking HB1 (with the time scale from $t = -55s$ to $t = 55s$ where $t = 0$ means the smallest gap to the lead vehicle). Across all events, a consistent pattern emerged with maximum pupil dilation

around $t = 0$, reflecting the smallest gap to the lead vehicle and closely corresponding to the peak in perceived risk. Figure A.6 compares the median pupil diameter in four different conditions (different participant groups and different events) over time windows of 10 seconds. We divided the period of one event into 10 time windows and checked the change of the median pupil diameter across all participants in these 10 time windows. A Kruskal-Wallis test showed significant variations in median pupil diameter over the 10 time windows within four different conditions ($p = 0.000$). We tested 5 extra participants without speaking and analysed the signals (the black curve in Figure A.6). The t-test result indicates no significant difference compared to the median series of the 22 original participants, demonstrating that speaking has no significant influence on pupil dilation.

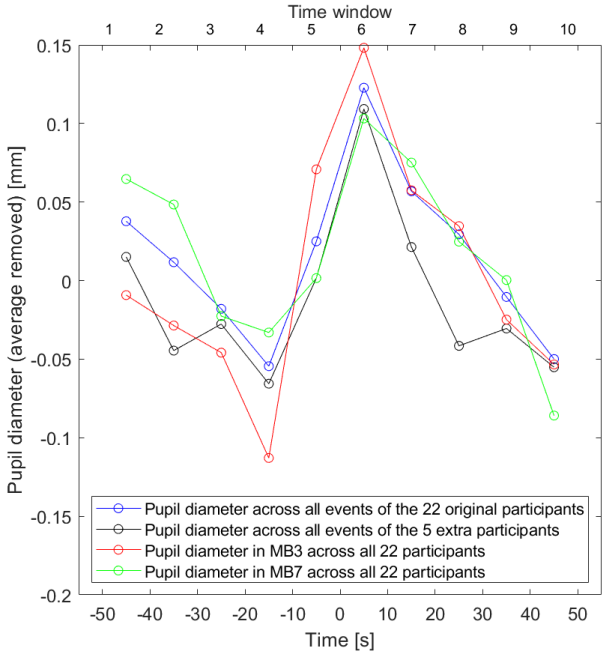


Figure A.6: Median pupil diameter (average removed) over 10 time windows of 10 seconds. The time scale is $t = -55s$ to $t = 55s$, where $t = 0$ means the smallest gap to the lead vehicle.

Table A.8: Correlation coefficients between participant-averaged pupil diameter signal and continuous perceived risk signal within different events. Signals were re-sampled with 10 Hz within time scope $t = -20s$ to $t = 10s$. Insignificant correlations are shaded in grey.

Event	a (first exposure)		b (second exposure)		Over all events		N
	r	p	r	p	R	p	
MB1	0.47	0.00	0.29	0.00	-	-	300
MB2	0.50	0.00	0.11	0.05	-	-	300
MB3	0.86	0.00	0.83	0.00	-	-	300
MB4	0.58	0.00	0.72	0.00	-	-	300
MB5	0.30	0.00	0.23	0.00	-	-	300
MB6	0.89	0.00	0.66	0.00	-	-	300
MB7	0.53	0.00	-0.28	0.00	-	-	300
MB8	0.31	0.00	0.10	0.07	-	-	300
MB9	0.64	0.00	0.45	0.00	-	-	300
HB1	-0.07	0.21	-0.26	0.00	-	-	300
Over all 397 events	-	-	-	-	0.88	0.00	300

A.4.3. ECG

Table A.9: ANOVA results of ECG metrics (significance not corrected for multiple comparisons. Greenhouse-Geisser correction was applied on each because of violation of the sphericity assumption; Correction factors were well below 0.75.)

	1st encounters	2nd encounters
IBI	$F(3.877, 81.4) = 2.467, p = 0.053$	$F(5.318, 106.368) = 4.032, p = 0.002$
RMSSD	$F(2.832, 59.471) = 1.419, p = 0.247$	$F(4.042, 80.836) = 0.945, p = 0.443$
HF	$F(2.039, 42.825) = 1.059, p = 0.357$	$F(3.655, 73.098) = 1.521, p = 0.209$

Table A.10: Repeated measure correlations between ECG and risk metrics

	N	Correlation: r	Significance: p
IBI vs. max_cont._risk	433	-0.053	0.272
RMSSD vs. max_cont._risk	433	0.006	0.894
HF vs. max_cont._risk	433	0.023	0.634
IBI vs. min_THW	433	-0.001	0.979
RMSSD vs. min_THW	433	0.064	0.185
HF vs. min_THW	433	0.071	0.140
max_cont._risk vs. min_THW	433	-0.383	< 0.001

Since these BPM increases do not correlate with risk, we evaluated whether the effect was caused by verbal responses. Bernardi *et al.* [254] show that free talking can increase heart rate by about 8 bpm compared to spontaneous, fast or slow breathing. The absence of the verbal rating tasks did not elicit a consistent change in average heart rate and a paired t-test does not yield a significant difference (see Figure A.7): $t(9) = -0.434, p = 0.674$. Hence the BPM increase is not an artefact of speaking.

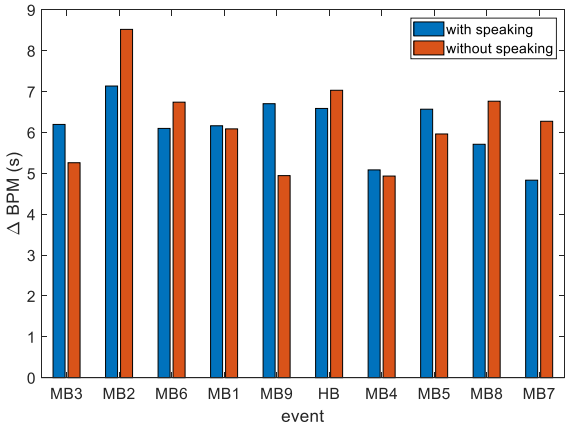


Figure A.7: Comparison of BPM increase following each event (difference between local extrema after smoothing) between the original 25 participants, and the extra 5 participants without speaking

B

APPENDIX FOR CHAPTER 3

B.1. RELATED PERCEIVED RISK MODELS

B.1.1. REGRESSION PERCEIVED RISK MODEL (RPR)

The Regression Perceived Risk model (RPR) is an event-based perceived risk model derived from our previous simulator experiment, where 18 merging events with various merging distances and braking intensities on a 2-lane highway were simulated. RPR predicts human drivers' event-based perceived risk ratings ranging from 0-10 regarding merging events based on the corresponding kinematic data from the simulator drive.

The RPR model builds on several assumptions:

- Perceived risk stems from the vehicles directly in front of the subject vehicle, which means the merging vehicles cause perceived risk only after entering the current lane.
- Drivers can accurately estimate the motion information (e.g., relative position, velocity, acceleration, etc.) with the human sensory system.

The initial model can predict event-based perceived risk [117], as shown in Equation (B.1)

$$R_{RPR\_event} = 9.384 - 2.473 \cdot \ln(gap_{min}) - 0.038 \cdot YDL - 0.201 \cdot BI_{max} + 0.470 \cdot GEN \quad \text{(B.1)}$$

where R_{RPR_event} is the event-based perceived risk ranging from 0-10; gap_{min} is the minimum clearance in metres to the leading vehicle during an event; YDL represents the years with a valid driving license; BI_{max} denotes the maximum braking intensity (m/s^2) of the merging vehicle; GEN represents the gender of the participants with $Female = 1$ and $Male = 0$. The model coefficients were originally calibrated based on a perceived risk dataset [117], detailed in Section 3.5.1.

We extend the model to compute real-time perceived risk by replacing min_gap and max_BI with the real-time values. YDL and GEN are omitted since they remain

constant for a certain group of participants, and can be accounted for by the intercept. In this way, RPR is formulated in the continuous time domain as

$$R_{RPR}(t) = C_0 + C_1 \cdot \ln(x_n(t) - x_s(t)) + C_2 \cdot a_{n,X}(t) \tag{B.2}$$

where $x_n(t)$ and $x_s(t)$ are the real-time longitudinal position (m) of the neighbouring vehicle and the subject vehicle; $a_{n,X}(t)$ is the current acceleration (m/s^2) of the neighbouring vehicle, which is the braking intensity in this model; According to the simulator experiment settings [117], the validity range of the model is that $x_n(t) - x_s(t) < 33\,\text{m}$ and $-8\,\text{m/s}^2 \leqslant a_{n,X}(t) \leqslant -2\,\text{m/s}^2$. Verification is required for the model outside this range. For enhanced performance, we recalibrate parameters C_0, C_1 and C_2 using two datasets.

B.1.2. PERCEIVED PROBABILISTIC DRIVING RISK FIELD MODEL (PPDRF)

Perceived Probabilistic Driving Risk Field Model (PPDRF) enhances the Probabilistic Driving Risk Field Model (PDRF) [40] by accounting for diverse traffic scenarios and driver individuality. The model is inspired by artificial potential field used in driving automation [119], [120], [255]. PDRF estimates collision risk by considering potential risk from non-moving vehicles/objects and kinetic risk from other road users. The former accounts for collision energy and probability with stationary obstacles, while the latter involves spatial overlap with neighbouring vehicles using predicted positions and stochastic accelerations. In stable highway driving, the longitudinal and lateral accelerations of a neighbour follow a Gaussian distribution [124], [125]. However, due to uncertainties and behavioural deviations, human drivers perceive risk differently, leading to a bias between objective and perceived risk. To address this, we introduce assumptions to extend PDRF into PPDRF for predicting perceived risk.

- The future acceleration in longitudinal and lateral directions of neighbouring vehicles follows independent Gaussian distributions with the current acceleration as the mean value, remaining constant over the prediction horizon;
- The subject vehicle maintains the current acceleration over the prediction horizon;

The two assumptions simplify road users' motion.

In PPDRF model, human drivers, at time t, perceive a total risk as a combination of kinetic risk and potential risk as follows

$$R_{PPDRF}(t) = R_{n,s}(t) + R_{o,s}(t) \tag{B.3}$$

The kinetic risk in PPDRF concerning moving neighbouring vehicles is given by

$$R_{n,s}(t) = 0.5 M_s \beta^2 \left|\Delta v_{s,n}(t+\tau)\right|^2 \cdot \tilde{p}(n,s \mid t) \tag{B.4}$$

where $R_{n,s}(t)$ is the kinetic collision risk between the subject vehicle s and a neighbouring vehicle n in Joules at time t. $\beta = \frac{M_n}{M_s + M_n}$ denotes the mass ratio. M_s and M_n are the mass of the subject vehicle and the neighbouring vehicle. $\Delta v_{s,n}(t+\tau)$ is the relative velocity between the subject vehicle and the neighbouring vehicle at time $t+\tau$. $\tilde{p}(n,s \mid t)$

is the collision probability to the neighbouring vehicle estimated by drivers ranging on [0,1].

The collision probability $\tilde{p}(n,s \mid t)$ to the neighbouring vehicle estimated by human drivers is constructed as Equation (B.5).

$$\tilde{p}(n,s \mid t) = \mathcal{N}\left(\frac{\Delta x(t) - \Delta v_X(t)\tau}{0.5\tau^2} \mid \mu_X(t), \tilde{\sigma}_X\right) \cdot \mathcal{N}\left(\frac{\Delta y(t) - \Delta v_Y(t)\tau}{0.5\tau^2} \mid \mu_Y(t), \tilde{\sigma}_Y\right) \tag{B.5}$$

where $\mathcal{N}$ is the assumed Gaussian collision probability density function (Figure B.1). $\mu_X(t)$ and $\mu_Y(t)$ represent the mean values for longitudinal and lateral acceleration distribution, while $\tilde{\sigma}X$ and $\tilde{\sigma}Y$ are the respective standard deviations. The relative spacing and velocities between the subject and neighbouring vehicles are denoted as $\Delta x(t)$, $\Delta y(t)$, $\Delta v_X(t)$, and $\Delta v_Y(t)$. PPDRF evaluates collision probability using multiple values of $\tau = 0.5\,\text{s}, 1\,\text{s}, 2\,\text{s}, 3\,\text{s}$, with the model employing all these values to maximise the computed collision probability. Using the constant acceleration assumption, the predicted position of the subject vehicle and stochastic positions of neighbouring vehicles are calculated over a prediction horizon. Spatial overlap and collision predictions are determined accordingly. The actual $\tilde{p}(n,s \mid t)$ is obtained through integration over the expected accelerations.

The potential risk posed to vehicle s by a static object o can be modelled as

$$R_{o,s}(t) = 0.5kM\left(\Delta v_{s,o}(t)\right)^2 \cdot \max\left(e^{-|r_{s,o}|/D}, 0.001\right) \tag{B.6}$$

where $R_{o,s}(t)$ denotes the potential risk caused by the static object o; M denotes the mass of s; $|r_{s,o}| = ||\boldsymbol{p}_s - \boldsymbol{p}_n||$ is the distance between the subject vehicle s and the non-moving object o; $V_{s,o}$ denotes the relative velocity; $0.5kM(V_{s,o})^2$ represents the expected crash energy scaled by the parameter k, with range $[0-1]$, which is set to 1 in this study representing the neighbour is immovable; the term $e^{-|r_{s,o}|/D}$ is the collision probability ranging between [0-1], where D determines the steepness of descent of the potential field, varying among different drivers.

It is noteworthy that $R_{PPDRF}(t)$ represents a probabilistic energy value, and can attain up to 3×10^4 J under stable motorway driving conditions [40].

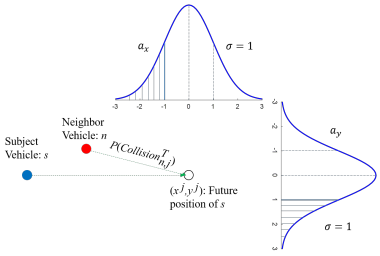


Figure B.1: The acceleration distribution of the neighbouring vehicle and the relative spacing between the subject vehicle and the neighbouring vehicle

B.1.3. DRIVING RISK FIELD MODEL (DRF)

The DRF represents human drivers' risk perception as a 2D field, combining the probability (probability field) and consequence (severity field) of an event [41], the product of which provides an estimation of driver's perceived risk. The DRF model was derived from a simulator experiment involving obstacle avoidance with 77 obstacles distributed on a 2D plane in front of the subject vehicle. During each drive, one obstacle was randomly chosen and suddenly appeared, after which participants needed to steer to avoid the obstacle and gave a non-negative number indicating required steering effort. Based on the position information of the obstacles, the maximum steering angle, and the subjective ratings, the DRF model was fitted to the data, and thereby it is essentially an empirical model. The DRF is based on the following assumptions:

- Perceived risk is the product of the probability of a hazardous event occurring estimated by drivers and the event severity;
- The perceived risk field widens as the longitudinal distance from the subject vehicle increases;
- The height of the perceived risk field decays as the lateral and longitudinal distance from the vehicle increases;

The DRF model quantifies overall perceived risk as

$$R_{DRF}(t) = \sum p(x(t), y(t)) \cdot sev(t) \tag{B.7}$$

where $p(x(t), y(t))$ is the probability of an event happening at position $(x(t), y(t))$; $sev(t)$ is the severity field of events. Specifically, in straight drive, the probability field can be simplified as

$$p(x(t), y(t)) = h \cdot \exp\left(\frac{-y(t)^2}{2\sigma^2}\right) \tag{B.8}$$

$$h = s \cdot \left(x(t) - v_{s,X}(t) \cdot t_{la}\right)^2 \tag{B.9}$$

$$\sigma = m \cdot x(t) + c \tag{B.10}$$

where the subject vehicle is at the origin $(0,0)$ with h and σ representing the height and the width of the Gaussian at longitudinal position $x(t)$; s defines the steepness of the height parabola; t_{la} is the human driver's preview time (s); m defines the widening rate of the 2D probability field; c is the quarter width of the subject vehicle (m). $v_{s,X}(t)$ is the subject vehicle's velocity (m/s). The lateral cross-section of the 2D probability field is a Gaussian. Note that the height of the Gaussian h and the width σ are separately modelled as a parabola and linear function of longitudinal distance x in front of the subject vehicle.

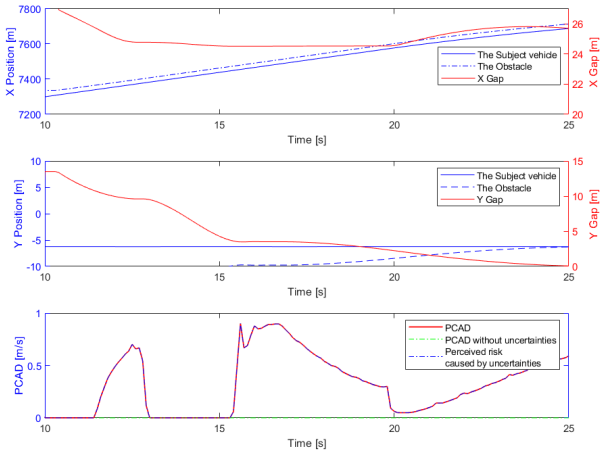
The severity field of the events in this study can be defined as

$$sev(t) = \begin{cases} C_{sev}, & (x(t), y(t)) \in A^O, \\ 0, & (x(t), y(t)) \notin A^O. \end{cases} \tag{B.11}$$

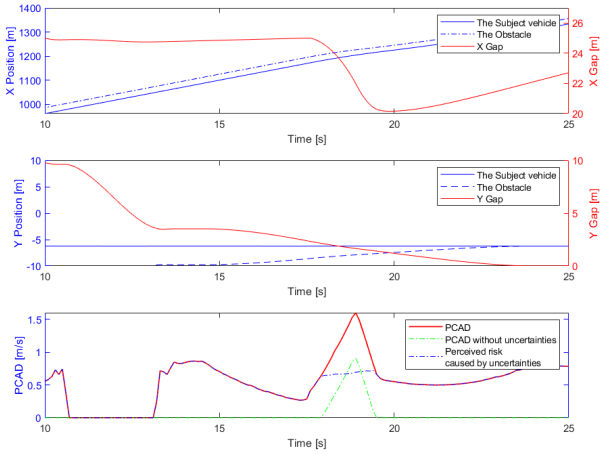
where C_{sev} is the severity value that is set empirically and A^O represents a neighbouring vehicle's spatial area.

B.2. PCAD TIME HISTORY OUTPUT

This Appendix presents the PCAD time history outputs in Dataset Merging (Figure B.2 and Dataset Obstacle Avoidance (Figure B.3).

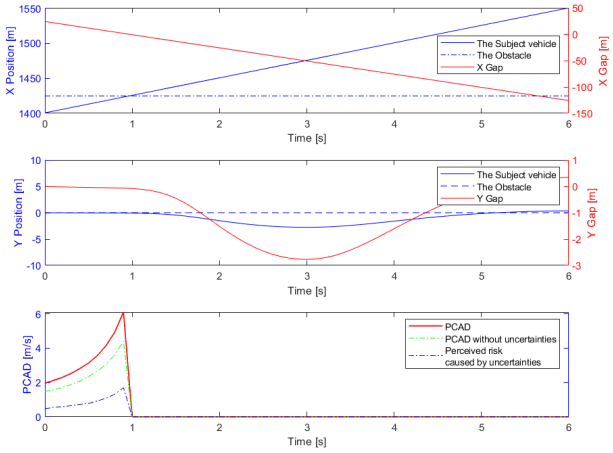


(a) Merging gap is 25 m and the deceleration of the merging vehicle is $-2\,m/s^2$

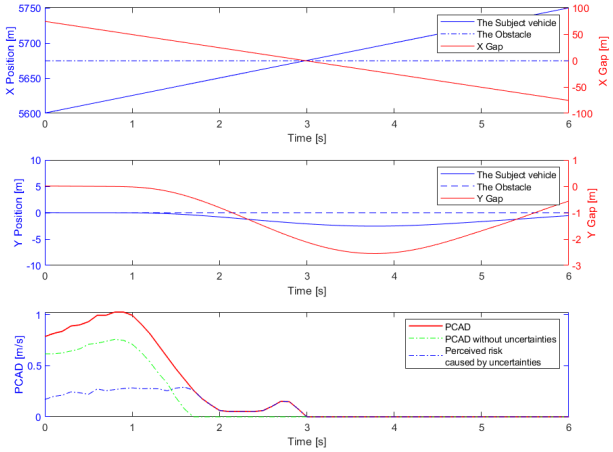


(b) Merging gap is 25 m and the deceleration of the merging vehicle is $-8\,m/s^2$

Figure B.2: PCAD historical output in merging and hard braking events



(a) The obstacle position is (25, 0), which means the obstacle is 25 m in front and the lateral offset is 0.



(b) The obstacle position is (75, 0), which means the obstacle is 75 m in front and the lateral offset is 0.

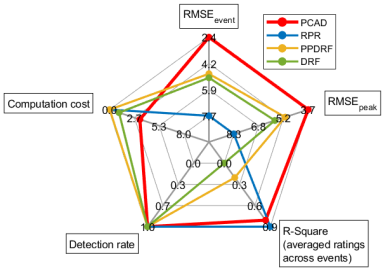
Figure B.3: PCAD historical output in obstacle avoidance events

B.3. CROSS VALIDATION

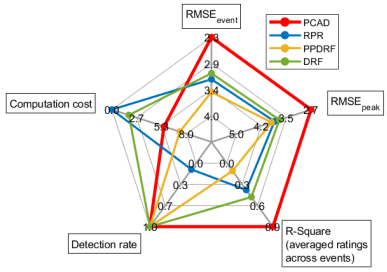
This Appendix presents the model performance in cross-validation between the two datasets.

Table B.1: Model performance indicators for the cross-validation

Dataset	Performance indicators	PCAD	RPR	PPDRF	DRF
Dataset Merging (with parameters calibrated with Dataset Obstacle Avoidance)	$RMSE_{event}$	2.37	7.72	4.86	5.14
	$RMSE_{peak}$	3.73	8.29	5.14	5.79
	Adjusted R-Square	0.79	0.88	0.20	0.00
	Detection rate	1.00	1.00	1.00	1.00
	Time consumption (ms)	3.25	2.09×10^{-4}	1.03×10^{-2}	1.01
Dataset Obstacle Avoidance (with parameters calibrated with Dataset Merging)	$RMSE_{event}$	2.28	3.20	3.48	3.19
	$RMSE_{peak}$	2.73	3.84	3.93	3.87
	Adjusted R-Square	0.90	0.38	0.11	0.42
	Detection rate	1.00	0.09	1.00	1.00
	Time consumption (ms)	5.58	1.98×10^{-4}	7.28	1.82

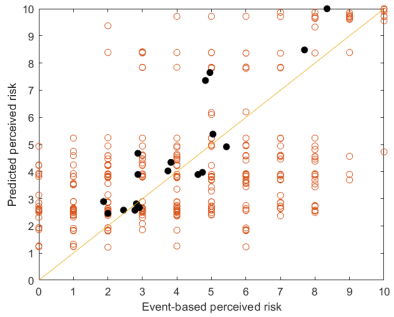


(a) Performances in Dataset Merging (with parameters calibrated with Dataset Obstacle Avoidance)

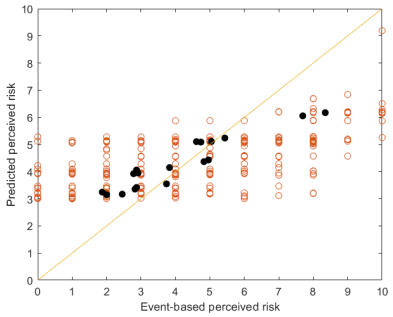


(b) Performances in Dataset Obstacle Avoidance (with parameters calibrated with Dataset Merging)

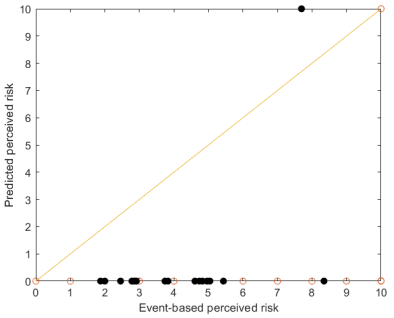
Figure B.4: Radar chart of model performance indicators for the cross-validation



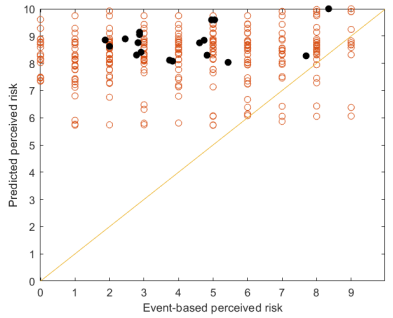
(a) PCAD (Adjusted R-Square = 0.79)



(b) RPR (Adjusted R-Square = 0.88)

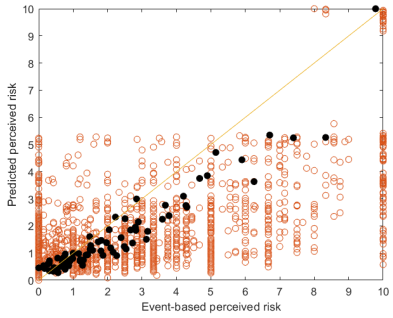


(c) PPDRF (Adjusted R-Square = 0.20)

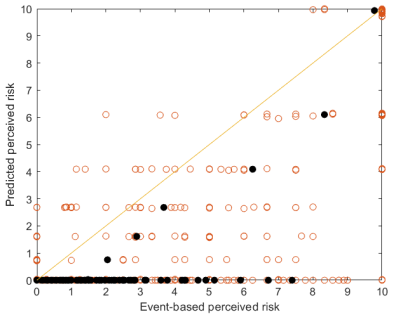


(d) DRF (Adjusted R-Square = 0.00)

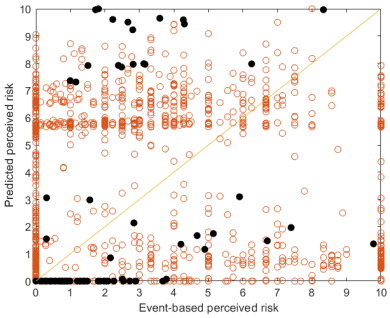
Figure B.5: Validation results in Dataset Merging with model parameters calibrated based on Dataset Obstacle Avoidance. '○' indicates raw event-based perceived risk and '●' indicates the averaged event-based perceived risk across the same event type.



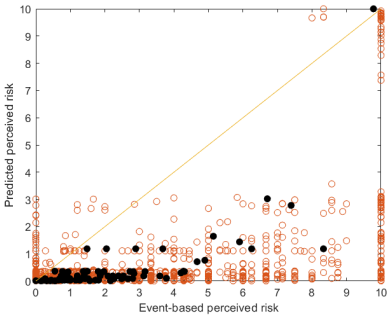
(a) PCAD (Adjusted R-Square = 0.90)



(b) RPR (Adjusted R-Square = 0.38)



(c) PPDRF (Adjusted R-Square = 0.11)



(d) DRF (Adjusted R-Square = 0.48)

Figure B.6: Validation results in Dataset Obstacle Avoidance with model parameters calibrated based on Dataset Merging. '○' indicates raw event-based perceived risk and '●' indicates the averaged event-based perceived risk across the same event type.

B

B.4. EXPLANATION OF THE UNCERTAIN VELOCITY DIRECTION

In this Appendix, we explain why the line connecting the subject vehicle and the neighbouring vehicle is selected as the direction of the uncertain velocity.

With the uncertain velocity and the perceived velocity derived from acceleration, Equation (3.3) and (3.6) are changed to

$$\begin{aligned}\dot{\boldsymbol{\theta}}'_{sj_1,nj_2} &= \frac{\left(\boldsymbol{p}_{sj_1}-\boldsymbol{p}_{nj_2}\right)\times\left(\boldsymbol{v}'_{sj_1}-\boldsymbol{v}'_{nj_2}\right)}{\left\|\boldsymbol{p}_{sj_1}-\boldsymbol{p}_{nj_2}\right\|^2}\\ &= \frac{\left(\boldsymbol{p}_{sj_1}-\boldsymbol{p}_{nj_2}\right)\times\left[\left(\boldsymbol{v}_{sj_1}+\Delta\boldsymbol{v}_{s,a}+\Delta\boldsymbol{v}_{s,u}\right)-\left(\boldsymbol{v}_{nj_2}+\Delta\boldsymbol{v}_{n,a}+\Delta\boldsymbol{v}_{n,u}\right)\right]}{\left\|\boldsymbol{p}_{sj_1}-\boldsymbol{p}_{nj_2}\right\|^2}\\ &= \dot{\boldsymbol{\theta}}_{sj_1,nj_2}+\Delta\dot{\boldsymbol{\theta}}_{sj_1,nj_2,a}+\Delta\dot{\boldsymbol{\theta}}_{sj_1,nj_2,u},\quad j_1,j_2\in\{l,r\}\end{aligned} \tag{B.12}$$

and

$$\begin{aligned}\dot{d}'_{s,n} &= \frac{1}{d_{s,n}}\left(\boldsymbol{p}_s-\boldsymbol{p}_n\right)^T\left(\boldsymbol{v}'_s-\boldsymbol{v}'_n\right)\\ &= \frac{1}{d_{s,n}}\left(\boldsymbol{p}_s-\boldsymbol{p}_n\right)^T\left[\left(\boldsymbol{v}_s+\Delta\boldsymbol{v}_{s,a}+\Delta\boldsymbol{v}_{s,u}\right)-\left(\boldsymbol{v}_n+\Delta\boldsymbol{v}_{n,a}+\Delta\boldsymbol{v}_{n,u}\right)\right]\\ &= \dot{d}_{s,n}+\dot{d}_{n,a}+\dot{d}_{s,n,u}\end{aligned} \tag{B.13}$$

where $\boldsymbol{v}'_s=\boldsymbol{v}_s+\Delta\boldsymbol{v}_{s,a}+\Delta\boldsymbol{v}_{s,u}$ and $\boldsymbol{v}'_n=\boldsymbol{v}_n+\Delta\boldsymbol{v}_{n,a}+\Delta\boldsymbol{v}_{n,u}$ are the perceived velocity with the uncertain velocity $\Delta\boldsymbol{v}_{s,u}$ and $\Delta\boldsymbol{v}_{n,u}$ for the subject vehicle and the neighbouring vehicle respectively.

In order to make the situation more dangerous, the uncertain velocity has to create a situation that is opposite to Equation (3.9) namely

$$\min\dot{\boldsymbol{\theta}}_{si,nj,u}\cdot\max\dot{\boldsymbol{\theta}}_{si,nj,u}\leqslant 0\ (i,j\in\{l,r\},)\text{ and }\dot{d}_{s,n,u}<0, \tag{B.14}$$

where all relative bearing rate and the distance changing rate are only caused by the uncertain velocity.

Comparing with Equation (3.9), the optimal direction for the uncertain velocity based on Equation (B.13) to create a negative distance changing rate should be

$$\begin{aligned}\begin{bmatrix}\frac{\partial\dot{d}_{s,n,u}}{\partial v_{s,X,u}}\\ \frac{\partial\dot{d}_{s,n,u}}{\partial v_{s,Y,u}}\end{bmatrix} &= \begin{bmatrix}\frac{\partial\dot{d}'_{s,n}}{\partial v_{s,X,u}}\\ \frac{\partial\dot{d}'_{s,n}}{\partial v_{s,Y,u}}\end{bmatrix} = \frac{1}{d_{s,n}}\begin{bmatrix}\Delta\boldsymbol{p}_X\\ \Delta\boldsymbol{p}_Y\end{bmatrix} = \frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}\\ \begin{bmatrix}\frac{\partial\dot{d}_{s,n,u}}{\partial v_{n,X,u}}\\ \frac{\partial\dot{d}_{s,n,u}}{\partial v_{n,Y,u}}\end{bmatrix} &= \begin{bmatrix}\frac{\partial\dot{d}'_{s,n}}{\partial v_{n,X,u}}\\ \frac{\partial\dot{d}'_{s,n}}{\partial v_{n,Y,u}}\end{bmatrix} = \frac{1}{d_{s,n}}\begin{bmatrix}-\Delta\boldsymbol{p}_X\\ -\Delta\boldsymbol{p}_Y\end{bmatrix} = -\frac{\boldsymbol{p}_s-\boldsymbol{p}_n}{||\boldsymbol{p}_s-\boldsymbol{p}_n||}\end{aligned} \tag{B.15}$$

where $\Delta\boldsymbol{p}_X=\boldsymbol{p}_{s,X}-\boldsymbol{p}_{n,X}$, $\Delta\boldsymbol{p}_Y=\boldsymbol{p}_{s,Y}-\boldsymbol{p}_{n,Y}$.

Equation (B.15) means that the distance direction is the optimal for the uncertain velocity of the subject vehicle and the neighbouring vehicle to create a negative distance changing rate in Equation (B.14).

Simultaneously, the uncertain velocity should not cause extra relative bearing rate which makes the situation less dangerous. In other words, the uncertain velocity should follow the normal direction of the relative bearing rate gradient. Hence, the direction should be

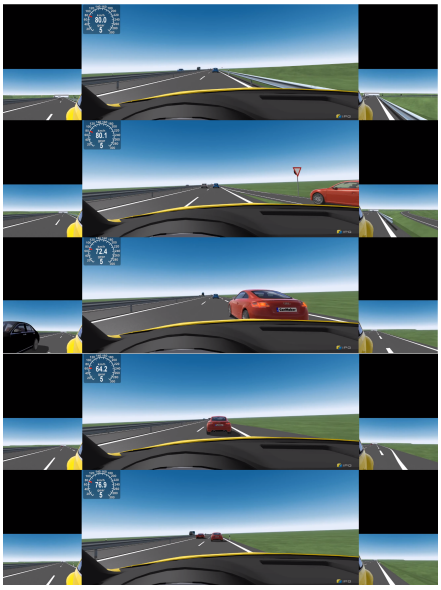
$$\begin{aligned}
\begin{bmatrix} \frac{\partial \dot{\theta}_{s,n,u}}{\partial v_{s,X,u}} \\ \frac{\partial \dot{\theta}_{s,n,u}}{\partial v_{s,Y,u}} \end{bmatrix} \Big|_{\perp} = \begin{bmatrix} \frac{\partial \dot{\theta}'_{s,n}}{\partial v_{s,X,u}} \\ \frac{\partial \dot{\theta}'_{s,n}}{\partial v_{s,Y,u}} \end{bmatrix} \Big|_{\perp} = \frac{1}{d_{s,n}^2} \begin{bmatrix} \Delta \boldsymbol{p}_Y \\ -\Delta \boldsymbol{p}_X \end{bmatrix} \Big|_{\perp} = \frac{\boldsymbol{p}_s - \boldsymbol{p}_n}{||\boldsymbol{p}_s - \boldsymbol{p}_n||^2} \\
\begin{bmatrix} \frac{\partial \dot{\theta}'_{s,n}}{\partial v_{n,X,u}} \\ \frac{\partial \dot{\theta}'_{s,n}}{\partial v_{n,Y,u}} \end{bmatrix} \Big|_{\perp} = \frac{1}{d_{s,n}^2} \begin{bmatrix} -\Delta \boldsymbol{p}_Y \\ \Delta \boldsymbol{p}_X \end{bmatrix} \Big|_{\perp} = -\frac{\boldsymbol{p}_s - \boldsymbol{p}_n}{||\boldsymbol{p}_s - \boldsymbol{p}_n||^2}
\end{aligned} \tag{B.16}$$

which are exactly the directions shown in Equation (B.15). That means the distance direction is the optimal direction for the uncertain velocity to create a negative distance changing rate and in the meantime not to cause less perceived risk.

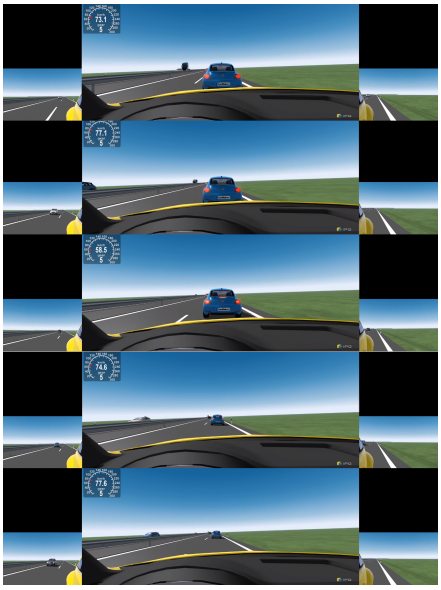
C

Appendix for Chapter 4

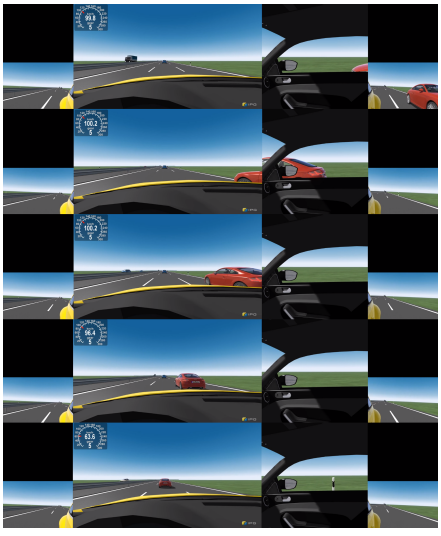
C.1. EXTENDED DATA



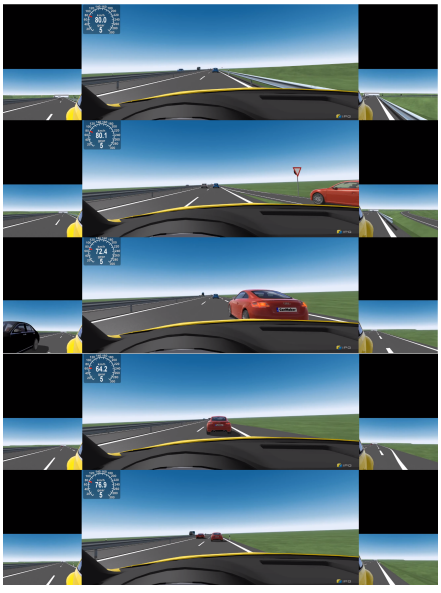
(a) Video stream of an MB event



(b) Video stream of an HB event



(c) Video stream of a LC event



(d) Video stream of an SVM event

Figure C.1: Video streams of MB, HB, LC and SVM events. Each image shows the last frame of a clip. After each clip the perceived risk is rated.

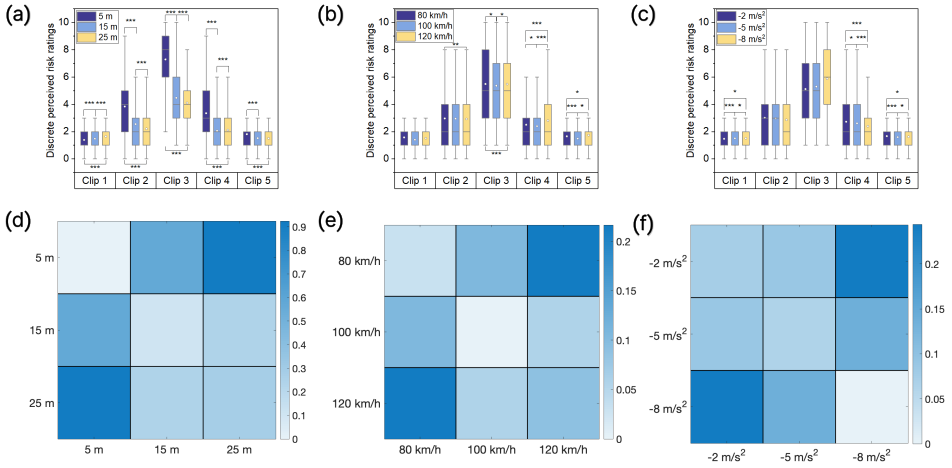


Figure C.2: Environment parameters influence in HB. **a**, Initial braking distance: Perceived risk ratings in 5 clips under condition A (5 m), B (15 m) and condition C (25 m). **b**, Cruising speed: Perceived risk ratings in 5 clips under condition A (80 km/h), B (100 km/h) and condition C (120 km/h). **c**, Braking Intensity of the leading vehicle: Perceived risk ratings in 5 clips under condition A ($-2\,m/s^2$), B ($-5\,m/s^2$) and condition C ($-8\,m/s^2$). **d**, J-S divergence of perceived risk ratings between different initial braking distance (5 m, 15 m and 25 m), **e**, J-S divergence of perceived risk ratings between different cruising speed (80 km/h, 100 km/h and 120 km/h), **e**, J-S divergence of perceived risk ratings between different cruising speed (80 km/h, 100 km/h and 120 km/h), **f**, J-S divergence of perceived risk ratings between different braking intensity of the leading vehicle ($-2\,m/s^2$, $-5\,m/s^2$ and $-8\,m/s^2$).

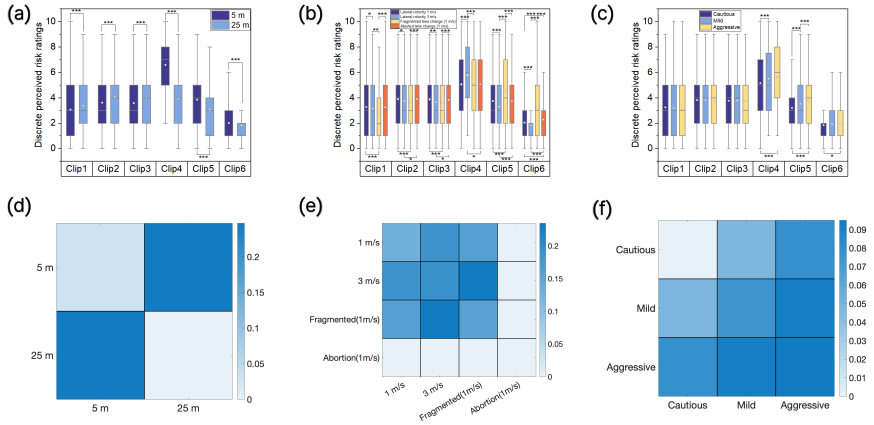


Figure C.3: Environment parameters influence in LC. **a**, Merging distance: Perceived risk ratings in 6 clips under condition A (5 m) and condition B (15 m). **b**, Lateral control: Perceived risk ratings in 6 clips under condition A (Lateral speed 1 m/s), B (Lateral speed 1 m/s), condition C (fragmented lane change with lateral speed 1 m/s) and condition D (aborted lane change with lateral speed Lateral speed 1 m/s). **c**, ACC category: Perceived risk ratings in 6 clips under condition A (Cautious ACC), B (Mild ACC) and C (Aggressive ACC). **d**, J-S divergence of perceived risk ratings between different merging distance (5 m and 15 m), **e**, J-S divergence of perceived risk ratings between different lateral behaviour of merging vehicle (normal lane change with lateral speed 1 m/s, normal lane change with lateral speed 3 m/s, fragmented lane change with lateral speed 1 m/s and aborted lane change with lateral speed 1 m/s). **f**, J-S divergence of perceived risk ratings between ACC categories (Cautious, mild and aggressive ACC).

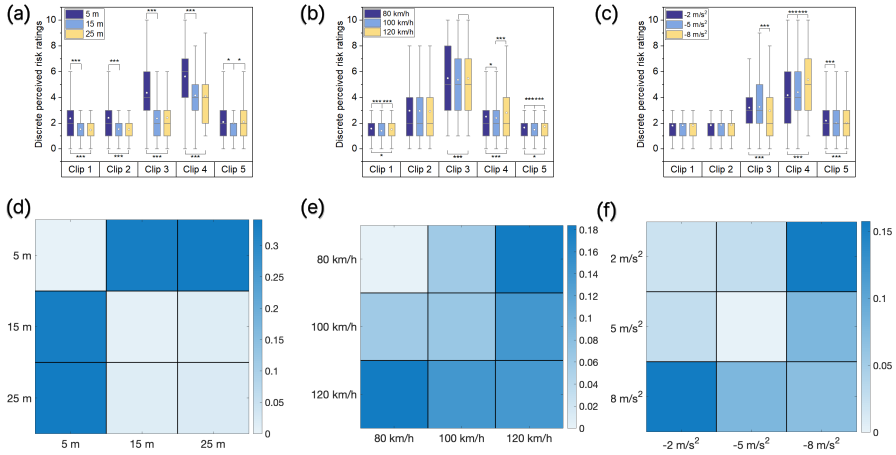


Figure C.4: Environment parameters influence in SVM. **a**, Initial merging distance: Perceived risk ratings in 5 clips under condition A (5 m), B (15 m) and condition C (25 m). **b**, Cruising speed: Perceived risk ratings in 5 clips under condition A (80 km/h), B (100 km/h) and condition C (120 km/h). **c**, Braking Intensity of the leading vehicle: Perceived risk ratings in 5 clips under condition A ($-2\,m/s^2$), B ($-5\,m/s^2$) and condition C ($-8\,m/s^2$). **d**, J-S divergence of perceived risk ratings between different initial merging distance (5 m, 15 m and 25 m), **e**, J-S divergence of perceived risk ratings between different cruising speed (80 km/h, 100 km/h and 120 km/h), **e**, J-S divergence of perceived risk ratings between different cruising speed (80 km/h, 100 km/h and 120 km/h), **f**, J-S divergence of perceived risk ratings between different braking intensity of the leading vehicle ($-2\,m/s^2$, $-5\,m/s^2$ and $-8\,m/s^2$).

Algorithm 1 Rating Moment Alignment Based on Empirical Data

procedure ALIGNRATINGMOMENTS(eventData)
    Initialise a matrix to store aligned rating moments
    for each event in eventsData **do**
        for each clip of the event **do**
            if the clip is the first clip **then**
                0 s is a rating moment of the first rating in the first clip
            else if the clip is the last clip **then**
                30 s or 36 s is a rating moment of the last rating in the last clip
            end if
            if there exists a stimulus (merging or braking) within this clip **then**
                Determine the onset time of the stimuli
                Determine the timing for perceived risk peak as the moment for the current perceived risk rating based on statistics
                Determine the timing for perceived risk returning to baseline as the rating moment for the next perceived risk rating
            else
                if there exists a determined timing within this clip **then**
                    Continue
                else
                    Determine the first and last moment of this clip as the rating moments for the current perceived risk rating.
                end if
            end if
            Store the aligned rating moment
        end for
    end for
    return the array of aligned rating moments
end procedure

Algorithm 2 Linear Interpolation Method

procedure LINEARINTERPOLATION(x, y, x_{new})
    Initialize an array y_{new} to store interpolated values
    for each x_i in x_{new} **do**
        Find the closest points x_a and x_b from x such that $x_a \leq x_i \leq x_b$
        Calculate the slope m between (x_a, y_a) and (x_b, y_b) from y
        Interpolate to find y_i using the equation $y_i = y_a + m \cdot (x_i - x_a)$
        Store y_i in y_{new}
    end for
    return y_{new}
end procedure

Algorithm 3 Quadratic Spline with Monotonicity Adjustment (Part 1)

```
1: procedure QUADRATICSPLINE(x_train, y_train)
2:     n ← length(x_train) − 1
3:     A ← zeros(2n, 3n)
4:     b ← zeros(2n, 1)
5:     coeffs ← array of 3n zeros
6:
7:     procedure BUILDSYSTEM(A, b, x_train, y_train)
8:         for i ← 1 to n do
9:             A(2i − 1, (3i − 2) : (3i)) ← [x_train(i)^2, x_train(i), 1]
10:            b(2i − 1) ← y_train(i)
11:            A(2i, (3i − 2) : (3i)) ← [x_train(i + 1)^2, x_train(i + 1), 1]
12:            b(2i) ← y_train(i + 1)
13:        end for
14:    end procedure
15:
16:    procedure APPLYMONOTONICITY(A, b, x_train, y_train)
17:        peaks, troughs ← IDENTIFYPEAKSTROUGHS(y_train)
18:        for all peak in peaks do
19:            Apply zero derivative constraint for segments adjacent to peak
20:        end for
21:        for all trough in troughs do
22:            if trough is not adjacent to peak then
23:                Apply zero derivative constraint for segments adjacent to trough
24:            end if
25:        end for
26:    end procedure
27:
28:    procedure SOLVESYSTEM(A, b)
29:        coeffs ← A\b
30:    end procedure
31: end procedure
```

C

Algorithm 3 Quadratic Spline with Monotonicity Adjustment (Part 2)

Continued from Part 1
procedure QUADRATICSPLINE CONTINUED()
    procedure EVALUATESPLINE($coeffs$, x_eval)
        $y\_spline \leftarrow$ array of size(x_eval) zeros
        for $i \leftarrow 1$ to n **do**
            $indices \leftarrow \text{find}(x\_eval \geq x\_train(i) \wedge x\_eval < x\_train(i+1))$
            $y\_spline(indices) \leftarrow coeffs(3i-2) \cdot x\_eval(indices)^2 + coeffs(3i-1) \cdot x\_eval(indices) + coeffs(3i)$
        end for
        return y_spline
    end procedure

    BUILDSYSTEM(A, b, x_train, y_train)
    APPLYMONOTONICITY(A, b, x_train, y_train)
    SOLVESYSTEM(A, b)
    $y\_quad \leftarrow$ EVALUATESPLINE($coeffs, x\_dense$)

    return y_quad
end procedure

Algorithm 4 PCHIP Improved Cubic Spline Interpolation (Part 1)

procedure PCHIPNEW(x, y, xq)
    (x, y, sizey) $\leftarrow$ CHECKANDADJUSTINPUT(x, y)
    $h \leftarrow$ COMPUTEINTERVALS(x)
    $\Delta \leftarrow$ COMPUTESLOPES(y, h)
    slopes $\leftarrow$ COMPUTEMODIFIEDPCHIPSLOPES(h, Δ)
    $yq \leftarrow$ COMPUTEPWCH(x, y, slopes, h, Δ)
    yq.dim $\leftarrow$ sizey
    if nargin = 3 **then**
        $yq \leftarrow$ EVALUATEPCHIP(yq, xq)
    end if
    return yq
end procedure

Algorithm 4 PCHIP Improved Cubic Spline Interpolation (Part 2)

Continued from Part 1
procedure CHECKANDADJUSTINPUT(x, y)
    Ensure x and y are of compatible sizes
    Remove NaNs from x and y
    Sort x and reorder y correspondingly
    Determine end conditions based on y's size
    return (x, y, sizey, endslopes)
end procedure

procedure COMPUTEINTERVALS(x)
    $h \leftarrow \text{diff}(x)$
    return h
end procedure

procedure COMPUTESLOPES(y, h)
    $\Delta \leftarrow \text{diff}(y)/\text{repmat}(h, \text{size}(y, 1), 1)$
    return Δ
end procedure

procedure COMPUTEMODIFIEDPCHIPSLOPES(h, Δ)
    $d \leftarrow \text{InitializeSlopes}(\text{size}(h))$
    $d \leftarrow \text{AssignZeroSlopes}(d)$
    $d \leftarrow \text{ComputeNonzeroSlopes}(h, \Delta, d)$
    return d
end procedure

procedure COMPUTEPWCH(x, y, slopes, h, Δ)
    Perform piecewise cubic Hermite interpolation
    return yq
end procedure

procedure EVALUATEPCHIP(yq, xq)
    Evaluate the piecewise cubic Hermite interpolant at xq
    return yq
end procedure

Table C.1: Rating moments in MB Scenario (in seconds). Within R1, R2 and R5, the rating for the second moment was copied.

Event number	R1		R2		R3	R4	R5	
1	0.00	5.90	11.00	13.90	16.85	20.78	24.00	30.00
2	0.00	5.90	10.40	13.40	16.25	20.18	24.00	30.00
3	0.00	5.90	10.20	13.00	15.55	19.48	24.00	30.00
4	0.00	5.90	11.20	13.80	16.95	20.88	24.00	30.00
5	0.00	5.90	11.00	13.30	16.05	19.98	24.00	30.00
6	0.00	5.90	10.40	12.80	16.75	20.68	24.00	30.00
7	0.00	5.90	11.00	13.80	17.25	21.18	24.00	30.00
8	0.00	5.90	11.00	13.30	16.65	20.58	24.00	30.00
9	0.00	5.90	10.10	12.90	16.25	20.18	24.00	30.00
10	0.00	5.90	10.80	12.90	15.55	19.48	24.00	30.00
11	0.00	5.90	10.30	12.40	14.85	18.78	24.00	30.00
12	0.00	5.90	9.90	12.20	14.55	18.48	24.00	30.00
13	0.00	5.90	10.70	12.90	15.65	19.58	24.00	30.00
14	0.00	5.90	10.50	12.50	14.95	18.88	24.00	30.00
15	0.00	5.90	10.20	12.10	16.65	20.58	24.00	30.00
16	0.00	5.90	10.80	12.90	15.95	19.88	24.00	30.00
17	0.00	5.90	10.40	12.40	16.25	20.18	24.00	30.00
18	0.00	5.90	10.20	12.10	15.95	19.88	24.00	30.00
19	0.00	5.90	12.20	14.20	16.55	20.48	24.00	30.00
20	0.00	5.90	12.00	13.80	16.05	19.98	24.00	30.00
21	0.00	5.90	12.00	13.60	15.75	19.68	24.00	30.00
22	0.00	5.90	12.50	14.20	16.55	20.48	24.00	30.00
23	0.00	5.90	12.00	13.80	16.05	19.98	24.00	30.00
24	0.00	5.90	12.00	13.60	15.75	19.68	24.00	30.00
25	0.00	5.90	12.60	14.20	16.85	20.78	24.00	30.00
26	0.00	5.90	12.20	13.80	16.05	19.98	24.00	30.00
27	0.00	5.90	12.00	13.60	17.95	21.88	24.00	30.00

Table C.2: Rating moments in HB Scenario (in seconds). Within R1, R2 and R5, the rating for the second moment was copied.

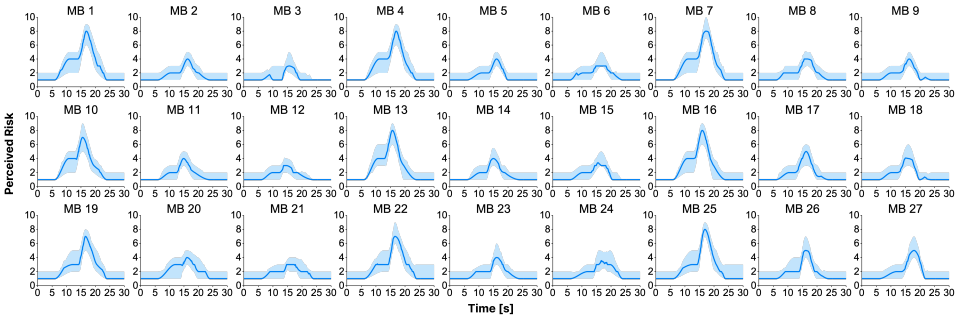
Event number	R1		R2		R3	R4	R5	
1	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
2	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
3	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
4	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
5	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
6	0.00	5.90	9.00	14.70	15.85	19.78	24.00	30.00
7	0.00	5.90	9.00	14.60	15.25	19.18	24.00	30.00
8	0.00	5.90	9.00	14.80	15.95	19.88	24.00	30.00
9	0.00	5.90	9.00	14.90	15.25	19.18	24.00	30.00
10	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
11	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
12	0.00	5.90	9.00	14.10	15.25	19.18	24.00	30.00
13	0.00	5.90	9.00	14.30	15.25	19.18	24.00	30.00
14	0.00	5.90	9.00	14.30	15.45	19.38	24.00	30.00
15	0.00	5.90	9.00	14.30	15.25	19.18	24.00	30.00
16	0.00	5.90	9.00	14.60	15.75	19.68	24.00	30.00
17	0.00	5.90	9.00	14.40	15.25	19.18	24.00	30.00
18	0.00	5.90	9.00	14.30	15.45	19.38	24.00	30.00
19	0.00	5.90	9.00	14.30	15.25	19.18	24.00	30.00
20	0.00	5.90	9.00	14.30	15.45	19.38	24.00	30.00
21	0.00	5.90	9.00	14.20	15.25	19.18	24.00	30.00
22	0.00	5.90	9.00	14.30	15.45	19.38	24.00	30.00
23	0.00	5.90	9.00	14.40	15.25	19.18	24.00	30.00
24	0.00	5.90	9.00	14.30	15.45	19.38	24.00	30.00
25	0.00	5.90	9.00	14.40	15.25	19.18	24.00	30.00
26	0.00	5.90	9.00	14.40	15.55	19.48	24.00	30.00
27	0.00	5.90	9.00	14.40	15.25	19.18	24.00	30.00

Table C.3: Rating moments in LC Scenario (in seconds). Within R1, R5 and R6, the rating for the second moment was copied.

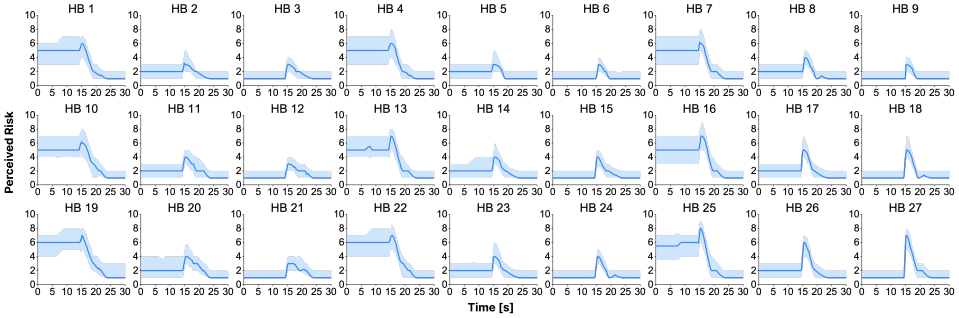
Event number	R1		R2	R3	R4	R5		R6	
1	0.00	5.90	8.00	14.40	21.55	25.48		30.00	36.00
2	0.00	5.90	8.00	14.40	21.80	25.73		30.00	36.00
3	0.00	5.90	8.00	14.40	21.35	25.28		30.00	36.00
4	0.00	5.90	8.00	14.40	23.05	26.98		30.00	36.00
5	0.00	5.90	8.00	14.40	22.55	26.48		30.00	36.00
6	0.00	5.90	8.00	14.40	22.95	25.95		30.00	36.00
7	0.00	5.90	8.00	14.40	19.95	23.88		30.00	36.00
8	0.00	5.90	8.00	14.40	21.60	24.55		30.00	36.00
9	0.00	5.90	8.00	14.40	20.35	24.28		30.00	36.00
10	0.00	5.90	8.00	14.40	23.55	27.48		30.00	36.00
11	0.00	5.90	8.00	14.40	19.65	23.58		30.00	36.00
12	0.00	5.90	8.00	14.40	22.95	26.88		30.00	36.00
13	0.00	5.90	8.00	14.40	19.85	27.25		31.18	36.00
14	0.00	5.90	8.00	14.40	20.05	27.25		31.18	36.00
15	0.00	5.90	8.00	14.40	20.05	27.15		31.08	36.00
16	0.00	5.90	8.00	14.40	22.95	26.88		32.25	36.00
17	0.00	5.90	8.00	14.40	19.65	28.75		32.68	36.00
18	0.00	5.90	8.00	14.40	22.95	26.88		32.25	36.00
19	0.00	5.90	8.00	14.40	22.75	26.68	29.30	33.23	36.00
20	0.00	5.90	8.00	14.40	19.75	23.68	26.20	30.13	36.00
21	0.00	5.90	8.00	14.40	23.05	26.98	29.30	33.23	36.00
22	0.00	5.90	8.00	14.40	19.75	23.68	26.20	30.13	36.00
23	0.00	5.90	8.00	14.40	22.95	26.88	29.30	33.23	36.00
24	0.00	5.90	8.00	14.40	19.75	23.68	26.20	30.13	36.00

Table C.4: Rating moments in SVM Scenario (in seconds). Within R1, R2 and R5, the rating for the second moment was copied.

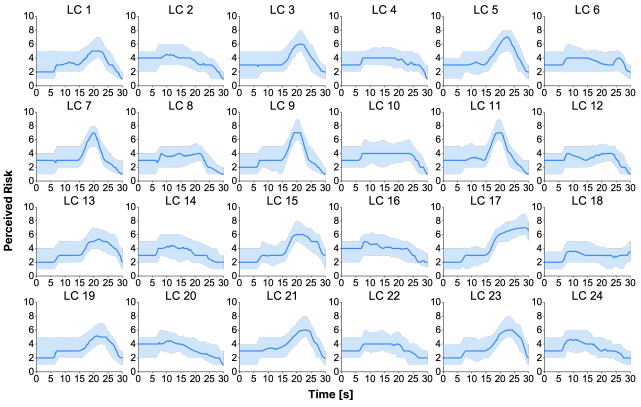
Event number	R1		R2		R3	R4	R5	
1	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
2	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
3	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
4	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
5	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
6	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
7	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
8	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
9	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
10	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
11	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
12	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
13	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
14	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
15	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
16	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
17	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
18	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
19	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
20	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
21	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
22	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
23	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
24	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
25	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
26	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00
27	0.00	5.90	12.00	12.70	16.20	19.15	24.00	30.00



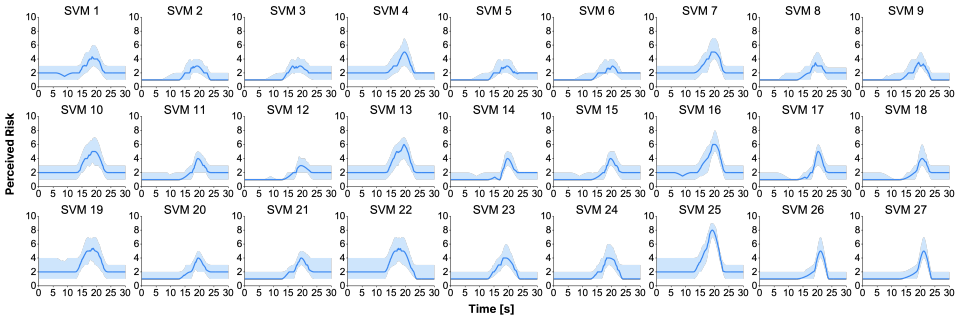
(a) Continuous perceived risk in MB scenarios.



(b) Continuous perceived risk in HB scenarios.



(c) Continuous perceived risk in LC scenarios.



(d) Continuous perceived risk in SVM scenarios.

Figure C.5: Continuous perceived risk in four scenarios. The solid blue curve represents the mean of perceived risk and the light blue areas represent the standard deviations.

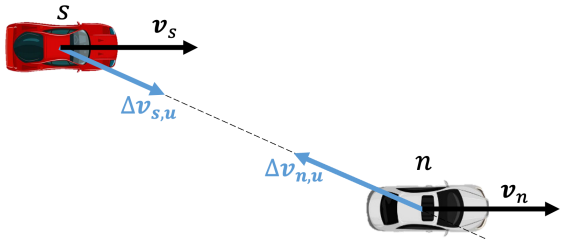
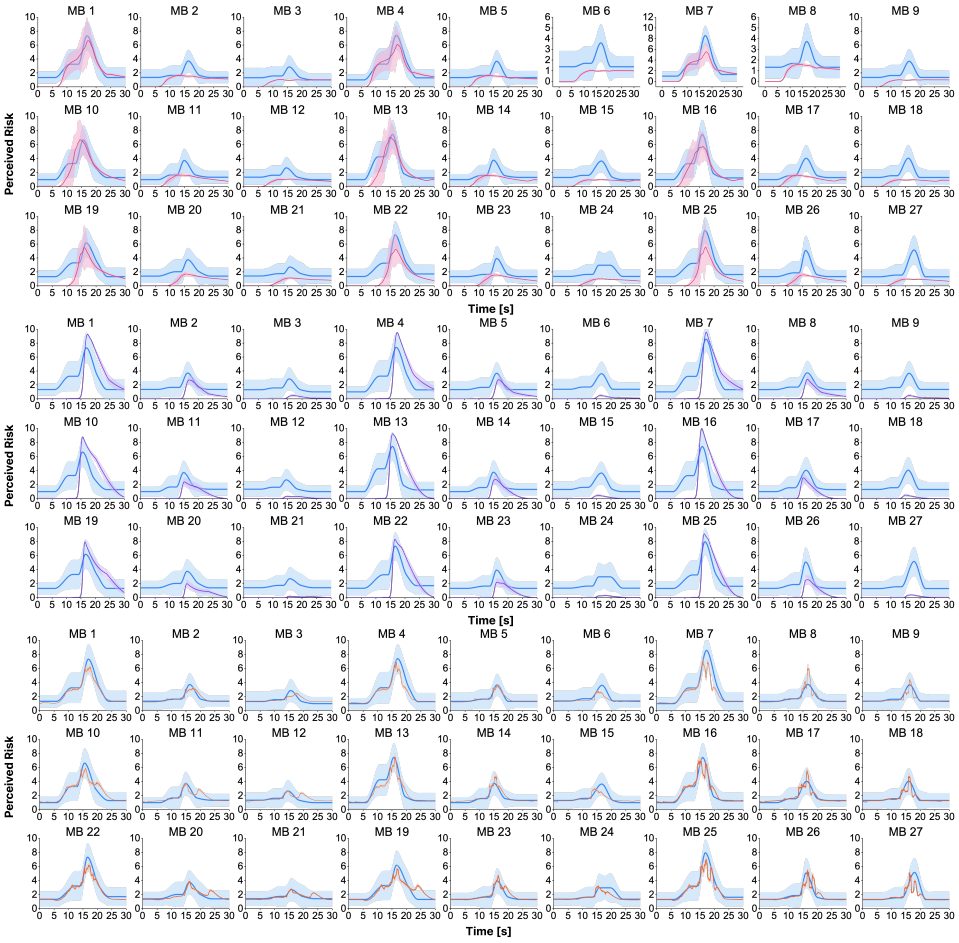


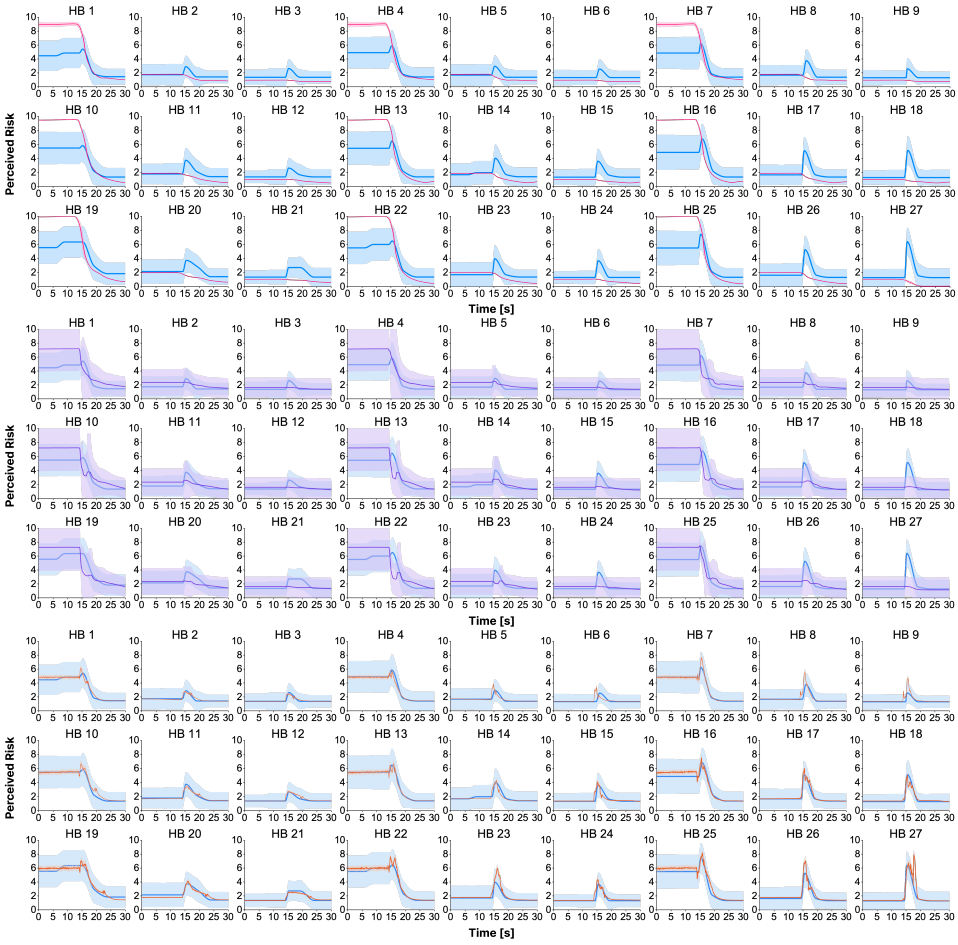
Figure C.6: The uncertain velocity $\Delta \boldsymbol{v}_{s,u}$ and $\Delta \boldsymbol{v}_{n,u}$ of the subject vehicle s and the neighbouring vehicle n. In this case, the subject vehicle (red) is passing by a neighbouring vehicle (white). The uncertain velocities $\Delta \boldsymbol{v}_{s,u}$ and $\Delta \boldsymbol{v}_{n,u}$ are pointing to each other.

Table C.5: Parameters to be calibrated of computational perceived risk models

Model	Parameters	Explanation
PCAD	$\sigma_{n,X}$	The standard deviation in X of the velocity Gaussian of a neighbouring vehicle
	$\sigma_{n,Y}$	The standard deviation in Y of the velocity Gaussian of a neighbouring vehicle
	$\sigma_{s,X}$	The standard deviation in X of the velocity Gaussian of the subject vehicle
	$\sigma_{s,Y}$	The standard deviation in Y of the velocity Gaussian of the subject vehicle
	$t_{s,a}$	The accumulation time for the acceleration-based velocity of the subject vehicle
	$t_{n,a}$	The accumulation time for the acceleration-based velocity of a neighbouring vehicle
	α	The exponent of the power function in weighting function
DRF	D	The steepness of descent of the potential field
	s	The steepness of the height parabola of the risk field
	t_{la}	Human driver's preview time
	m	The rate of the risk field width expanding
	c	The initial width of the DRF

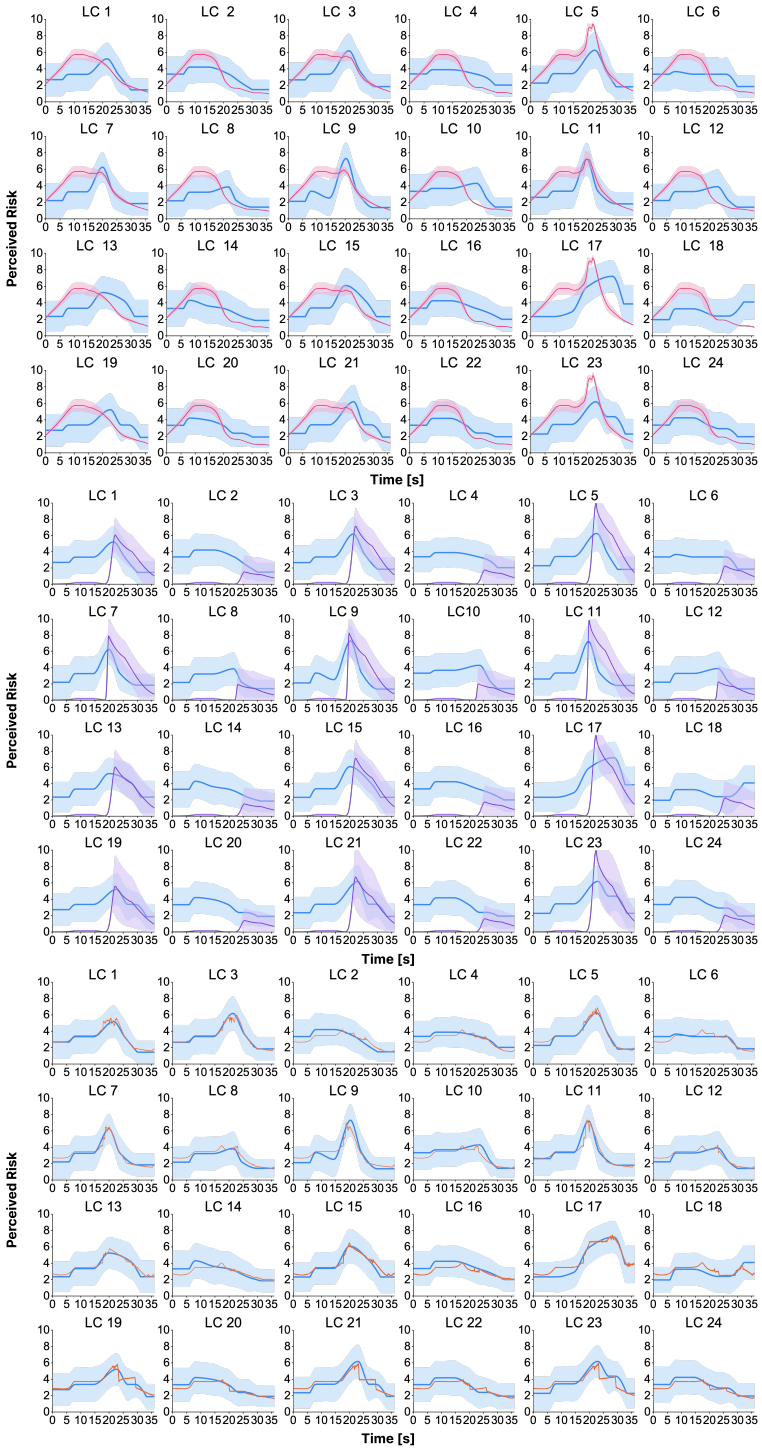


(a) Prediction results of PCAD, DRF and neural network from the top to bottom in MB scenario. Solid curves represent the mean of perceived risk and model output and the light coloured area represent the epistemic uncertainty of the model.

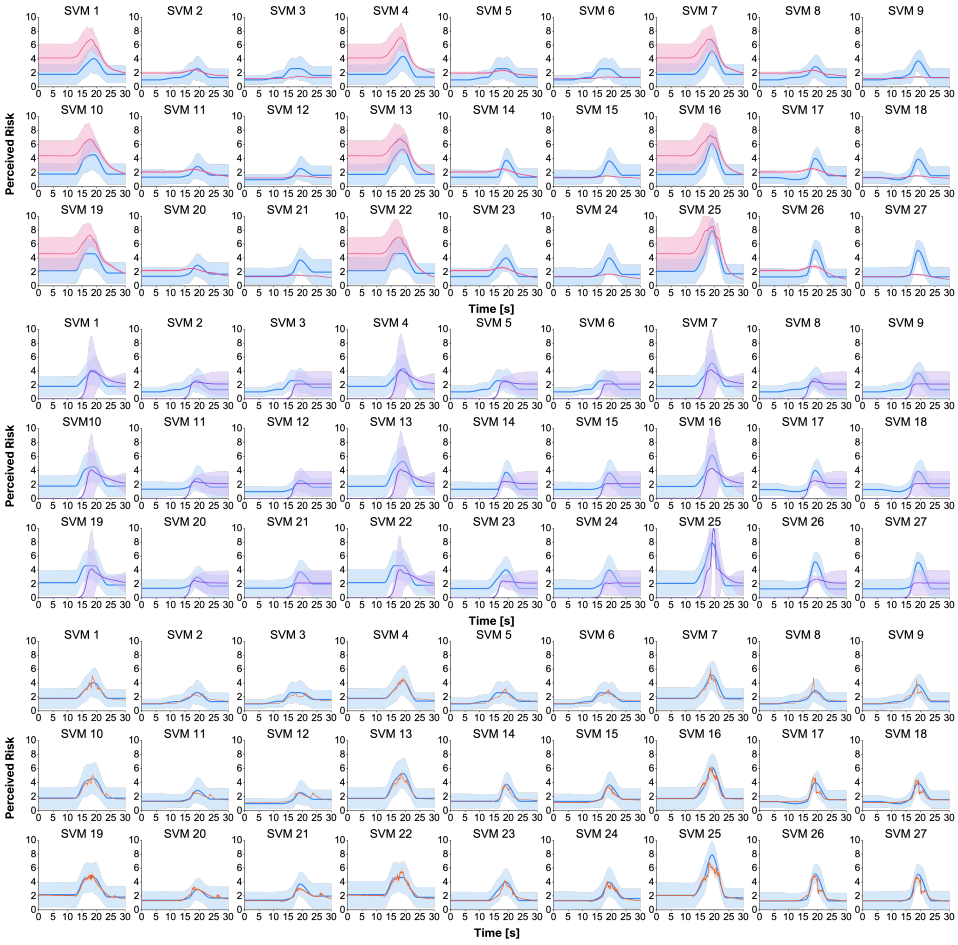


(b) Prediction results of PCAD, DRF and neural network from the top to bottom in HB scenario. Solid curves represent the mean of perceived risk and model output and the light coloured area represent the epistemic uncertainty of the model.

C



(c) Prediction results of PCAD, DRF and neural network from the top to bottom in LC scenario. Solid curves represent the mean of perceived risk and model output and the light coloured area represent the epistemic uncertainty of the model.



(d) Prediction results of PCAD, DRF and neural network from the top to bottom in SVM scenario. Solid curves represent the mean of perceived risk and model output and the light coloured area represent the epistemic uncertainty of the model.

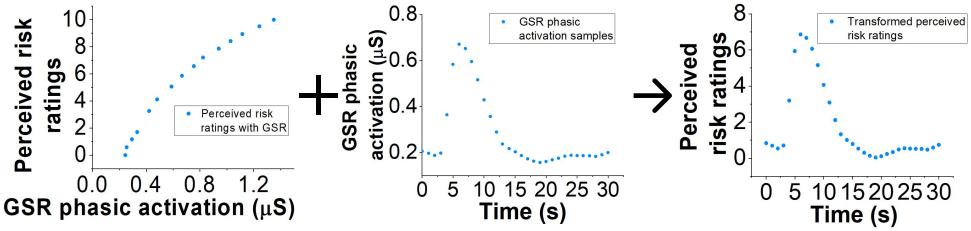
Figure C.7: Prediction results of PCAD, DRF and neural network in four scenarios.

C.2. THE SELECTION OF THE INTERPOLATION METHOD

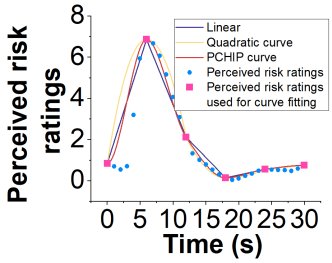
This section clarifies the selection of the method to interpolate discrete perceived risk ratings. The interpolation methods were selected to satisfy certain criteria: traversing all known data points, preserving monotonicity between adjacent samples, and ensuring that the first derivative at the first, last, and peak points is zero, where applicable. Exempt from this derivative condition is the Linear Interpolation method. We checked three interpolation strategies to learn the most suitable interpolation method to obtain continuous perceived risk ratings: Linear Interpolation, Quadratic Spline Interpolation with Monotonicity Adjustment, and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) (Algorithms 2-4 in Appendix C.1).

We used the data from two pivotal studies: Ref. [153] that clarified the relationship between time and GSR in response to stimuli, and Ref. [154] that explained the correlation between GSR and perceived risk ratings. Based on this, the relationship between time and perceived risk faced with a stimulus was derived, represented in Figure C.8a. To establish time-continuous perceived risk ratings, we strategically selected the 1st, 7th, 13th, 19th, 25th, and 31st data points from a total of 31 samples of a time-continuous perceived risk signal for interpolation using linear, quadratic, and PCHIP methodologies. The accuracy of these interpolations was evaluated by calculating the deviation of the remaining samples from the interpolated curve, thereby determining the most accurate method, which is shown in Figure C.8b.

The comparison focused on their effectiveness in capturing the temporal change in perceived risk. The analysis revealed differences in how each method interpolated the discrete data, providing insights into the most accurate and representative methodologies for continuous risk evaluation. The *RMSE* between perceived risk ratings and the fitted perceived risk ratings for Linear Interpolation, Quadratic Spline Interpolation with Monotonicity Adjustment and Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) are 0.94, 1.42, and 0.88 separately. As graphically represented in Figure C.8b, these interpolations show how each method captures the temporal change in perceived risk ratings. Specifically, PCHIP offers the closest fit to the trajectory of perceived risk ratings, indicated by the lowest *RMSE*.



(a). Deriving the relationship between perceived risk ratings and time by combining the relationship between GSR phasic activation (μS) and perceived risk ratings (first plot) with the GSR temporal response (second plot), resulting in the derived temporal profile of perceived risk (third plot).



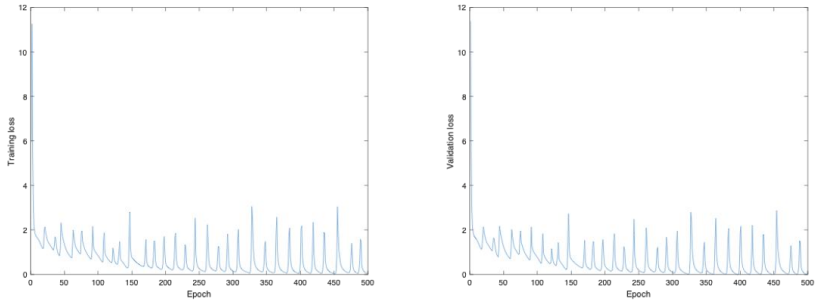
(b) Interpolating time continuous perceived risk signal. Different fitting methods were used to interpolate these discrete perceived risk ratings, where the PCHIP curve provided the best fit for generating a continuous signal over time.

Figure C.8: Deriving a time-continuous perceived risk signal from discrete perceived risk ratings.

C.3. DNN OVERFITTING EVALUATION

In this evaluation, we firstly check the training and validation loss curves to confirm if there are any significant signals of overfitting. Taking MB scenario as the example, Figure C.9 indicates that during the training the loss on the validation data is decreasing along with the loss on the training data which is not a signal of overfitting.

Additionally, we trained a simplified DNN with fewer neurons (referred as DNN B, with $21 \times 50 \times 2$ neurons) and compared the original DNN (referred as DNN A, with $21 \times 500 \times 2$ neurons) with the simplified version in specific scenario types. Both types of models were trained and validated under the same conditions to assess overfitting risks and robustness against variations due to random factors during training, such as neuron dropout. For both types of DNNs, a learning rate of 0.001 and a dropout rate of 50% were employed, and we randomly used 80% of the data for training with the remaining 20% for validation. Training was repeated 10 times to evaluate the impact of training randomness on model performance. Table C.6 shows the loss on the 80% training data and loss on the 20% validation data. The original DNN A has smaller loss on both training and validation data with less deviation across 10 times of training compared to DNN B with less neurons, which is also not a signal of overfitting.



(a) Loss on training data

(b) Loss on validation data

Figure C.9: Training and validation loss curves.

Table C.6: Training and validation loss (with RMSE mean (std) in each cell)

	Loss (RMSE) on 80% training data	Loss (RMSE) on 20% validation data
DNN A	0.6724 (0.045)	0.6815 (0.037)
DNN B	0.9748 (0.052)	0.9837 (0.061)

C.4. SUPPLEMENTARY MATERIALS

C.4.1. ONLINE QUESTIONNAIRE

The online questionnaire used in the study can be accessed at Qualtrics.

C.4.2. LOCAL FEATURE CONTRIBUTIONS TO PERCEIVED RISK OVER TIME

The local feature contributions to perceived risk over time in all 105 events can be accessed at https://surfdrive.surf.nl/files/index.php/s/N3HQxZ4ocsXBuzV.

D

APPENDIX FOR CHAPTER 5

D.1. PRELIMINARY EXPERIMENT

D.1.1. METHOD

EXPERIMENT CONDITIONS

The five interface conditions were the same as in Main Experiment, but the experimental design was between-subjects; each participant experienced one of five user interfaces.

SCENARIO

In Table D.1, the scenarios of the preliminary experiment are described, where seven different event types were repeated twice, and the order was randomised but fixed.

Table D.1: Event types of the preliminary experiment

Manoeuvre	**Merging gap (m)**	**Automation action**	**Event**	
			Exposure 1	**Exposure 2 (Repetition)**
Braking	5	-2 m/s^2 deceleration	E1a	E1b
		-8 m/s^2 deceleration	E2a	E2b
	15	-2 m/s^2 deceleration	E3a	E3b
		-8 m/s^2 deceleration	E4a	E4b
	25	-2 m/s^2 deceleration	E5a	E5b
		-8 m/s^2 deceleration	E6a	E6b
Lane Change	25	Lane change	E7a	E7b

APPARATUS

The apparatus was the same as in the main text experiment, but there were no side mirrors on the simulator. To prevent feeling unsafe, drivers were informed that there were no other vehicles on the left lane when changing the lane as a manoeuvre [256], [257].

MEASUREMENTS

During the experiment, brake pedal signal and eye gaze behaviours were recorded. In addition, trust and perceived risk questions using a 10-point Likert scale [117] were collected on a tablet on the left side of the steering wheel in English.

PARTICIPANTS

In each UI group, seventeen drivers participated, with a total of eighty-five participants (twenty-one females) holding a driving license for more than one year. The average age of participants was 30.7 years (SD= 13.1 years). Twenty-four participants had experienced adaptive cruise control (ACC), seventeen lane keeping assist (LKA), and eleven combination of ACC and LKA.

D

D.1.2. RESULTS

BRAKING BEHAVIOUR

Participants' brake pedal usage of each interface condition is shown in Figure D.1. In the No UI condition, most participants used the brake pedal, and the least number of participants used the brake pedal in the SM-VA UI condition. Participants who used the brake pedal all used it in the most critical events EI2a and EI2b, slowing down with $-8\,\mathrm{m/s^2}$ deceleration and 5 m merging gap.

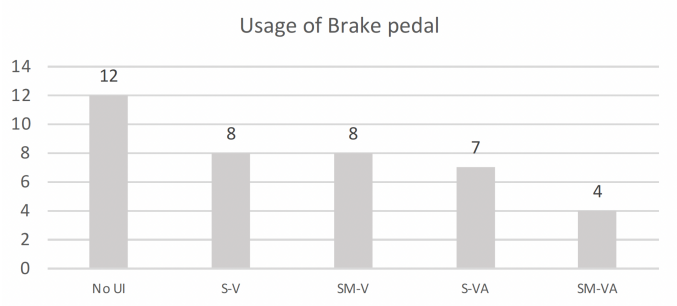


Figure D.1: Number of participants who used brake pedal on each user interface

TRUST AND PERCEIVED RISK

Average trust and perceived risk for each user interface are shown in Figure D.2. There was an effect of UI on trust, but no effect on perceived risk. We analysed the effect of UI on trust and perceived risk using repeated-measure ANOVA. Participants' trust and perceived risk on each UI were used as the dependent variable and the interface condition as an independent factor. The main effect of user interface conditions on trust was significant ($F(4,84) = 4.23 p = 0.004, \eta^2 = 0.18$). The post-doc analysis indicated that No UI condition is lower than SM-V, S-VA, and SM-VA. Participants had the highest perceived risk in No UI, but there was no significant effect of user interface condition on perceived risk ($F(4,84) = 1.48 p = 0.217, \eta^2 = 0.07$).

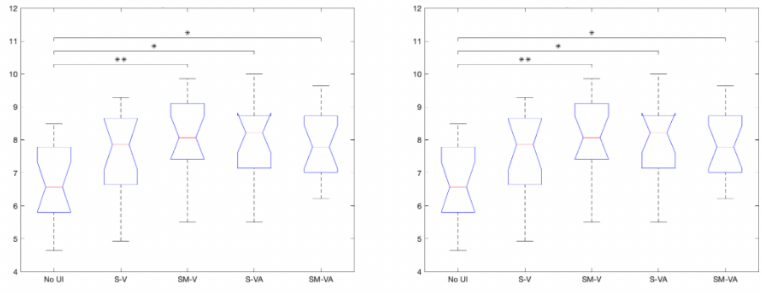


Figure D.2: Trust (Left) and Perceived Risk (Right) in each user interface condition ($*p < 0.05, **p < 0.01, ***p < 0.001$)

EYE GAZE BEHAVIOUR

Participants gazed almost (i.e., more than 90% of the simulator drive duration to the road and the UI display area. Hence, it is reasonable only to analyse the gaze fixation duration and the gaze transitions between these two areas.

Eye-tracking results did not show a significant effect of UI conditions on the fixation duration towards the road ($F(4,72) = 1.46, p = 0.225$) and the gaze transition frequency between the road and the display ($F(4,72) = 1.47, p = 0.221$), but the gaze fixation duration towards the display area with S-V UI is significantly higher than that with No UI ($F(4,72) = 3.32, p = 0.015$). However, we cannot conclude that the surrounding information via visual UI can attract drivers' attention because of the between-subject experiment design, which will be discussed later. In general, the average fixation duration towards the display and the average transition frequency are higher in all UI conditions than No UI conditions, as shown in Figure D.3.

We merged the two exposures of specific event types to analyse their effects. In most of the cases, event types did not significantly influence the fixation duration towards the road, the display area and the transition number between the road and the display. However, the 50 percentile of the boxes in Figure D.4 generally indicates that drivers concentrated more on display in non-critical events with a $-2\,\mathrm{m/s^2}$ deceleration but focused more on the road in critical events with a $-8\,\mathrm{m/s^2}$ deceleration.

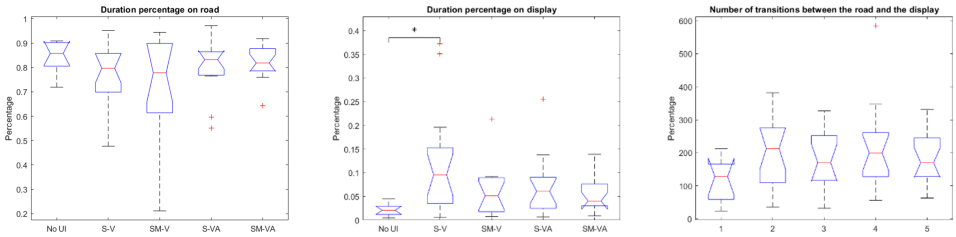


Figure D.3: Left: Fixation duration ratio on the road; Middle: fixation duration ratio on display ($*p < 0.05$); Right: Transition numbers of fixations between the road and the display.

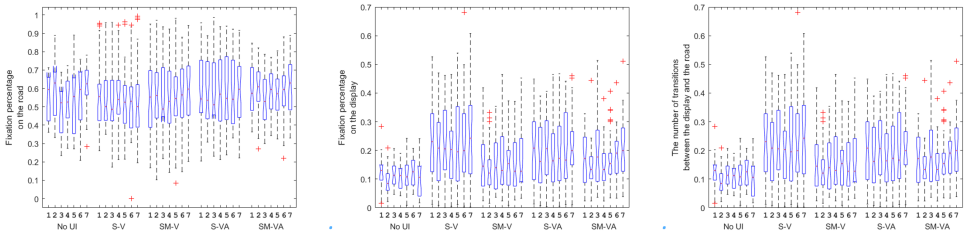


Figure D.4: Left: Fixation duration ratio on the road (per event); Middle: fixation duration ratio on display (per event); Right: Transition numbers of fixation between the road and the display (per event). Numbers 1 7 represent EI1 EI7, including both exposures.

D

D.2. SUPPLEMENTARY AUDIO

- Supplementary audio 1
  https://ars.els-cdn.com/content/image/1-s2.0-S1369847824000305-mmc1.mp3
- Supplementary audio 2
  https://ars.els-cdn.com/content/image/1-s2.0-S1369847824000305-mmc2.mp3
- Supplementary audio 3
  https://ars.els-cdn.com/content/image/1-s2.0-S1369847824000305-mmc3.mp3

BIBLIOGRAPHY

[1] ERTRAC, *Automated driving roadmap*, 2017.

[2] Waymo, "Waymo safety report," Waymo, Tech. Rep., Mar. 2021. [Online]. Available: https://downloads.ctfassets.net/sv23gofxcuiz/4gZ7ZUxd4SRj1D1W6z3rpR/2ea16814cdb42f9e8eb34cae4f30b35d/2021-03-waymo-safety-report.pdf.

[3] NPR. "Waymo's robotaxi service set to expand into Los Angeles." (Mar. 2024), [Online]. Available: https://www.npr.org/2024/03/14/1238489046/waymo-robotaxi-los-angeles.

[4] BBC. "Driverless bus coming to Belfast Harbour Estate." (Feb. 2023), [Online]. Available: https://www.bbc.com/news/uk-northern-ireland-64468998.

[5] FT. "California clears way for driverless taxis in San Francisco." (2023), [Online]. Available: https://www.ft.com/content/a8de16bd-ffd0-41a8-b4af-f9b69fb4e304.

[6] BBC. "Cruise self-driving cars investigated after two accidents." (Oct. 2023), [Online]. Available: https://www.bbc.com/news/technology-67133409.

[7] CNN. "A woman was found trapped under a driverless car. It wasn't the first car to hit her." (Oct. 2023), [Online]. Available: https://edition.cnn.com/2023/10/03/tech/driverless-car-pedestrian-injury/index.html.

[8] CNN. "GM self-driving car subsidiary withheld video of a crash, California DMV says." (Oct. 2023), [Online]. Available: https://edition.cnn.com/2023/10/24/business/california-dmv-cruise-permit-revoke/index.html.

[9] Z. Xu, K. Zhang, H. Min, Z. Wang, X. Zhao, and P. Liu, "What drives people to accept automated vehicles? Findings from a field experiment," *Transportation research part C: emerging technologies*, vol. 95, pp. 320–334, 2018. DOI: 10.1016/j.trc.2018.07.024.

[10] I. Nastjuk, B. Herrenkind, M. Marrone, A. B. Brendel, and L. M. Kolbe, "What drives the acceptance of autonomous driving? An investigation of acceptance factors from an end-user's perspective," *Technological Forecasting and Social Change*, vol. 161, p. 120 319, 2020.

[11] S. Nordhoff, J. Stapel, X. He, A. Gentner, and R. Happee, "Perceived safety and trust in SAE Level 2 partially automated cars: Results from an online questionnaire," *Plos one*, vol. 16, no. 12, 2021. DOI: 10.1371/journal.pone.0260953.

[12] J. H. Lee and C. H. Song, "Effects of trust and perceived risk on user acceptance of a new technology service," *Social Behavior and Personality: an international journal*, vol. 41, no. 4, pp. 587–597, 2013.

[13] T. Zhang, D. Tao, X. Qu, X. Zhang, R. Lin, and W. Zhang, "The roles of initial trust and perceived risk in public's acceptance of automated vehicles," *Transportation research part C: Emerging Technologies*, vol. 98, pp. 207–220, 2019.

[14] M. T. Pascale, D. Rodwell, P. Coughlan, S.-A. Kaye, S. Demmel, S. G. Dehkordi, A. Bond, I. Lewis, A. Rakotonirainy, and S. Glaser, "Passengers' acceptance and perceptions of risk while riding in an automated vehicle on open, public roads," *Transportation research part F: traffic psychology and behaviour*, vol. 83, pp. 274–290, 2021.

[15] ISO, *ISO 26262 Road vehicles- functional safety- part 1: vocabulary*, 2018.

[16] S. Nordhoff, V. Malmsten, B. van Arem, P. Liu, and R. Happee, "A structural equation modeling approach for the acceptance of driverless automated shuttles based on constructs from the unified theory of acceptance and use of technology and the diffusion of innovation theory," *Transportation research part F: traffic psychology and behaviour*, vol. 78, pp. 58–73, 2021. DOI: 10.1016/j.trf.2021.01.001.

[17] I. Y. Noy, D. Shinar, and W. J. Horrey, "Automated driving: Safety blind spots," *Safety science*, vol. 102, pp. 68–78, 2018. DOI: 10.1016/j.ssci.2017.07.018.

[18] S. Motamedi, P. Wang, T. Zhang, and C.-Y. Chan, "Acceptance of full driving automation: Personally owned and shared-use concepts," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 62, no. 2, pp. 288–309, Mar. 2020. DOI: 10.1177/0018720819870658.

[19] C. Peng, S. Horn, R. Madigan, C. Marberger, J. D. Lee, J. Krems, M. Beggiato, R. Romano, C. Wei, E. Wooldridge, R. Happee, M. Hagenzieker, and N. Merat, "Conceptualising user comfort in automated driving: Findings from an expert group workshop," *Transportation Research Interdisciplinary Perspectives*, vol. 24, p. 101 070, Mar. 2024. DOI: 10.1016/j.trip.2024.101070.

[20] J. C. Zoellick, A. Kuhlmey, L. Schenk, D. Schindel, and S. Blüher, "Amused, accepted, and used? Attitudes and emotions towards automated vehicles, their relationships, and predictive value for usage intention," *Transportation research part F: traffic psychology and behaviour*, vol. 65, pp. 68–78, 2019. DOI: 10.1016/j.trf.2019.07.009.

[21] J. K. Choi and Y. G. Ji, "Investigating the importance of trust on adopting an autonomous vehicle," *International Journal of Human-Computer Interaction*, vol. 31, no. 10, pp. 692–702, 2015. DOI: 10.1080/10447318.2015.1070549.

[22] N. Du, D. Tilbury, L. Robert, X. J. Yang, and A. Pradhan, "A cross-cultural study of trust building in autonomous vehicles," 2018.

[23] B. Herrenkind, I. Nastjuk, A. B. Brendel, S. Trang, and L. M. Kolbe, "Young people's travel behavior – Using the life-oriented approach to understand the acceptance of autonomous driving," *Transportation Research Part D: Transport and Environment*, vol. 74, pp. 214–233, Sep. 2019. DOI: 10.1016/j.trd.2019.07.023.

[24] I. Roche-Cerasi, "Public acceptance of driverless shuttles in Norway," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 162–183, Oct. 2019. DOI: 10.1016/j.trf.2019.09.002.

[25] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of management review*, vol. 23, no. 3, pp. 393–404, 1998.

[26] S. Castaldo, *Trust in market relationships*. Edward Elgar Publishing, 2007.

[27] J. B. Rotter, "Interpersonal trust, trustworthiness, and gullibility.," *American psychologist*, vol. 35, no. 1, p. 1, 1980.

[28] J. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human factors*, vol. 46, no. 1, pp. 50–80, 2004. DOI: 10.1518/hfes.46.1.50.30392.

[29] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Human factors*, vol. 57, no. 3, pp. 407–434, 2015. DOI: 10.1177/0018720814547570.

[30] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Academy of management review*, vol. 20, no. 3, pp. 709–734, 1995.

[31] S. Osswald, D. Wurhofer, S. Trösterer, E. Beck, and M. Tscheligi, "Predicting information technology usage in the car: Towards a car technology acceptance model," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2012, pp. 51–58. DOI: 10.1145/2390256.2390264.

[32] E. A. Prasetio and C. Nurliyana, "Evaluating perceived safety of autonomous vehicle: The influence of privacy and cybersecurity to cognitive and emotional safety," *IATSS research*, vol. 47, no. 2, pp. 160–170, 2023.

[33] M. Jin, G. Lu, F. Chen, and X. Shi, "How driving experience affect trust in automation from Level 3 automated vehicles? An experimental analysis," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2020, pp. 1–6. DOI: 10.1109/ITSC45102.2020.9294247.

[34] N. L. Tenhundfeld, E. J. de Visser, A. J. Ries, V. S. Finomore, and C. C. Tossell, "Trust and distrust of automated parking in a Tesla Model X," *Human factors*, vol. 62, no. 2, pp. 194–210, 2020. DOI: 10.1177/0018720819865412.

[35] M. Li, B. E. Holthausen, R. E. Stuck, and B. N. Walker, "No risk no trust: Investigating perceived risk in highly automated driving," *Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, pp. 177–185, 2019. DOI: 10.1145/3342197.3344525.

[36] S. Kolekar, J. de Winter, and D. Abbink, "Which parts of the road guide obstacle avoidance? Quantifying the driver's risk field," *Applied ergonomics*, vol. 89, p. 103 196, 2020. DOI: 10.1016/j.apergo.2020.103196.

[37] J. de Winter, J. Hoogmoed, J. Stapel, D. Dodou, and P. Bazilinskyy, "Predicting perceived risk of traffic scenes using computer vision," *Transportation research part F: traffic psychology and behaviour*, vol. 93, pp. 235–247, 2023.

[38] P. Ping, Y. Sheng, W. Qin, C. Miyajima, and K. Takeda, "Modeling driver risk perception on city roads using deep learning," *IEEE Access*, vol. 6, pp. 68 850–68 866, 2018. DOI: 10.1109/ACCESS.2018.2879887.

[39] R. J. Kiefer, D. J. LeBlanc, and C. A. Flannagan, "Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments," *Accident Analysis & Prevention*, vol. 37, no. 2, pp. 295–303, 2005. DOI: 10.1016/j.aap.2004.09.003.

[40] F. A. Mullakkal-Babu, M. Wang, X. He, B. van Arem, and R. Happee, "Probabilistic field approach for motorway driving risk assessment," *Transportation research part C: emerging technologies*, vol. 118, p. 102 716, 2020. DOI: 10.1016/j.trc.2020.102716.

[41] S. Kolekar, J. De Winter, and D. Abbink, "Human-like driving behaviour emerges from a risk-based driver model," *Nature Communications*, no. September, pp. 1–19, 2020. DOI: 10.1038/s41467-020-18353-4.

[42] J. J. Gibson and L. E. Crooks, "A theoretical field-analysis of automobile-driving," *The American journal of psychology*, vol. 51, no. 3, pp. 453–471, 1938.

[43] P. G. Tzouras, H. Farah, E. Papadimitriou, N. van Oort, and M. Hagenzieker, "Tram drivers' perceived safety and driving stress evaluation. a stated preference experiment," *Transportation research interdisciplinary perspectives*, vol. 7, p. 100 205, 2020.

[44] C. Chen, Z. Lan, G. Zhan, Y. Lyu, B. Nie, and S. E. Li, "Quantifying the individual differences of driver'risk perception with just four interpretable parameters," *arXiv preprint arXiv:2211.10907*, 2022.

[45] P. Bazilinskyy, Y. Eisma, D. Dodou, and J. C. de Winter, "Risk perception: A study using dashcam videos and participants from different world regions," *Traffic injury prevention*, vol. 21, no. 6, pp. 347–353, 2020.

[46] I. Tusseyeva, A. Oleinikov, A. Sandygulova, and M. Rubagotti, "Perceived safety in human–cobot interaction for fixed-path and real-time motion planning algorithms," *Scientific Reports*, vol. 12, no. 1, p. 20 438, Nov. 2022. DOI: 10.1038/s41598-022-24622-7.

[47] S. Sheng, E. Pakdamanian, K. Han, Z. Wang, J. Lenneman, D. Parker, and L. Feng, "Planning for automated vehicles with human trust," *ACM Transactions on Cyber-Physical Systems*, vol. 6, no. 4, pp. 1–21, 2022.

[48] A. Basantis, M. Miller, Z. Doerzaph, and M. L. Neurauter, "Assessing alternative approaches for conveying automated vehicle "intentions"," *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 6, pp. 622–631, 2021. DOI: 10.1109/Thms.2021.3106892.

[49] L. Oliveira, C. Burns, J. Luton, S. Iyer, and S. Birrell, "The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle," *Transportation Research Part F-Traffic Psychology and Behaviour*, vol. 72, pp. 280–296, 2020. DOI: 10.1016/j.trf.2020.06.001.

[50] T. V. Sawitzky, P. Wintersberger, A. Riener, and J. L. Gabbard, "Increasing trust in fully automated driving: Route indication on an augmented reality head-up display," in *Pervasive Displays 2019 - 8th ACM International Symposium on Pervasive Displays*, 2019.

[51] SAE, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," SAE International, Warrendale, PA, USA, Tech. Rep., 2021.

[52] X. Dong, M. DiScenna, and E. Guerra, "Transit user perceptions of driverless buses," *Transportation*, vol. 46, no. 1, pp. 35–50, 2019. DOI: 10.1007/s11116-017-9786-y.

[53] W. Griffin, N. Haworth, and D. Twisk, "Patterns in perceived crash risk among male and female drivers with and without substantial cycling experience," *Transportation Research Part F-Traffic Psychology and Behaviour*, vol. 69, pp. 1–12, 2020. DOI: 10.1016/j.trf.2019.12.013.

[54] J. A. Cox, V. Beanland, and A. J. Filtness, "Risk and safety perception on urban and rural roads: Effects of environmental features, driver age and risk sensitivity," *Traffic injury prevention*, vol. 18, no. 7, pp. 703–710, 2017. DOI: 10.1080/15389588.2017.1296956.

[55] A. D. Kaplan, T. T. Kessler, and P. A. Hancock, "How trust is defined and its use in human-human and human-machine interaction," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, vol. 64, 2020, pp. 1150–1154. DOI: 10.1177/1071181320641275.

[56] J. Kraus, D. Scholz, D. Stiegemeier, and M. Baumann, "The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency," *Human factors*, vol. 62, no. 5, pp. 718–736, 2020. DOI: 10.1177/0018720819853686.

[57] P. A. Hancock, T. T. Kessler, A. D. Kaplan, J. C. Brill, and J. L. Szalma, "Evolving trust in robots: Specification through sequential and comparative meta-analyses," *Human factors*, 2020. DOI: 10.1177/0018720820922080.

[58] K. Schaefer, "The perception and measurement of human-robot trust," Ph.D. dissertation, University of Central Florida, 2013.

[59] A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. A. Hancock, "Trust in artificial intelligence: Meta-analytic findings," *Human Factors*, 2021. DOI: 10.1177/00187208211013988.

[60] J. Kraus, D. Scholz, and M. Baumann, "What's driving me? Exploration and validation of a hierarchical personality model for trust in automated driving," *Human factors*, 2020. DOI: 10.1177/0018720820922653.

[61] C. Gold, M. Körber, C. Hohenberger, D. Lechner, and K. Bengler, "Trust in automation–before and after the experience of take-over scenarios in a highly automated vehicle," *Procedia Manufacturing*, vol. 3, pp. 3025–3032, 2015. DOI: 10.1016/j.promfg.2015.07.847.

[62] F. Walker, A. Boelhouwer, T. Alkim, W. B. Verwey, and M. H. Martens, "Changes in trust after driving Level 2 automated cars," *Journal of advanced transportation*, 2018. DOI: 10.1155/2018/1045186.

[63] K. Garidis, L. Ulbricht, A. Rossmann, and M. Schmäh, "Toward a user acceptance model of autonomous driving," in *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 2020.

[64] V. D. Pyrialakou, C. Gkartzonikas, J. D. Gatlin, and K. Gkritza, "Perceptions of safety on a shared road: Driving, cycling, or walking near an autonomous vehicle," *Journal of safety research*, vol. 72, pp. 249–258, 2020. DOI: 10.1016/j.jsr.2019.12.017.

[65] J. Stapel, A. Gentner, and R. Happee, "Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving," *Transportation research part F: traffic psychology and behaviour*, vol. 60, pp. 590–605, 2019. DOI: 10.1016/j.trf.2018.11.006.

[66] R. E. Yagoda and D. J. Gillan, "You want me to trust a robot? the development of a human–robot interaction trust scale," *International Journal of Social Robotics*, vol. 4, no. 3, pp. 235–248, 2012. DOI: 10.1007/s12369-012-0144-0.

[67] P. Rossner and A. C. Bullinger, "How do you want to be driven? Investigation of different highly-automated driving styles on a highway scenario," in *International Conference on Applied Human Factors and Ergonomics*, Springer, Cham, 2019, pp. 36–43.

[68] M. Beggiato, F. Hartwich, and J. Krems, "Physiological correlates of discomfort in automated driving," *Transportation research part F: traffic psychology and behaviour*, vol. 66, pp. 445–458, 2019. DOI: 10.1016/j.trf.2019.09.018.

[69] M. Saffarian, R. Happee, and J. D. Winter, "Why do drivers maintain short headways in fog? A driving-simulator study evaluating feeling of risk and lateral control during automated and manual car following," *Ergonomics*, vol. 55, no. 9, pp. 971–985, 2012. DOI: 10.1080/00140139.2012.691993.

[70] D. Cleij, J. Venrooij, P. Pretto, D. M. Pool, M. Mulder, and H. H. Bülthoff, "Continuous subjective rating of perceived motion incongruence during driving simulation," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 1, pp. 17–29, 2018. DOI: 10.1109/THMS.2017.2717884.

[71] X. Xiong, M. Wang, Y. Cai, L. Chen, H. Farah, and M. Hagenzieker, "A forward collision avoidance algorithm based on driver braking behavior," *Accident Analysis & Prevention*, vol. 129, pp. 30–43, 2019. DOI: 10.1016/j.aap.2019.05.004.

[72] D. H. Taylor, "Drivers' galvanic skin response and the risk of accident," *Ergonomics*, vol. 7, no. 4, pp. 439–451, 1964. DOI: 10.1080/00140136408930761.

[73] D. M. Morris, J. M. Erno, and J. J. Pilcher, "Electrodermal response and automation trust during simulated self-driving car use," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 61, 2017, pp. 1759–1762. DOI: 10.1177/1541931213601921.

[74] I. B. Ajenaghughrure, S. C. D. C. Sousa, and D. Lamas, "Psychophysiological modelling of trust in technology: Comparative analysis of algorithm ensemble methods," in *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, IEEE, 2021, pp. 000 161–000 168. DOI: 10.5220/0010237701610173.

[75] R. Tang, J. H. Kim, R. Parker, and Y. J. Jeong, "Indicating severity of vehicle accidents using pupil diameter in a driving simulator environment," in *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, 2018, pp. 647–656.

[76] J. Perello-March, C. Burns, M. Elliott, and S. Birrell, "Integrating trust in automation into driver state monitoring systems," in *International Conference on Human Interaction and Emerging Technologies*, 2019, pp. 344–349. DOI: 10.1007/978-3-030-25629-6.

[77] S. F. Varotto, H. Farah, T. Toledo, B. van Arem, and S. P. Hoogendoorn, "Modelling decisions of control transitions and target speed regulations in full-range adaptive cruise control based on risk allostasis theory," *Transportation Research Part B: Methodological*, vol. 117, pp. 318–341, 2018. DOI: 10.1016/j.trb.2018.09.007.

[78] R. Fuller, "Driver control theory: From task difficulty homeostasis to risk allostasis," in *Handbook of traffic psychology*, 2011, pp. 13–26.

[79] D. N. Lee, "A theory of visual control of braking based on information about time-to-collision," *Perception*, vol. 5, no. 4, pp. 437–459, 1976.

[80] S. Marsh and M. R. Dibben, "The role of trust in information science and technology," *Annual Review of Information Science and Technology (ARIST)*, vol. 37, pp. 465–98, 2003.

[81] C. Hu and J. Wang, "Trust-based and individualizable adaptive cruise control using control barrier function approach with prescribed performance," *IEEE Transactions on Intelligent Transportation Systems*, 2021. DOI: 10.1109/TITS.2021.3066154.

[82] F. A. Dreger, J. C. de Winter, and R. Happee, "How do drivers merge heavy goods vehicles onto freeways? A semi-structured interview unveiling needs for communication and support," *Cognition, Technology & Work*, vol. 22, no. 4, pp. 825–842, 2020. DOI: 10.1007/s10111-019-00601-3.

[83] Y. R. Khusro, Y. Zheng, M. Grottoli, and B. Shyrokau, "MPC-based motion-cueing algorithm for a 6-DOF driving simulator with actuator constraints," *Vehicles*, vol. 2, no. 4, pp. 625–647, 2020. DOI: 10.3390/vehicles2040036.

[84] F. A. Mullakkal-Babu, M. Wang, B. Van Arem, and R. Happee, "Design and analysis of full range adaptive cruise control with integrated collision avoidance strategy," in *2016 IEEE 19th International conference on intelligent transportation systems (ITSC)*, 2016, pp. 308–315. DOI: 10.1109/ITSC.2016.7795572.

[85] L. Meyer-Waarden and J. Cloarec, ""Baby, you can drive my car": Psychological antecedents that drive consumers' adoption of AI-powered autonomous vehicles," *Technovation*, vol. 109, p. 102 348, 2022. DOI: 10.1016/j.technovation.2021.102348.

[86] H. Sedghamiz, "Biosigkit: A MATLAB toolbox and interface for analysis of biosignals," *Journal of Open Source Software*, vol. 3, no. 30, p. 671, 2018. DOI: 10.21105/joss.00671.

[87] H. Sedghamiz and D. Santonocito, "Unsupervised detection and classification of motor unit action potentials in intramuscular electromyography signals," in *2015 e-health and bioengineering conference (EHB)*, IEEE, 2015, pp. 1–6. DOI: 10.1109/EHB.2015.7391510.

[88] H. J. Baek, C. H. Cho, J. Cho, and J. M. Woo, "Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability," *Telemedicine and e-Health*, vol. 21, no. 5, pp. 404–414, 2015. DOI: 10.1089/tmj.2014.0104.

[89] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 258, 2017. DOI: 10.3389/fpubh.2017.00258.

[90] M. E. Kret and E. E. Sjak-Shie, "Preprocessing pupil size data: Guidelines and code," *Behavior research methods*, vol. 51, no. 3, pp. 1336–1342, 2019. DOI: 10.3758/s13428-018-1075-y.

[91] A. D. Astin, "Finger force capability: Measurement and prediction using anthropometric and myoelectric measures," Ph.D. dissertation, Virginia Polytechnic Institute and State University, USA, 1999.

[92] L. Huang, Q. Xia, F. Xie, H. L. Xiu, and H. Shu, "Study on the test scenarios of Level 2 automated vehicles," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2018, pp. 49–54. DOI: 10.1109/IVS.2018.8500600.

[93] D. Zhao, Y. Guo, and Y. J. Jia, "Trafficnet: An open naturalistic driving scenario library," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2017, pp. 1–8.

[94] A. Sharma, Z. Zheng, J. Kim, A. Bhaskar, and M. M. Haque, "Is an informed driver a better decision maker? A grouped random parameters with heterogeneity-in-means approach to investigate the impact of the connected environment on driving behaviour in safety-critical situations," *Analytic methods in accident research*, vol. 27, p. 100 127, 2020. DOI: 10.1016/j.amar.2020.100127.

[95] ISO, *Intelligent transport systems — adaptive cruise control systems — performance requirements and test procedures*, 2018.

[96] M. Beggiato, M. Pereira, T. Petzoldt, and J. Krems, "Learning and development of trust, acceptance and the mental model of ACC. a longitudinal on-road study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 35, pp. 75–84, 2015. DOI: 10.1016/j.trf.2015.10.005.

[97] C. Wang, Y. Xie, H. Huang, and P. Liu, "A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling," *Accident Analysis & Prevention*, vol. 157, p. 106 157, 2021. DOI: 10.1016/j.aap.2021.106157.

[98] X. Ma, Z. Feng, X. Zhu, and Z. Ma, "Driver risk perception model under critical cut-in scenarios," SAE Technical Paper, Tech. Rep. 2018-01-1626, 2018. DOI: 10.4271/2018-01-1626.

[99] J. Kraus, D. Scholz, E. M. Messner, M. Messner, and M. Baumann, "Scared to trust? – predicting trust in highly automated driving by depressiveness, negative self-evaluations and state anxiety," *Frontiers in psychology*, vol. 10, p. 2917, 2020. DOI: 10.3389/fpsyg.2019.02917.

[100] N. Rhodes and K. Pivik, "Age and gender differences in risky driving: The roles of positive affect and risk perception," *Accident Analysis & Prevention*, vol. 43, no. 3, pp. 923–931, 2011. DOI: 10.1016/j.aap.2010.11.015.

[101] J. Ayoub, X. J. Yang, and F. Zhou, "Modeling dispositional and initial learned trust in automated vehicles with predictability and explainability," *Transportation research part F: traffic psychology and behaviour*, vol. 77, pp. 102–116, 2021. DOI: 10.1016/j.trf.2020.12.015.

[102] J. Z. Bakdash and L. R. Marusich, "Repeated measures correlation," *Frontiers in psychology*, vol. 8, p. 456, 2017. DOI: 10.3389/fpsyg.2017.00456.

[103] T. Kondoh, T. Yamamura, S. Kitazaki, N. Kuge, and E. Boer, "Identification of visual cues and quantification of drivers' perception of proximity risk to the lead vehicle in car-following situations," *Journal of Mechanical Systems for Transportation and Logistics*, vol. 1, no. 2, pp. 170–180, 2008. DOI: 10.1299/jmtl.1.170.

[104] G. Lu, B. Cheng, Q. Lin, and Y. Wang, "Quantitative indicator of homeostatic risk perception in car following," *Safety Science*, vol. 50, no. 9, pp. 1898–1905, 2012. DOI: 10.1016/j.ssci.2012.05.007.

[105] A. D. Kaplan, T. T. Kessler, T. L. Sanders, J. Cruit, J. C. Brill, and P. A. Hancock, "A time to trust: Trust as a function of time in human-robot interaction," in *Trust in Human-Robot Interaction*, Elsevier Inc., 2021. DOI: 10.1016/b978-0-12-819472-0.00006-x.

[106] D. He and B. Donmez, "Influence of driving experience on distraction engagement in automated vehicles," *Transportation research record*, vol. 2673, no. 9, pp. 142–151, 2019. DOI: 10.1177/0361198119843476.

[107] A. Borowsky and T. Oron-Gilad, "Exploring the effects of driving experience on hazard awareness and risk perception via real-time hazard identification, hazard classification, and rating tasks," *Accident Analysis & Prevention*, vol. 59, pp. 548–565, 2013. DOI: 10.1016/j.aap.2013.07.008.

[108] A. Feldhütter, C. Gold, A. Hüger, and K. Bengler, "Trust in automation as a matter of media influence and experience of automated vehicles," in *Proceedings of the human factors and ergonomics society annual meeting*, Los Angeles, vol. 60, 2016, pp. 2024–2028. DOI: 10.1177/1541931213601460.

[109] L. M. Hulse, H. Xie, and E. R. Galea, "Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age," *Safety science*, vol. 102, pp. 1–13, 2018. DOI: 10.1016/j.ssci.2017.10.001.

[110] J. Stapel, A. Gentner, and R. Happee, "On-road trust and perceived risk in level 2 automation," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 89, pp. 355–370, July 2022. DOI: 10.1016/j.trf.2022.07.008.

[111] J. D. Lee, S. Y. Liu, J. Domeyer, and A. DinparastDjadid, "Assessing drivers' trust of automated vehicle driving styles with a two-part mixed model of intervention tendency and magnitude," *Human factors*, vol. 63, no. 2, pp. 197–209, 2021. DOI: 10.1177/0018720819880363.

[112] A. P. Afghari, S. Washington, M. M. Haque, and Z. Li, "A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk," *Analytic Methods in Accident Research*, vol. 18, pp. 1–14, 2018. DOI: 10.1016/j.amar.2018.03.002.

[113] World Health Organization, *Road traffic injuries*, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

[114] N. Nadimi, H. Behbahani, and H. R. Shahbazi, "Calibration and validation of a new time-based surrogate safety measure using fuzzy inference system," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 1, pp. 51–58, 2016. DOI: 10.1016/j.jtte.2015.09.004.

[115] L. Eboli, G. Mazzulla, and G. Pungillo, "How to define the accident risk level of car drivers by combining objective and subjective measures of driving style," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 49, pp. 29–38, 2017. DOI: 10.1016/j.trf.2017.06.004.

[116] H. Summala, "Risk control is not risk adjustment: The zero-risk theory of driver behaviour and its implications," *Ergonomics*, vol. 31, no. 4, pp. 491–506, 1988. DOI: 10.1080/00140138808966694.

[117] X. He, J. Stapel, M. Wang, and R. Happee, "Modelling perceived risk and trust in driving automation reacting to merging and braking vehicles," *Transportation Research Part F: Psychology and Behaviour*, vol. 86, no. February, pp. 178–195, 2022. DOI: 10.1016/j.trf.2022.02.016.

[118] T. Kondoh, N. Furuyama, T. Hirose, and T. Sawada, "Direct Evidence of the Inverse of TTC Hypothesis for Driver's Perception in Car-Closing Situations," *International Journal of Automotive Engineering*, vol. 5, no. 4, pp. 121–128, 2014. DOI: 10.20485/jsaeijae.5.4{_}121.

[119] J. Wang, J. Wu, X. Zheng, D. Ni, and K. Li, "Driving safety field theory modeling and its application in pre-collision warning system," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 306–324, 2016. DOI: 10.1016/j.trc.2016.10.003.

[120] L. Li, J. Gan, Z. Yi, X. Qu, and B. Ran, "Risk perception and the warning strategy based on safety potential field theory," *Accident Analysis & Prevention*, vol. 148, no. December, 2020. DOI: 10.1016/j.aap.2020.105805.

[121] R. Fuller, "A conceptualization of driving behaviour as threat avoidance," *Ergonomics*, vol. 27, no. 11, pp. 1139–1155, 1984. DOI: 10.1080/00140138408963596.

[122] M. Sivak, "The information that drivers use: Is it indeed 90% visual?" *Perception*, vol. 25, no. 9, pp. 1081–1089, 1996.

[123] J. R. Ward, G. Agamennoni, S. Worrall, A. Bender, and E. Nebot, "Extending Time to Collision for probabilistic reasoning in general traffic scenarios," *Transportation Research Part C: Emerging Technologies*, vol. 51, pp. 66–82, 2015. DOI: 10.1016/j.trc.2014.11.002.

[124] P. Wagner, R. Nippold, S. Gabloner, and M. Margreiter, "Analyzing human driving data an approach motivated by data science methods," *Chaos, Solitons and Fractals*, vol. 90, pp. 37–45, 2016. DOI: 10.1016/j.chaos.2016.02.008.

[125] J. Ko, R. Guensler, and M. Hunter, "Analysis of effects of driver/vehicle characteristics on acceleration noise using GPS-equipped vehicles," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 13, no. 1, pp. 21–31, 2010. DOI: 10.1016/j.trf.2009.09.003.

[126] J. Jansson, "Collision avoidance theory: With application to automotive collision mitigation," Ph.D. dissertation, Linköping University Electronic Press, 2005.

[127] R. E. Chandler, R. Herman, and E. W. Montroll, "Traffic dynamics: Studies in car following," *Operations research*, vol. 6, no. 2, pp. 165–184, 1958.

[128] B. Sultan, M. Brackstone, and M. McDonald, "Drivers' use of deceleration and acceleration information in car-following process," *Transportation Research Record*, vol. 1883, no. 1, pp. 31–39, 2004.

[129] J. Duan, Z. Li, and G. Salvendy, "Risk illusions in car following: Is a smaller headway always perceived as more dangerous?" *Safety science*, vol. 53, pp. 25–33, 2013.

[130] J. Ding, J. Wang, C. Liu, M. Lu, and K. Li, "A Driver steering behavior model based on lane-keeping characteristics analysis," *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, pp. 623–628, 2014. DOI: 10.1109/ITSC.2014.6957759.

[131] G. J. Blaauw, "Driving experience and task demands in simulator and instrumented car: A validation study," *Human Factors*, vol. 24, no. 4, pp. 473–486, 1982.

[132] F. Kochi, Y. Saito, N. Uchida, and M. Itoh, "Task difficulty, risk feeling, and safety margin in the determination of driver behavior to prepare for traffic conflicts," *Accident Analysis & Prevention*, vol. 192, p. 107 284, 2023.

[133] Crashdashes, *How many people in comments could do better*, https://www.instagram.com/reel/CsKIpjSr78Z/?utm_source=ig_web_copy_link, 2023.

[134] R. N. Jazar, *Vehicle dynamics*. Springer, 2008, vol. 1.

[135] L. Aarts and I. Van Schagen, "Driving speed and the risk of road crashes: A review," *Accident Analysis & Prevention*, vol. 38, no. 2, pp. 215–224, 2006.

[136] D. Finch, P. Kompfner, C. Lockwood, and G. Maycock, "Speed, speed limits and crashes. crowthorne, berkshire," *Transport Research Laboratory TRL, Project Report PR*, vol. 58, 1994.

[137] R. Fuller, "The task-capability interface model of the driving process," *Recherche-Transports-Sécurité*, vol. 66, pp. 47–57, 2000.

[138] R. Näätänen and H. Summala, "Road-user behaviour and traffic accidents," *Publication of: North-Holland Publishing Company*, 1976.

[139] F. P. McKenna, "The human factor in driving accidents an overview of approaches and problems," *Ergonomics*, vol. 25, no. 10, pp. 867–877, 1982.

[140] T. Rundmo and T. Nordfjærn, "Does risk perception really exist?" *Safety Science*, vol. 93, pp. 230–240, 2017. DOI: 10.1016/j.ssci.2016.12.014.

[141] S. Kim, X. He, R. van Egmond, and R. Happee, "Designing user interfaces for partially automated vehicles: Effects of information and modality on trust and acceptance," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 103, pp. 404–419, 2024. DOI: 10.1016/j.trf.2024.02.009.

[142] A. El Jouhri, A. El Sharkawy, H. Paksoy, O. Youssif, X. He, S. Kim, and R. Happee, "The influence of a color themed hmi on trust and take-over performance in automated vehicles," *Frontiers in psychology*, vol. 14, p. 1 128 285, 2023.

[143] N. Science, T. Council, and U. S. D. of Transportation, "Ensuring American leadership in automated vehicle technologies," National Science, Technology Council, and United States Department of Transportation, Tech. Rep., Jan. 2020.

[144] S. G. Charlton, N. J. Starkey, J. A. Perrone, and R. B. Isler, "What's the risk? a comparison of actual and perceived driving risk," *Transportation research part F: traffic psychology and Behaviour*, vol. 25, pp. 50–64, 2014.

[145] T. Evans, R. Stuckey, and W. Macdonald, "Young drivers' perceptions of risk and difficulty: Day versus night," *Accident Analysis & Prevention*, vol. 147, p. 105 753, 2020.

[146] F. W. Siebert and F. L. Wallis, "How speed and visibility influence preferred headway distances in highly automated driving," *Transportation research part F: traffic psychology and behaviour*, vol. 64, pp. 485–494, 2019.

[147] C. Chen, X. Zhao, H. Liu, G. Ren, and X. Liu, "Influence of adverse weather on drivers' perceived risk during car following based on driving simulations," *Journal of modern transportation*, vol. 27, pp. 282–292, 2019.

[148] L. Jing, W. Shan, and Y. Zhang, "Risk preference, risk perception as predictors of risky driving behaviors: The moderating effects of gender, age, and driving experience," *Journal of Transportation Safety & Security*, vol. 15, no. 5, pp. 467–492, 2023.

[149] A.-M. Sourelli, R. Welsh, and P. Thomas, "Objective and perceived risk in overtaking: The impact of driving context," *Transportation research part F: traffic psychology and behaviour*, vol. 81, pp. 190–200, 2021.

[150] S. Kolekar, B. Petermeijer, E. Boer, J. D. Winter, and D. Abbink, "A risk field-based metric correlates with driver ' s perceived risk in manual and automated driving : A test-track study," *Transportation Research Part C*, vol. 133, no. November, p. 103 428, 2021. DOI: 10.1016/j.trc.2021.103428.

[151] J. Parvizi and S. Kastner, "Promises and limitations of human intracranial electroencephalography," *Nature neuroscience*, vol. 21, no. 4, pp. 474–483, 2018.

[152] D. Taylor, "Drivers'galvanic skin response and the risk of accident," *Ergonomics*, vol. 7, no. 4, pp. 439–451, 1964.

[153] N. Du, X. J. Yang, and F. Zhou, "Psychophysiological responses to takeover requests in conditionally automated driving," *Accident Analysis & Prevention*, vol. 148, p. 105 804, 2020.

[154] J. Petit, C. Charron, and F. Mars, "Risk assessment by a passenger of an autonomous vehicle among pedestrians: Relationship between subjective and physiological measures," *Frontiers in neuroergonomics*, vol. 2, p. 682 119, 2021.

[155] M. Zhang, M. Jipp, and K. Ihme, "The novelty appraisal of the feeling of risk in vehicles," *International journal of environmental research and public health*, vol. 19, no. 21, p. 14 259, 2022.

[156] F. Vintila, T. C. Kübler, and E. Kasneci, "Pupil response as an indicator of hazard perception during simulator driving," *Journal of eye movement research*, vol. 10, no. 4, 2017.

[157] J. A. Thomas and D. Walton, "Measuring perceived risk: Self-reported and actual hand positions of suv and car drivers," *Transportation research part F: traffic psychology and behaviour*, vol. 10, no. 3, pp. 201–207, 2007.

[158] M. Fourie, D. Walton, and J. Thomas, "Naturalistic observation of drivers' hands, speed and headway," *Transportation research part F: traffic psychology and behaviour*, vol. 14, no. 5, pp. 413–421, 2011.

[159] D. Walton and J. A. Thomas, "Naturalistic observations of driver hand positions," *Transportation research part F: traffic psychology and behaviour*, vol. 8, no. 3, pp. 229–238, 2005.

[160] D. De Waard, T. G. Van den Bold, and B. Lewis-Evans, "Driver hand position on the steering wheel while merging into motorway traffic," *Transportation research part F: traffic psychology and behaviour*, vol. 13, no. 2, pp. 129–140, 2010.

[161] A. Vance, B. B. Anderson, C. B. Kirwan, and D. Eargle, "Using measures of risk perception to predict information security behavior: Insights from electroencephalography (EEG)," *Journal of the Association for Information Systems*, vol. 15, no. 10, p. 2, 2014.

[162] L. Zhu, Q. Ma, X. Bai, and L. Hu, "Mechanisms behind hazard perception of warning signs: An EEG study," *Transportation research part F: traffic psychology and behaviour*, vol. 69, pp. 362–374, 2020.

[163] R. B. Brandt, "The definition of an "ideal observer" theory in ethics," *Philosophy and Phenomenological Research*, vol. 15, no. 3, pp. 407–413, 1955.

[164] W. S. Geisler, "Ideal observer theory in psychophysics and physiology," *Physica Scripta*, vol. 39, no. 1, p. 153, 1989.

[165] A. Francl and J. H. McDermott, "Deep neural network models of sound localization reveal how perception is adapted to real-world environments," *Nature human behaviour*, vol. 6, no. 1, pp. 111–133, 2022.

[166] L. Ziebell, C. Collin, S. Rainville, M. Mazalu, and M. Weippert, "Using an ideal observer analysis to investigate the visual perceptual efficiency of individuals with a history of non-suicidal self-injury when identifying emotional expressions," *PloS one*, vol. 15, no. 2, e0227019, 2020.

[167] X. He, R. Happee, and M. Wang, "A new computational perceived risk model for automated vehicles based on potential collision avoidance difficulty (PCAD)," *Transportation Research Part C: Emerging Technologies*, vol. 166, no. 104751, 2024. DOI: 10.1016/j.trc.2024.104751.

[168] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[169] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim, *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature Biomedical Engineering*, vol. 2, no. 10, p. 749, 2018.

[170] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.

[171] Vimeo, *Vimeo//a video sharing platform*, https://www.vimeo.com, 2023.

[172] Qualtrics, *Qualtrics xm//the leading experience management software*, https://www.qualtrics.com, 2023.

[173] Prolific, *Prolific//quickly find research participants you can trust.* https://www.prolific.co, 2023.

[174] IBM Corp., *IBM SPSS Statistics for Windows*, Software, version 29.0.1.0, Available from IBM, Armonk, NY, 2020.

[175] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information theory*, vol. 37, no. 1, pp. 145–151, 1991.

[176] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transportation research part F: traffic psychology and behaviour*, vol. 8, no. 2, pp. 97–120, 2005.

[177] D. D. Salvucci and R. Gray, "A two-point visual control model of steering," *Perception*, vol. 33, no. 10, pp. 1233–1248, 2004.

[178] MathWorks, *fmincon - Documentation*, https://de.mathworks.com/help/optim/ug/fmincon.html, 2024.

[179] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[180] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[181] D. F. Cooper and N. Ferguson, "Traffic studies at T-Junctions. 2. A conflict simulation record," *Traffic Engineering & Control*, vol. 17, no. Analytic, 1976.

[182] C. Hydén, "The development of a method for traffic safety evaluation: The Swedish traffic conflicts technique," *Bulletin Lund Institute of Technology, Department*, no. 70, 1987.

[183] D. Gettman and L. Head, "Surrogate safety measures from traffic simulation models," *Transportation research record*, vol. 1840, no. 1, pp. 104–115, 2003.

[184] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[185] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[186] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[187] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" In *Advances in neural information processing systems*, 2017.

[188] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods," *Machine learning*, vol. 110, no. 3, pp. 457–506, 2021.

[189] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, 2016, pp. 1050–1059.

[190] M.-Y. Moon, K. Choi, N. Gaul, and D. Lamb, "Treating epistemic uncertainty using bootstrapping selection of input distribution model for confidence-based reliability assessment," *Journal of Mechanical Design*, vol. 141, no. 3, p. 031 402, 2019.

[191] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowledge and information systems*, vol. 41, pp. 647–665, 2014.

[192] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1135–1144.

[193] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International conference on machine learning*, PMLR, 2017, pp. 3145–3153.

[194] A. Datta, S. Sen, and Y. Zick, "Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems," in *2016 IEEE symposium on security and privacy (SP)*, IEEE, 2016, pp. 598–617.

[195] V. Branzi, L. Domenichini, and F. La Torre, "Drivers' speed behaviour in real and simulated urban roads–a validation study," *Transportation research part F: traffic psychology and behaviour*, vol. 49, pp. 1–17, 2017.

[196] O. Shechtman, S. Classen, K. Awadzi, and W. Mann, "Comparison of driving errors between on-the-road and simulated driving assessment: A validation study," *Traffic injury prevention*, vol. 10, no. 4, pp. 379–385, 2009.

[197] D. R. Mayhew, H. M. Simpson, K. M. Wood, L. Lonero, K. M. Clinton, and A. G. Johnson, "On-road and simulated driving: Concurrent and discriminant validation," *Journal of safety research*, vol. 42, no. 4, pp. 267–275, 2011.

[198] G. Li, Z. Li, V. Knoop, and H. van Lint, "How far ahead should autonomous vehicles start reacting to conflicts? exploring the uncertainty-based efficiency-safety trade-off," *Exploring the Uncertainty-Based Efficiency-Safety Trade-Off*, 2023.

[199] X. Wang, Z. Li, J. Alonso-Mora, and M. Wang, "Reachability-based confidence-aware probabilistic collision detection in highway driving," *Engineering*, vol. 33, pp. 90–107, 2024.

[200] J. Wang, J. Wu, and Y. Li, "The driving safety field based on driver–vehicle–road interactions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2203–2214, 2015.

[201] L. Li, J. Gan, X. Ji, X. Qu, and B. Ran, "Dynamic driving risk potential field model under the connected and automated vehicles environment and its application in car-following modeling," *IEEE transactions on intelligent transportation systems*, vol. 23, no. 1, pp. 122–141, 2020.

[202] A. Prieto, B. Prieto, E. M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, and I. Rojas, "Neural networks: An overview of early research, current frameworks and new challenges," *Neurocomputing*, vol. 214, pp. 242–268, 2016.

[203] P. P. Balestrassi, E. Popova, A. d. Paiva, and J. M. Lima, "Design of experiments on neural network's training for nonlinear time series forecasting," *Neurocomputing*, vol. 72, no. 4-6, pp. 1160–1178, 2009.

[204] H. Farah, G. B. Piccinini, M. Itoh, and M. Dozza, "Modelling overtaking strategy and lateral distance in car-to-cyclist overtaking on rural roads: A driving simulator experiment," *Transportation research part F: traffic psychology and behaviour*, vol. 63, pp. 226–239, 2019.

[205] E. Stefansson, F. J. Jiang, E. Nekouei, H. Nilsson, and K. H. Johansson, "Modeling the decision-making in human driver overtaking," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 15 338–15 345, 2020.

[206] X. Tang, K. Yang, H. Wang, W. Yu, X. Yang, T. Liu, and J. Li, "Driving environment uncertainty-aware motion planning for autonomous vehicles," *Chinese Journal of Mechanical Engineering*, vol. 35, no. 1, p. 120, 2022.

[207] C. Wei, R. Romano, N. Merat, Y. Wang, C. Hu, H. Taghavifar, F. Hajiseyedjavadi, and E. R. Boer, "Risk-based autonomous vehicle motion control with considering human driver's behaviour," *Transportation research part C: Emerging Technologies*, vol. 107, pp. 1–14, 2019.

[208] T. Litman, *Autonomous vehicle implementation predictions*, Victoria Transport Policy Institute Victoria, Canada, 2017.

[209] R. Shabanpour, N. Golshani, A. Shamshiripour, and A. Mohammadian, "Eliciting preferences for adoption of fully automated vehicles using best-worst analysis," *Transportation Research Part C-Emerging Technologies*, vol. 93, pp. 463–478, 2018. DOI: 10.1016/j.trc.2018.06.014.

[210] R. Krueger, T. H. Rashidi, and J. M. Rose, "Preferences for shared autonomous vehicles," *Transportation Research Part C-Emerging Technologies*, vol. 69, pp. 343–355, 2016. DOI: 10.1016/j.trc.2016.06.015.

[211] S. Nordhoff, J. de Winter, M. Kyriakidis, B. van Arem, and R. Happee, "Acceptance of driverless vehicles: Results from a large cross-national questionnaire study," *Journal of Advanced Transportation*, 2018. DOI: 10.1155/2018/5382192.

[212] L. M. Cysneiros, M. A. Raffi, and J. C. S. D. Leite, "Software transparency as a key requirement for self-driving cars," in *2018 IEEE 26th International Requirements Engineering Conference (RE 2018)*, 2018, pp. 382–387. DOI: 10.1109/Re.2018.00-21.

[213] H. Detjen, S. Faltaous, B. Pfleging, S. Geisler, and S. Schneegass, "How to increase automated vehicles' acceptance through in-vehicle interaction design: A review," *International Journal of Human-Computer Interaction*, vol. 37, no. 4, pp. 308–330, 2021. DOI: 10.1080/10447318.2020.1860517.

[214] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997. DOI: 10.1518/001872097778543886.

[215] K. M. Wilson, S. Y. Yang, T. Roady, J. Kuo, and M. G. Lenne, "Driver trust & mode confusion in an on-road study of level-2 automated vehicle technology," *Safety Science*, vol. 130, 2020. DOI: 10.1016/j.ssci.2020.104845.

[216] S. Kim, R. van Egmond, and R. Happee, "Effects of user interfaces on take-over performance: A review of the empirical evidence," *Information*, vol. 12, no. 4, 2021. DOI: 10.3390/info12040162.

[217] M. Ghazizadeh, J. D. Lee, and L. N. Boyle, "Extending the technology acceptance model to assess automation," *Cognition Technology & Work*, vol. 14, no. 1, pp. 39–49, 2012. DOI: 10.1007/s10111-011-0194-3.

[218] S. M. Merritt, D. Lee, J. L. Unnerstall, and K. Huber, "Are well-calibrated users effective users? associations between calibration of trust and performance on an automation-aided task," *Human Factors*, vol. 57, pp. 34–47, 2015.

[219] J. Haspiel, N. Du, J. Meyerson, L. P. Robert, D. Tilbury, X. J. Yang, and A. K. Pradhan, "Explanations and expectations: Trust building in automated vehicles," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI'18)*, 2018, pp. 119–120. DOI: 10.1145/3173386.3177057.

[220] S. O'Kane, *Tesla autopilot, distracted driving to blame in deadly 2018 crash*, The Verge, 2020.

[221] P. R. Center, *Automation in everyday life*, 2017.

[222] J. B. Lyons, K. S. Koltai, N. T. Ho, W. B. Johnson, D. E. Smith, and R. J. Shively, "Engineering trust in complex automated systems," *Ergonomics in Design*, vol. 24, no. 1, pp. 13–17, 2016. DOI: 10.1177/1064804615611272.

[223] D. A. Norman, "The problem with automation - inappropriate feedback and interaction, not over-automation," *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, vol. 327, no. 1241, pp. 585–593, 1990. DOI: 10.1098/rstb.1990.0101.

[224] M. R. Endsley, B. Bolte, and D. G. Jones, *Designing for situation awareness : an approach to user-centered design*. Taylor & Francis, 2003, https://www.taylorfrancis.com/books/9781135729493.

[225] V. Alonso and P. de la Puente, "System transparency in shared autonomy: A mini review," *Frontiers in Neurorobotics*, vol. 12, 2018, https://doi.org/10.3389/fnbot.2018.00083.

[226] O. Carsten and M. H. Martens, "How can humans understand their automated cars? HMI principles, problems and solutions," *Cognition Technology & Work*, vol. 21, no. 1, pp. 3–20, 2019. DOI: 10.1007/s10111-018-0484-0.

[227] J. Gao and J. D. Lee, "Effect of shared information on trust and reliance in a demand forecasting task," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2006.

[228] J. B. Lyons, G. G. Sadler, K. Koltai, H. Battiste, N. T. Ho, L. C. Hoffmann, D. Smith, W. Johnson, and R. Shively, "Shaping trust through transparent design: Theoretical and experimental guidelines," in *Advances in Human Factors in Robots and Unmanned Systems*, 2017, pp. 127–136. DOI: 10.1007/978-3-319-41959-6_11.

[229] J. E. Mercado, M. A. Rupp, J. Y. Chen, M. J. Barnes, D. Barber, and K. Procci, "Intelligent agent transparency in human–agent teaming for Multi-UxV management," *Human Factors*, vol. 58, no. 3, pp. 401–415, 2016.

[230] E. Visser, M. Cohen, A. Freedy, and R. Parasuraman, "A design methodology for trust cue calibration in cognitive agents," in *Proceedings of the International Conference on Virtual, Augmented and Mixed Reality*, 2014.

[231] C. Chang, R. A. Grier, J. Maynard, J. Shutko, M. Blommer, R. Swaminathan, and R. Curry, "Using a situational awareness display to improve rider trust and comfort with an AV taxi," in *Proceedings of the Human Factors and Ergonomics Society 2019 Annual Meeting*, 2019.

[232] P. Hock, J. Kraus, M. Walch, N. Lang, and M. Baumann, "Elaborating feedback strategies for maintaining automation in highly automated driving," in *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, Association for Computing Machinery, Inc, vol. 16, 2016, pp. 105–112.

[233] J. Koo, J. Kwac, W. Ju, M. Steinert, L. Leifer, and C. Nass, "Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance," *International Journal of Interactive Design and Manufacturing - IJIDeM*, vol. 9, no. 4, pp. 269–275, 2015. DOI: 10.1007/s12008-014-0227-2.

[234] R. H. Y. Ma, A. Morris, P. Herriotts, and S. Birrell, "Investigating what level of visual information inspires trust in a user of a highly automated vehicle," *Applied Ergonomics*, vol. 90, 2021. DOI: 10.1016/j.apergo.2020.103272.

[235] A. Mackay, I. Fortes, C. Santos, D. Machado, P. Barbosa, V. Boas, J. Ferreira, N. Costa, C. Silva, and E. Sousa, "The impact of autonomous vehicles' active feedback on trust," in *Advances in Safety Management and Human Factors. AHFE 2019.*, 2020.

[236] Y. C. Liu, "Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems," *Ergonomics*, vol. 44, no. 4, pp. 425–442, 2001. DOI: 10.1080/00140130010011369.

[237] J. Edworthy, "Does sound help us to work better with machines? A commentary on rauterberg's paper 'about the importance of auditory alarms during the operation of a plant simulator'," *Interacting with Computers*, vol. 10, no. 4, pp. 401–409, 1998.

[238] E. Özcan and R. V. Egmond, "Basic semantics of product sounds," *International Journal of Design*, vol. 6, no. 2, pp. 41–54, 2012.

[239] L. Avetisyan, J. Ayoub, and F. Zhou, "Investigating explanations in conditional and highly automated driving: The effects of situation awareness and modality," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 89, pp. 456–466, 2022. DOI: 10.1016/j.trf.2022.07.010.

[240] J. Dong, E. Lawson, J. Olsen, and M. Jeon, "Female voice agents in fully autonomous vehicles are not only more likeable and comfortable, but also more competent," in *Proceedings of the 2020 HFES 64th International Annual Meeting*, 2021.

[241] A. Waytz, J. Heafner, and N. Epley, "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle," *Journal of Experimental Social Psychology*, vol. 52, pp. 113–117, 2014.

[242] N. Du, F. Zhou, D. Tilbury, L. P. Robert, and X. J. Yang, "Designing alert systems in takeover transitions: The effects of display information and modality," in *AutomotiveUI '21: 13th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2021, pp. 173–180. DOI: 10.1145/3409118.3475155.

[243] X. Li, A. Vaezipour, A. Rakotonirainy, S. Demmel, and O. Oviedo-Trespalacios, "Exploring drivers' mental workload and visual demand while using an in-vehicle HMI for eco-safe driving," *Accident Analysis & Prevention*, vol. 146, 2020.

[244] R. C. Goncalves, T. L. Louw, R. Madigan, M. Quaresma, R. Romano, and N. Merat, "The effect of information from dash-based human-machine interfaces on drivers' gaze patterns and lane-change manoeuvres after conditionally automated driving," *Accident Analysis and Prevention*, vol. 174, 2022. DOI: 10.1016/j.aap.2022.106726.

[245] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989. DOI: 10.2307/249008.

[246] L. Buckley, S. Kaye, and A. K. Pradhan, "A qualitative examination of drivers' responses to partially automated vehicles," *Transportation Research Part F-Traffic Psychology and Behaviour*, vol. 56, pp. 167–175, 2018.

[247] S. Kim, R. van Egmond, and R. Happee, "How manoeuvre information via auditory (spatial and beep) and visual UI can enhance trust and acceptance in automated driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 100, pp. 22–36, 2024.

[248] N. Ferson, E. Pizzigoni, F. Garnier-Follet, and C. Val, "A new methodology to model driver behaviour accounting for the variation in driving manner using naturalistic driving data," in *Proceedings: 26th International Technical Conference on The Enhanced Safety of Vehicles, Eindhoven, Netherlands*, 2019.

[249] G. Cohen-Lazry, A. Borowsky, and T. Oron-Gilad, "The effects of continuous driving-related feedback on drivers' response to automation failures," in *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting*, 2017.

[250] D. Albers, J. Radlmayr, N. Grabbe, S. Hergeth, F. Naujoks, Y. Forster, and K. Bengler, *Human-machine interfaces for automated driving: Development of an experimental design for evaluating usability*. Springer International Publishing, 2021.

[251] T. Hashimoto, A. Knauss, T. Aderum, O. Bostrom, T. Matsushita, D. Wag, E. K. Chung, and T. Hirose, "Benefits of intuitive auditory cues for blind spot in supporting personalization," National Highway Traffic Safety Administration (NHTSA), Tech. Rep., 2019.

[252] P. Ulleberg and T. Rundmo, "Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers," *Safety Science*, vol. 41, no. 5, pp. 427–443, 2003, ISSN: 09257535. DOI: 10.1016/S0925-7535(01)00077-7.

[253] D. Hennessy, *Social, personality, and affective constructs in driving*. Elsevier, 2011, Publication Title: Handbook of Traffic Psychology, ISBN: 978-0-12-381984-0. DOI: 10.1016/B978-0-12-381984-0.10012-8. [Online]. Available: http://dx.doi.org/10.1016/B978-0-12-381984-0.10012-8.

[254] L. Bernardi, J. Wdowczyk-Szulc, C. Valenti, S. Castoldi, C. Passino, G. Spadacini, and P. Sleight, "Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability," *Journal of the American College of Cardiology*, vol. 35, no. 6, pp. 1462–1469, 2000. DOI: 10.1016/S0735-1097(00)00595-7.

[255] D. Ni, "A Unified Perspective on Traffic Flow Theory Part I : The Field Theory," *Applied Mathematical Sciences*, vol. 7, no. 39, pp. 1929–1946, 2013.

[256] T. Kohn, M. Gottlieb, M. Schermann, and H. Krcmar, "Improving take-over quality in automated driving by interrupting non-driving tasks," in *Proceedings of IUI 2019*, 2019, pp. 510–517. DOI: 10.1145/3301275.3302323.

[257] L. Lorenz, P. Kerschbaum, and J. Schumann, "Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop?" In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2014.

ACKNOWLEDGEMENTS

I still clearly recall the day more than five years ago, anxiously participating in my PhD interview on Skype in a small rented room in Shanghai. I admit it was the worst interview I had ever given but amazingly it brought me to this point, right here, writing these words. Over the past five years, I have received so much care, encouragement, and support from the people around me. This PhD thesis would not have been possible without all of you. I am truly and deeply grateful.

First and foremost, I want to express my gratitude to my promotor, Riender Happee. I feel very lucky to have had you as my supervisor. Your extensive experience as a researcher greatly influenced my work, making it more solid. You are great at guiding your students and understand what we need. You were always there to encourage me, especially during my difficult times, often saying, "you are not the worst PhD I have had." Luckily, I did not know you like joking around at that time. For all of this and more, thank you, Riender.

I want to thank my copromotor Meng Wang. True to his name, he's both mighty and talented. Trust me, if you spoke Chinese, you would get it. I owe him a huge thanks for his patience with me who is anything but gifted. Meng made sure I was always on track, both in my research and in life, even when I determined to go off the rails. He was not just my guide through my PhD journey, but also helped me a lot when it came to my career, making sure I did not end up in trouble. Thank you, Meng, and all the best for what is next.

Let me express my sincere gratitude to my committee members: Gustav, Hans, Haneen, Jonas, and Joost. I truly appreciate that you agreed to join my committee, took time out of their busy schedules to read my thesis, and provided me with invaluable feedback.

I would like to extend my gratitude to a few individuals who have been not only collaborators but true mentors. Barys, thank you for your hands-on guidance in teaching me how to use the DAVSi simulator and for your unwavering support in refining the experimental setup, especially during the challenging COVID times. Jork, you are incredibly talented. I deeply appreciate your dedication to my experiments, teaching me to use the equipment, and helping me design the tools I needed. Freddy, thank you for helping me quickly get into the research groove, and for giving me the opportunity to gain experience in publishing in a top journal. Sina, I have always admired your research style, writing skills, and efficiency, qualities that have pushed me towards becoming an independent researcher. Chang, I still vividly remember those countless late-night discussions and the relay-like writing marathons, complete with ringing phone messages deep into the night. And of course, who could forget the elderly man next to HEMA? Soyeon, I feel very fortunate to have had you as part of my PhD journey. Thank you for your selfless dedication during our collaboration, especially for taking on almost all the work when I was absent. Conversations with you were always a breath of fresh air. And, I must say, I am even luckier to still be alive after that car ride you drove to Den Haag five

years ago. Xinwei, your contributions to our collaborative research have been immense. But beyond that, you have always acted like an older brother, offering care and encouragement when I needed it most. Zirui, I do not need to elaborate on how much you have helped me academically. Thank you so much for everything. I still remember vividly all the places we explored together around the world. And next time, please do check if the flight is to Dublin, not to Berlin!

Next, I must thank my friends and colleagues at the Intelligent Vehicle group. First off, a shout-out to my dear office mates in E-0-290. The Plank Street Boys and my unofficial English trainers Alberto, Hidde, and Wilbert. You were always so full of energy, making E-0-290 a place of joy and cultural exchange. My English got better thanks to you, and eventually, I even became the noisiest one in the office. Thank you all for the laughter and the memories. András, our resident entrepreneur, wish you all the success, though I still firmly believe psychology is a science! Ewoud, you were already an experienced hand when I was just starting my PhD. Thanks for being a great role model. Zimin, possibly the most focused person in the office (at least from what I could see from my angle). Thank you for all the dinners we had together on campus, the countless late nights spent together in the office, and all those much-needed talks. All the best to you. Guopeng, though we never managed to be in the office at the same time, we were 'time-space companions.' Thank you for all the casual talks during the final stage of my PhD. I also want to acknowledge the rest of my close colleagues in this group. Yanggu, I still miss our daily remote talks during COVID. Your outlook on current affairs and life left me truly impressed. I miss those fried chicken and barbecue days. Wishing you all the best! Varun, you are such a kind and easygoing person; I always felt a sense of warmth after talking with you. Vishrut, you are an optimistic and fun guy, and every moment spent with you has been full of laughter. Chrys, you are a warm and open-hearted person and also our first female PhD candidate, a milestone worth celebrating! Shiming, please allow me to include Shuhan's name here as well. Thank you for your endless care, be it in life, health, or spirit. Thank you for always taking me in during holidays and even my own birthday when I was "homeless". I believe you will handle any challenges with the same strength and resilience that you have always shown. Tuğrul, a fantastic researcher, a great presenter, and above all, a person of unbreakable spirit. You always filled me with motivation. Farzam, thank you for your invaluable advice during the final phase of my thesis. Oscar, those lunchtime and after-work talks helped me learn a lot more about the Netherlands. Yancong, you are such a dedicated worker and researcher. I truly miss the times we went out for drinks, travelled, and watched football games together. Wishing you all the best in your career. Mubariz, you are an outstanding researcher, leader, and also a professional badminton player who destroyed my confidence on the court. Thank you for all the insightful conversations. Ted, wishing you all the best, both in your PhD journey and your weightlifting career. Ksander, Joris, Mojtaba, Nick, thank you for your contributions during the early days of my simulator experiments. Jetze, it was clear you are full of great ideas. Thank you for seamlessly taking over the section meeting coordinator role from me. Raj, a man of few words but great strength, you taught me that actions always speak louder than words. Marko, you are always a pleasure to talk to, and I admire your photography skills. Thank you for all the enjoyable conversations and for bringing so much positivity to the group. Ronald, I learned a lot about Dutch cul-

ture from our conversations, and you happen to know more Chinese than anyone else around, aside from the native speakers, so go ahead and eat more! Mario, your insights after the section meeting genuinely inspired me regarding my research. Last but not least, many thanks to Dariu, Georgios, Holger, and Julian for your contributions to the group. You have all made it a fantastic place to work.

I want to express my gratitude to the SHAPE-IT family. Special thanks go to Jonas and Jacqueline, who were there for their exceptional coordinating and project management. To the Supervisory Board, Prof. Jonas Bärgman, Prof Klaus Bengler, Natasha Merat, Prof. Joost de Winter, Prof. Marjan Hagenzieker, Prof. Natasha Merat, Prof. Gustav Markkula, Prof. Martin Baumann, Prof. Olga Pollatos, Prof. Marco Dozza, Prof. Christian Berger, Dr. Yee Mun Lee, Dr. Chongfeng Wei, and Dr. Eric Knauss, thank you for your care and invaluable feedback on my work. To my fellow ESRs: Nikol, Naomi, Chi, Yue, Chen, Yuan-Cheng, Amna, Wilbert, Siri, Sarang, Amir Hossein, Ali, and Xiaomi, I feel so lucky to have spent some of the most important years of my life alongside you all in the SHAPE-IT project. I cherished every moment we travelled, hiked, attended conferences, and trained together. You are all incredibly talented people, and I wish you nothing but success in everything you pursue in the future.

I also want to thank my TTS lab colleagues: Marjan and Haneen, as the directors of the TTS lab, and also my Go/NoGo committee, you treated me as if I were your own PhD student, always offering incredible support and care. I am deeply grateful for your guidance, understanding, and the kindness you showed me throughout my PhD journey. To Siri, Sina, Nagarjun, Yongqi, Johan, Jinyang, Samir, Yiyun, Dennis, Narayana, and Solmaz, thank you for being there and making the lab such a welcoming place.

To the friends I made here in Delft: Meiqi, Jing, Ruochen, Yuxia, Zi'ang, Peng, and Chun-Ting, thank you for being there and for all the amazing memories we made. Your company has been invaluable during this journey.

To my colleagues at TME during my secondment: Audrey, Felipe, Niels, Daniel, Maminirina, Davide, Marco, Kevin, Luca, Stefano, Bastien, and David, thank you for giving me warmth during the winter of 2022 in Brussels. You truly made it feel like home away from home.

I am also incredibly grateful to my colleagues from TU Dresden during my visit there: Heike, Lutz, Xinyu, Sven, Yikai, Henning, Sebastian, Runhao, Lei, Xiaocong, Menglin, Liying, Django, Angelika, Jyotirmaya, Matthias, Tobias, Azizur, and Susanne. You made me feel welcome and made my three months in Dresden unforgettable.

To my new group and office mates in the Department of Transportation and Planning, Ali, Srinath, Jonah, Wouter, Mariko, Yiru, Xue, Kexin, Samir, Saeed, Yiyun, Lucas, and Chaopeng, thank you for giving me a new home and for making me feel part of the group. A special thanks to Hans for taking me in when I was on the brink of unemployment and for bring me into a new research field, rescuing someone who was about to become truly jobless.

I also want to thank all my participants who took part in my experiments. This thesis wouldn't have been possible without you.

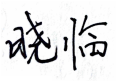
To my old friends: Hongfeng and Jiawei, I am grateful for the care and concern you have always shown me, and for the countless talks we had over the internet. Wenqi, Qiyu, and Xiao, thank you for taking care of my parents while I was away. Your support

during that difficult time meant the world to me, and I am forever grateful. Bingxin, thank you for the late-night chats over video calls that helped keep me sane and happy. Weijian, thank you for your unwavering support, all your suggestions for visualisations, and designing the cover for this thesis.

To my parents, thank you for your endless support. I know my choices and life path have been not conventional. As your only son, I could not be there with you, and as the years go by, you face challenges far greater than mine, yet you never complained and always supported me wholeheartedly. I hope I can make you proud and repay even a small part of all that you have given me.

Forgive me for not being able to mention every single person who has helped me along the way. Thank you all.

And finally, I want to thank myself for never giving up, for persevering when it would have been easier to quit. May this thesis be not the end of my journey in exploring the world, but rather a beginning. Even more, I wish for this to be more than just a hope.



October, 2024

CURRICULUM VITÆ

Xiaolin HE

21-10-1993 Born in Linyi, Shandong Province, China.

EDUCATION

2009–2012 High School
Yishui County No. 2 High School

2012–2016 Bachelor of Science in Vehicle Engineering
Jilin University

2016–2019 Master of Science in Vehicle Engineering
Tongji University

2020–2024 PhD in Human Factors in Automated Vehicles
Marie Sklodowska-Curie Early Stage Researcher in SHAPE-IT project
Delft University of Technology
Project title: AV occupants' perception of safety in relation to AV behaviour
Thesis: Driver's perceived risk in relation to automated vehicle behaviour
Promotor: Prof. dr. ir. R. Happee
Promotor: Prof. dr. M. Wang

EXPERIENCES

2013–2016 Student Engineer
JLU Conception Eco-racing Team of Jilin University - Changchun, China

2019-2020 Parts and Testing Engineer for Automated Parking Systems
SAIC Volkswagen Automotive Co., Ltd. - Shanghai, China

2022 Research Intern
Toyota Motor Europe - Brussels, Belgium

2023 Visiting Researcher
Technische Universität Dresden - Dresden, Germany

PROFESSIONAL MEMBERSHIPS AND RESPONSIBILITIES

2023	Vice-Chair IEEE Intelligent Vehicle Risk Assessment Working Group 3439 (ITSS/SC/IVREWG)
2023	Workshop (WS07) Co-Organiser IEEE Intelligent Vehicles Symposium
2020–2023	Lecturer and Teaching Assistant "Human Factors of Automated Driving" Master Course, TU Delft
2021	Co-Supervisor Bachelor Graduation Project, TU Delft: "The Influence of a Colour-Themed HMI on Trust and Take-Over Performance in Automated Vehicles"

AWARDS

2024	The 3rd Place Prize in ROAD mode of TRA VISIONS 2024 Young Researcher Competition

List of Publications

Journal Papers

10. **X. He**, R. Happee, M. Wang, "A new computational perceived risk model for automated vehicles based on potential collision avoidance difficulty (PCAD)," *Transportation Research Part C: Emerging Technologies* **166**, 104751, DOI: 10.1016/j.trc.2024.104751 (2024). (Chapter 3 of this thesis)

9. S. Kim, **X. He***, R. van Egmond, R. Happee, "Designing user interfaces for partially automated vehicles: effects of information and modality on trust and acceptance," *Transportation Research Part F: Traffic Psychology and Behaviour* **103**, 404-419, DOI: 10.1016/j.trf.2024.02.009 (2024). * Co-first author (Chapter 5 of this thesis)

8. **X. He**, J. Stapel, M. Wang, R. Happee, "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, DOI: 10.1016/j.trf.2021.08.002 (2022). (Chapter 2 of this thesis)

7. **X. He**, Z. Li, X. Wang, R. Happee, M. Wang, "Reading minds on the road: Decoding perceived risk in automated vehicles through 140K ratings". In revision. (Chapter 4 of this thesis)

6. C. Lu, **X. He***, H. van Lint, H. Tu, R. Happee, M. Wang, "Performance evaluation of surrogate measures of safety with naturalistic driving data," *Accident Analysis & Prevention* **162**, 106403, DOI: 10.1016/j.aap.2021.106403 (2021). * Co-first author

5. S. Nordhoff, J. Stapel, **X. He**, R. Happee, "Do driver's characteristics, system performance, perceived safety, and trust influence how drivers use partial automation? A structural equation modelling analysis," *Frontiers in Psychology* **14**, 1125031, DOI: 10.3389/fpsyg.2023.1125031 (2023).

4. A. El Jouhri, A. El Sharkawy, H. Paksoy, O. Youssif, **X. He**, S. Kim, R. Happee, "The influence of a colour themed HMI on trust and take-over performance in automated vehicles," *Frontiers in Psychology* **14**, 1128285, DOI: 10.3389/fpsyg.2023.1128285 (2023).

3. S. Nordhoff, J. Stapel, **X. He**, A. Gentner, R. Happee, "Exploring the Factors of Perceived Safety and Trust in SAE Level 2 Partially Automated Cars Using Principal Component Analysis," *SSRN 4156405*.

2. S. Nordhoff, J. Stapel, **X. He**, A. Gentner, R. Happee, "Perceived safety and trust in SAE Level 2 partially automated cars: Results from an online questionnaire," *PLOS One* **16**(12), e0260953, DOI: 10.1371/journal.pone.0260953 (2021).

1. F. A. Mullakkal-Babu, M. Wang, **X. He**, B. van Arem, R. Happee, "Probabilistic field approach for motorway driving risk assessment," *Transportation research part C: emerging technologies* **118**, 102716, DOI: 10.1016/j.trc.2020.102716 (2020).

CONFERENCE PAPERS

2. **X. He**, R. Happee, M. Wang, "Computational perceived risk models in SAE Level 2 driving automation: formulation, validation and comparison with experimental data," Poster paper presented at *2023 TRB Annual Meeting*, Washington DC, USA.

1. **X. He**, H. Chen, J. Chen, W. Ran, Y. Nishimura, K. Ando, "Evaluation and Optimization of Driver Steering Override Strategy for LKAS Based on Driver's Acceptability," *2018 Intelligent and Connected Vehicles Symposium*, Kunshan, Jiangsu, China.

PROPOSITIONS ACCOMPANYING THE DISSERTATION

1. Perceived risk in automated driving is much more influenced by interactions with other vehicles than by differences in user's personal characteristics. (Chapter 2)
2. Modelling perceived risk in automated driving based on modelling collision risk directly addresses the core concern of people during driving. (Chapter 3)
3. Perceived risk in automated driving arises not only from imminent collisions but also from uncertainties in the manoeuvres of neighbouring vehicles and the control of the ego vehicle. (Chapter 4)
4. Enhancing the transparency of automated driving systems can reduce perceived risk but cannot eliminate it. (Chapter 5)
5. Enforcing a 40-hour workweek limit at universities enhances scientific quality.
6. The distinction between scientists and artists lies in their objectives for processing data.
7. Preconception is a form of overfitting in the brain and is difficult to detect.
8. Dutch authorities believe that highly skilled migrants are excellent at driving.
9. Epidemic diseases have a more profound impact on human society compared with wars.
10. Removing propositions from TU Delft doctorate regulations slows down science.

These propositions are regarded as opposable and defendable, and have been approved as such, by the promotors prof. dr. ir. R. Happee and prof. dr. M. Wang.