On-road trust and perceived risk in Level 2 automation

Stapel, Jork; Gentner, Alexandre; Happee, Riender

DOI
10.1016/j.trf.2022.07.008

Publication date
2022

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Transportation Research Part F: Psychology and Behaviour 89 (2022) 355–370

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

On-road trust and perceived risk in Level 2 automation

Jork Stapel [a], Alexandre Gentner [b], Riender Happee [a]

[a] TU Delft University of Technology, the Netherlands
[b] Toyota Motor Europe, Brussels, Belgium

ARTICLE INFO

Keywords:
Driver state monitoring
Driver trust
Perceived safety
Automated driving
2000 MSC:
0000
1111

ABSTRACT

To encourage appropriate use of driving automation, we need to understand and monitor driver's trust and risk perception. We examined (1) how trust and perceived risk are affected by automation, driving conditions and experience and (2) how well perceived risk can be inferred from behaviour and physiology at three levels: over traffic conditions, aggregated risk events, and individual risk events.

30 users with and without automation experience drove a Toyota Corolla with driving support. Safety attitude, subjective ratings, behaviour and physiology were examined.

Driving support encouraged a positive safety attitude and active driver involvement. It reduced latent hazards while maintaining saliently perceived risks. Drivers frequently overruled lane centring (3.1 times/minute) and kept their feet on or above the pedals using ACC (65.8% of time). They comfortably used support on curvy motorways and monotonic and congested highways but less in unstable traffic and on roundabouts. They trusted the automation 65.4%, perceived 36.0% risk, acknowledged the need to monitor and would not engage in more secondary tasks than during manual driving.

Trust-in situation reduced 2.0% when using automation. It was 8.2% higher than trust-in-automation, presumably due to driver self-confidence. Driving conditions or conflicts between driver and automation did not affect trust-in-automation.

At the traffic condition level, physiology showed weak and partially counter-intuitive effects. For aggregated risk events, skin conductance had the clearest response but was discernible from baseline in $< 50\%$. Pupil dilation and heart rate only increased with strong braking and active lane departure assist. For individual risk events, a CNN classifier could not identify risk events from physiology. We conclude that GSR, heart rate and pupil dilation respond to perceived risk, but lack specificity to monitor it on individual events.

1. Introduction

Trust is a vital factor in the successful adoption of automated driving (Ghazizadeh et al., 2012). Trust is defined 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, irrespective of the ability to monitor or control that other party'* (Mayer and Davis, 1995). If the automation is given too much trust, drivers may misuse it or become complacent which negatively affects monitoring and take over performance (Brookhuis and de Waard, 2001). Under-trust may conversely lead to disuse (van Huysduynen et al., 2018; Lee et al., 2021), reducing benefits in road safety, sustainability and comfort. Trust calibration can therefore be an effective strategy to encourage appropriate use (Cabrall et al., 2019) and understanding a system's trust propensity is essential to evaluate its overall safety implications.

Trust calibration requires an appropriate understanding of how trust develops, what influences it and how it affects behaviour. Marsh and Dibben (2003) identified three layers of trust: dispositional, situational and learned trust. Hoff and Bashir (2015) extended

https://doi.org/10.1016/j.trf.2022.07.008
Received 6 May 2022; Received in revised form 29 June 2022; Accepted 10 July 2022
Available online 19 July 2022

this to trust in automated vehicles and considered the influence of personal characteristics in the three layers. Empirical models have also been developed. Kraus et al. (2020) investigated trust development and recovery following system faults in two driving automation studies, demonstrating that trust increased along with knowledge accumulation. Hu and Wang (2021) propose a dynamic model of trust in adaptive cruise control (ACC) which integrates errors between desired and actual ACC behaviour. The driver is expected to engage or disengage ACC when certain trust thresholds are exceeded. Kolekar et al. (2020) presented the driving risk field capturing perceived risk induced by static and dynamic obstacles. The model was calibrated through driver's verbal ratings and steer responses to obstacles in manual simulator driving.

Active, closed-loop trust calibration requires momentary trust to be estimated or monitored. Several reliable trust measurement techniques exist. Subjective ratings have been regarded as the ground truth for trust measurement and can be queried through surveys, rated verbally or continuously through handset controls (Beggiato et al., 2019; Cleij et al., 2018). Driver state monitoring in consumer vehicles however requires less intrusive techniques. Behaviours such as monitoring, intervening or engaging in other activities can reflect automation reliance and compliance. Naturalistic braking profiles have also been used to cluster near-miss events to different risk levels (Xiong et al., 2019). Minimum time to collision (TTC) or maximum looming can show the driver's acceptance threshold of perceived risk (Kiefer et al., 2005).

Physiological indicators have also been used to monitor trust and related constructs such as anxiety or activity in the parasympathetic nervous system. Morris et al. (2017) compared driver trust to galvanic skin responce (GSR) in different automation driving modes, finding lower skin conductance and higher trust when the automation drives safely compared to when taking risks. Ajenaghughrure et al. (2021) examined GSR, electrocardiogram (ECG), electroencefalogram (EEG) and electrooculogram (EOG) in simulated driving and identified brain activity and gaze as the most robust indicators of trust in driving automation. Healey and Picard (2005) found that an increase in heart rate indicates stress (though Beggiato et al. (2019) found the opposite). Tang et al. (2018) showed that pupil diameter increases significantly faster in more severe crashes. For further information see a review by Lohani et al. (2019) on established relations between various physiological metrics and driver states.

While physiology is promising for driver state monitoring, there are some concerns about its utility outside carefully controlled experiments (Jorna, 1992; Fairclough, 2009; Heikoop et al., 2015). Most literature that identified relations between trust and physiology is performed in a controlled setting which minimises confounding factors through design and averaging across repeated events. Even though several effects are found reliably across studies (Panicker and Gayathri, 2019), similar responses to different constructs are also demonstrated (Khan and Lee, 2019). It therefore remains challenging to apply physiology in trust monitoring applications because of the lack of construct specificity of physiological signals (Byrne and Parasuraman, 1996). It also remains unclear if physiology is best examined over longer periods, aggregated over repeated events or can be processed in real-time.

In this study, we examine the safety attitude towards a driver support system of a commercially available vehicle, evaluate how automation use, driving conditions and prior automation experience affect trust and risk perception, and analyse the usefulness and limitations of physiology in on-road monitoring of risk perception. We explore how well perceived risk can be inferred from behaviour and physiology at three levels. At the first level, we compare physiological metrics averaged over periods of driving in conditions characterised by different levels of trust. At the second level we compare the average physiological response to aggregated risk events to periods without such events. At the third level, we examine the discriminative power of the same metrics for individual risk events.

We investigate on-road trust and perceived risk in SAE Level 2 driver support, answering the following questions:

1. How do trust and perceived risk depend on driving conditions, automation use and automation experience?
2. How do behavioural indicators and physiological measures reflect trust and perceived risk during critical events and automation conflicts?

The paper is structured as follows. The experimental procedure is described in Section 2 with further details in Appendix A to Appendix C on data pre-processing, questionnaires and forms. For the results in Section 3, we first describe the drivers' safety attitude and perceived safety. We then examine how trust and perceived risk were affected by automation use, traffic conditions and automation experience in Section 3.3, and the physiological response to these conditions in Section 3.4. Section 3.5 provides results on the sensitivity of physiological indicators to aggregated and individual risk events with further details in Appendix E. The findings are discussed in Section 4 and we reflect on our research questions in Section 5.

2. Methods

2.1. Equipment

Participants drove a 2020 Toyota Corolla hybrid with Toyota Safety Sense 2 (TSS2) driver support. TSS2 includes full-range adaptive cruise control (ACC) for speeds between 0 km/h and 180 km/h and steering support, featuring lane departure warning (LDW; alerting the driver when crossing or about to cross lane markings unintentionally), lane departure assist (LDA; applies momentary steer input to avoid unintended lane departure) and lane trace assist (LTA; applying continuous, limited steering forces to guide the driver towards the centre of the lane). LDW and LDA can be enabled when driving approximately 50 km/h or faster. LTA can be enabled whenever ACC is enabled. All steer support systems require the vehicle to detect lane markings, though LTA can also operate in car following mode. The steer assist features are not able to keep the vehicle inside the lanes without the driver's continuous aid. LTA disengages temporarily when driver steering inputs disagree with the LTA reference. All steering support can be easily overruled by the driver and disengages while the turn indicator is operational. When the vehicle does not detect steering wheel inputs,

LTA will start to swerve and ultimately turns off. Once enabled, LDW, LDA and LTA function whenever conditions are met. ACC has to be explicitly re-activated whenever it turns off, for instance after driver braking.

2.2. Instrumentation

Physiological measures were selected based on personal curiosity, relevance in literature and the authors' confidence in correct recording and interpretation within the scope of this study. Heart rate (chest; lead II) and GSR (left foot) were logged with a MindMedia NeXus 4 and disposable electrodes. Pupil dilation was observed using a Tobii Pro Glasses 2 head mounted eye tracker.

Following the same philosophy as Hartwich et al. (2015), risk ratings were recorded continuously, where they are expected to be less influenced by memory effects compared to post-drive ratings. The continuous rating was provided using a small force sensing resistor connected to the driver's right thumb, allowing continuous operation while holding the steering wheel. Resistance was mapped to a risk scale from 0 to 10, with weak forces suppressed to zero to filter out forces naturally occurring while holding the steering wheel. Squeezing the pressure sensor signified a higher perceived risk, in accordance to the natural tendency in uncomfortable situations (Hartwich et al., 2015). Visual feedback on this scale was provided to the participant with an LED bar consisting of 10 levels (Fig. 1). We have used this setup successfully in a prior driving simulator study (He et al., 2022).

A 360 action camera mounted on the roof of the vehicle recorded the vehicle surroundings at 0.3 Hz. A webcam observed right foot placement in the pedal bay. The webcam was connected to the experimenter's laptop which also logged vehicle speed using an OBD-II logger, physiological data and ratings of trust and perceived risk.

Vehicle CAN bus data was logged and decoded by a proprietary CAN logger provided by Toyota.

Further details on signal pre-processing can be found in Appendix A.

2.3. Questionnaires

Participants filled an informed consent form (Appendix C). A pre-drive questionnaire collected participant demographics, prior experience with driver-support features, 10-item big 5 personality (Rammstedt and John, 2007), ADSES self-efficiency (George et al., 2007), a-priory willingness to use (Payre et al., 2014) driving automation and a priori trust in automation (Jian et al., yyyy).

A post-drive questionnaire collected trust in automation (as above), reliance and reliability (Choi and Ji, 2015; Nordhoff et al., 2021), perceived risk (Nordhoff et al., 2021), willingness to use ACC and steer support for various conditions, understanding of the instrument panel (Nordhoff et al., 2021) and continuous risk rating device. The full questionnaires are in Appendix B. Five or seven-point Likert scales anchored between strongly disagree and strongly agree were converted to a percentage of agreement for analysis and presentation.

2.4. Experimental procedure

The experiment was approved by the TU Delft Human Research Ethics Committee.30 participants (7 female; age 33 (std. 14); mileage past 12 months 6000 km (std. 6000 km)) took part in this experiment. 18 had prior experience using ACC and will be further regarded as automation experienced. Within this group, 14 had prior experience using steer support.

Before the drive, participants filled out the pre-drive questionnaire and were equipped with the physiological sensors and continuous rating device.

A short motorway section was used to familiarize the participants with operating the automation. Participants were instructed to drive manually or automated, with ACC heading set to 1.8s. An experimenter verbally asked for ratings on trust-in-automation (when used) and trust-in situation (always) every 40 s. The distinction was made to explore how trust is attributed to the automation and other factors. Because of Covid19, the experimenter was seated in the rear-right seat to maximise distance with the participant and

Fig. 1. Overview of physiological sensors, eye tracker, pressure sensor for continuous subjective rating and LED bar for visual feedback of pressure sensor.

both were wearing face masks at all times.

The experimental route consisted of 37 min of highway driving and 20 min on a curved 80 km/h two-lane motorway illustrated in Fig. 2. On the highway, 17 min classified as monotonic, with low to moderate traffic intensity, two lanes per size and only few on/off ramps; and 15 min classified as engaging, with high traffic intensity, 3–5 lanes per side and several on/off ramps and interchanges). The curvy motorway provided low traffic complexity but imposed stricter margins on steering performance with one narrow lane per direction (2.6 m wide), and with several curves and roundabouts (on which participants were encouraged not to use the automation). Environments were always visited in the order monotonic - engaging - curvy motorway, and the order for driving manual and automated was randomised. One participant drove the route in reverse order to avoid severe congestion, and two inexperienced drivers were conditionally placed in the manual-first condition because of elevated anxiety while entering the highway.

The authors also planned to investigate the relation between driver trust and compliance behaviour by examining how trust changes when drivers are encouraged to comply to the automation's steering inputs as much as safely possible. We hypothesised that natural compliance is higher than would be required for complete trust in situation, and that there is an interaction where trust reduces and more risk is perceived only when drivers comply more than what they are naturally inclined to. A condition with instructed compliance was included on the curved motorway. However, the hypotheses could not be tested because the automation is designed to reduce steering performance when steering complacency is detected. The condition was hence excluded from analysis.

2.4.1. Statistics

ANOVAs were performed applying Mauchly's test for the sphericity assumption of within-subject effects. If sphericity was significant, we corrected the degrees of freedom with Greenhouse-Geisser when $\varepsilon < 0.75$ and with Huynh-Feldt otherwise; following Girden's guidelines (Girden, 1992). For significant effects and interactions, we also report the relevant differences in estimated marginal means (diff. EEM) as effect size.

3. Results

CAN data was successfully recorded for 17 automation experienced and 11 automation inexperienced participants. All drivers successfully followed the instructions on when to use and not use driver support. The eye tracker was not used for 10 participants who indicated they required their own glasses to drive safely. Pedal bay video was lost for 5 participants. 11 participants experienced traffic congestion, which was analysed as a separate condition.

Many participants reported difficulties with providing the continuous risk. 5 participants reported unintentional presses, 15 occasionally forgot to press and 15 occasionally pressed late, 6 found the rating task distractive and only 8 indicated they had no difficulty performing this task.

We describe the participant's safety attitude towards the automation in Section 3.1 and provide descriptives their trust and the occurrence risk events and driver-automation conflict in Section 3.2. We then describe these descriptives and their statistical effects based on conditions in Section 3.3, aggregated event-related potential in Section 3.4 and detection performance of individual events based on physiological recordings in 3.5. Verbal and questionnaire ratings (as well as their differences and measures of variance) are expressed as a percentage on the item's range.

3.1. Safety attitude towards automation

After the experiment, participants indicated whether they want to use driving automation. Many disagreed with the statement "I

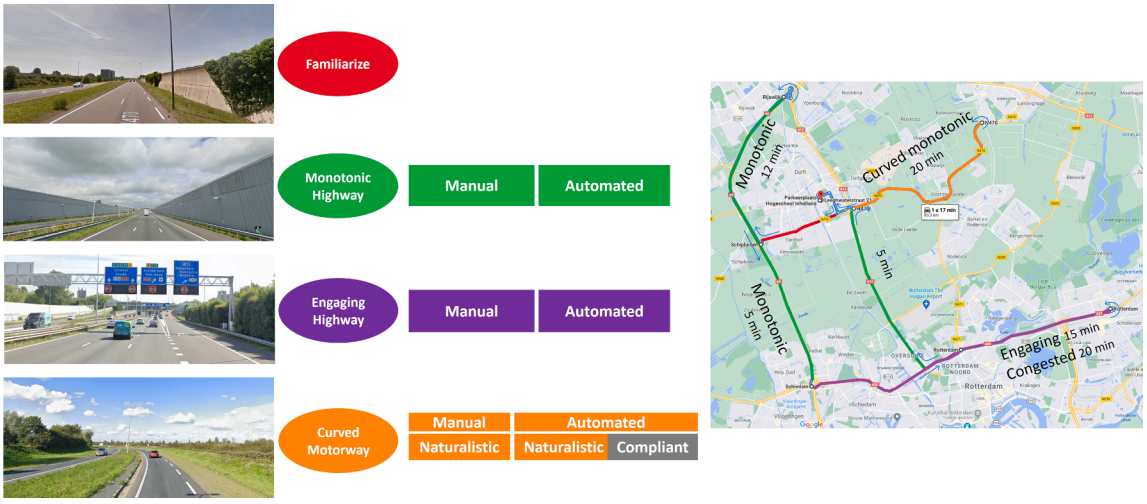

Fig. 2. Overview of the route (83 km near Delft and Rotterdam) and driving conditions. Images and map adopted from google maps and Streetview.

would prefer manual over automated driving in every occasion" (E 39%, P25 17%, P75 67%). Many were willing to use automation while tired (E 71%, P25 67%, P75 83%) or bored (E 67%, P25 50%, P75 83%). When asked to rate their willingness to use the automation while intoxicated, 9 said they would never use it and 8 would always use it, all with the remark they would never take the wheel while intoxicated. The remaining 12 participants were 68% willing to use the automation while intoxicated.

Participants also felt that the automation was not without fault as they only marginally agreed that "automated driving improves safety" (E 59%, P25 50%, P75 67%). They disagreed that "the automation always acted appropriately" (E 34%, P25 17%, P75 33%) and neither agreed nor disagreed with the statements "the automation worked reliably" (E 54%, P25 33%, P75 67%) and "the automation perceived the environment correctly" (E 49%, P25 33%, P75 67%).

Participants indicated they would monitor the road equally or slightly more compared to manual driving (E 56%, P25 50%, P75 67%; where 50% was anchored as "Same as manual driving"). The automation did not change the willingness to engage in mentally distracting secondary tasks such as a phone call (E 47%, P25 50%, P75 50%) and slightly reduced the willingness to engage in visually distracting tasks such as texting (E 44%, P25 46%, P75 50%).

After the experiment, participants indicated their comfort using automation in various conditions. As shown in Fig. 3, drivers were (6.95%) less comfortable using LTA compared to ACC ($F(1,26)=6.604$, $p=.016$). They were generally comfortable using both systems in monotonic and congested highways as well as on curvy motorways, and less comfortable using the systems in unstable traffic and on roundabouts.

3.2. Trust and perceived risk ratings

Table 1 shows pre- and post-drive ratings of trust and perceived risk. Verbal real-time trust ratings are summarized in Fig. 4.

Fig. 5 shows the occurrence rate of conflicts between driver and automation. Occurrence rate of indicated risk and potentially risky events are shown in Fig. 6. The procedures for annotating risk events and behaviours are described in Appendix A. The intensity and duration of risk events as indicated by participants through the pressure sensor are shown in Fig. 7. There were only few moments where participants indicated risk through the pressure sensor. Fig. 8 shows the driver's right foot placement in the pedal bay.

3.3. Condition effects

We examined if the experimental conditions elicited a change in trust and perceived risk.

3.3.1. Effects of automation

Verbal real-time ratings of trust in situation were higher during manual (85.4%) compared to automated conditions (83.7%) ($F(1, 26)=5.845$, $p=.023$; diff. EMM = 2.0%, SE = 0.7%, p=.023) – see Fig. 4.

Participants did not indicate perceived risk to differ between automation use compared to manual driving ($F(1, 21)=1.171$, $p=.291$). However, automation did affect the occurrence of short time headway (THW < 0.8s) ($F(1,21)=19.421$, $p<.001$), which interacted with road type ($F(1.42, 29.75)=10.152$, $p=.001$). Automation reduced short THW from 80.6 to 11.4 times/hour (diff. EMM = 58.3, SE = 12. 8, $p<.001$) on monotonic highways and from 66.3 to 42.2 times/hour (diff. EMM = 24.0, SE = 9.1, $p=.016$) on

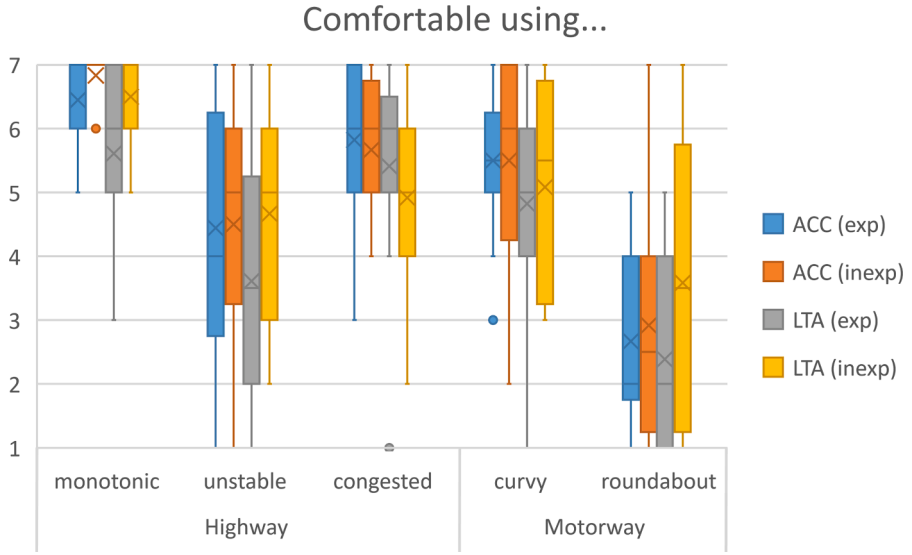

Fig. 3. Post-drive comfort using ACC and LTA in various conditions. Scale ranges from 1 (strongly disagree) to 7 (strongly agree) with being comfortable. Boxes indicate the 1st and 3rd quartiles, x marks averages and whiskers indicate the extrema, excluding the outliers marked as dots (ratings outside the box by 1.5 times the interquartile range). Outliers were not excluded from analysis.

Table 1
Overall trust and perceived risk ratings from the pre-drive and post-drive questionnaires expressed as a percentage on the 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Standard deviation between brackets.

	Trust in automation		Perceived risk	
	Pre drive	Post drive	Pre drive	Post drive
Experienced	61.7% (13.4%)	60.7% (17.7%)	38.5% (14.2%)	35.4% (14.5%)
Inexperienced	66.3% (7.4%)	71.8% (16.8%)	37.8% (11.7%)	32.5% (11.2%)

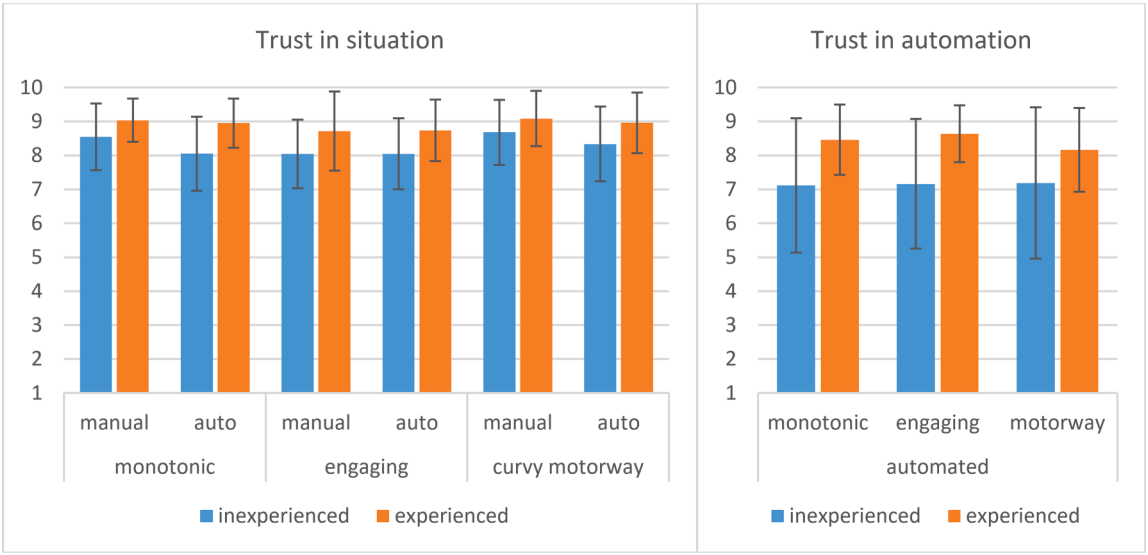

Fig. 4. Verbal real-time ratings for trust in situation and trust in automation indicated on a scale from 1 to 10. Whiskers indicate standard deviation across all ratings.

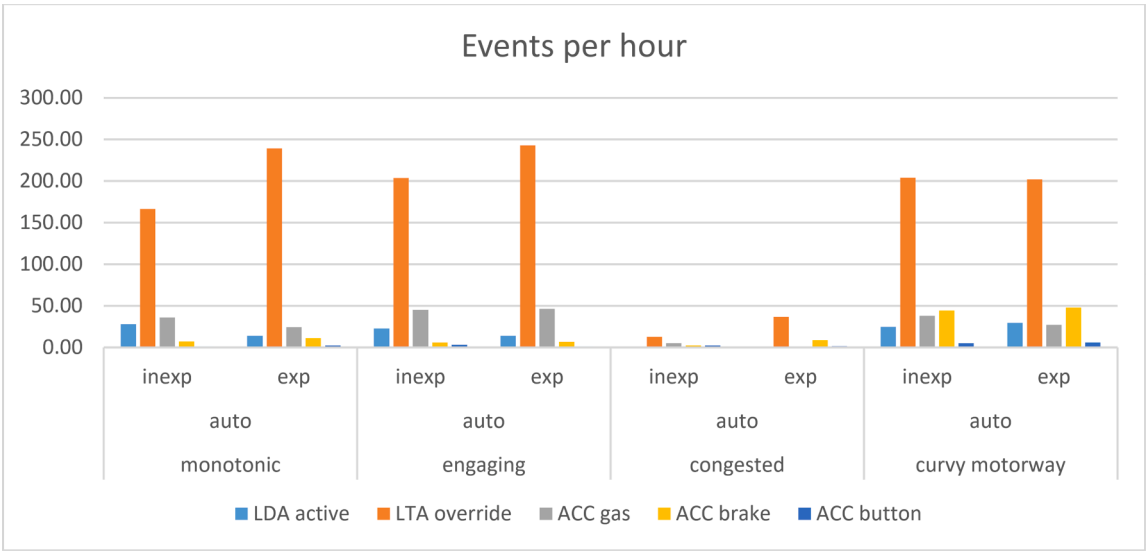

Fig. 5. Occurrence of conflicts between driver and automation. LDA active indicates how often lane departure assist overruled the driver to keep the vehicle in its lane. LTA override indicates how often the reference steer input of lane trace assist differed from driver steering inputs. ACC overrides occur by using gas pedal, disengage by braking and disengage by button.

engaging highways. The reduction was not statistically significant on curvy motorways (diff. EMM = 7.9; SE = 4.9; p=.118). Occurrence rate of TTC < 3s was not affected by automation ($F(1,21)$=0.341, p=.566).

The occurrence of strong braking was also significantly affected by automation ($F(1,21)$=13.067, p=.002) and interacted with road

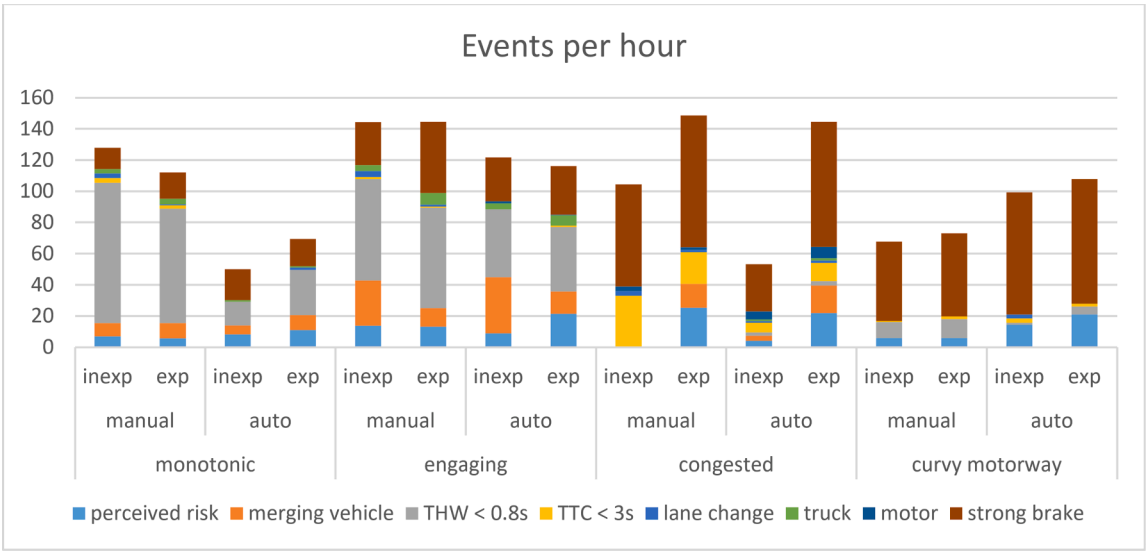

Fig. 6. Risk related event rates for various event types per condition and experience, perceived risk is measured through the pressure sensor, other events are annotated as described in Appendix A.

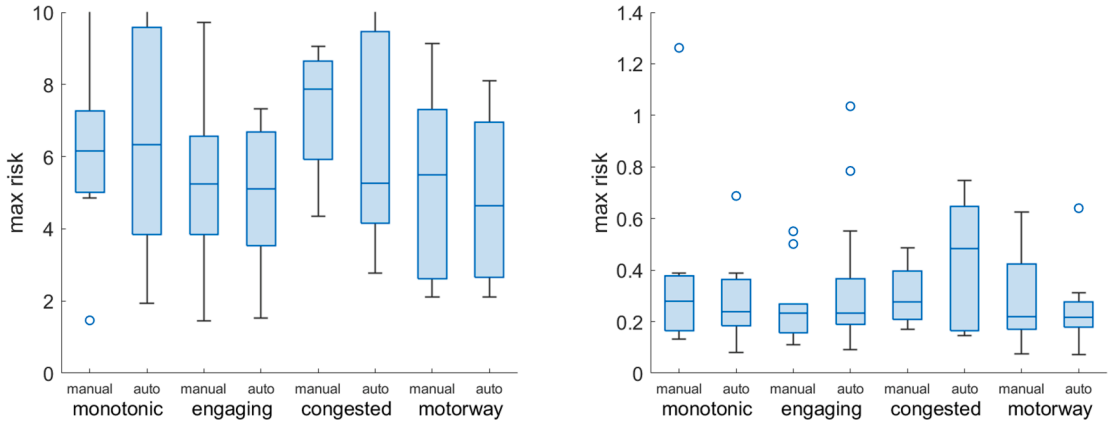

Fig. 7. Participant-averaged peak perceived risk as indicated through pressure sensor peak (left) and duration (right). Boxes indicate the 1st and 3rd quartiles, x marks averages and whiskers indicate the extrema, excluding the outliers marked as dots (ratings outside the box by 1.5 times the interquartile range). Outliers were not excluded from analysis.

type (F(2,42)=15.129, $p<.001$). Automation increased the occurrence of strong braking by from 44.3 to 68.6 events/hour (diff. EMM = 24.3, SE = 3.176, $p<.001$) on curvy motorways, because automation was disengaged through braking while approaching roundabouts. No automation effect occurred on the monotonic (increased 6.0/h, SE = 4.0, $p=.145$) or engaging (decreases 5.7/h, SE = 4.5, $p=.216$) highway.

3.3.2. Effects of environment

Road type significantly affected trust in situation (F(1.93,50.2)=5.597, $p=.007$) but not trust in automation (F(1.48, 38.47)=0.838 $p=.408$). Trust in situation was significantly higher in the monotonic condition (84.9%) compared to the engaging condition (82.1%) (diff. EMM = 2.8%, SE = 1.2%, p=.031) and was not significantly larger for curvy motorways compared to monotonic highways (diff. = 1.8%, SE = 1.4%, $p=.239$).

Perceived risk did not depend on road type (F(2,42)=0.386, $p=.682$). However, risky situations were encountered at different rates between road types (merging vehicle: F(1.69, 35.44)=41.547, $p<.001$; THW < 0.8s: F(1.33, 27.94)=28.261; $p<.001$; strong brake (>2.5 m/s2): F(2,42)=115.685, $p<.001$). Merges occurred more often in the engaging compared to the monotonic highway condition (diff. EMM = 13.3 per hour, SE = 2.85, $p<.001$). Short time headways did not occur significantly less on the monotonic than engaging highways (diff. EMM 2.9, SE 9.0, $p=.749$) but occurred less on curvy motorways (diff. EMM = 44.0, SE 4.5, $p<.001$). Strong braking occurred more often on curvy motorways compared to monotonic highways (diff. EMM = 48.0, SE 2.9, $p<.001$) and more often in engaging compared to monotonic highways (diff. EMM = 8.9, SE 3.5, $p=.020$). Event rate of TTC < 3s was not affected (F(2,42)=

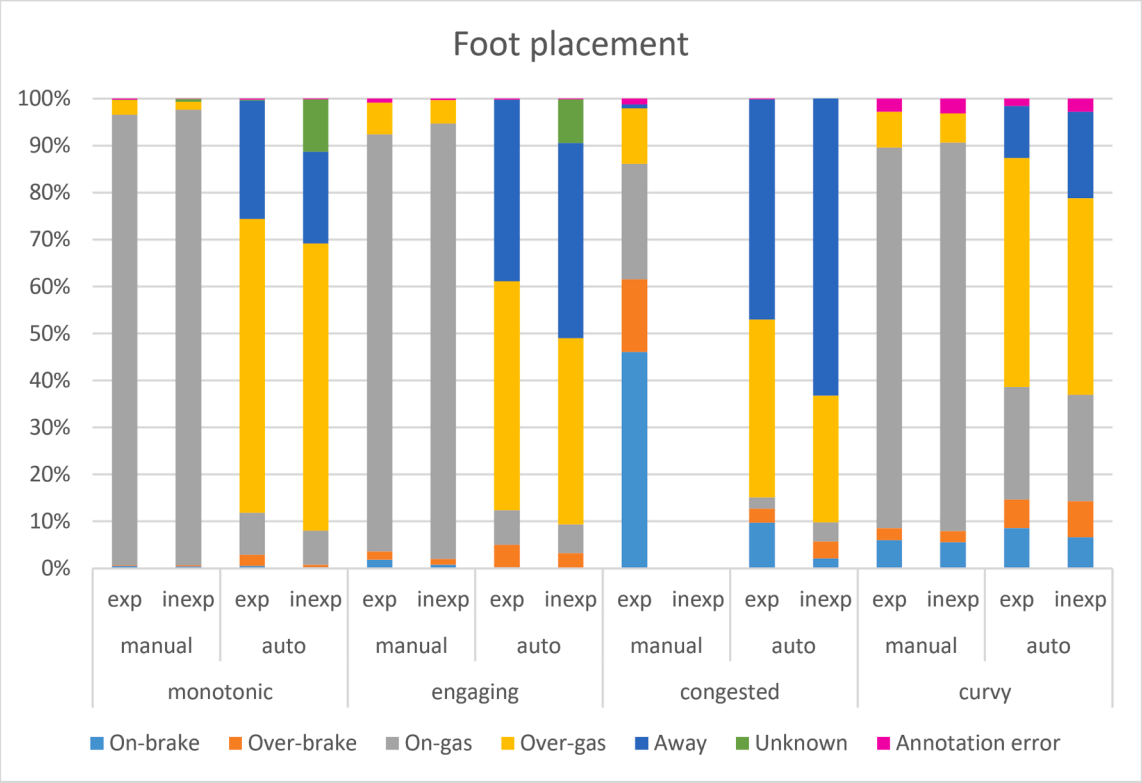

Fig. 8. Foot placement in the pedal bay for each condition. On-brake and on-gas correspond to pedal activity registered on the CAN bus. Over-brake and over-gas correspond to foot observed on or hovering above the pedal while not registered on CAN bus. Away corresponds to foot observed to be away from or below a pedal. Unknown corresponds to periods where foot placement could not be ascertained from footage. Annotation error corresponds to moments where video annotation is in conflict with CAN bus, e.g. foot observed above gas pedal while brake activity registered on CAN. Possible causes include synchronisation accuracy (up to 1 s latency for the pedal-bay video feed), lost video frames and annotation errors.

0.762, p=.473).

Participants encountered a high conflict rate (3.1/minute) between their own steering inputs and LTA, but road type did not affect the occurrence rate (F(2, 42)=1.389, p=.261), nor that of LDA activation (F(1.19, 25.06)=0.137, p=.735) or ACC intervention through gas override (F(2,42)=0.591, p=.559). ACC deactivation rate was significantly higher on the curvy motorway compared to the monotonic highway, both through button (F(2, 42)=4.321, p=.021; diff. EMM = 4.2 per hour, SE = 1.59, p=.015) and through braking (F(2,42)=73.270, p<.001; diff. EMM = 36.4, SE = 4.5, p<.001), but this is explained by the need to slow down for roundabouts on this road. Deactivation did not differ between monotonic and engaging highway.

While using ACC, drivers often preferred to keep their feet on or above the gas pedal (65.8% of the time; Fig. 8), where feet were above the pedals the most during monotonic highway driving (76.6%) and least during congestion (50.5%). Full-range ACC did not significantly reduce brake pedal use (F(1,21)=1.738, p=.202) but reduced gas pedal use by 76.9% (F(1,21)=2845, p<.001).

Drivers kept their feet closer to the pedals during the (automated) monotonic condition compared to the engaging highway condition (F(2, 42)=9.326, p<.001; diff. EMM 12.4%, SE = 4.8%, p=.018). Since the monotonic conditions always preceded the engaging conditions, this can indicate an increased reliance, or a stronger desire to adopt a comfortable posture later in the drive.

3.3.3. Effects of automation experience

No significant effects of prior automation experience on trust and perceived risk were found throughout this study.

Prior experience did not affect questionnaire trust ratings (F(1,28)=0.199, p=.659), nor was there an interaction between pre-post experiment and prior experience (F(1,28)=1.280, p=.267). For the real-time verbal trust ratings, there also were no significant experience effects for trust in automation (F(1,26)=3.735, p=.064) or trust in situation (F(1, 26)=3.027, p=.094) where the experienced group tended to give higher ratings especially for trust in automation.

The experiment also did not change participants' perceived risk between pre-drive and post-drive questionnaire (F(1, 28)=2.176; p=.151), and questionnaire perceived risk was not affected by prior experience (F(1, 28)=0.199; p=.659), interaction between pre-post and prior experience (F(1,28)=0.139, p=.712), or real-time ratings (F(1,21)<0.001, p=.991).

The inexperienced group encountered 6.3 merges/h more (SE = 1.950, p=.004) compared to the experienced group (F(1,21)= 10.592, p=.004), but there was no difference in short headway (F(1,21)=0.290, p=.596), short TTC (F(1,21)=0.092, p=.765) or strong braking rate (F(1,21)=0.727, p=.404). Experience also had no significant effects on conflicts with LDA (F(1,21)=2.675, p=.117),

steering support (LTA:F(1,21)=0.688, p=.416;) or overruling of ACC (disable rate: F(1,21)=0.070, p=.794; gas pedal override: F(1,21)=0.634, p=.435).

Experienced and inexperienced drivers also did not adopt different foot placement strategies (F(1,21)<0.001, p=.969).

3.3.4. Further effects between subjective ratings

The experiment did not change participants' trust ratings between pre-drive and post-drive questionnaire (F(1, 28)=0.607; p=.422). Verbal real-time trust ratings were higher compared to questionnaire ratings. For automated conditions, trust in automation (75.4%) was rated significantly lower compared to trust in situation (83.7%) (F(1,26)=6.753, p=.015; diff. EMM = 8.3%, SE = 2.0% p=.015) and a significant interaction between trust attribution (to situation or automation) and road type (F(1.727, 44.900)=4.705, p=.018) suggests this distinction was present on monotonic highways (diff. EMM = 7.8%, SE = 2.9%, p=.012) and curvy motorways (diff. EMM = 11.4%, SE = 3.8%, p=.006) but not in the engaging highway condition (diff. = 5.6%, SE = 3.3%, p=.102). Trust in automation also varied more compared to trust in situation (Fig. 4).

There was no within-subject relation between changes in trust and foot-away time according to a repeated measures correlation (r(47)=-.023, p=.875). There were also no significant correlations between post-drive trust and the amount of conflicts experienced (LTA steer conflict: r(23)=.219, p=.316; LDA: r(23)=.176, p=.422; ACC_gas: r(22)=-.015, p = 946; ACC_brake: r(23)=-.268, p=.217) or perceived risk (LTA steer conflict: r(23)=.123, p=.577; LDA: r(23)=.055, p=.802; ACC_gas: r(22)=.322, p = 144; ACC_brake: r(23)=-.288, p=.182).

3.4. Condition effects on physiological measures

Mixed ANOVAs were performed to examine if the experimental conditions affected the physiological metrics. To minimise possible confounding factors, results exclude periods of congested driving and pupil data excludes tunnel driving. Effects are summarised in Table 2. Additional descriptives are in Appendix D.

While real-time trust ratings were significantly affected by environment and automation, most physiological metrics were not. GSR phasic drive suggests participants had more emotional arousal during automation use (diff. EMM = 0.04 μS, SE 0.01, p=.001). Cardiovascular power in the HF band was higher in manual than in automated driving (diff. EMM = 45, SE 18.5, p=.023) which suggests a stronger parasympathetic (rest and digest) activity during manual driving.

Table 2
Mixed ANOVAs for various physiological signals. Significant effects are highlighted. 3X represents the three way interaction condition*automation*experience. Tunnel sections are excluded when analysing pupil dilation.

factor	environment	automation	experience	condition *automation	experience *automation	experience *condition	3x
GSR phasic drive	F(2, 50)=1.89, p=.161	F(1, 25)=13.61, p=.001	F(1, 25)=3.18, p=.087	F(2, 50)=2.77, p=.073	F(1, 25)=0.02, p=.891	F(2, 50)=0.23, p=.798	F(2, 50)=0.26, p=.769
Pupil	F(1.1, 13.7)=1.21, p=.295	F(1, 13)=0.97, p=.342	F(1, 13)=11.07, p=.005	F(1.2, 15.0)=0.43, p=.552	F(1, 13)=0.23, p=.640	F(1.1, 13.7)=0.92, p=.359	F(1.2, 15.0)=0.87, p=.382
ECG							
HR	F(2, 50)=13.14, p<.001	F(1, 25)=2.82, p=.106	F(1, 25)=0.45, p=.508	F(1.3, 33.3)=0.35, p=.622	F(1, 25)=0.09, p=.764	F(2, 50)=0.55, p=.578	F(1.3, 33.3)=0.64, p=.472
HRV	F(2, 50)=0.31, p=.736	F(1, 25)=2.33, p=.140	F(1, 25)=0.01, p=.936	F(1.5, 37.4)=1.38, p=.260	F(1, 25)=1.23, p=.279	F(2, 50)=0.56, p=.573	F(1.5, 37.4)=3.22, p=.064
IBI	F(1.7, 42.1)=11.23, p<.001	F(1, 25)=2.65, p=.116	F(1, 25)=0.44, p=.514	F(1.8, 44.7)=0.21, p=.790	F(1, 25)=0.09, p=.765	F(1.7, 42.1)=0.51, p=.575	F(1.8, 44.7)=1.61, p=.213
SDNN	F(2, 50)=2.35, p=.106	F(1, 25)=0.17, p=.683	F(1, 25)=0.17, p=.683	F(2, 50)=0.82, p=.444	F(1, 25)<0.01, p=.960	F(2, 50)=0.50, p=.610	F(2, 50)=2.41, p=.100
SDSD	F(2, 50)=3.00, p=.059	F(1, 25)=0.01, p=.936	F(1, 25)=0.02, p=.894	F(2, 50)=0.24, p=.790	F(1, 25)=0.38, p=.543	F(2, 50)=1.22, p=.303	F(2, 50)=2.54, p=.089
RMSSD	F(2, 50)=2.35, p=.106	F(1, 25)=0.05, p=.832	F(1, 25)=0.03, p=.868	F(2, 50)=0.22, p=.801	F(1, 25)=0.09, p=.772	F(2, 50)=0.68, p=.513	F(2, 50)=2.98, p=.060
PNN20	F(1.7, 43.5)=2.29, p=.120	F(1, 25)=2.11, p=.159	F(1, 25)<0.01, p=.996	F(2, 50)=0.54, p=.584	F(1, 25)=0.01, p=.938	F(1.7, 43.5)=0.30, p=.714	F(2, 50)=2.50, p=.092
PNN50	F(1.2, 29.6)=1.88, p=.180	F(1, 25)=0.94, p=.341	F(1, 25)=0.11, p=.742	F(2, 50)=0.50, p=.610	F(1, 25)=0.34, p=.568	F(1.2, 29.6)=0.04, p=.880	F(2, 50)=1.84, p=.170
LF	F(2, 50)=0.61, p=.546	F(1, 25)=0.49, p=.492	F(1, 25)=0.25, p=.622	F(2, 50)=1.85, p=.168	F(1, 25)=4.70, p=.040	F(2, 50)=0.88, p=.421	F(2, 50)=0.80, p=.457
HF	F(1.7, 41.7)=0.73, p=.463	F(1, 25)=5.82, p=.023	F(1, 25)=0.03, p=.873	F(2, 50)=0.84, p=.438	F(1, 25)=0.17, p=.681	F(1.7, 41.7)=3.28, p=.056	F(2, 50)=2.22, p=.119
LF/HF	F(2, 50)=0.50, p=.608	F(1, 25)=0.51, p=.483	F(1, 25)=0.02, p=.901	F(2, 50)=0.18, p=.839	F(1, 25)=0.54, p=.467	F(2, 50)=0.50, p=.610	F(2, 50)=1.31, p=.280

The remaining effects did not indicate clear relations. The inexperienced group had 0.37 mm larger pupils compared to the experienced group (SE 0.11, p=.005), which is likely related to individual differences. Heart rate reduced by 1.2 BPM from monotonic to engaging highways (SE 0.35, p=.002) and 0.8 BPM from engaging highways to curvy motorways (SE 0.35, p=.033). Since anxiety should have caused the opposite, this trend is more likely a time-on-task effect. This is supported by repeated measures correlation between time in trip and heart rate (r(120)=.406, p<.001). The LF power interaction suggests that automation increased LF power for experienced drivers and decreased it for inexperienced drivers, which in classical beliefs would suggest that experienced users were stressed while inexperienced users were comforted by the automation. However, as discussed by Shaffer et al. (2014) LF power is no longer believed to reflect sympathovagal balance. It is more likely to indicate blood pressure regulation and respiration under slow breathing rates.

3.5. *Effects of risk events on perceived risk and physiological response*

3.5.1. *Responses to risk-events and non-events*

To examine if perceived risk causes physiological responses, we examined event-averaged responses for each event type and compared these to a baseline period when no events took place. Event averaging improves the signal to noise ratio of any responses

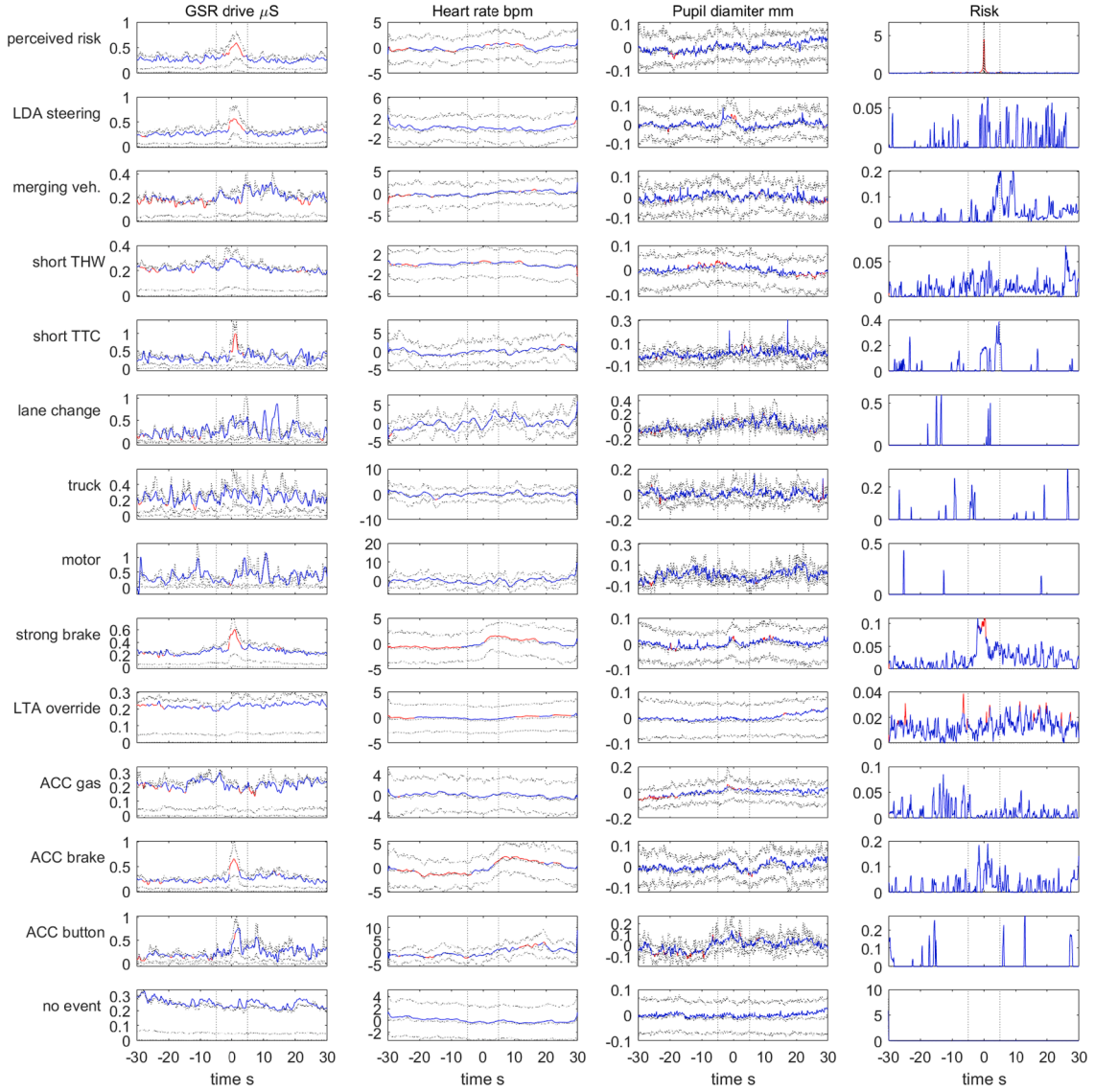

Fig. 9. Physiological responses to various aggregated events and non-events. Columns indicate from left to right: GSR, heart rate, pupil diameter, and continuously perceived risk. Averages are in blue, and in red where differing significantly (Welch t, α = .01) from non-events (bottom row). Dashed lines indicate the 25, 50 and 75 percentile response. Dashed vertical lines mark the 10 s surrounding the event.

which are time-locked to the event while reducing unrelated physiological activity. The same principles (and limitations) are used in the analysis of event-related potential in EEG studies (de Tommaso et al., 2020).

Fig. 9 provides the average and percentile responses to each event. Heart variability metrics were not considered since they require longer samples (Baek et al., 2015). Sample mean heart rate and pupil dilation were removed before averaging. Since a large number of evaluations is made across the time domain, a stricter significance criterion of $\alpha = .01$ is used to visualise where responses differ from the non-event baseline.

For GSR phasic drive, the 25 percentile remains close to zero for all event types, which means at least 25% of all events elicit no GSR response. The average response tends to be close to the 75 percentile, suggesting that at any time only 25% to 50% of responses contribute phasic activity to the observed event-related potential.

Perceived risk gives a timed response in GSR phasic drive. During this response, the 50 percentile reaches only 0.29 μS, which is comparable to the 75 percentile response in non-events. No response to risk perception is visible in heart rate or pupil dilation.

Moments of strong braking increased GSR phasic drive, heart rate, pupil dilation and perceived risk (risk was reported in 6.5% of all strong braking events). Average perceived risk peaked 2.35 s before maximum deceleration, followed by the GSR response 1.25 s before maximum deceleration. This early response suggests GSR to reflect arousal from perceiving risk rather than the physical activity of resolving it. Heart rate responded only after reaching maximum deceleration. The lower-than-baseline heartrate preceding the event is an artefact from removing sample means prior to averaging. The responses of disengaging ACC through braking are similar to those of hard braking (because 50% of these events include hard braking) disengaging ACC through button did not elicit any physiological response.

Short time headways did not affect physiology and risk was indicated in 3.2% of the events. Short time to collision co-occurred with risk perception in 5.3% of events and gave the largest response in GSR phasic drive. While the 25 percentile did not deviate much from zero, the 50 percentile increased to a relatively high 0.42 μS. Short THW and TTC did not elicit a discernible response on heart rate or pupil dilation. Lane departure assist increased GSR response and pupil dilation, but the latter only differs from baseline at $\alpha = .05$.

Merging vehicles, performing a lane change or being near to a truck or motorcycle did not give a physiological response. There were only few of these events (Fig. 6), making it harder to attain statistical significance. Moments where driver and LTA disagree also did not lead to a physiological response and risk was perceived for only 2.9% of these occurrences. Similarly, moments where the driver pressed the gas pedal while ACC was active did not lead to perceived risk (0.6%) or a physiological response, though a brief increase in pupil diameter occurred two seconds before using the gas pedal.

3.5.2. *Event classification by response features*

To evaluate if perceived risk can be identified from physiological signals, we trained a classifier to discriminate events from non-events. We require a successful classifier to at least separate moments where participants indicated risk from non-events. Ideally, the classifier should be able to also identify situations where a timed physiological response was visible in Fig. 9. Positive events were therefore drawn from perceived risk, LTA, short TTC and hard braking. Negatives were drawn from the non-events.

As a classifier, we designed a CNN which takes a 10 s sample of physiological data consisting of 100 samples per channel (GSR phasic drive, heart rate and pupil dilation) and outputs a class for each (non) event. The architecture operates similarly to a template matching procedure, except the templates are learned rather than manually crafted. The convolution allows the network to learn characteristic signal patterns regardless of their latency within the sample. This way, the network accounts for variance in time difference between event occurrence and event perception. Further details on designing the classifier is given in Appendix E.

Fig. 10 shows the confusion matrix for training and validation sets of the best-performing classifier. Per-class validation accuracy was 46.2%. Events were identified as risky 58.6% of the time while non-events are correctly identified 63.7% of the time. 36.2% of perceived risk events were labelled as non-events. Even on the training set, overall accuracy was only 56.1%. The classifier was therefore unable to meaningfully discriminate between events and non-events.

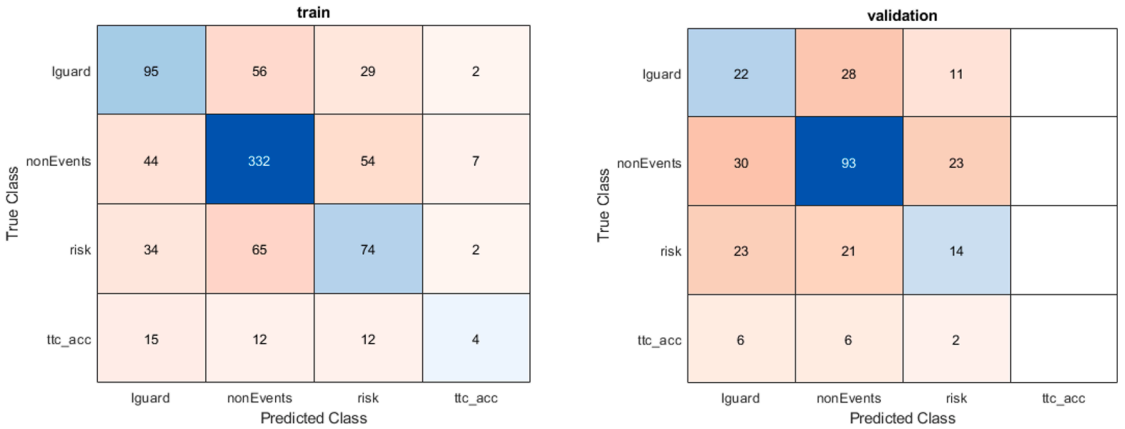

Fig. 10. Best performing classifier: 32 learned templates of 1.5s in length.

4. Discussion

4.1. Automation effects on safety

Participants were aware of automation limitations and the need for monitoring. Despite this awareness, drivers remained willing to use the automation even when tired or bored, and would perform secondary tasks similarly to manual driving. They were comfortable using it in monotonic and congested highways as well as on curvy motorways, but less comfortable in unstable traffic and on roundabouts. Disuse in unstable traffic is concerning and calls for further study. Disuse on roundabouts is expected and desirable as these are not part of the operational design domain (ODD). Overall, the automation seems to encourage a safety attitude appropriate for its level of automation (SAE2 driver support). Drivers were (6.95%) less comfortable using LTA compared to ACC. A similar observation was made in an on-line survey among 629 users of driving automation (Nordhoff et al., under review) This could relate to the high conflict rate (3.1/minute) between their own steering inputs and LTA, which suggest a lower perceived autonomy or trustworthiness of LTA compared to ACC, which may have further contributed to the driver's safety attitude.

Automation reduced short time headway occurrence (THW < 0.8s) on highways compared to manual driving, but did not significantly reduce the occurrence of short TTC (<3s) or merge events. Interestingly, short THW were not perceived as risky while short TTC often were. However, this discrepancy does not imply that automation is addressing the wrong risks. The increased headway improves occupant safety by reducing a hidden, latent hazard through safe distance keeping, while short TTC is a more salient, imminent hazard. In a way, the ACC helps to reduce risks which the driver is less aware of.

These findings align with the conclusion by Heikoop et al. (2017) that real-world use of driving automation may not be as detrimental to monitoring performance as is suggested by some classical and simulation-based vigilance literature (Mackworth, 1948; Greenlee et al., 2018). It is possible that these issues have to some extent been addressed successfully in the design of real-world systems which support but not automate steering, though further analysis is required to verify this in naturalistic driving.

4.2. Implications on trust and risk perception

Pre- and post-drive ratings were not affected by the experiment or prior automation experience, nor did participants converge to unanimous levels of trust or perceived risk, which confirms that these attitudes are individual. While automation-experienced drivers likely already stabilised their views (Lee and See, 2004), also the automation-inexperienced drivers did not adapt their ratings. This could mean that the one hour exposure was too short for changing trust, or that this experience did not necessitate a change in trust. An alignment of pre-drive expectations and post-drive reflection may have been facilitated by the experiment's description and introduction to the support features.

Trust-in situation was affected by road conditions, but trust-in-automation was not. Trust-in-automation changed from moment to moment regardless of environment, which supports that trust-in-automation is situational and should be evaluated within the context of the situation at hand.

Trust attribution was examined by comparing ratings of trust-in situation and trust-in-automation. Trust was rated higher for situation than for automation. This distinction demonstrates that overall trust is comprised of more than automation alone. Trust-in situation was lower during automation use, which means that automation is either a part of, or has an influence on the situation from a trust perspective. The higher trust-in situation may have several causes. Drivers may have considered highway traffic to be relatively predictable, or felt they were in control of the situation while driving support was active. Alternatively, drivers may have found it intuitive to learn the system functionality and limitations. Khastgir et al. (2018) also demonstrated that trust can be high even for poorly performing systems as long as drivers understand the system limitations.

A significant interaction showed that the advance of trust-in situation over trust-in-automation disappeared in the engaging condition. Lee and See (2004) already stipulated that uncertainty and complexity challenges trust.

There was a strong positive correlation between individual post-drive trust and perceived risk. This strengthens beliefs that trust and perceived safety are closely dependent (Nordhoff et al., 2021).

4.3. Behaviour and compliance

Automation-experienced users had more LTA steering conflicts but experience did not affect LDA event rate. This suggests that automation-inexperienced drivers had a higher compliance compared to experienced drivers, without affecting lateral safety. This compliance may be caused by initial over-trust, which is a common bias before encountering sufficient system limitations to accurately gauge limitations and performance variability (Muir, 1987). An alternative explanation is that the automation-experienced participants were unaccustomed with the lower authority of the steering support in the Toyota vehicle, since they reported familiarity with systems of different brands, most of which provided steering automation (e.g. Tesla, Mercedes, Volkswagen) rather than steering support (e.g. BMW, Toyota).

The high conflict rate between driver and LTA provides a salient and frequent clue about the autonomy of the steering support system. This may explain the seemingly well-calibrated expectations towards the system's capabilities. Interestingly, post-drive trust did not correlate with the amount of conflicts detected. This could mean that the effect of conflict rate on trust was already saturated, or that larger or more consistent conflicts are needed to affect trust. Alternatively, the automation's strategy of conditionally reducing LTA performance to encourage engaged driving behaviour may have negated the effect on trust. Conveying trustworthiness of support systems through conflict rate remains feasible, but could not be demonstrated by this study.

Drivers kept their feet on or above the gas pedal while using ACC (65.8% of the time), though the proportion of time with feet away from the pedals increased from the (automated) monotonic to the engaging highway condition. This behaviour increases brake reaction time and suggests drivers maintained postural readiness to intervene (Deo and Trivedi, 2020). Trust-in situation and comfort to use automation showed the opposite effect and trust-in-automation was unaffected. This contradicts the idea that reliance behaviour reflects changes in trust. Since the engaging condition always followed the monotonic, the effect is better explained by a gradually increasing desire for postural comfort. Participants with high trust-in-automation tended to spend more time with their feet away from the pedals, but there was no within-subject relation between momentary trust and foot placement. Our sample size is too small to explore if this results from a personality trait like trust propensity.

4.4. Risk perception and physiology

At a mesoscopic level, only 5 of the 13 physiological metrics differed significantly between conditions, with little consistency on the factors of influence. The GSR phasic drive and heart rate HF power differed significantly between manual and automated driving. Between the driving environments, only heart rate (and its inverse IBI) differed significantly. Effects for heart rate and variability on arousal and stress have been somewhat inconsistent in preceding works, in particular for on-road conditions (Lohani et al., 2019).Pupil dilation differed between groups and LF component of heart rate variability interacted between experience and automation use, but these findings may not generalise, since groups samples were small and the relation between LF and anxiety has been debated (Shaffer et al., 2014). The small number of affected dependent variables and their partially counter-intuitive effects (e.g. automation increasing arousal according to GSR but reducing it according to the heart's HF component) indicate limited utility of physiological sensing in on-road driving when averaged over road sections. However, it should be noted that the difference in trust between these conditions was complex, with clear differences in risk event rates and trust-in situation, but no significant differences in trust-in-automation.

At a smaller time scale, physiological responses were more insightful. Moments of perceived-risk, LDA activation, short TTC, strong braking and disabling ACC by braking resulted in a low-latency response in GSR phasic drive when averaged over multiple encounters, and these responses occurred across all driving environments. Pupil dilation only responded to LDA and hard braking, while heart rate only responded to hard braking. These findings agree with Healey and Picard (2005) who also found best sensitivity of GSR, followed by heart rate and its variability. Beggiato et al. (2019) found more consistent pupil dilation among their 4 most risky situations. Their simulated encounters were experienced more consistently among 40 participants, resulting in better signal to noise ratio. They also found that heart rate declined during their encounters, reproducing their earlier work with similar sensors (Beggiato et al., 2018) but contrasting a large body of stress, anxiety and workload literature (Lohani et al., 2019).

Our findings suggest that of all examined physiological signals, GSR phasic drive was the most indicative of perceived risk. However, responses existed for 75% of the events, and were discernible from baseline variation in $< 50\%$ of the events. It is likely that not all risk events (especially when no risk was indicated) were perceived as risky, for instance when anticipated or occurring in a non-threatening manner. The extent to which events were risky can however not be inferred reliably from the continuous risk ratings, in part because the rating task was not continuously performed, and in part because more mundane risks may not have been indicated, for instance because of its familiarity (Geller, 2001). A CNN was unable to distinguish between risky and baseline events based on physiological data alone. This can be a consequence of the same rating difficulties, in particular the occurrence of late and unintentional presses. Alternatively some risky events may not cause physiological anxiety, for instance when the situation is only marginally risky, anticipated or easily resolved by the driver. However, most likely, the physiological indicators lacked specificity to perceived risk.

4.5. Limitations and future work

The compliance condition could not be analysed because of confounding factors. Further research is thus needed to evaluate the relation between trust, perceived risk and compliance behaviour. We hypothesise that the large amount of steering conflict and continuation of salient risks during automation use contributed to the well calibrated safety attitude among participants, but we were not able to test this in this study. A direct comparison between different steering automation philosophies is required to further investigate this.

Many participants had difficulty providing the continuous risk rating in this on-road study. This difficulty was surprising, since the setup has been applied successfully in preceding work in a simulator (He et al., 2022), and others have demonstrated similar devices to be effective as well (Cleij et al., 2018; Rossner and Bullinger, 2019). Since the rating technique has only been used in simulated autonomous driving, it is possible that combining the continuous rating with manual/supervised driving on public roads caused a high workload where drivers prioritise safe driving over the rating task. It is possible that drivers normally do not consider risk consciously while driving. This would make the rating task more artificial than originally anticipated, and hence more likely to be forgotten while multitasking. Alternatively, it is possible that only the reporting aspect is easily forgotten and conscious risk assessment is still taking place. We recommend caution when considering continuous subjective rating in an on-road setting, or in combination with many other tasks.

We recommend that design for trust calibration in supervised automation should not only consider calibration of trust-in-automation, but also calibration of self-confidence and other trust agents. It may be worthwhile to investigate how perceived usefulness of different use cases relate these different elements of trust. We hypothesise that safety and comfort benefits are fully reliant on trust-in-automation, while a fun-factor of using supervised automation requires trust-in-self. It may be interesting to explore if moderate distrust-in-automation provides positive excitement as a remedy for drowsy driving.

5. Conclusions

5.1. Was the automation safe? Was it accepted?

The observed safety attitude towards the ACC and LTA in the tested vehicle was appropriate for a driving support system. Participants believed that the automation improved safety, but also that it needs to be monitored. They were willing to use the automation, and indicated they would engage less often in mentally and visually distracting tasks while using the automation. They were generally comfortable using ACC and LTA in monotonic and congested highways as well as on curvy motorways, but less comfortable in unstable traffic and on roundabouts.

ACC improved occupant safety by reducing latent risks that were not perceived by the drivers (short headways), but did not affect the occurrence of salient risks (short TTC). Safety aspects of steering support were not evaluated in this study.

These findings indicate that user safety attitudes are not as detrimentally affected by driving support systems as suggested by some classical and simulation-based literature.

5.2. How were trust and perceived risk affected by the experimental conditions?

Automation use decreased trust-in situation by 2.0% but did not affect perceived risk. While using automation, drivers distinguished between trust-in-automation and trust-in situation, where trust-in situation was 8.3% higher on average, presumably due to driver self-confidence.

Driving environment affected the driver's trust-in situation but not their trust-in-automation, though trust-in-automation did vary from moment to moment. Dynamic aspects of trust-in-automation should therefore be examined situationally. Environment also affected the willingness to use automation, which was lowest during unstable traffic and on roundabouts.

No effects of prior automation experience were found in this study. Ratings of trust and perceived risk were strongly, negatively correlated.

These insights may help improve the calibration of trust in driving automation implementations.

5.3. How well can physiology and behaviour indicate trust and perceived risk in on-road driving conditions?

Averaged over test conditions, physiological metrics showed weak and partially counter-intuitive effects, which could not be explained by trust, perceived risk, driving conditions or automation use. For aggregated risk events, GSR phasic drive had the clearest response but was discernible from baseline in less than half the events. Pupil dilation and heart rate only increased during strong braking and LDA activity. A CNN was unable to distinguish between individual risk events and non-events based on physiology, which suggests that physiology lacks the specificity and reliability required for on-road, real-time monitoring of risk perception. For such applications, alternative measures have to be explored.

Driver behaviour demonstrated active participation in the driving task, frequently overruling lane centring support (3.1 times/minute), and keeping feet on or above the pedals while using ACC (65.8% of time). The amount of conflict between driver and automation did not correlate with trust. The automation may have effectively conveyed its trustworthiness through conflict rate, but this hypothesis was not tested by this study. Participants with high trust-in-automation ratings spent more time with their feet away from the pedals, but there was no moment-to-moment relation between trust and foot placement.

These findings may aid the development of trust monitoring models and systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was funded by Toyota Motor Europe, NV/SA.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.trf.2022.07.008.

References

Ajenaghughrure, I., Da Costa, Cláudia, Sousa, Sónia, & Lamas, D. (2021). Psychophysiological modelling of trust in technology: Comparative analysis of psychophysiological signals. *VISIGRAPP, 2*, 161–173. https://doi.org/10.5220/0010237701610173

Baek, H. J., Cho, C. H., Cho, J., & Woo, J. M. (2015). Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemedicine journal and e-health: the official journal of the American Telemedicine Association, 21*, 404–414. https://doi.org/10.1089/tmj.2014.0104

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using smartbands, pupillometry and body motion to detect discomfort in automated driving. *Frontiers in Human Neuroscience, 12*, 338. https://doi.org/10.3389/fnhum.2018.00338

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*, 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Brookhuis, K. A., & de Waard, D. (2001). Behavioural impacts of advanced driver assistance systems - an overview. *European Journal of Transport and Infrastructure Research*, 245–253.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*, 249–268. https://doi.org/10.1016/0301-0511(95)05161-9

Cabrall, C. D. D., Eriksson, A., Dreger, F., Happee, R., & de Winter, J. (2019). How to keep drivers engaged while supervising driving automation? a literature survey and categorisation of six solution areas. *Theoretical Issues in Ergonomics Science, 20*, 332–365. https://doi.org/10.1080/1463922X.2018.1528484

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction, 31*, 692–702. https://doi.org/10.1080/10447318.2015.1070549

Cleij, D., Venrooij, J., Pretto, P., Pool, D. M., Mulder, M., & Bulthoff, H. H. (2018). Continuous subjective rating of perceived motion incongruence during driving simulation. *IEEE Transactions on Human-Machine Systems, 48*, 17–29. https://doi.org/10.1109/THMS.2017.2717884

Deo, N., & Trivedi, M. M. (2020). Looking at the driver/rider in autonomous vehicles to predict take-over readiness. *IEEE Transactions on Intelligent Vehicles, 5*, 41–52. https://doi.org/10.1109/TIV.2019.2955364

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers, 21*, 133–145. https://doi.org/10.1016/j.intcom.2008.10.011

Geller, E.S., 2001. Chapter five: Sensation, perception and perceived risk, in: Geller, E.S. (Ed.), The Psychology of Safety Handbook. CRC Press LLC, pp. 76–84.

George, S., Clark, M., & Crotty, M. (2007). Development of the adelaide driving self-efficacy scale. *Clinical rehabilitation, 21*, 56–61. https://doi.org/10.1177/0269215506071284

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

Girden, E.R., 1992. ANOVA: Repeated measures. volume no. 07–084 of Sage university papers Quantitative applications in the social sciences. SAGE publications, Newbury Park Calif.

Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2018). Driver vigilance in automated vehicles: Hazard detection failures are a matter of time. *Human Factors and Ergonomics Society, 18720818761711*. https://doi.org/10.1177/0018720818761711

Hartwich, F., Beggiato, M., Dettmann, A., & Krems, J. F. (2015). Drive me comfortable: Customized automated driving styles for younger and older drivers. *VDI Berichte*.

He, X., Stapel, J., Wang, M., & Happee, R. (2022). Modelling risk and trust in driving automation: reacting to merging and braking vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 86*, 178–195. https://doi.org/10.1016/j.trf.2022.02.016

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE transactions on intelligent transportation systems, 6*. https://doi.org/10.1109/TITS.2005.848368

Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2015). Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science, 17*, 284–303. https://doi.org/10.1080/1463922X.2015.1101507

Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study. *Applied ergonomics, 60*, 116–127. https://doi.org/10.1016/j.apergo.2016.10.016

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

Hu, C., & Wang, J. (2021). Trust-based and individualizable adaptive cruise control using control barrier function approach with prescribed performance. *IEEE Transactions on Intelligent Transportation Systems, 1–11*. https://doi.org/10.1109/TITS.2021.3066154

Jian, J.Y., Bisantz, A.M., Drury, C.G., Llinas, J., Foundations for an empyrically determined scale of trust in automated systems.

Jorna, P. (1992). Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. *Biological Psychology*, 237–257.

Khan, M. Q., & Lee, S. (2019). A comprehensive survey of driving monitoring and assistance systems. *Sensors (Basel, Switzerland), 19*. https://doi.org/10.3390/s19112574

Khastgir, S., Birrell, S., Dhadyalla, G., & Jennings, P. (2018). Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation Research Part C: Emerging Technologies, 96*, 290–303. https://doi.org/10.1016/j.trc.2018.07.001

Kiefer, R. J., LeBlanc, D. J., & Flannagan, C. A. (2005). Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments. *Accident; analysis and prevention, 37*, 295–303. https://doi.org/10.1016/j.aap.2004.09.003

Kolekar, S., de Winter, J., & Abbink, D. (2020). Which parts of the road guide obstacle avoidance? quantifying the driver's risk field. *Applied ergonomics, 89*, 103196. https://doi.org/10.1016/j.apergo.2020.103196

Kraus, J., Scholz, D., Stiegemeier, D., & Baumann, M. (2020). 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, 62*, 718–736. https://doi.org/10.1177/0018720819853686

Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior research methods, 51*, 1336–1342. https://doi.org/10.3758/s13428-018-1075-y

Lee, J., Abe, G., Sato, K., & ITOH, M. (2021). Developing human-machine trust: Impacts of prior instruction and automation failure on driver trust in partially automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 81*, 384–395. https://doi.org/10.1016/j.trf.2021.06.013

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors, 46*, 50–80.

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in human neuroscience, 13*, 57. https://doi.org/10.3389/fnhum.2019.00057

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1*, 6–21. https://doi.org/10.1080/17470214808416738

Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology, 37*, 465–498. https://doi.org/10.1002/aris.1440370111

Mayer, R. C., & Davis, J. H. (1995). An interactive model of organizational trust. *Acadamy of management review, 20*, 709–734.

Morris, D. M., Erno, J. M., & Pilcher, J. J. (2017). Electrodermal response and automation trust during simulated self-driving car use. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*, 1759–1762. https://doi.org/10.1177/1541931213601921

Muir, B. M. (1987). *Trust between humans and machines, and the design of decision aids. international journal man-machine studies* (pp. 527–539).

Nordhoff, S., Stapel, J., Gentner, A., Happee, R., under review. Exploring the factors of perceived safety and trust in sae level 2 partially automated cars using principal component analysis.

Nordhoff, S., Stapel, J., He, X., Happee, R., & Gentner, A. (2021). Perceived safety and trust in sae level 2 partially automated cars: Results from an online questionnaire. *PloS one, 16*, e0260953. https://doi.org/10.1371/journal.pone.0260953

Panicker, S. S., & Gayathri, P. (2019). A survey of machine learning techniques in physiology based mental stress detection systems. *Biocybernetics and Biomedical Engineering, 39*, 444–469. https://doi.org/10.1016/j.bbe.2019.01.004

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 252–263. https://doi.org/10.1016/j.trf.2014.04.009

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality, 41*, 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rossner, P., & Bullinger, A. C. (2019). Do you shift or not? influence of trajectory behaviour on perceived safety during automated driving on rural roads. *International Conference on Human-Computer Interaction, 245–254*. doi:10.1007/978-3-030-22666-4{\textunderscore }18.

Sedghamiz, H. (2018). Biosigkit: A matlab toolbox and interface for analysis of biosignals. *Journal of Open Source Software, 3*, 671. https://doi.org/10.21105/joss.00671

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health, 5*, 258. https://doi.org/10.3389/fpubh.2017.00258

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology, 5*, 1040. https://doi.org/10.3389/fpsyg.2014.01040

Tang, R., Kim, J. H., Parker, R., & Jeong, Y. J. (2018). Indicating severity of vehicle accidents using pupil diamiter in a driving simulator environment. *International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management, 647–656*. doi:10.1007/978-3-319-91397-1{\textunderscore }53.

de Tommaso, M., Betti, V., Bocci, T., Bolognini, N., Di Russo, F., Fattapposta, F., Ferri, R., Invitto, S., Koch, G., Miniussi, C., Piccione, F., Ragazzoni, A., Sartucci, F., Rossi, S., Arcara, G., Berchicci, M., Bianco, V., Delussi, M., Gentile, E., Giovannelli, F., Mannarelli, D., Marino, M., Mussini, E., Pauletti, C., Pellicciari, M. C., Pisoni, A., Raggi, A., & Valeriani, M. (2020). Pearls and pitfalls in brain functional analysis by event-related potentials: A narrative review by the italian psychophysiology and cognitive neuroscience society on methodological limits and clinical reliability-part i. *Neurological sciences: official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, 41*, 2711–2735. https://doi.org/10.1007/s10072-020-04420-7

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018). Heart rate analysis for human factors: development and vaidation of an open source toolkit for noisy naturalistic heart rate data. *Humanist Conference, 6*.

van Huysduynen, H. H., Terken, J., & Eggen, B. (2018). Why disable autopilot? *AutomotiveUI*. https://doi.org/10.1145/3239060.3239063

Xiong, X., Wang, M., Cai, Y., Chen, L., Farah, H., & Hagenzieker, M. (2019). A forward collision avoidance algorithm based on driver braking behavior. *Accident; Analysis and Prevention, 129*, 30–43. https://doi.org/10.1016/j.aap.2019.05.004