Transportation Research Part F: Psychology and Behaviour 87 (2022) 149–163

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving

Vishnu Radhakrishnan [a,*], Natasha Merat [a], Tyron Louw [a], Rafael Cirino Gonçalves [a], Guilhermina Torrao [a,1], Wei Lyu [a,2], Pablo Puente Guillen [b,3], Michael G. Lenné [c]

[a] Institute of Transport Studies, University of Leeds, Leeds LS2 9JT, United Kingdom
[b] Toyota Motor Europe NV/SA, Hoge Wei 33, 1930 Zaventem, Belgium
[c] Seeing Machines Ltd., Canberra 2609, Australia

ARTICLE INFO

Keywords:
Workload
Car-following
Psychophysiology
Heart-rate variability (HRV)
Electrodermal activity (EDA)
Highly automated driving (HAD)

ABSTRACT

This driving simulator study, conducted as a part of Horizon2020-funded L3Pilot project, investigated how different car-following situations affected driver workload, within the context of vehicle automation. Electrocardiogram (ECG) and electrodermal activity (EDA)-based physiological metrics were used as objective indicators of workload, along with self-reported workload ratings. A total of 32 drivers were divided into two equal groups, based on whether they engaged in a non-driving related task (NDRT) during automation (SAE Level 3) or monitored the drive (SAE Level 2). Drivers in both groups were exposed to two counterbalanced experimental drives, lasting ~ 18 min each, of *Short* (0.5 s) and *Long* (1.5 s) Time Headway conditions during automated car-following (*ACF*), which was followed by a takeover that happened with or without a lead vehicle. Results showed that driver workload due to the NDRT was significantly higher than both monitoring the drive during *ACF* and manual car-following (*MCF*). Furthermore, the results indicated that a lead vehicle maintain a shorter THW can significantly increase driver workload during takeover scenarios, potentially affecting driver safety. This warrants further research into understanding safe time headway thresholds to be maintained by automated vehicles, without placing additional cognitive or attentional demands on the driver. Our results indicated that ECG and EDA signals are sensitive to variations in workload, which warrants further investigation on the value of combining these two signals to assess driver workload in real-time, to help future driver monitoring systems respond appropriately to the limitations of the driver, and predict their performance in the driving task, if and when they have to resume manual control of the vehicle after a period of automated driving.

* Corresponding author.
E-mail addresses: mn16vr@leeds.ac.uk (V. Radhakrishnan), N.Merat@its.leeds.ac.uk (N. Merat), T.L.Louw@leeds.ac.uk (T. Louw), tsrg@leeds.ac.uk (R.C. Gonçalves), guilhermina.torrao@vtt.fi (G. Torrao), pablo.puenteguillen@aecom.com (P. Puente Guillen), mike.lenne@seeingmachines.com (M.G. Lenné).
[1] VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Finland.
[2] School of Business Administration, Northeastern University, Shenyang, China.
[3] AECOM, 20 Colmore Circus Queensway, Birmingham B4 6AT, United Kingdom.

https://doi.org/10.1016/j.trf.2022.04.002
Received 9 June 2021; Received in revised form 25 March 2022; Accepted 7 April 2022
Available online 18 April 2022

1. Introduction

In recent years, we have seen a gradual increase in the implementation of Advanced Driving Assistance Systems, such as lane-keeping assistance or adaptive cruise control, in vehicles, with manufacturers striving to attain higher levels of vehicle automation capabilities. Highly Automated Driving (HAD) systems such as Traffic Jam Pilot, are currently available on the market (IEEE, 2019) and the use of the first SAE Level 3 (L3; SAE, 2021), Automated Lane Keeping System became legal in the UK in 2021 (Kinnear et al., 2021). Unlike SAE Level 2 (L2) systems, SAE Level 3 does not require constant monitoring of the drive, and drivers can engage in other non-driving related tasks (NDRTs), and activities. The main rationale for the implementation of HAD is its hypothesised provision of increased comfort and safety (ERTRAC, 2017). However, in terms of the human driver, one of the unwanted consequences of vehicle automation is the out of the loop (OOTL) phenomenon, which, according to Merat et al. (2018), refers to a state where, when automation is engaged, the driver is not monitoring the driving environment, and may or may not be in physical control of the vehicle. Studies have shown that when drivers are out of the loop, a decrement in performance is observed (Endsley & Kiris, 1995) once they are required to resume control of the vehicle, which can lead to reduced safety, for example, if the driver is required to avoid an impending collision (Louw et al., 2017; Wandtner et al., 2018; Zeeb et al., 2016).

Driving performance and safety, after resumption of manual control from HAD, is also affected by driver workload, which describes the relationship between physical and cognitive resources demanded by a task, and those resources available to be supplied by the driver (Dogan et al., 2019; Parasuraman et al., 2008). The relationship between workload and task performance is complex, and follows an 'inverted U-shape' relationship (Bruggen, 2015; de Waard, 1996). For HAD, both high workload (overload - where the task demand exceeds the available resources) and low workload (underload - where the attentional capacity of the driver is reduced due to low task demand and monotony during automation) can result in a performance decrement, and increases the chances of driver error (Parasuraman et al., 2008; Young & Stanton, 2002). Parasuraman et al. (2008) have also suggested that workload is a better predictor of drivers' future performance, than their current performance. For example, research has shown that a driver's ability to safely resume control from automation is likely to be affected if they are engaged in a high workload task during HAD, such as a demanding NDRT, with worse performance observed compared to a no task period during HAD (Gold et al., 2015; Zeeb et al., 2016). This effect of workload on performance can be especially problematic if there is a sudden increase or change in task demand (such as an unexpected takeover scenario due, e.g. to avoid an obstacle in the lane), compromising safety. For optimal performance in the driving task, reducing the likelihood of errors, drivers are required to maintain a moderate level of workload (Bruggen, 2015; de Waard, 1996). As engagement in NDRTs is likely to increase with higher levels of automation (Carsten et al., 2012; NTSB, 2017), especially in L3, it is important to understand how driver workload changes during different stages of HAD, in order to provide appropriate mitigation strategies, and reduce performance decrements during transitions of control (Merat et al., 2012; Meteier et al., 2021).

An example of obstacle avoidance after resuming control from automation is preventing a rear-end collision during car following scenarios. Rear-end collisions account for over 31% of all collisions in the US (National Highway Traffic Safety Administration, 2009, p. 56). Car-following refers to the longitudinal following of a lead vehicle by drivers. Given that car-following is a pre-cursor to rear-end collisions (Li et al., 2017), it is of value to understand how different car-following situations in HAD can affect driver workload. Time headway (THW), and time-to-collision, are two safety indicators used in car-following situations to understand drivers' longitudinal driving behaviour (Vogel, 2003). THW is defined as the elapsed time between the front of the lead vehicle passing a point on the roadway and the front of the following vehicle passing the same point (Evans, 1991, p. 313).

Research has shown that driver workload can be influenced by the THW maintained by a vehicle. For example, in a manual driving study, conducted in a driving simulator, Liu et al. (2019) found that subjective ratings for workload were significantly higher for the 0.5 and 1 s THW conditions, compared to the 2, 2.5 and 3 s THW conditions. In their study on HAD, Siebert & Wallis (2019) reported that drivers were significantly more uncomfortable during THWs under 1.5 s, as reported by their subjective ratings of riskiness due to the THW, compared to longer THWs, while driving at 50 km/h, 100 km/h and 150 km/h respectively. However, participants' subjective ratings of riskiness due to the THW maintained by the vehicle was also dependent on environmental factors such as visibility (fog), and traffic conditions such as following a truck, with driver discomfort increasing with lower visibility, increased traffic, and when following a truck, as opposed to a car. In Louw et al. (2020), we observed, via subjective ratings, that shorter headways maintained by HAD are perceived as riskier or unsafe by drivers, especially when they are not in control of the driving task. Resuming control from automation in the presence of a closer lead vehicle is also likely to be more demanding, especially following engagement in an NDRT (Mehler et al., 2009), further exacerbating the OOTL effect, with studies on HAD showing that engagement in NDRTs increases driver workload, and negatively affects their driving performance after takeovers (Du et al., 2020; Wandtner et al., 2018; Zeeb et al., 2016).

Given the high inter-individual variability in how people are affected by, or perceive, workload, it can be challenging to accurately measure and interpret it on a moment-to-moment basis. However, the ability to objectively measure workload in real-time during different stages of HAD is crucial, as it can provide insights into drivers' capabilities and limitations, when they are required to resume control of the vehicle, ultimately helping to improve the safety of the automated system. Real-time, minimally-intrusive, and continuous assessment of driver workload can be used to assist the driver, for example, to warn them of dangerous overload or underload situations (Merat et al., 2012). Therefore, in the current paper, we investigated the added value of using physiological signals to objectively measure driver workload, in HAD.

Electrocardiogram (ECG)-based physiological metrics such as heart rate (HR), heart rate variability (HRV), ECG-derived respiration rate (EDR), and metrics derived from electrodermal activity (EDA) signals, have been used extensively to understand and measure workload, in both manual driving and HAD (Biondi et al., 2018; Du et al., 2020; Hidalgo-Muñoz et al., 2019; Mehler et al., 2009).

A general finding is that an increase in drivers' workload is associated with an increase in HR and EDR, the latter of which is the number of breaths a person takes in a minute, as derived from an ECG signal (Hidalgo-Muñoz et al., 2019; Mehler et al., 2009). However, an increase in workload results in a decrease in HRV, which is the physiological phenomenon of variation in time interval between heartbeats (Mehler et al., 2009). A decrease in HRV is reflected by a reduction in drivers' root mean square of successive differences in R-R intervals (RMSSD) (Orsila et al., 2008). An EDA signal consists of a slowly evolving tonic component, called skin conductance level (SCL) and a rapidly evolving phasic component, called skin conductance response (SCR) (Braithwaite et al., 2015; Cacioppo et al., 2007). SCRs generally have a faster decay time, compared to other EDA and ECG-based metrics, thus making them more sensitive to changes in stimuli that are constantly evolving and/or of short duration (Braithwaite et al., 2015). Both SCL and SCRs are shown to increase with stress and workload during driving (Du et al., 2020; Foy & Chapman, 2018; Mehler et al., 2009), and the number of SCRs per minute (nSCR/min) has been shown to increase in situations that involve high stress/workload or discomfort for drivers (Foy & Chapman, 2018; Radhakrishnan et al., 2020).

1.1. *Current study*

This study investigated how manipulations of workload, during different stages of L2 and L3 HAD affects drivers' psychophysiological metrics. Workload was manipulated by introducing two lead vehicle conditions (*Lead/No Lead*) in an urban driving environment, and two THW conditions when a lead vehicle was present (*Short* and *Long*). To understand how the different vehicle automation states affected driver workload in the context of car-following, our study exposed drivers to automated car-following (*ACF*) segments, manual car-following (*MCF*) segments, and *Takeover* segments, with the latter involving transitions of control from automated to manual driving. To study the effect of NDRT on drivers' workload levels, a between-participant design was used, with one group of drivers (*L2*) asked to monitor the driving environment at all times, and another (*L3*) asked to engage in an NDRT when automation was engaged. ECG- and EDA-based physiological data were collected as objective measures of workload, and compared with drivers' self-reported workload ratings. The following research questions were addressed:

How is L2 drivers' workload, as measured by changes in their physiological state, and self-reported workload ratings, affected by the two THW conditions (*Short* vs *Long*) when monitoring the drive during *ACF*?
Is drivers' workload during *ACF* affected by an NDRT (*L2* vs *L3*)?
Does drivers' workload vary between *ACF* and *MCF*?
Is drivers' workload during the takeover affected by the THWs maintained by the automated controller (*Short* vs *Long*)?
Is drivers' workload during the takeover affected by engagement in an NDRT during automation (*L2* vs *L3*)?

2. Materials and methods

2.1. *Participants*

A total of 32 participants (16 for each level of automation), each with a valid UK driving licence, took part in this driving simulator-based study. A total of 6 participants (3 each from *L2* and *L3*) were excluded from the analysis. Of this, 3 participants did not adhere to the instruction to follow the lead vehicle, and 3 others were excluded due to missing physiological data. For the 13 (4 female, 9 male) participants considered in the *L2* group, the mean age of the participants was 42 ± 17 years, with a mean driving experience of 22 ± 16 years. The 13 participants (3 female, 11 male) of the *L3* group had a mean age of 33 ± 8 years, with a mean driving experience of 14 ± 8 years. Prior to the experiment, participants were instructed to avoid caffeinated products, consumption of alcohol, and engagement in extreme exercise, to control for their effect on physiological data, as recommended in Laborde, Mosley, & Thayer (2017). All participants gave consent to take part in the study, in accordance with the rules and regulations of the University of Leeds ethics committee (LTTRAN-054), and were compensated with £25 for taking part in the study.

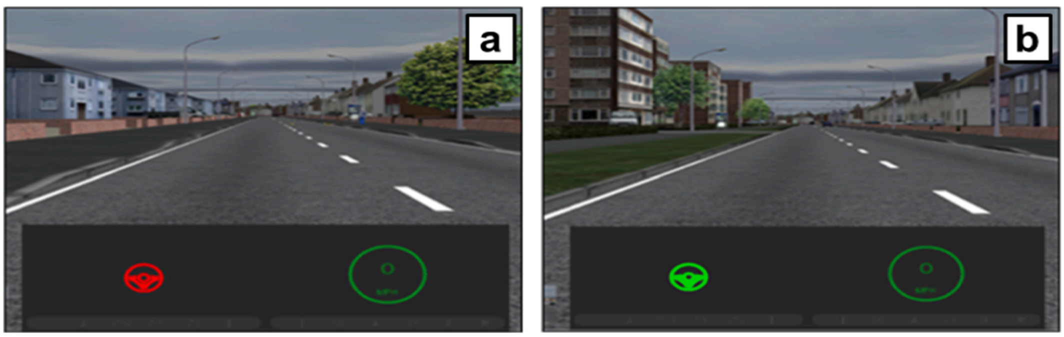

Fig. 1. HMI Interface on the dashboard: (a) when automation was disengaged; (b) Automation was engaged.

2.2. Apparatus

The experiment was conducted in the full motion-based University of Leeds Driving Simulator (UoLDS), which consists of a Jaguar S-type cab housed in a 4 m diameter spherical projection dome with a 300° field of view projection system. The simulator also incorporates 8 degrees of freedom electrical motion system. This consists of a 500 mm stroke-length hexapod motion platform, carrying the 2.5 T payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees of freedom of the Cartesian inertial frame. Additionally, the platform is mounted on a railed gantry that allows a further 5 m of effective travel in surge and sway. Drivers' physiological data were collected using Biopac MP35 data acquisition system at 500 Hz, which consisted of ECG electrodes and an EDA sensor.

The Automated Driving System (ADS) was designed to control lateral and longitudinal operation of the vehicle at a speed of 40 mph. The HMI interface on the dashboard showed a red steering wheel symbol when the ADS was inactive (Fig. 1a), and a green steering wheel symbol when the ADS was active (Fig. 1b).

2.3. Study design

This study incorporated a mixed design, with within-participant factors of Time Headway (*Short*, *Long*), Drive Mode (*ACF*, *MCF*) and Lead Vehicle (*Lead*, *No Lead*), and a between-participant factor of Level of Automation (*L2*, *L3*). All factors, with the exception of Drive Mode, were counterbalanced.

Following a ~10-min practice drive, each participant experienced two experimental drives. All drives were completed in a single-carriageway urban environment, with a speed limit of 40 mph, and low-density oncoming traffic. For all drivers, their first experimental drive consisted of free driving for ~2 min. After around 2 min, a lead vehicle joined the driving lane and drivers were instructed to follow the lead vehicle for about 5 min, in what we termed as the manual baseline drive. Drivers' baseline car-following behaviour, including their preferred headway, was collected during this segment. Except for the manual baseline segment, which was only present in the first experimental drive, the two experimental drives were the same. There were 4 segments in each experimental drive, which were experienced in the following order: Automated drive 1, Manual drive 1, Automated drive 2, Manual drive 2 (Fig. 2).

During automation, the participants experienced one of the two THW conditions (0.5 s for the *Short* THW and 1.5 s for the *Long* THW, derived from the 25th and 75th percentile of a driver behaviour model by Ferson et al. (2019), based on naturalistic driving studies incorporating drivers' instantaneous aggressiveness during car-following scenarios), in a counterbalanced order. After the manual baseline drive, or at the start of the second experimental drive, drivers experienced a ~1-min free drive, after which the ADS was available. Drivers received a verbal-audio prompt: "Attention, engage automation" through the car's speakers, upon which, they could engage the ADS feature by pressing a button on the steering wheel. About a minute into the automated drive, a lead vehicle moved into the driving lane, starting the *ACF* segment.

ACF was followed by an auditory-verbal takeover request: "Attention, get ready to takeover". The takeover request was presented when the ego vehicle reached a section of the road with faded lane markings, representing a system limitation condition for the ADS. The takeover request was followed by a short acoustic tone (1000 Hz, lasting 0.2 s), with increasing frequency, until the driver resumed manual control. The ADS could be disengaged by either pulling the left-hand stalk, rotating the steering wheel by more than 2°, or pressing the brake or accelerator pedals.

To study how the time headway of a lead vehicle affected workload, we introduced two kinds of takeover scenarios in this study, one with the lead vehicle present (*Lead*) and one without the lead vehicle (*No Lead*). In the *No Lead* condition, the lead vehicle exited the road at an intersection shortly before a takeover request was given, and subsequently, a new lead vehicle joined the driving lane from the next intersection, ~10 s after participants resumed manual control of the vehicle. In the *Lead* condition, the lead vehicle continued in the driving lane when the takeover request was issued.

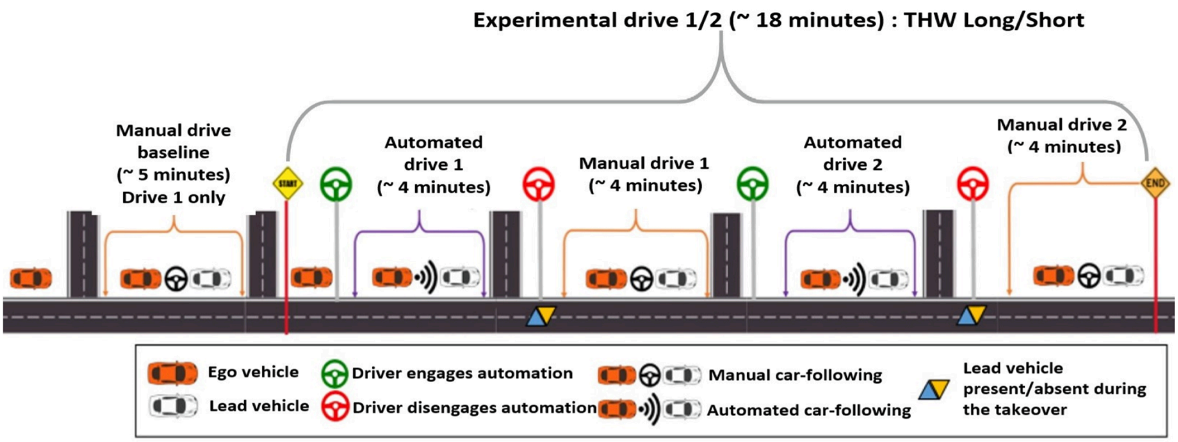

Fig. 2. Schematic representation of the experimental drives.

Fig. 3. (a) A representation of the arrows task with the upward facing arrow circled in red; (b) A participant engaging in arrows task in L3 group during automation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Each takeover was followed by a period of manual car-following (*MCF*). Each experimental drive consisted of *ACF*, *Takeover* and *MCF*, in that order, repeated twice, and incorporating the Lead Vehicle factor (see Fig. 2 and Fig. 4).

Level of Automation determined whether or not participants could engage in NDRTs during automation. In the *L3* group, participants were not required to monitor the drive and were asked to engage in an 'Arrows Task' (Jamson & Merat, 2005) during automation. This task required participants to search and select the upward-facing arrow, in a 4x4 grid of arrows displayed on a touchscreen (Fig. 3a), placed in the centre console, near the gearshift (Fig. 3b). The screen showed participants' cumulative score, as well as a "score to beat". To ensure full engagement in the Arrows task, participants were told they would receive an additional £5 if they beat this score, but, for ethical reasons, every participant was paid an additional £5 at the end of the study. This task was only available after engagement of automation, until when the takeover request was given. The instructions for participants in L2 automation was to monitor the road scene at all times, although they removed their hands from the steering wheel and foot off the pedals when automation was engaged.

2.4. Self-reported workload ratings

Each participant was asked to rate their workload 13 times (7 times in the first experimental drive, including once during the manual baseline drive and 6 times during the second experimental drive). Response was provided verbally on a scale of 1–10, with 10 denoting highest workload. When they were engaged in either *ACF* or *MCF* (two each per experimental drive), they were prompted with the following verbal-auditory message ~2.5 min after starting *ACF* or *MCF* (roughly halfway through the drive): "Please rate your workload now". Similarly, when they resumed manual control of the vehicle (two times per experimental drive), they were prompted with the following message 10 s after the takeover: "Please rate your workload during the takeover".

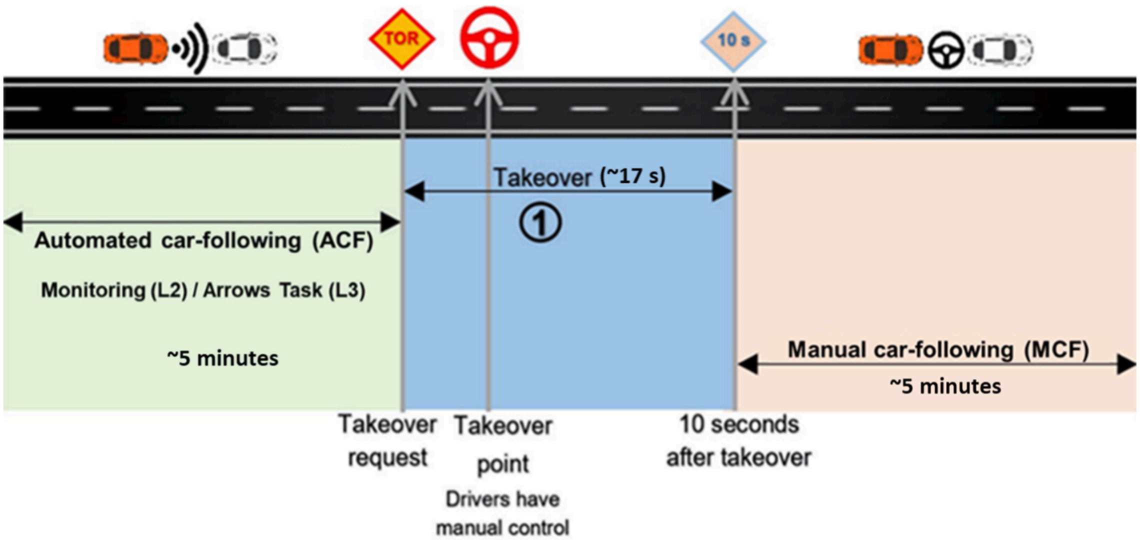

Fig. 4. Schematic depicting the Time windows used for data analysis.

2.5. Procedure

Upon arrival, participants were briefed with a description of the study, after which they were invited to sign a consent form, with an opportunity to ask questions about the study. Three ECG electrodes were then attached to the participant's chest, and 2 EDA electrode bands were attached to the index and middle finger of their non-dominant hand. Once the participant was seated in the simulator cab, we collected the physiological baseline data, where participants were asked to relax for a period of 7 min with their eyes closed, palms on their laps facing upwards (Braithwaite et al., 2015; Laborde et al., 2017). This was used to standardise the experimental physiological data. Participants then performed a practice drive, which included both automated and manual driving. During the practice drive, participants were talked through the various aspects of the vehicle HMI, how to engage and disengage automation and those in the *L3* group practised the Arrows task. After the practice drive, participants experienced the two experimental drives, which lasted ~18 min each.

For all manual driving segments, participants were instructed to adhere to the posted speed limit of 40 mph and drive in the centre of the lane. They were also asked not to overtake the lead vehicle, but otherwise, follow the normal rules of the road, ensuring the safe operation of the vehicle, and maintaining their desired distance from the lead vehicle. After each experimental drive, the participants were given a 10-min break, during which they were asked to complete a set of questionnaires, including Arnett's Sensation Seeking Questionnaire (Arnett, 1994), traffic locus of control (Özkan & Lajunen, 2005) and the Driver Style Questionnaire (French et al., 1993), see Louw et al. (2020). However, results from these questionnaires are not reported here, since they did not include questions about driver workload.

2.6. Data analysis

To analyse drivers' workload during *ACF*, *MCF* and *Takeover*, the physiological data was first segmented into appropriate time windows. Participants provided verbal self-reported workload ratings, for each of the *ACF*, *MCF* and *Takeover* windows, as mentioned in Section 2.4, which aligned with the segments used for physiological data collection. The time window for *ACF* was established as the time from when the lead vehicle entered the driving lane during the automated drive, until when the takeover request was given to the driver. The verbal response provided by the driver during this time window was considered as the self-reported workload rating for *ACF*.

Drivers' data from when they received the takeover request, until 10 s after they resumed manual control of the vehicle was classified as data for the *Takeover* window (see Fig. 4). The subjective workload ratings for the *Takeover* window were provided 10 s after drivers resumed manual control of the vehicle. We took 10 s after takeover as the cut-off point, as previous research has shown that the peak in driving performance decrement is observed within 10–15 s after takeover of control (Merat et al., 2014). For the purpose of this study, we classified the time window from 10 s after resuming manual control until re-engaging automation as *MCF*, and the verbal response given by the participant during this window was considered as self-reported workload rating during *MCF*.

Drivers experienced two experimentally similar *ACF* and *MCF* scenarios, in each of the experimental drives, as seen in Fig. 2. A set of t-tests applied to the physiological data, and self-reported workload ratings revealed no significant differences between the two *ACF* and two *MCF* scenarios, within each experimental drive. Therefore, in order to study changes in workload during automated and manual car-following, data for the two *ACF* scenarios, and the two *MCF* scenarios, in each experimental drive, were aggregated to a single representation. This was applied for both physiological data, and self-reported workload ratings.

2.7. Data analysis tools

The ECG data was processed on Kubios HRV premium software (Tarvainen et al., 2014). EDA signals were pre-processed, and artefacts were removed using custom algorithms based on recommendations from Braithwaite et al. (2015) and Kikhia et al. (2016), using MATLAB R2016a. The EDA data was analysed using Ledalab v3.9 (Benedek & Kaernbach, 2010), a MATLAB-based software package. The EDA signal was decomposed into tonic and phasic components, using continuous decomposition analysis (CDA; Benedek & Kaernbach, 2010). To identify a phasic event as an SCR, an amplitude threshold of 0.01 μS was used (Braithwaite et al., 2015). nSCR/min was computed as total number of SCRs in the window (above the amplitude threshold), divided by the time duration of the window (in seconds), which was then multiplied by 60, to get the number of SCRs per minute.

2.8. Statistical analysis

Statistical analysis was conducted with IBM SPSS Statistics 26. A Shapiro Wilk's test showed that the majority (>75%) of the group-level estimates were normally distributed for each of the dependent variables used, for every ANOVA test we conducted. For statistical significance, an α-value of 0.05 was used as a limiting criterion, and partial eta-squared was computed as an effect size statistic. Degrees of freedom were Greenhouse-Geisser corrected when Mauchly's test showed a violation of sphericity. Homogeneity of data was tested using Levene's test. In cases when the data was heterogeneous, the differences in group sizes were mostly equal (largest/smallest < 1.5), and the ANOVA was sufficiently robust to handle such heterogeneity in the data (Pituch & Stevens, 2016). Data from a participant was identified as an outlier if it was 3 times the interquartile range (IQR) above the 3rd quartile or below the 1st quartile, of the dataset. Due to a technical error with the voice recorder, self-reported workload ratings were missing for 5 participants, and self-reflected workload ratings data from a participant was identified as an outlier, across all the analyses done in this study. Therefore, data from these participants was excluded from the self-reported workload rating analysis. Data from 2 participants (one each from the

L2 and the *L3* group) in the EDR metric, nSCR/min metric and the RMSSD metric, were identified as outliers and excluded from the analysis. For analysis of RQ 3, data from 4 additional participants were identified as outliers, and removed from nSCR/min analysis.

For the self-reported workload ratings, only 50% of the dataset was normally distributed, and Levene's test revealed that workload ratings for the *Long* THW condition during *ACF* across the *L2* and *L3* groups were only slightly heterogeneous (p =.046), with equal group size. The skewness and kurtosis of the non-normally distributed data were within the acceptable range ±2 (George & Mallery, 2010). The one way ANOVA and the mixed ANOVA were robust enough to accommodate this violation of normality and homogeneity, with only a small effect on Type I error (Blanca et al., 2017; Pituch & Stevens, 2016).

3. Results

Since there is high inter-individual variability in physiological data, and this study incorporated a between-participant design, the physiological data during each of the *ACF*, *MCF* and *Takeover* time windows, for each participant, were standardised by representing the physiological values during *ACF*, *MCF* and *Takeover*, for each participant as a percentage of their respective physiological baseline data, collected before the experiment (see Section 2.5).

3.1. The effect of time Headway on driver workload during ACF in the L2 group (RQ 1)

To understand how drivers' workload was affected by the two THW conditions, during the monitoring phase in the *L2* group, we performed a one-way ANOVA with repeated measures on drivers' RMSSD, mean HR, EDR, nSCR/min and self-reported workload ratings, with a within-participant factor of Time Headway (*Short*, *Long*).

As shown in Table 1, there were no main effects of Time Headway, across all the physiological metrics and self-reported workload ratings, with participants showing similar physiological activity, and self-reported workload ratings, in the *Short* and the *Long* THW conditions.

3.2. The effect of an NDRT on driver workload during ACF (RQ 2)

To understand how drivers' workload during *ACF* was affected by engagement in an NDRT, compared to just monitoring the drive, we performed a one-way ANOVA on drivers' RMSSD, mean HR, EDR, nSCR/min and self-reported workload ratings, with a between-participant factor of Level of Automation (*L2*, *L3*). As seen in Section 3.1, in the *L2* group, participants exhibited similar physiological activity and self-reported workload ratings across the two THW conditions during *ACF*. In the *L3* group, participants were engaged in the visual Arrows task during *ACF,* without paying attention to the road and hence, did not observe the two THW conditions. Therefore, we combined the physiological metrics and self-reported workload ratings during *ACF*, for *Short* and *Long* THW conditions, into a single representation, for each group.

Our results indicated that there was a significant main effect of Level of Automation, across all the physiological metrics (Fig. 5, Table 2), during *ACF*, with drivers in the L3 group having significantly higher physiological activation, and hence, workload, when engaged in an NDRT, compared to those in the L2 group who were just monitoring the drive. Although statistically insignificant, a similar trend was observed in drivers' self-reported workload ratings, with drivers reporting higher workload when engaged in an NDRT (*L3* group), compared to monitoring the drive (*L2* group), during *ACF* (Fig. 6).

3.3. The effect of drive Mode on driver workload (RQ3)

To understand how drivers' workload varied between *ACF* and *MCF*, we conducted a one-way ANOVA with repeated measures on drivers' RMSSD, mean HR, EDR, nSCR/min values, and self-reported workload ratings, with a within-participant factor of Drive Mode (*ACF*, *MCF*). This was done separately for the *L2* and *L3* groups, since drivers in the *L3* group were only exposed to NDRT during *ACF* and not during *MCF*. Since drivers had similar workload across the two THW conditions, for all the physiological metrics and self-reported workload ratings, we combined the values for *Short* and *Long* THW conditions into a single representation, for both *ACF* and *MCF*.

There was an effect of Drive Mode on drivers' EDR values, in the *L2* group (see Table 3 and Fig. 8a), with drivers having significantly higher respiration rates during *MCF*, compared to *ACF*. However, there was no statistically significant effect of Drive Mode on any other physiological metrics, or self-reported workload ratings, for the *L2* group (see Table 3). On the other hand, a significant effect

Table 1
Results of the one-way ANOVA with repeated measures (RQ1) across various physiological measures and subjective ratings, for the Time Headway condition, in the L2 group.

Predictor	df1	df2	F	p	η_p^2
1. RMSSD	1	11	0.154	0.702	0.014
2. Mean HR	1	12	0.001	0.976	0.000
3. EDR	1	11	1.77	0.211	0.138
4. nSCR/min	1	11	0.041	0.843	0.004
5. Self-reported workload ratings	1	9	4.06	0.075	0.311

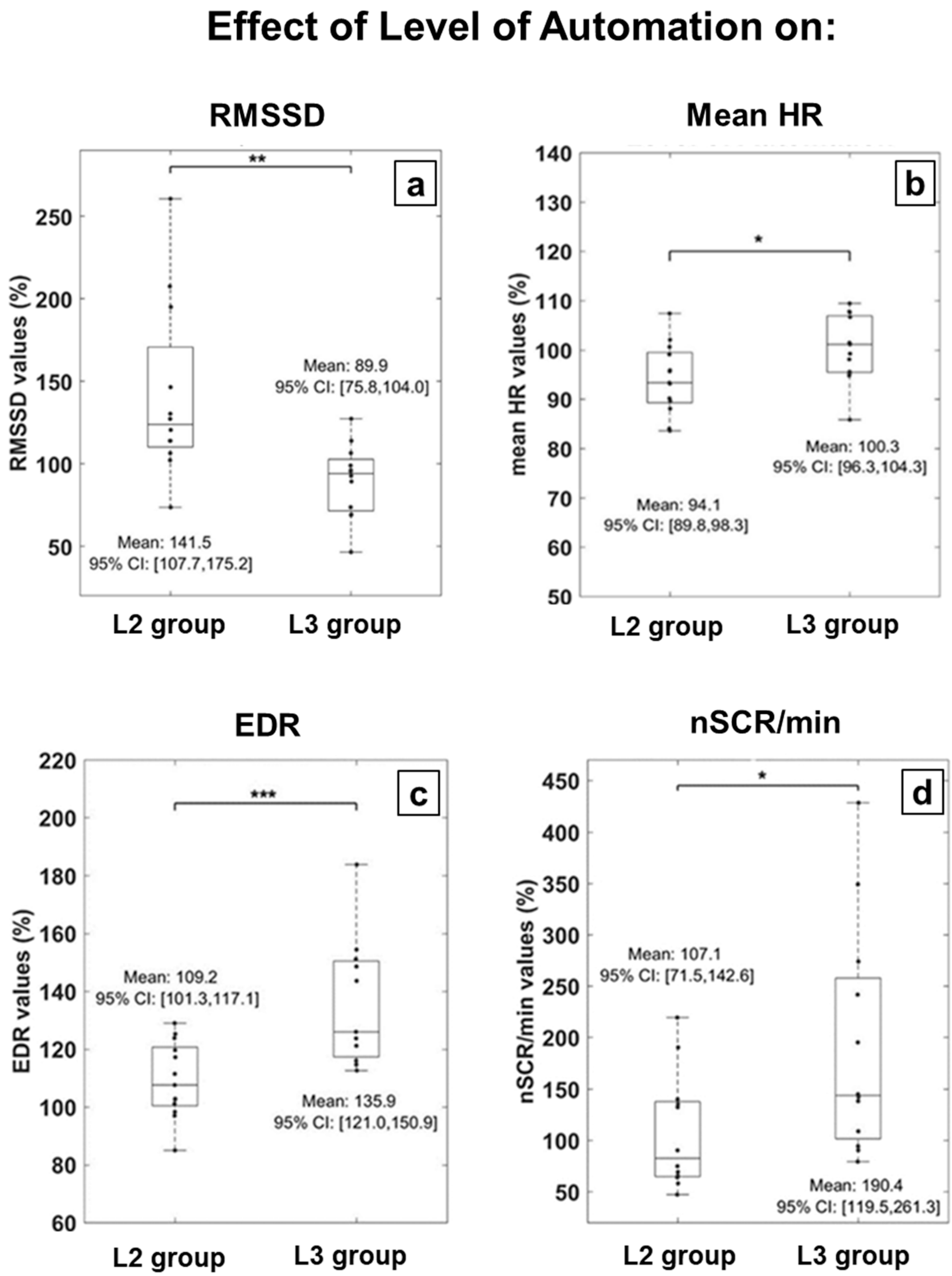

Fig. 5. Effect of Level of Automation on workload during ACF as reflected by (a) RMSSD (b) Mean HR (c) EDR and (d) nSCR/min metric. *$p \leq 0.05$, **$p \leq 0.01$ and ***$p \leq 0.001$.

of Drive Mode was seen across all the physiological metrics evaluated in the *L3* group (see Table 3). Drivers had significantly higher physiological activity during *ACF,* compared to *MCF* (see Table 3, Fig. 7, Fig. 8b, Fig. 9a), suggesting that a higher level of workload was imposed by engaging in the NDRT. Results from the ANOVA did not reveal any statistically significant differences in self-reported workload ratings between *ACF* and *MCF* in the *L3* group (see Table 3 and Fig. 9b).

Table 2
Results of the one-way ANOVAs across various physiological measures and self-reported workload ratings during ACF (RQ 2), for the Level of Automation condition.

Predictor	df1	df2	F	p	η_p^2
1. RMSSD	1	22	9.616	0.005	0.304
2. Mean HR	1	24	5.41	0.029	0.184
3. EDR	1	22	13.34	0.001	0.377
4. nSCR/min	1	22	5.35	0.030	0.196
5. Self-reported workload ratings	1	18	1.258	0.277	0.065

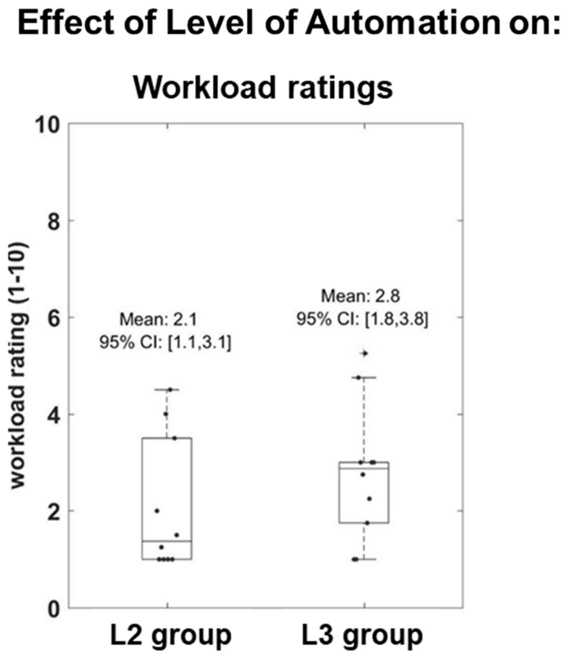

Fig. 6. Effect of Level of Automation on drivers' self-reported workload ratings during ACF.

3.4. *The effect of takeover on driver workload (RQ 4 and RQ 5)*

ECG-based metrics, such as RMSSD, mean HR and EDR, were not analysed around takeovers, as accurate analysis of such metrics requires a minimum time window of 5 min (Malik et al., 2012), and in this study, the takeover time windows (shown in light blue, Fig. 4) only lasted 15–20 s. As mentioned in Section 2.6, nSCR/min reported in this section was computed using the data for Takeover Window (shown in light blue, Fig. 4), that is, the time period from when they received the takeover request, until 10 s after they resumed manual control of the vehicle.

To understand how driver workload was affected by the time headway of the lead vehicle during takeovers, and whether their workload level during the takeover was affected by their engagement in an NDRT during automation (which occurred prior to the takeover), we performed a 2x2 mixed ANOVA, including a within-participant factor of Time Headway (*Short*, *Long*) and between-participant factor of Level of Automation (*L2*, *L3*), on drivers' nSCR/min, and self-reported workload ratings. Only the *Lead* vehicle condition was considered here, as the Time Headway factor is irrelevant when there is no lead vehicle. For the nSCR/min metric, there was missing data from 5 participants, and data from 2 other participants were classified as outliers and excluded from the analysis.

As shown in Table 3, ANOVA results revealed that the effect of Time Headway on drivers' nSCR/min was nearing statistical significance (p =.056) with drivers of both groups having higher mean nSCR/min values during the *Short* THW condition, compared to the *Long* THW condition (see Fig. 10a and Table 4). There was also an effect of Time Headway on drivers' self-reported workload ratings. Drivers had significantly higher self-reported workload ratings during the *Short* THW condition, compared to the *Long* THW condition (see Fig. 10b and Table 4), similar to the trend observed in nSCR/min metric. There was no effect of Level of Automation or interaction effects, for either nSCR/min or self-reported workload ratings, suggesting that drivers of both group experienced similar levels of workload during the takeovers.

4. Discussion and conclusions

This study investigated how changes in driver workload, imposed by different demands from a simulator-based automated driving

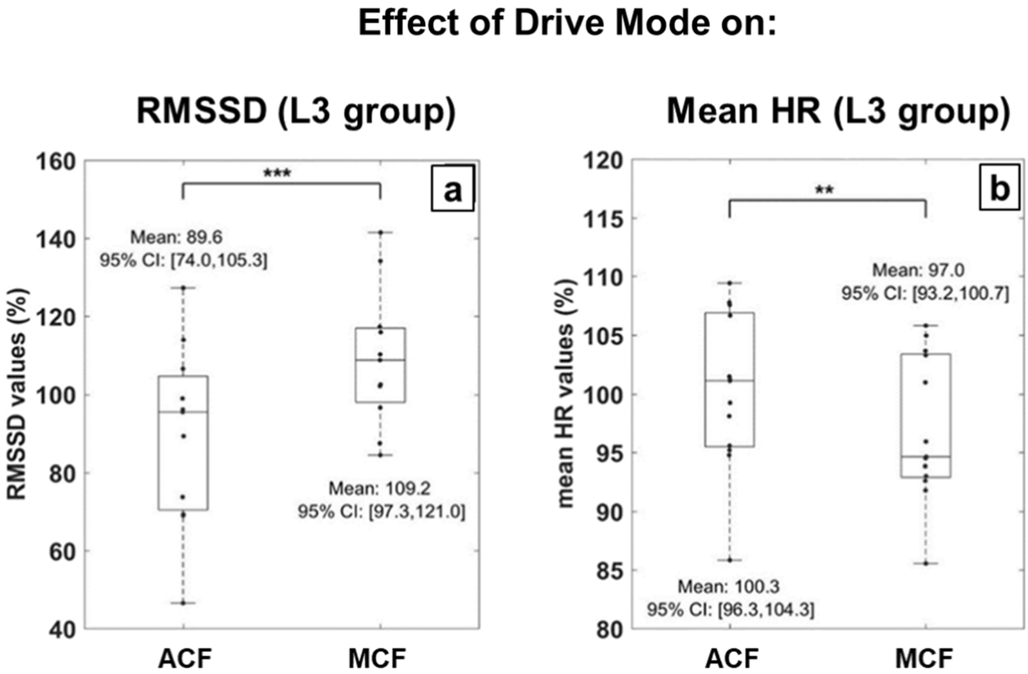

Fig. 7. Effect of Drive Mode on workload in the L3 Group, as reflected in (a) RMSSD metric and (b) mean HR metric. $^{**}p \le 0.01$ and $^{***}p \le 0.001$.

Table 3
Results of one-way ANOVA with repeated measures across various physiological measures, on Drive Mode, in the L2 and the L3 group.

Predictor	df1	df2	*F*	*p*	η_p^2
1. RMSSD					
L2 group	1	10	0.140	0.716	0.014
L3 group *	1	10	21.99	0.001	0.687
2. Mean HR					
L2 group	1	12	0.383	0.548	0.031
L3 group *	1	12	15.99	0.002	0.571
3. EDR					
L2 group *	1	10	6.08	0.030	0.336
L3 group *	1	10	18.98	0.001	0.655
4. nSCR/min					
L2 group	1	12	0.069	0.797	0.006
L3 group *	1	8	7.39	0.026	0.480
5. Self-reported workload ratings					
L2 group	1	9	0.421	0.533	0.045
L3 group	1	9	0.040	0.845	0.004

* statistically significant at $p \le 0.05$.

study, with car-following scenarios, affected drivers' physiological state and self-reported workload ratings. Two groups of participants were recruited to study how the presence of a lead vehicle, and its Time Headway (THW), affected workload during different stages of *L2* and *L3* automated driving. We also investigated whether engagement in an NDRT during *L3* automation increased driver workload.

Both physiological (ECG and EDA-based) metrics, and self-reported workload ratings indicated that the *L2* drivers experienced a similar level of workload when monitoring the lead vehicle during automation, whether this was in the *Short* (0.5 s) or *Long* (1.5 s) THW condition. This result is in contrast to results from a manual driving simulator study by Liu et al. (2019), who found significantly higher subjective feedback of workload for short THW conditions (0.5 and 1 s), compared to longer THWs of 2, 2.5 and 3 s. The absence of an effect of headway in the current study may be because the perceptual difference in the two levels of THW used was not prominent enough to affect drivers' workload levels, especially when they were simply monitoring the driving environment. Therefore, future studies should compare workload experienced at different THWs by both manual and automated driving, to understand if the workload experienced by short THWs is directly related to drivers' responsibility for the driving task.

When automation was engaged, drivers in the *L3* group (who were asked to conduct the Arrows task during automation) had significantly higher levels of workload, as illustrated by all the physiological metrics, when compared to the *L2* group. This suggests that the workload associated with both the physical movement (hand/finger), and the cognitive elements of the Arrows task in the *L3* group, was higher than that experienced by the simple monitoring of the drive by the *L2* group, although the between participant

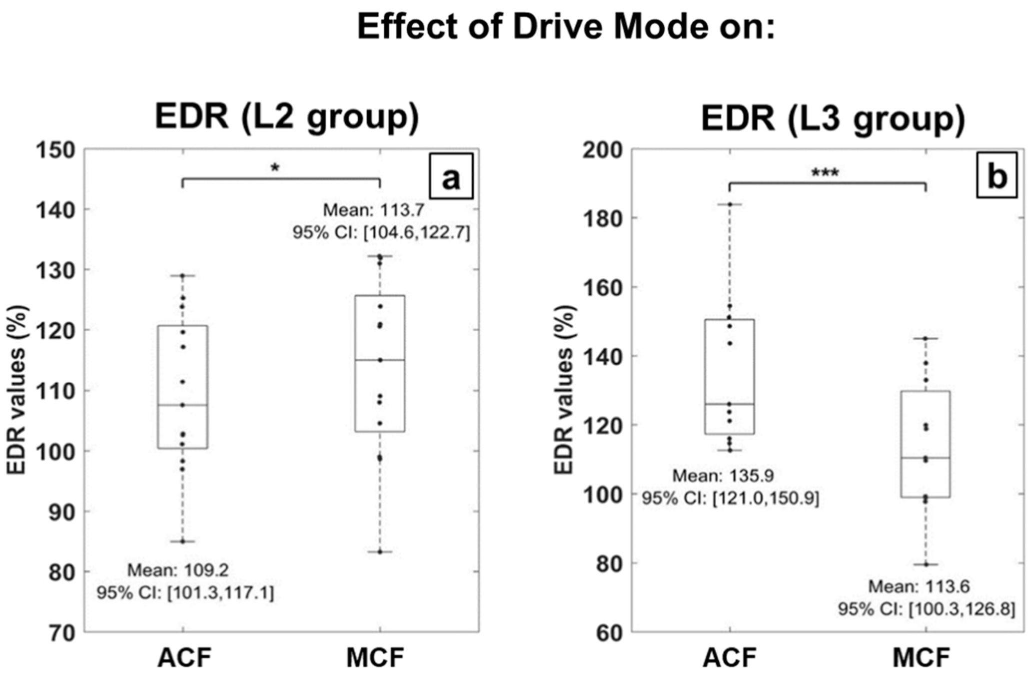

Fig. 8. Effect of Drive Mode on workload, as reflected in (a) EDR metric in the L2 group and (b) EDR metric in the L3 group. *p ≤ 0.05 and **p ≤ 0.001.

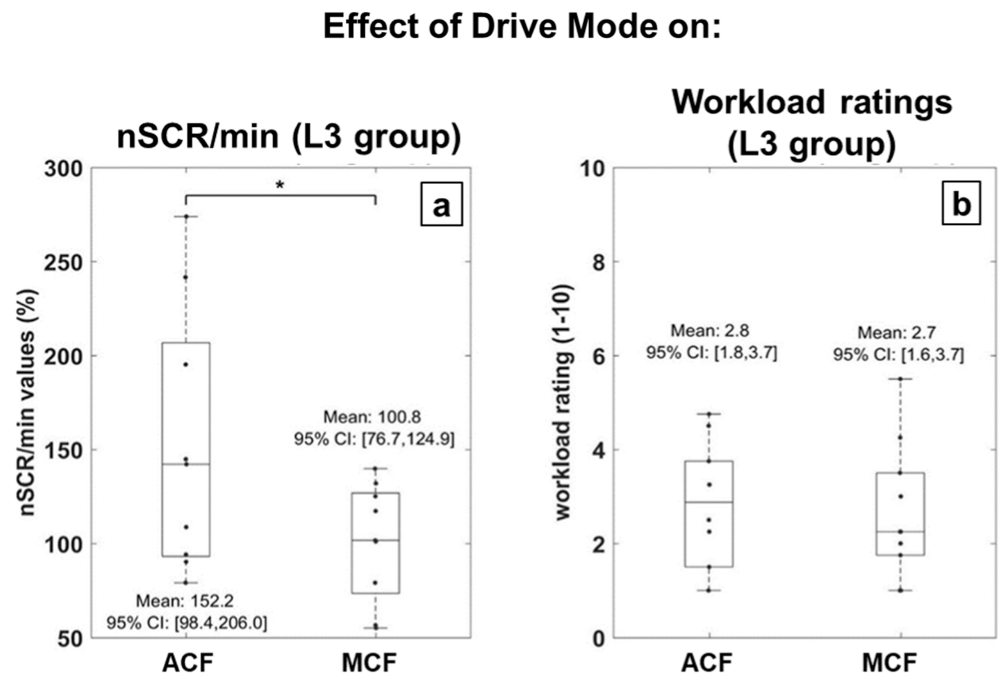

Fig. 9. Effect of Drive Mode on workload in the L3 group, as reflected in (a) nSCR/min metric and (b) Self-reported workload ratings. *p ≤ 0.05.

nature of this study must also be taken into account, when considering the implications of these results. These results are in agreement with Müller et al. (2021), who found a significant increase in drivers' workload levels, as reflected in their RMSSD values, when they were engaged in texting, using a touch-screen phone, compared to monitoring the drive, during an automated driving simulator study. Our participants' self-reported workload ratings during this period of automation were found to be similar for the *L2* and *L3* groups. Hart & Wickens (1990) argue that such self-reported workload ratings reflect subjective impressions of workload, and may differ from the workload reflected by task performance, which may suggest that drivers underestimated their workload levels when performing

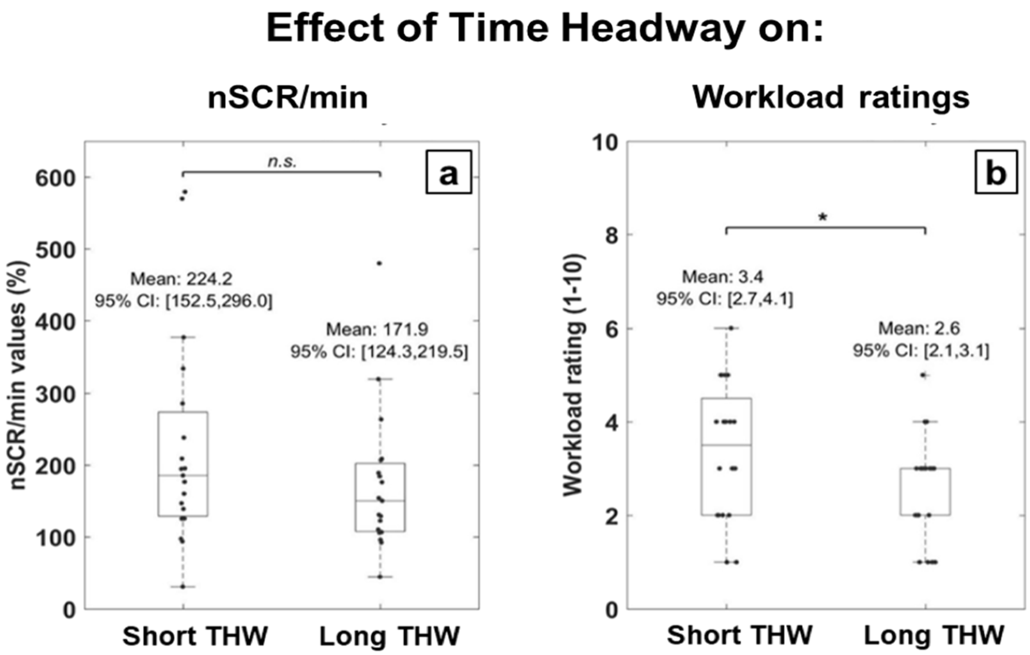

Fig. 10. Effect of Time Headway condition when a lead vehicle was present, during takeover, (a) on drivers' nSCR/min metric and (b) self-reported workload ratings during. n.s. nearing significance (p =.056), *p ≤ 0.05.

Table 4
Results of mixed ANOVAs across nSCR/min and self-reported workload ratings, during takeovers.

Predictor	df1	df2	*F*	*p*	η_p^2
1. nSCR/min					
Time Headway [n.s.]	1	17	4.21	0.056	0.198
Level of Automation	1	17	0.830	0.375	0.047
2. Self-reported workload ratings					
Time Headway	1	18	5.36	0.033	. 229
Level of Automation	1	18	0.045	0.834	0.003

[n.s.] nearing significance.

the Arrows task.

When comparing the physiological metrics observed in manual and automated driving in the *L2* group, results were only significant for the ECG based EDR-metric, with an in increase in respiratory activity during manual driving, compared to automated driving. This is likely due to the sensitivity of the EDR-metric to the physical act of steering and pedal control, required in manual driving (Hidalgo-Muñoz et al., 2019; Omlin et al., 2016). All other physiological metrics, and also the self-reported workload ratings, indicated that drivers had similar cognitive workload levels when monitoring the drive during automation, compared to when they were engaged in manual driving. This result is similar to that of Lohani et al. (2021), and Stapel et al. (2019), who observed that drivers exhibited similar levels of cognitive workload, as indicated by their RMSSD values (Lohani et al., 2021), or subjective ratings (Stapel et al., 2019), during L2 automation and manual driving.

For the *L3* group, drivers showed significantly higher workload, as indicated by all the physiological metrics, when they were engaged in the Arrows task during automated driving, compared to the workload levels observed in manual driving. While it is difficult to accurately separate the physical and cognitive demands of the Arrows task, it is assumed that the physical demand of responding to this simple touchscreen task was lower than that of manual driving. Therefore, higher levels of physiological activity during automation in the *L3* group are more likely to be linked to the cognitive demand from the Arrows task.

To understand the sensitivity of different physiological indices to physical and cognitive demands within the driving context, our study incorporated three distinct scenarios: manual driving, the monitoring task in *L2*, and conducting the Arrows task in *L3*, each of which had varying levels of physical and cognitive demands. In terms of the physical demand, we proposed that the monitoring task had the lowest physical demand, with manual driving likely being more physically demanding than the Arrows task. Given the similar driving environment across all the scenarios, we argue that the monitoring task and manual driving likely had similar cognitive demands (Lohani et al., 2021; Stapel et al., 2019), with the highest cognitive demand imposed by the Arrows task. The EDR was the only metric that picked up the changes in physical demand between monitoring and manual driving in the *L2* group, along with the differences in cognitive demand between the Arrows task and manual driving in the *L3* group, suggesting that it may be a better indicator of drivers' overall (physical and cognitive) workload levels in this study.

When considering how different Time Headways affected driver workload during takeovers across the *L2* and the *L3* groups, EDA-based nSCR/min were found to be marginally higher for the short THW condition, with the differences between the two THWs only approaching significance ($p = .056$). However, a significant difference was found for the self-reported ratings, with higher values for the *Short* THW condition. Therefore, although simple monitoring of a lead vehicle, which maintained a short headway during automation, did not seem to affect *L2* drivers' workload, an actual resumption of control while closely following a lead vehicle does seem to have increased drivers' perceived and actual workload levels. In terms of how workload at the takeover was affected by the previous activity (i.e. engaging in the Arrows task for the *L3* group, versus monitoring the drive in the *L2* group), results were similar for both physiological and self-reported workload ratings. This might suggest that the takeover was extremely demanding in itself, masking the effect of any activities that took place before the takeover request. On the other hand, since this was a non-critical takeover scenario, and drivers had adequate time to stop engaging in the Arrows task before resuming control, any additional workload imposed by the Arrows task may have been eliminated by this stage. Further research into safe and acceptable Time Headways maintained by AV controllers is, therefore, warranted, to ensure safe transition of control to the driver when required, without placing additional demand on the driver. Our results share some similarities with the findings of Dogan et al. (2019), who observed an increase in driver workload, measured by subjective ratings, with increases in the criticality of takeovers, influenced by shorter takeover lead time, and presence of an obstacle in the driving lane.

To conclude, our results illustrate the added value of psychophysiological metrics in identifying driver workload, during different stages of an automated drive. Our findings suggest that a demanding NDRT such as the Arrows task, can significantly increase drivers' workload levels, compared to just monitoring the drive, or manual driving. While the time headway conditions did not have an effect on drivers' workload levels when monitoring the drive during automation, we observed that the presence of a lead vehicle maintaining a shorter time headway significantly increased drivers' workload levels during the takeover, when they had to resume manual control of the vehicle. Although physiological signals are susceptible to motion artefacts, we were able to filter these out and objectively monitor drivers' workload levels, in a fully motion-based simulator environment, which emulates motion experienced during real-world driving. However, it is well known that physiological metrics are sensitive to a wide range of stimuli (Backs & Boucsein, 2000). To clearly interpret driver state using physiological metrics, especially in real-world scenarios, it is important to know the specific conditions and context which induced such physiological changes in drivers (Beggiato et al., 2019). Combining physiological metrics with camera-based sensors, and eye tracking data, can help to identify the cause of such physiological changes, and further improve driver state predictions. Therefore, further work is warranted to assess the real-time use of such physiological signals as part of future driver monitoring systems, to support drivers' safe operation of automated vehicles. Finally, it should be noted that in our study, participants in the *L3* group were always required to engage in an NDRT during automation. However, this would not necessarily be the case in real-world SAE L3 driving, where drivers could also choose to observe the driving environment. Our aim here, was to understand the impact of sustained eyes-off-road behaviour on physiological trends, making a clear distinction between the *L2* and *L3* groups. Future studies should investigate the impact of more naturalistic, self-regulated behaviours in SAE L3 driving.

Funding

This study was supported by funding from the European Commission Horizon 2020 program under the project L3Pilot, grant agreement number 723051. The main author's PhD is co-funded by Seeing Machines Pty Ltd, Canberra, Australia.

CRediT authorship contribution statement

Vishnu Radhakrishnan: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Natasha Merat:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Tyron Louw:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision. **Rafael Cirino Gonçalves:** Conceptualization, Investigation. **Guilhermina Torrao:** Conceptualization, Investigation, Writing – review & editing. **Wei Lyu:** Conceptualization, Investigation, Writing – review & editing. **Pablo Puente Guillen:** Conceptualization, Resources, Project administration, Funding acquisition. **Michael G. Lenné:** Resources, Writing – review & editing, Supervision.

Acknowledgement

The authors would like to thank all partners within L3Pilot and the University of Leeds Driving Simulator for their cooperation and valuable contribution. Responsibility for the information and views set out in this publication lies entirely with the authors. Additionally, the authors would like to thank Ms. Chen Peng, University of Leeds, for providing valuable insights regarding statistical methods used in this paper.

References

Arnett, J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences, 16*(2), 289–296. https://doi.org/10.1016/0191-8869(94)90165-1

Backs, R. W., & Boucsein, W. (2000). *Engineering psychophysiology: Issues and applications*. Lawrence Erlbaum.

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*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., & Strayer, D. L. (2018). 80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62*(1), 1878–1882. https://doi.org/10.1177/1541931218621427

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Datos no normales: ¿es el ANOVA una opción válida? *Psicothema, 29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing electrodermal activity & skin conductance responses (SCRs) for psychophysiological experiments*. https://doi.org/10.1017.S0142716405050034.

Bruggen, A. (2015). An empirical investigation of the relationship between workload and performance. *Management Decision, 53*(10), 2377–2389. https://doi.org/10.1108/MD-02-2015-0063

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *The handbook of psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Human Factors, 54*(5), 747–761. https://doi.org/10.1177/0018720812460246

de Waard, D. (1996). *The measurement of drivers' mental workload.* s.n.

Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 62*, 494–504. https://doi.org/10.1016/j.trf.2019.02.010

Du, N., Yang, X. J., & Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving. *Accident Analysis and Prevention, 148*. https://doi.org/10.1016/j.aap.2020.105804

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(2), 381–394. https://doi.org/10.1518/001872095779064555

ERTRAC (2017). *Automated driving roadmap*. European Road Transport Research Advisory Council. http://www.ertrac.org/uploads/documentsearch/id48/ERTRAC_Automated_Driving_2017.pdf.

Evans, L. (1991). *Traffic safety and the driver*. Van Nostrand Reinhold.

Ferson, N., Pizzigoni, E., Garnier-Follet, F., & Val, C. (2019). A new methodology to model driver behaviour accounting for the variation in driving manner using naturalistic driving data. In *26th international technical conference on the enhanced safety of vehicles* (pp. 1–8). https://www.researchgate.net/publication/338549089_A_new_methodology_to_model_driver_behaviour_accounting_for_the_variation_in_driving_manner_using_naturalistic_driving_data.

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics, 73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

French, D. J., West, R. J., Elander, J., & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergonomics, 36*(6), 627–644. https://doi.org/10.1080/00140139308967925

George, D., & Mallery, P. (2010). SPSS for Windows step by step: A simple guide and reference, 17.0 update. In *Boston: Pearson Education, Inc* (10th ed.). Allyn & Bacon.

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime - Performing non-driving related tasks while driving highly automated. *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931215591360

Hart, S. G., & Wickens, C. D. (1990). Workload assessment and prediction. In H. R. Booher (Ed.), *Manprint* (pp. 257–296). Springer Netherlands. https://doi.org/10.1007/978-94-009-0437-8_9.

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude, C. (2019). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Frontiers in Human Neuroscience, 12*, 525. https://doi.org/10.3389/fnhum.2018.00525

IEEE. (2019). *New level 3 autonomous vehicles hitting the road in 2020*. IEEE Innovation at Work. https://innovationatwork.ieee.org/new-level-3-autonomous-vehicles-hitting-the-road-in-2020/.

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2005.04.002

Kikhia, B., Stavropoulos, T. G., Andreadis, S., Karvonen, N., Kompatsiaris, I., Sävenstedt, S., Pijl, M., & Melander, C. (2016). Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors (Switzerland), 16*(12). https://doi.org/10.3390/s16121989

Kinnear, N., Stuttard, N., Hynd, D., Helman, S., & Edwards, M. (2021). *Safe performance of other activities in conditionally automated vehicles Automated Lane Keeping System Report details*. UK: Transport Research Laboratory.

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research – Recommendations for experiment planning, data analysis, and data reporting. In *Frontiers in psychology* (Vol. 8, p. 213). Frontiers. https://doi.org/10.3389/fpsyg.2017.00213.

Li, Y., Wang, H., Wang, W., Xing, L., Liu, S., & Wei, X. (2017). Evaluation of the impacts of cooperative adaptive cruise control on reducing rear-end collision risks on freeways. *Accident Analysis and Prevention, 98*, 87–95. https://doi.org/10.1016/j.aap.2016.09.015

Liu, K., Green, P., & Liu, Y. (2019). Traffic and ratings of driver workload: The effect of the number of vehicles and their distance headways. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 2134–2138. https://doi.org/10.1177/1071181319631051

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., & Strayer, D. L. (2021). No difference in arousal or cognitive demands between manual and partially automated driving: A multi-method on-road study. *Frontiers in Neuroscience, 15*. https://doi.org/10.3389/fnins.2021.577418

Louw, T., Goncalves, R., Torrao, G., Radhakrishnan, V., Lyu, W., Puente Guillen, P., & Merat, N. (2020). Do drivers change their manual car-following behaviour after automated car-following? *Cognition, Technology and Work*. https://doi.org/10.1007/s10111-020-00658-5

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2017). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention, 23*, 281–286. https://doi.org/10.1136/injuryprev-2016-042155

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (2012). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal, 17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board, 2138*, 6–12. https://doi.org/10.3141/2138-02

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors, 54*(5), 762–771. https://doi.org/10.1177/0018720812442087

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2018). The "Out-of-the-Loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work, 1–12*. https://doi.org/10.1007/s10111-018-0525-8

Meteier, Q., Capallera, M., Ruffieux, S., Angelini, L., Abou Khaled, O., Mugellini, E., & Sonderegger, A. (2021). Classification of drivers' workload using physiological signals in conditional automation. *Frontiers in Psychology, 12*, 596038. https://doi.org/10.3389/fpsyg.2021.596038

Müller, A. L., Fernandes-Estrela, N., Hetfleisch, R., Zecha, L., & Abendroth, B. (2021). Effects of non-driving related tasks on mental workload and take-over times during conditional automated driving. *European Transport Research Review, 13*(1), 1–15. https://doi.org/10.1186/s12544-021-00475-5

National Highway Traffic Safety Administration. (2009). *Traffic safety facts 2009: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system*. U.S. Department of Transportation.

NTSB (2017). *Highway accident report: Collision between a car operating with automated vehicle control systems and a tractor-semitrailer truck near Williston, Florida, May 7, 2016*. https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1702.pdf%0Ahttps://trid.trb.org/view/1485316.

Omlin, X., Crivelli, F., Heinicke, L., Zaunseder, S., Achermann, P., & Riener, R. (2016). Effect of rocking movements on respiration. *PLoS ONE, 11*(3). https://doi.org/10.1371/journal.pone.0150581

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., & Savinainen, M. (2008). Perceived mental stress and reactions in heart rate variability—A pilot study among employees of an electronics company. *International Journal of Occupational Safety and Ergonomics, 14*(3), 275–283. https://doi.org/10.1080/10803548.2008.11076767

Özkan, T., & Lajunen, T. (2005). Multidimensional traffic locus of control scale (T-LOC): Factor structure and relationship to risky driving. *Personality and Individual Differences, 38*(3), 533–545. https://doi.org/10.1016/j.paid.2004.05.007

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Pituch, K. A., & Stevens, J. P. (2016). Applied multivariate statistics for the social sciences. In *Applied multivariate statistics for the social sciences* (Sixth). Routledge, Taylor and Francis group. www.routledge.com/9780415836661.

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., & Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort. *Information (Switzerland), 11*(8), 390. https://doi.org/10.3390/INFO11080390

SAE International (2021). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*. https://www.sae.org/standards/content/j3016_202104/.

Siebert, F. W., & Wallis, F. L. (2019). How speed and visibility influence preferred headway distances in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 485–494. https://doi.org/10.1016/j.trf.2019.06.009

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 60*, 590–605. https://doi.org/10.1016/j.trf.2018.11.006

Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine, 113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident Analysis and Prevention, 35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human Factors, 60*(6), 870–881. https://doi.org/10.1177/0018720818768199

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention, 92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002