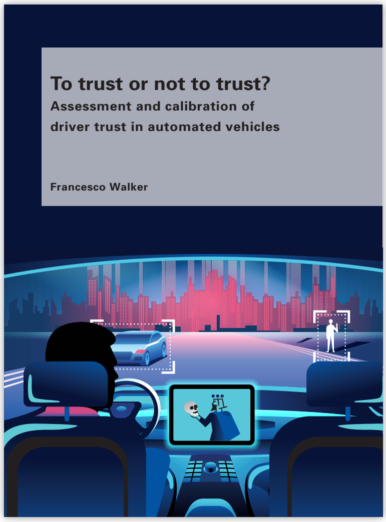
# To trust or not to trust?

**Assessment and calibration of driver trust in automated vehicles**

The studies reported in this dissertation were conducted as part of i-CAVE (Integrated Cooperative Automated Vehicle) – an NWO Perspectief programme in which a large consortium of universities, research institutes and industrial partners worked together to design, develop and test a highly automated passenger vehicle for use on the Eindhoven University of Technology campus. The theme of the thesis is driver trust in such automated vehicles and the way this affects driver-vehicle interaction.




**Naam:** **Dr. Francesco Walker**
**Promotie:** **Transport Engineering and Management (TEM), University of Twente**
**Promotors:** **prof. dr. Marieke Martens, prof. dr. ir. Willem Verwey**
**E-mail:** **f.walker@fsw.leidenuniv.nl**

Trust plays a key role in determining if and how drivers will use automated technology: under-trust may lead to drivers not using an automated driving system, preventing them from taking full advantage of its comfort and safety features. Conversely, over-trust may lead to drivers using the system beyond its capabilities. Failure to use automation and its misuse can be seen as two extremes of a spectrum. Drivers' position on this spectrum may change, depending on factors such as knowledge and expectations, and the perceived reliability of the vehicle. Ideally, trust should be well-calibrated, meaning that drivers' position on the disuse-misuse spectrum should be continuously aligned with the true reliability of the automated system. Yet, few studies have addressed how trust calibration can be achieved and measured, or identified situations leading to increases or decreases in trust. The studies reported in this thesis aim to partially fill these gaps, using on-road studies and simulated driving scenarios to investigate the way drivers calibrate their trust.

## The trust cycle

What is trust? In the literature, trust towards automated systems is commonly described as multilayered (Ghazizadeh et al., 2012; Hoff & Bashir, 2015; Lee & See, 2004). Nonetheless, there is still confusion and disagreement on the definition of the layers and their interactions. To shed light on this issue, we created a conceptual model integrating the most influential trust frameworks proposed in the literature. In this model, stable predispositions (dispositional trust), and preliminary information about the system (initial learned trust)both impact drivers' situational trust – the trust shown by a user in specific situations at specific times. However, engineering interventions can also influence the trust cycle by modifying vehicle behavior and the information provided by the Human Machine Interface (HMI). The use of the system in different situations determines drivers' dynamic learned trust, which develops continuously as users experience the behavior of the automated vehicle in different driving scenarios. One possible result is that the driver learns to trust the system (dynamic learned trust), but only in certain situations and under certain circumstances (situational trust).

Overall, the development of an appropriate level of trust should be seen as a cycle, in which dynamic learned trust and vehicle behavior feed back into situational trust, continuously impacting user's interaction with the system (Figure 1). In the model, each round of the cycle ends with a trust calibration assessment, verifying whether there is a mismatch between situational trust and vehicle behavior. The presence of such a mismatch is a sign that users' trust towards the automated driving system is not well calibrated.

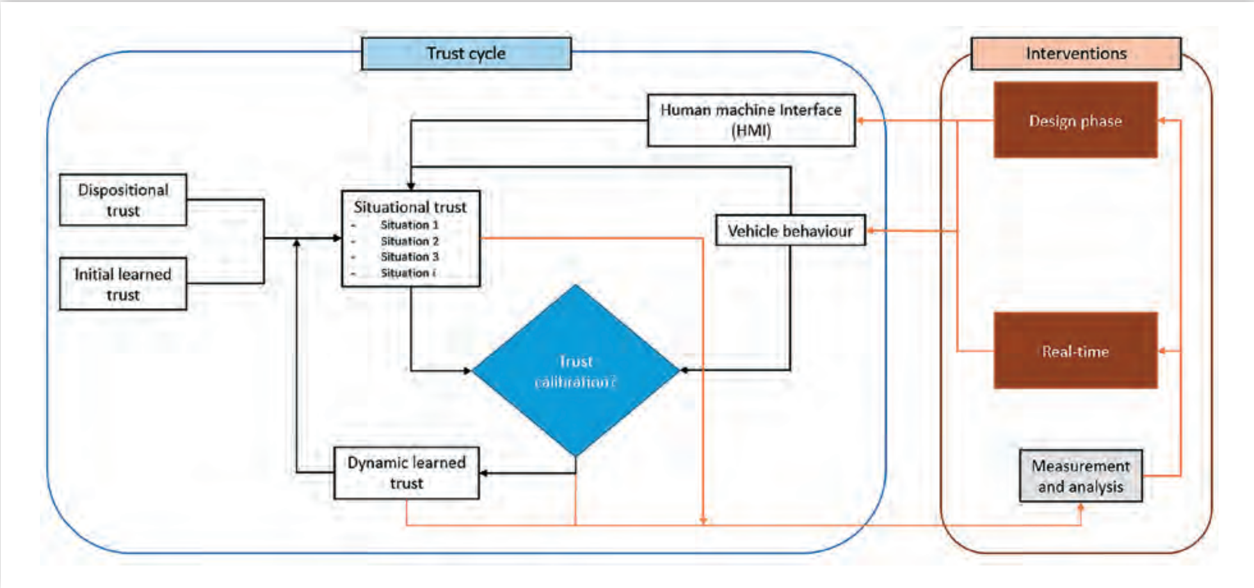



Figure 1. A conceptual model of the dynamic development of calibrated trust (based on Ghazizadeh et al., 2012; Hoff & Bashir, 2015; Lee & See, 2004) and the role of possible engineering interventions.

### Related research questions

Situational trust and dynamic learned trust are key elements in the model. The dissertation addressed four related research questions.

#### What is the effect of on-road experience with Level 2 systems on drivers' trust?

To answer this question, we gave drivers the opportunity to drive cars equipped with commercially available automated driving technology, on-road. We found that before driving the vehicles, drivers tended to overestimate the vehicles' capabilities (Figure 2). On-road experience gave drivers a better understanding of what the vehicles could and could not do. In particular, users learned to assign a certain level of situational trust to specific driving scenarios. In most cases, driver trust was lower after their experience than before. However, it was also better calibrated with the actual reliability of the

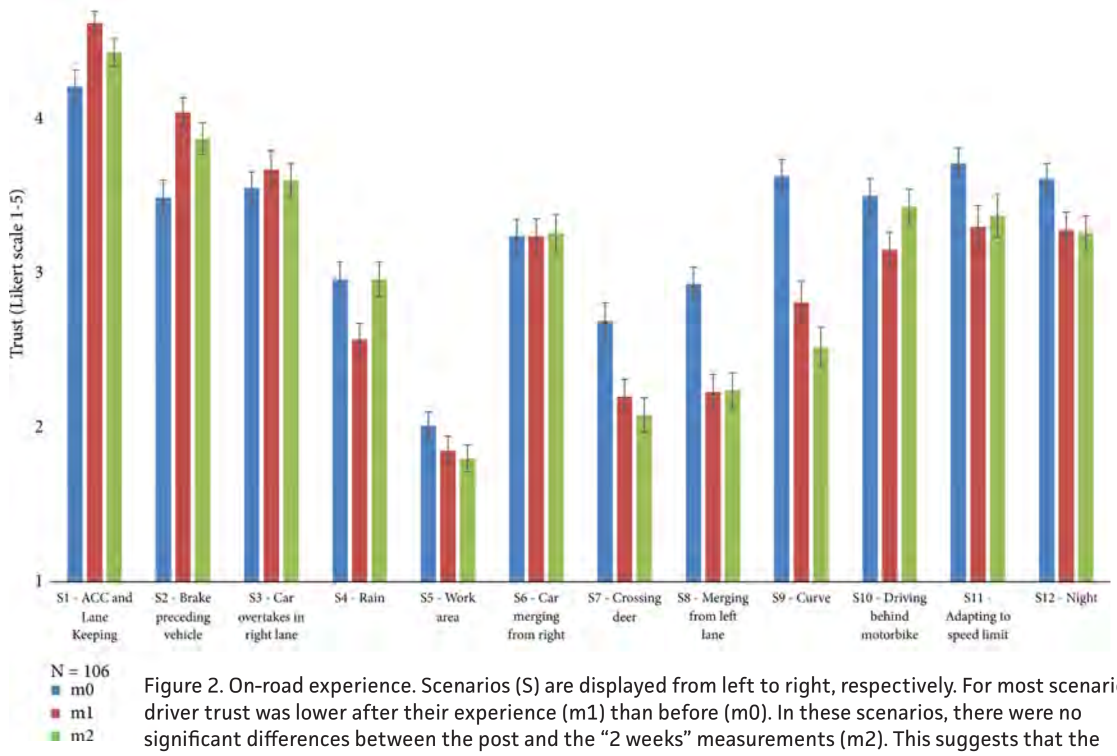



Figure 2. On-road experience. Scenarios (S) are displayed from left to right, respectively. For most scenarios, driver trust was lower after their experience (m1) than before (m0). In these scenarios, there were no significant differences between the post and the "2 weeks" measurements (m2). This suggests that the impact of the driving experience on drivers' trust was relatively stable.

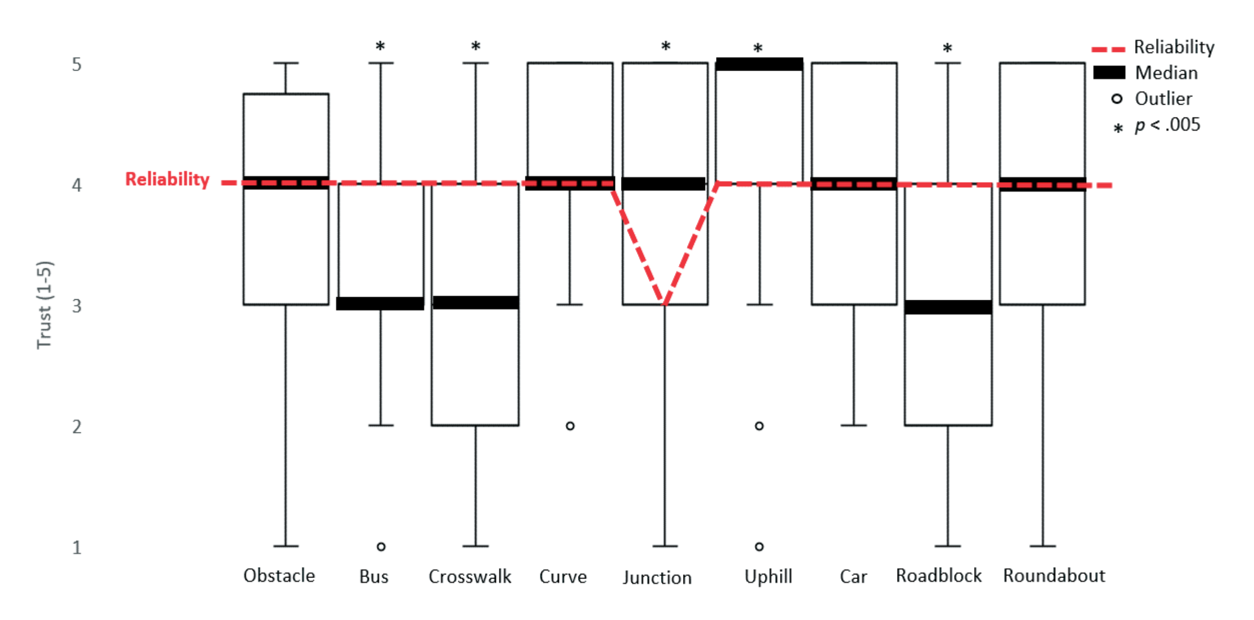



Figure 3. Comparison between scores for user (trust) and engineer (reliability). In specific situations, users' trust levels do not match engineer evaluations of vehicle reliability. The box and whiskers represent user trust in different scenarios. The dashed red line shows engineers' assessment of vehicle reliability in the same scenarios.

automated driving system. This shows the impact of experience in different situations (situational trust) on drivers' dynamic learned trust. Importantly, trust calibration improved even in scenarios that drivers never encountered during their experience on the road. This suggests that drivers build a mental model of vehicle capability that they update continuously based on their experience. This mental model allows them to form expectations about vehicle behavior, without having to experience all the possible situations that an automated system may encounter. Though we did not observe this in our study, it is nonetheless possible that in some circumstances (i.e. when the vehicle is unable to handle what appears to the driver as a "simple scenario") expectations could also be incorrect.

## Can driving simulator studies provide valid information about user behavior?

Driving simulator studies make it possible to investigate user behavior and trust in vehicles that are not ready for road-testing, and situations that cannot be safely investigated on the road. However, such studies create no physical risk for the driver. It is plausible, therefore, that drivers' behavior in the simulator will differ from their behavior on the road. The dissertation reports a simulator-based study that addresses this concern. In the study, participants reported a strong sense of presence – the feeling of truly belonging in the virtual environment – whether or not they were exposed to risk (i.e. the threat of receiving a mild electric shock in case of accident). Presence is a necessary requirement for validity and suggests that driving simulators can indeed be used to explore user behavior, despite the absence of actual (physical) risk.

## How can studies of user trust in specific driving situations inform vehicle design?

In another driving simulator study, we showed that users' trust levels in specific situations often fail to match engineers' evaluations of vehicle reliability (Figure 3). We also found that user responses included suggestions for specific changes in the design of the HMI. For example, users suggested that the HMI should provide a visual representation of the objects detected by the automated driving system in its surroundings and real-time feedback on vehicle performance. Some users also complained about specific aspects of vehicle behavior (e.g., speed, acceleration, lateral control), suggesting opportunities for improvement. These results demonstrate that studies of trust can provide useful input to vehicle design and support the well-recognized need for deeper collaboration between engineers and Human Factors researchers.

## How can trust in automated vehicles be measured objectively and in real-time?

Good trust calibration requires that a driver's position on the trust spectrum aligns with the actual reliability of the system in the situation the driver is experiencing. However, it is difficult to detect adjustments in trust in continuously changing circumstances using just questionnaires. A study reported in the dissertation demonstrates the possibility of monitoring trust in real-time using measures of driver electrodermal activity (EDA) and gaze behavior. We found that participants' monitoring behavior and EDA correlate with their self-reported trust in the automated system (Figure 4). We also show that the combination of the two measures provides a better measure of trust than either indicator taken singly.

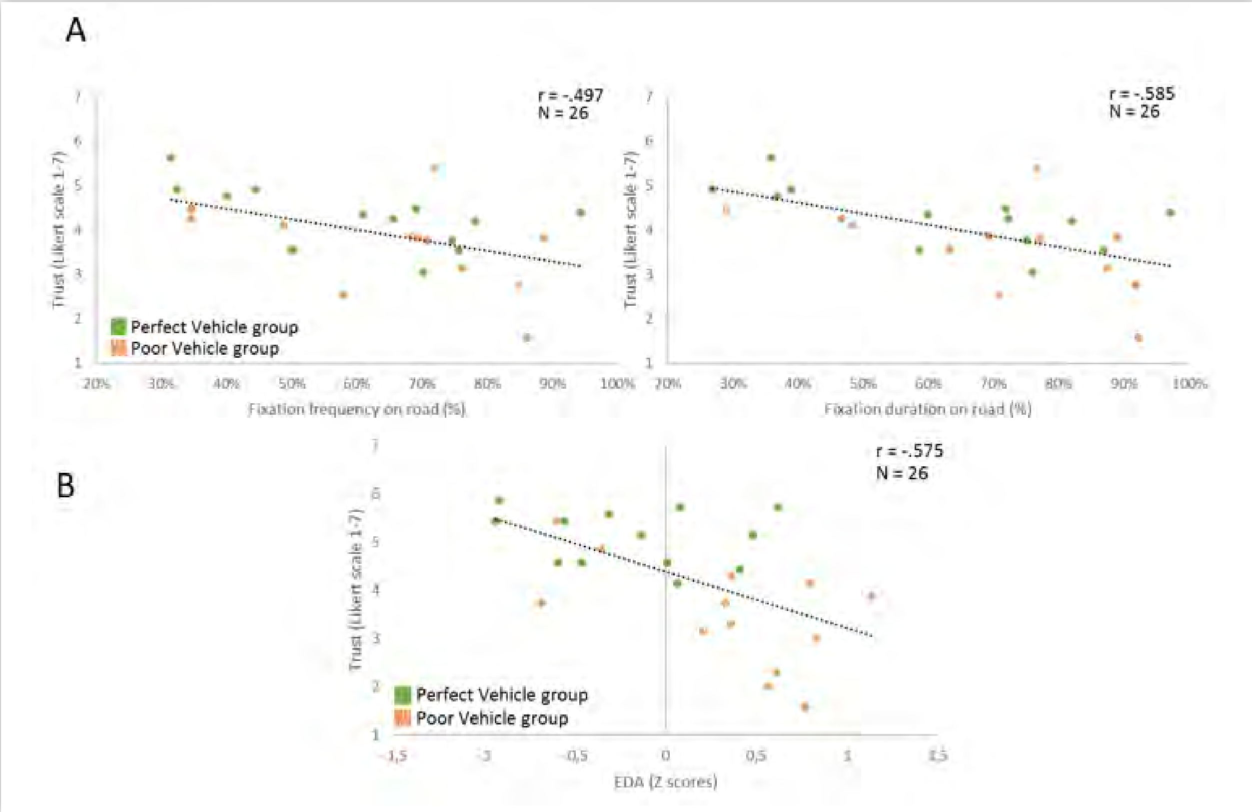



Figure 4. A: Correlation between self-reported trust and monitoring behavior. The higher the trust, the less participants monitored the road. B: Correlation between self-reported trust and Electrodermal Activity (EDA). The higher the trust, the lower participants' EDA.

Interestingly, observed changes in trust seem to have been driven not so much by the objective reliability of the system, but by its predictability. Participants who experienced a system that predictably behaved in an unreliable way (i.e., constant swerving, hard brakes) reported increased trust in the automated vehicle. Conversely, participants who experienced unpredicted unreliable behavior showed reduced trust. These results show that the predictability of automation failures strongly influences users' dynamic learned trust when interacting with automated driving systems.

## Conclusion

In summary, our findings show that:

1. user perceptions concerning the reliability of automated driving systems change significantly after on-road experience;
2. driving simulators can provide an effective sense of presence in their users. This is evidence that simulators can be validly use to investigate trust;
3. user trust is often misaligned with engineer assessments of vehicle reliability;
4. combined measures of gaze behavior and EDA can lead to an effective assessment of user trust in the automated driving system.

On the basis of these results, we suggest that reliable technology is not sufficient for automated vehicles to fulfil their promise. Safety and market acceptance of automated vehicles depends strongly on the attitudes and behaviors of their potential users. In particular, it is crucial that users develop appropriate levels of trust towards the technology. Cooperation between engineers, vehicle designers and Human Factors researchers can make a vital contribution to achieving this goal.

## References

Ghazizadeh, M., Lee, J.D., & Boyle, L.N. (2012). Extending the Technology Acceptance. Model to assess automation. *Cognition, Technology and Work, 14*(1), 39-49. https://doi.org/10.1007/s10111-011-0194-3.

Hoff, K.A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407-434. https://doi.org/10.1177/0018720814547570.

Lee, J.D., & See, K.A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392.

***Uit het juryrapport***
The results make a good contribution to application through recommendations for self-driving/automated cars.