

**Effects of platooning on signal-detection performance, workload, and stress
A driving simulator study**

Heikoop, Daniel; de Winter, Joost; van Arem, Bart; Stanton, N.A.

DOI

[10.1016/j.apergo.2016.10.016](https://doi.org/10.1016/j.apergo.2016.10.016)

Publication date

2017

Document Version

Final published version

Published in

Applied Ergonomics: human factors in technology and society

Citation (APA)

Heikoop, D., de Winter, J., van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study. *Applied Ergonomics: human factors in technology and society*, 60, 116-127. <https://doi.org/10.1016/j.apergo.2016.10.016>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

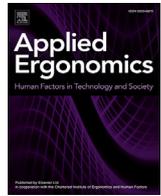
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' – Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.



Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study



Daniël D. Heikoop^{a,*}, Joost C.F. de Winter^b, Bart van Arem^c, Neville A. Stanton^a

^a Transportation Research Group, Faculty of Engineering and the Environment, Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton, SO16 7QF, UK

^b Department of BioMechanical Engineering, Delft University of Technology, The Netherlands

^c Department of Transport & Planning, Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:

Received 15 April 2016

Received in revised form

19 October 2016

Accepted 31 October 2016

Available online 22 November 2016

Keywords:

Psychophysiology

Automated driving

Vigilance

Monitoring

Supervisory control

Out-of-the-loop

ABSTRACT

Platooning, whereby automated vehicles travel closely together in a group, is attractive in terms of safety and efficiency. However, concerns exist about the psychological state of the platooning driver, who is exempted from direct control, yet remains responsible for monitoring the outside environment to detect potential threats. By means of a driving simulator experiment, we investigated the effects on recorded and self-reported measures of workload and stress for three task-instruction conditions: (1) No Task, in which participants had to monitor the road, (2) Voluntary Task, in which participants could do whatever they wanted, and (3) Detection Task, in which participants had to detect red cars. Twenty-two participants performed three 40-min runs in a constant-speed platoon, one condition per run in counter-balanced order. Contrary to some classic literature suggesting that humans are poor monitors, in the Detection Task condition participants attained a high mean detection rate (94.7%) and a low mean false alarm rate (0.8%). Results of the Dundee Stress State Questionnaire indicated that automated platooning was less distressing in the Voluntary Task than in the Detection Task and No Task conditions. In terms of heart rate variability, the Voluntary Task condition yielded a lower power in the low-frequency range relative to the high-frequency range (LF/HF ratio) than the Detection Task condition. Moreover, a strong time-on-task effect was found, whereby the mean heart rate dropped from the first to the third run. In conclusion, participants are able to remain attentive for a prolonged platooning drive, and the type of monitoring task has effects on the driver's psychological state.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The concept of a platoon—an actively coordinated, tightly spaced group of vehicles traveling together (Bergenheim et al., 2012; Ren and Green, 1994)—has been studied for several decades (e.g., Fenton et al., 1968; Thorpe et al., 1998). Because the vehicles in a platoon are driving with short yet constant headways, substantial benefits are achieved in terms of safety, traffic flow efficiency, and energy consumption (Hochstädter and Cremer, 1997; Karaaslan et al., 1991; Kunze et al., 2011; Tsugawa et al., 2011). Now that sensor, computer, and communication technologies are advancing rapidly, platooning is gaining interest among engineers (e.g., Larson et al., 2015; Ploeg et al., 2014) and Human Factors scientists (e.g., Gouy et al., 2014; Skottke et al., 2014).

Platooning often entails both longitudinal and lateral automation (e.g., Bergenheim et al., 2012), and hence no direct inputs by the driver are required. According to current legal frameworks, the driver must always be able to resume manual control (Kim et al., 2016; United Nations, 1968). Thus, the role of the driver in a platoon is, at present, ill-defined with, on the one hand, an exemption from control duties and, on the other, the ever-present requirement to be able to reclaim control (see also Norman, 2015). Unless the automated driving technology is legally allowed to drive in *all* environmental circumstances and is *perfectly* capable and reliable (or can always bring itself to a minimal-risk condition; Society of Automotive Engineers, 2014), the possibility remains that the driver has to take over control or modify the automation mode, set-points, or control laws (see also Sheridan, 2011).

Researchers have expressed concerns about the effects of platooning on the driver's psychological state (e.g., Levitan et al., 1998; Saffarian et al., 2012). Because the driver in a platoon is supervising the automation rather than manually controlling the car, there is

* Corresponding author.

E-mail address: d.d.heikoop@soton.ac.uk (D.D. Heikoop).

the risk of becoming drowsy, mentally underloaded, and fatigued (Cha, 2003; De Waard et al., 1999; Saxby et al., 2013; Young and Stanton, 2007). Although automated driving is experienced as effortless, at the same time the drivers are subjected to pressure because they have to remain alert in order to be able to intervene in a critical scenario (Banks et al., 2014; Casner et al., 2016). In fact, the notion that the vehicle is in control but the driver remains responsible for accidents that may occur has been said to be “a formula for extreme stress” (Hancock, 2015, p. 138). Furthermore, research has shown that when participants are tasked to monitor a machine in order to detect irregular events, they become frustrated and stressed (Scerbo, 2001; Szalma et al., 2004; Warm et al., 2008).

A common adage within the Human Factors domain is that humans are poor monitors (Hancock and Parasuraman, 1992; Kibler, 1965; Harris, 2002; Pritchett and Lewis, 2010; Sheridan, 1996; Wiener and Curry, 1980), or as Wiener (1985) put it: “After three decades of highly prolific research on human vigilance (Mackie, 1977), we are still making the seemingly contradictory statement: a human being is a poor monitor, but that is what he or she ought to be doing.” (p. 87). Farber (1999) pointed out that platooning drivers are unable to remain attentive for prolonged periods and will invariably engage in non-driving tasks. Empirical evidence concurs that drivers of automated vehicles are likely to engage in tasks such as calling on the phone, reading, interacting with a smartphone, or grabbing something from the rear compartment, making them unable to react in time if an emergency happens (Llaneras et al., 2013; Omae et al., 2005). It is for this reason that Google removed the steering wheel from their driverless cars (Teller, 2015). However, it is yet unknown whether Google's form of function allocation, in which the human is engineered out of the control loop, is tenable or legally acceptable (Kim et al., 2016). It certainly runs at odds with how automation has been deployed in complex systems such as aviation, water transport, and process control (see Sheridan, 2002).

Thus far, there appears to be no empirical evidence regarding the psychological state of platooning drivers as a function of monitoring task conditions. Moreover, much of what has been said of humans being poor monitors is based on experiments in which subjects sat in an isolated booth and responded to irregular stimuli having a low signal-to-noise ratio (cf. the highly-cited vigilance experiments by Mackworth, 1948). It is unclear to what extent the results of the classical vigilance paradigm generalize to complex supervisory tasks (Kibler, 1965; Stearman and Durso, 2016). According to a literature review by Cabrall et al. (2016), there is little overlap between the features of classic vigilance research and published experimental tasks of driving vigilance. A driving simulator study by Funke et al. (2007) found that drivers of a semi-automated vehicle actually performed *better* in a pedestrian-detection task than drivers in a manual control condition. Similarly, an on-road study by Davis et al. (2008) showed a performance improvement in target-detection performance for automated convoy driving as compared to manual convoy driving.

1.1. Present research

The aim of the present research was to investigate how the monitoring task of drivers in a platoon influences dimensions of stress, workload, and signal-detection performance. Participants were told that a critical situation may occur and that they had to intervene when needed. Three task instructions were compared: (1) ‘No Task’ (NT), in which no extra task was to be performed, (2) ‘Voluntary Task’ (VT), in which it was emphasized to the participants that they were free to do whatever they wanted, and (3) ‘Detection Task’ (DT), in which participants were asked to detect red cars among other traffic in the road environment. The NT condition

assessed the effects of monitoring demands that are similar to those that occur with modern forms of highly automated driving in which drivers should be vigilant for events that the automation cannot handle. The DT condition added extra task demands on top of the baseline monitoring demands, requiring the participant to scan cars in the environment. Conversely, the VT condition created a less demanding situation, allowing the driver to engage in non-driving tasks. The experiment was conducted in a driving simulator, providing a safe and controlled environment in which the traffic behaves identically for all participants.

Based on the aforementioned literature, we expected that the DT condition would yield the highest and the VT condition the lowest scores on stress and workload. In our study, stress dimensions (engagement, distress, & worry) were operationalized with the multi-dimensional Dundee Stress State Questionnaire (DSSQ; Matthews et al., 1999), whereas aspects of workload were assessed with the NASA Task Load Index (TLX). Additionally, we used cardiovascular measures, whereby heart rate was regarded as an indication of stress (Healey and Picard, 2004), and heart rate variability was regarded as an indication of workload (Brookhuis and De Waard, 2010; Cinaz et al., 2013; Fallahi et al., 2016; Jorna, 1992; Luque-Casado et al., 2016; Suriya-Prakash et al., 2015). Moreover, considering the literature about human vigilance performance, we expected that participants in the DT condition would miss a substantial number of red cars. An eye tracker was used to record the percentage of eye-closure as an indicator of task engagement (cf. Lal and Craig, 2002; Körber et al., 2015; Wierwille et al., 1994).

2. Method

2.1. Participants

Twenty-two participants (13 male) aged between 19 and 45 years ($M = 29.6$; $SD = 6.8$) with at least 1 year of driving experience ($M = 10.0$; $SD = 6.7$) were recruited. Most participants were from the University of Southampton community, with 14 participants being students, researchers, or lecturers at the university, a further four holding an engineering qualification, two being administrators, one being a medicine student, and one a police officer who indicated that driving is part of his profession. In order to retain a typical driving population, we did not apply exclusion criteria regarding personal characteristics that are known to be associated with heart rate variability, such as being a smoker (Barutcu et al., 2005) or general fitness level (Corrales et al., 2012; Luque-Casado et al., 2013). However, being healthy and having 20/20 vision were inclusion criteria, and given the acute effects of smoking on heart-rate variability (Karakaya et al., 2007; Manzano et al., 2011), we verified that none of the participants engaged in smoking in between the experimental sessions. Five participants indicated they drove less than once a month, five once a month, three 1–3 days a week, three 4–6 days a week, and six every day in the past 12 months. Seven participants indicated they drove 1–1000 miles, three 1001–5000 miles, six 5001–10,000 miles, four 10,001–20,000 miles, one 20,001–30,000 miles, and one over 50,000 miles in the past 12 months.

All participants in this experiment read and signed a consent form. The study was approved by the Ethics and Research Governance Online of the University of Southampton under submission ID number 13967.

2.2. Apparatus

The experiment was performed in the Southampton University Driving Simulator (SUDS; Fig. 1), a Jaguar XJ Saloon. The simulator ran on STISIM Drive[®] 3, which is a widely used driving simulator

software that allows for custom scenario building (Mets et al., 2011). The simulation was shown on three screens in the front creating a 135-degree field-of-view, one screen at the back for a rear view image, and two side mirror displays.

Seeing Machines FaceLab 5 eye tracker captured the participants' eye movements, and AD Instruments PowerLab26T (consisting of three MLA2505 biopotential electrodes, lead wires with disposable ECG electrode patches, and LabChart 8 software) was used for ECG measurements. 'Normal to Normal' (NN) intervals were extracted by the LabChart 8 software using the standard human ECG mode, with the default 2 standard deviation threshold for detection. Furthermore, a low-pass filter (cut-off frequency 50 Hz) was selected to filter extraneous noise.

2.3. Environment

The simulation showed an eight-lane highway (four lanes in either direction) with mild curves and hills. Participants were automatically transported within a five-car platoon, with the third car being the ego car. The environment consisted of sparsely distributed buildings, blank overhead signs, and trees. The starting position of the platoon was on the slow (left) lane of the highway, and the centre-to-centre distance between cars was 9 m (translating into a time headway of about 0.3 s). The longitudinal and lateral movements of all cars of the platoon, including the ego car, were synchronous and fully automated. At the start of each run, the platoon accelerated to 120 km/h and maintained that speed for the entire run. The platoon's lateral movement involved seven overtaking manoeuvres per run by means of a single lane change back and forth.

2.4. Procedure

Participants were asked to bring some entertainment they could

use while sitting down, like a book or e-reader, something to eat or drink, a laptop, or a mobile phone. These items could be used during one of the conditions of the experiment.

Upon arrival, participants received paper instructions explaining that they would be driving three 40-min runs along a highway in an automated platoon. Furthermore, information on the procedures of the experiment, a consent form, a figure depicting electrode placement, a demographics questionnaire, and the first part of the DSSQ were provided. This pre-task DSSQ queried the participants' current state prior to the simulator runs.

Participants read the instructions and filled out the questionnaires. In addition, the eye tracker was calibrated and the ECG electrodes were attached. The three electrodes were placed in a triangular configuration, with two electrodes placed below the far ends of the collar bones and one electrode over the xiphoid process (males), or one electrode at the top of the sternum and two electrodes below the ribs on both sides (females) (see e.g., Shaffer and Combatalade, 2013).

To indicate readiness to begin the experiment, participants pressed a handheld button, after which the run was started. After each run, the participants received the post-task DSSQ and the TLX. The post-task DSSQ asked the participants about their state while performing the task.

2.5. Conditions

The experiment consisted of three 40-min runs, one task condition per run in counterbalanced order. Although no automation malfunction actually occurred, for this is likely to be a rare event in real-life automated platooning, for all three conditions, participants were told that a critical situation may occur and that they had to intervene when needed. Specifically, before each run, participants received paper instructions, which differed per experimental condition:



Fig. 1. The Southampton University Driving Simulator (SUDS) during the experiment.

- (1) 'No Task' (NT), in which participants were asked to monitor the road, to intervene whenever a critical situation appears, and to avoid accidents at all times.
- (2) 'Voluntary Task' (VT), in which next to the tasks of the NT condition, participants were told they could do whatever they wanted, including reading, drinking, eating, detecting red cars, or even sleeping, as long as they felt able to intervene at all times.
- (3) 'Detection Task' (DT), in which, in addition to the tasks of the NT condition, participants were asked to detect red cars by pressing the handheld button.

Note that the driving task was fully automated, meaning that it was both hands-free and feet-free. Thus, although participants were instructed to intervene if the situation demanded it, they could not intervene at any point during the experiment.

The traffic density in the participants' direction of travel was low, with 23 vehicles distributed across the four lanes that were either overtaking or being overtaken by the platoon during the run. The opposite side of the road contained 5179 vehicles over the duration of the run, comprising a flow of traffic distributed across all four lanes with decreasing density towards the fast lane. The average capacity flow per lane was 1942 cars per hour, which is in line with typical free-flow traffic (Knoop et al., 2008). Each half run contained either a low rate (LR) or high rate (HR) of red cars, counterbalanced across participants. The red cars could be driving in the participants' own lane (1 being overtaken by the platoon), in the three other lanes (10 overtaking the platoon), as well as in the four lanes among the traffic driving in the opposite direction (59 in total), on predetermined randomly distributed times during the drive. The LR half contained 20 red cars per 20 min, whereas the HR half contained 50 red cars per 20 min, totalling 70 red cars per run. The 70 red cars were conspicuous with respect to the remaining traffic. That is, the colours of the remaining vehicles were never dark orange or brown, but, for example, black, grey, white, or blue.

2.6. Dependent measures

The following dependent measures were calculated per run:

- Hit rate (% of 70 red cars). The hit percentage was calculated automatically by assessing for each of the 70 red cars whether the button was pressed. In the calculation of the hit rate, a two-stage approach was used by distinguishing between the 59 approaching cars and the 11 overtaken/overtaking red cars that could occur simultaneously with the approaching cars. The time that approaching cars were visible was approximately 8 s, including the time they were visible in the mirrors. A hit for an approaching car was defined as a button press between 2 s before until 10 s after the red car could first be seen. This 2-s time buffer before and after appearance of the red cars was adopted to account for possible synchronisation discrepancies in the data logging. After determining the hits for the 59 approaching cars, it was assessed whether the remaining button presses could be assigned to the overtaken red car (20 s visibility interval) or the 10 overtaking red cars (40 s visibility interval). If a participant pressed a button more than once for the same red car, this counted as a single hit (cf. Mueller and Piper, 2014).
- The false alarm rate (% of 53 time intervals). A false alarm was defined as a button press in a time interval in which a red car could not be seen. The maximum possible number of false alarms was 53, and the total time interval in which there was an opportunity for generating a false alarm was 1409 s (corresponding to 58.5% of the eight 300-s intervals).

- Heart rate (bpm), a measure of stress.
- SDNN (ms). This time-domain measure of heart rate variability is defined as the mean of the standard deviation of all NN intervals. A decrease in SDNN is an indication of an increase of workload (e.g., Fallahi et al., 2016; Suriya-Prakash et al., 2015).
- LF/HF ratio. This frequency-domain measure of heart rate variability is defined as the power of the NN interval in the low-frequency (LF) 0.04–0.15 Hz range relative to the high frequency (HF) 0.15–0.40 Hz range. An increase in the LF/HF ratio is an indication of an increase in workload (Cinaz et al., 2013; Suriya-Prakash et al., 2015). Both the SDNN and the LF/HF ratio were calculated from the NN intervals after a default NN artefact filter, using an open-source MATLAB program provided by Vollmer (2015).
- Percentage eyes closed (PERCLOS; %), defined as the percentage of time that the eyes were practically closed across a moving time window of fixed size, excluding blinks.
- DSSQ, a self-report measure of stress states. Version 1.3 (Matthews et al., 2000) was used in this experiment. Standardized change scores for each scale of the DSSQ were calculated as follows: (post-score – pre-score)/(standard deviation of the pre-score) (Helton et al., 2002). The scores for the three DSSQ scales (engagement, distress and worry) were calculated as the means of four subscales (based on Fairclough and Venables, 2005; Heikoop et al., 2016b; Matthews et al., 2002; Matthews, 2014). Specifically, engagement consists of the subscales (1) energetic arousal, (2) success motivation, (3) intrinsic motivation, and (4) concentration, distress consists of (5) tense arousal, (6) hedonic tone, (7) control and confidence, and (8) anger/frustration, and worry consists of (9) self-focused attention, (10) self-esteem, (11) task-relevant interference, and (12) task-irrelevant interference. The internal reliability (Cronbach's alpha) of the subscales of the DSSQ ranges from 0.77 to 0.89 (Matthews et al., 2002).
- TLX, as a self-report measure to assess workload. The TLX is the most widely used measure of self-reported workload (De Winter 2014) and has shown a test-retest reliability of 0.83 across a four-week period (Hart and Staveland, 1988).

The hit rate, false alarm rate, cardiovascular measures, and eye-closure measure were calculated per 5-min interval to be able to assess time-on-task effects within a session, as well as the overall effect of a session. Specifically, each 2413-s long run was divided into eight 300-s segments (10–310 s, 310–610 s, ..., 2110–2410 s; the first 10 s and the last 3 s were discarded). The measures were calculated per segment and subsequently averaged across the eight segments. For the SDNN variable, this approach is equivalent to the SDNN index, a measure that is robust to low-frequency drifts in the data (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996).

2.7. Statistical analyses

The three conditions were compared with a repeated measures analysis of variance (ANOVA) with a significance level α of 0.05. The ANOVA was followed by paired comparisons between the three conditions using a Bonferroni correction. Hence, in the pairwise comparisons, a result was declared statistically significant if the p value was smaller than 0.05/3. Effect sizes between pairs of conditions were calculated as Cohen's d_z for matched pairs. With a sample size of 22, the power to detect medium ($d_z = 0.5$) and large effects ($d_z = 0.8$) at a significance level of 0.05/3 was 42% and 86%, respectively (Faul et al., 2007). In order to assess the validity of the

workload measures, a Spearman rank-order matrix was constructed between personal characteristics (age, gender, driving experience, driving frequency, mileage) and measures of workload (heart rate variability, self-reported workload).

3. Results

3.1. Button presses, hits, and false alarms

In the VT condition, 11 out of 22 participants took up the detection task, of which 6 performed this task throughout the entire run and 5 partially. In the three runs combined, the button was pressed 2118 times, of which 11 were false alarms, 2073 were hits, and 34 were cases in which a participant pressed the button more than once in a red-car interval. For example, some participants pressed the button twice when a red car was overtaking them: once when the red car appeared in the rear-view mirror and once again when it appeared in the frontal view. Participants never pressed the button twice in a no-red-car interval.

The participants' mean (*SD*) hit rate was 39.9% (42.1%) and 94.7% (4.2%), whereas the mean (*SD*) false alarm rate was 0.2% (0.6%) and 0.8% (1.7%) for the VT and DT conditions, respectively. For the DT condition, the hit rate and false alarm rate translate into a perceptual sensitivity (d') of 4.04 and a response bias (β) of 5.10 (Stanislaw and Todorov, 1999). In other words, participants were exceptionally well able to discriminate the red cars from the remaining traffic and were highly conservative in responding. Across the 22 participants, the hit rate ranged between 0% and 98.6% for the VT condition, and between 87.1% and 100% for the DT condition. The false alarm rate ranged between 0% and 1.9% for the VT condition, and between 0% and 7.5% for the DT condition.

Fig. 2 shows the hit rate per 5-min interval. It can be seen that the hit rate remained approximately constant with time, indicating there was no substantial vigilance decrement.

3.2. Self-report questionnaires: DSSQ and TLX

The results of the DSSQ show that in all three conditions, participants had lost substantial task engagement with respect to the pre-task score (i.e., standardized change scores below zero, see Fig. 3). Several statistically significant effects between the three conditions were observed for the three DSSQ dimensions. Specifically, the VT condition yielded a lower Distress score than the VT condition. Moreover, NT resulted in higher Worry than VT and DT.

The pairwise comparisons of the TLX showed that the DT condition was rated as more mentally demanding than the other two conditions (Fig. 4). Additionally, participants rated DT as significantly more effortful and frustrating compared to the VT condition.

3.3. Heart rate

The mean (*SD*) heart rate for the NT, VT, and DT conditions was 70.3 bpm (12.1), 69.4 bpm (10.9), and 70.9 bpm (12.5), respectively. According to a repeated measures ANOVA, the mean heart rate between the three conditions was not significantly different, $F(2,42) = 0.84$, $p = 0.438$. The pairwise comparisons were not significantly different from each other either (NT vs. VT: $d_z = 0.16$, $p = 0.462$; NT vs. DT: $d_z = -0.12$, $p = 0.577$; VT vs. DT: $d_z = -0.26$, $p = 0.242$).

A follow-up analysis revealed that there was a strong time-on-task effect, $F(2,42) = 14.1$, $p < 0.001$, with the heart rate dropping significantly from Run 1 ($M = 72.5$, $SD = 11.0$) to Run 3 ($M = 67.8$, $SD = 11.4$), $d_z = 0.96$, $p < 0.001$ (see Fig. 5).

3.4. Heart rate variability

The mean (*SD*) SDNN for the NT, VT, and DT conditions was 75.6 ms (29.5), 75.5 ms (30.7), and 75.0 ms (34.9), respectively. The three conditions were not significantly different ($F(2,42) = 0.01$, $p = 0.992$), and the pairwise comparisons were not significant either (NT vs. VT: $d_z = 0.00$, $p = 0.985$; NT vs. DT: $d_z = 0.02$, $p = 0.910$; VT vs. DT: $d_z = 0.02$, $p = 0.921$). As with heart rate, a time-on-task effect was found: $F(2,42) = 11.7$, $p < 0.001$; the SDNN rose from Run 1 ($M = 65.6$, $SD = 27.7$) to Run 3 ($M = 83.1$, $SD = 33.8$), $d_z = -0.82$, $p < 0.001$.

The mean (*SD*) LF/HF ratio was 1.15 (0.27), 1.07 (0.30), and 1.18 (0.30) for the NT, VT, and DT conditions, respectively. The three conditions were significantly different ($F(2,42) = 5.78$, $p = 0.006$), and the pairwise comparisons showed that the VT condition had a lower LF/HF ratio than the DT condition (NT vs. VT: $d_z = 0.51$, $p = 0.025$; NT vs. DT: $d_z = -0.17$, $p = 0.429$; VT vs. DT: $d_z = -0.76$, $p = 0.002$). There was no significant effect of run number on the LF/HF ratio ($F(2,42) = 1.03$, $p = 0.365$).

3.5. Percentage eyes closed (PERCLOS)

The eye-tracker was often unable to track the participants' eyes. First, we performed a data quality check to examine whether the amount of missing data differed between the three task conditions. For the eye-closure variable, 42% ($SD = 39\%$), 63% ($SD = 35\%$), and 42% ($SD = 33\%$) of data points were missing for the NT, VT, and DT conditions, respectively ($F(2,42) = 7.09$, $p = 0.002$; NT vs. VT: $d_z = -0.59$, $p = 0.012$; NT vs. DT: $d_z = -0.02$, $p = 0.927$; VT vs. DT: $d_z = 0.73$, $p = 0.003$). The particularly high amount of missing data for the VT condition may be explained by the fact that participants in this condition were told that they could do whatever they wanted, and therefore strayed from the eye tracker's field-of-view. When selecting the 16 participants who had available eye-closure data for each of their three runs, the mean (*SD*) eye-closure percentages were 5.2% (8.6%), 5.8% (8.9%) and 3.1% (3.7%) for the NT, VT, and DT conditions, respectively, $F(2,30) = 1.36$, $p = 0.271$.

There were no statistically significant differences between Runs 1, 2, and 3, $F(2,30) = 0.03$, $p = 0.973$, but a further analysis indicated that eye-closure increased with time-on-task. Specifically, averaged across the three task conditions, the mean eye-closure in the first and second half of the session was 5.1% and 6.9%, respectively. The amount of missing data increased as well, from 47.1% in the first half of the session to 51.0% in the second half of the session.

3.6. Correlation analysis

Table 1 shows the correlation matrix among the personal characteristics, cardiovascular measures, and TLX items. Although interpretation should be done with caution due to the small sample size, several patterns emerge. First, it can be seen that older participants showed indications of lower heart rate variability, which is in line with the literature (Voss et al., 2015). Second, the heart rate variability measures are not interpretably related to self-reported workload, with one scale of self-reported workload (frustration) in fact showing a statistically significant positive correlation with SDNN. Third, heart rate, SDNN, and the LF/HF ratio are strongly correlated with each other, indicating substantial redundancy of these cardiovascular measures. Fourth, signal detection performance is not significantly related to driving experience and mileage, but it is related to workload, which may be because people who tried harder on the task obtained a better detection performance.

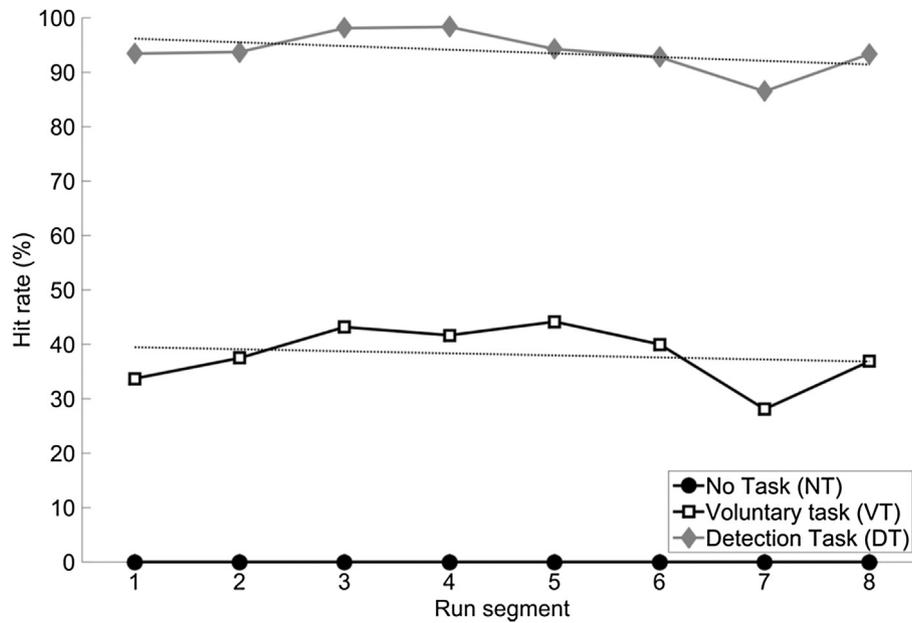


Fig. 2. Hit rate per 5-min segment during the run, for each of the three experimental conditions. The dotted lines represent linear trend lines. Note: The fluctuation in the hit rate between segments (e.g., low hit rate in Segment 7) is partly attributable to variations in the visibility of the cars (e.g., a red car which was partly/completely covered by another vehicle, or the redness of a car was visible from close by only).

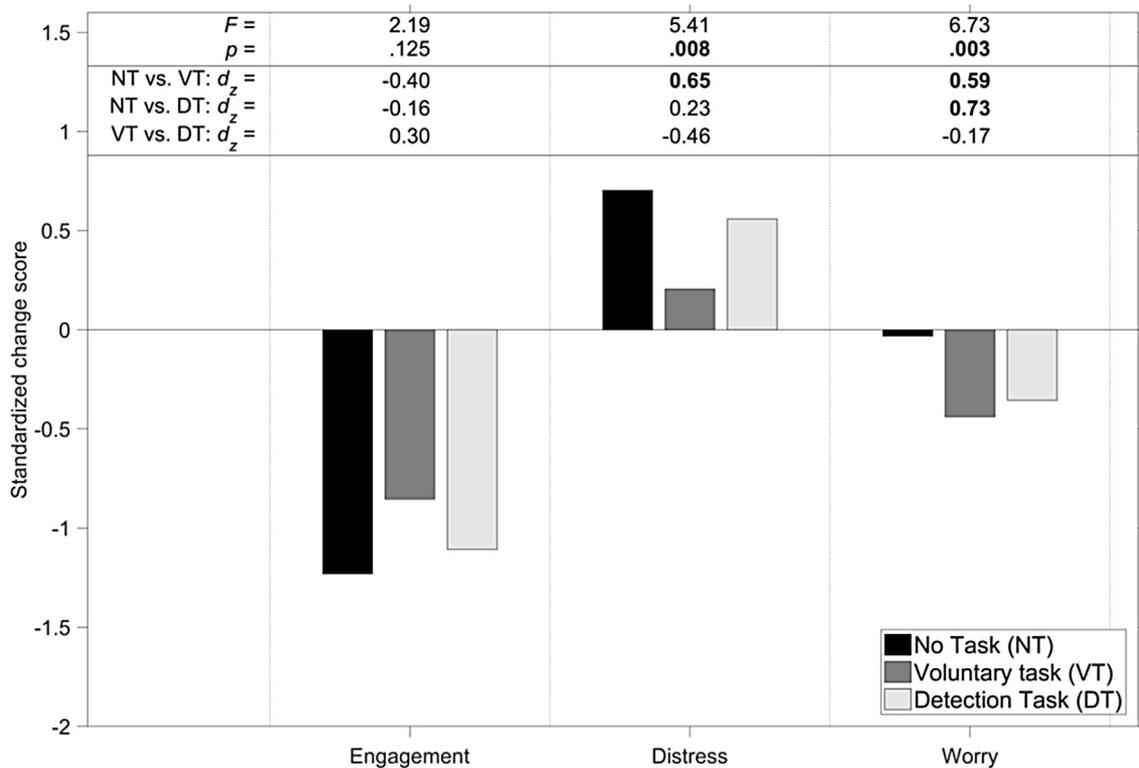


Fig. 3. Standardized change scores for the DSSQ for the three experimental conditions. Also shown are the F statistic and p value for the repeated measures ANOVA and the Cohen's d_z effect size between the three pairs of conditions. For the ANOVA, the p value is shown in boldface if $p < 0.05$. For the pairwise comparisons, the d_z is shown in boldface if $p < 0.05/3$.

4. Discussion

The aim of this study was to assess the effect of task instructions on dimensions of workload and stress (engagement, distress, worry) in an automated platoon for extended periods. For each task

condition, participants were informed that they had to monitor the road, intervene whenever a critical situation appeared, and avoid accidents at all times. In the Voluntary Task (VT) condition participants were further informed that they could do whatever they wanted. In the No Task (NT) condition, participants had no

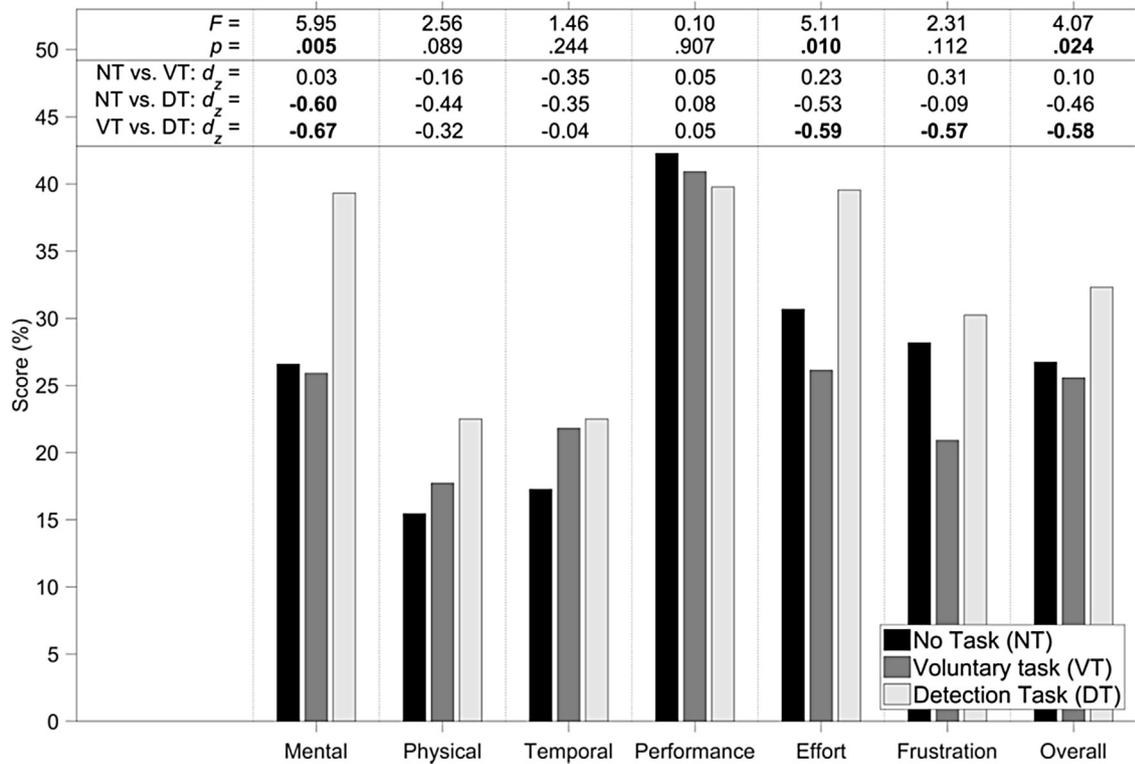


Fig. 4. Scores on the TLX for the three experimental conditions. The scores are expressed as a percentage and range from Very low (0%) to Very high (100%) for the mental demand, physical demand, temporal demand, effort, and frustration items, and from perfect (0%) to failure (100%) for the performance item. Also shown are the F statistic and p value for the repeated measures ANOVA and the Cohen's d_z effect size between the three pairs of conditions. For the ANOVA, the p value is shown in boldface if $p < 0.05$. For the pairwise comparisons, the d_z is shown in boldface if $p < 0.05/3$.

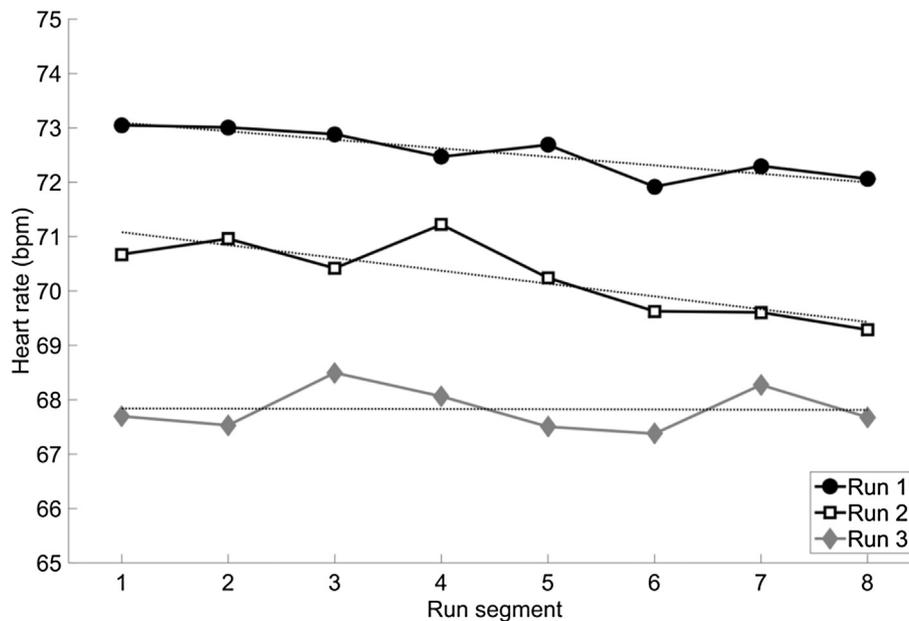


Fig. 5. Mean heart rate per 5-min segment during the run, for the three runs in chronological order. The dotted lines represent linear trend lines.

additional task, and in the Detection Task (DT) condition, participants had the task to detect red cars during the run.

Compared to other automated driving simulator experiments our research is extensive, with three 40-min sessions (2 h of driving in total), giving insights into the longer-term psychological effects of automated platooning. Previous driving simulator research on

automated driving involved a total driving time of an average of 1.05 h ($SD = 0.63$) (for an overview see De Winter et al., 2014). In comparison, typical vigilance research, the largest vigilance decrement occurs in the first 15 min (e.g., Mackworth, 1948), and the average trip duration in the U.S. and Europe is 20–30 min (McKenzie and Rapino, 2011; Pasaoglu et al., 2014).

Table 1Spearman rank-order correlations among personal characteristics and workload measures ($N = 22$, averaged across the three task conditions).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Age (years)	–															
2. Gender (1 = female, 2 = male)	0.00	–														
3. Driving experience (years)	0.87	–0.12	–													
4. Driving frequency (1–5)	0.31	–0.23	0.51	–												
5. Mileage (1–7)	0.48	–0.11	0.63	0.79	–											
6. Heart rate (bpm)	0.04	–0.18	0.00	0.35	0.08	–										
7. SDNN (ms)	–0.32	–0.02	–0.13	–0.14	–0.04	–0.77	–									
8. LF (ms ²)	–0.09	–0.04	0.05	0.42	0.22	0.48	–0.15	–								
9. HF (ms ²)	–0.32	–0.27	–0.10	–0.06	–0.06	–0.48	0.77	–0.01	–							
10. LF/HF ratio (no unit)	0.21	0.28	0.10	0.30	0.25	0.60	–0.73	0.50	–0.82	–						
11. TLX – Mental demand (%)	–0.15	–0.19	0.10	0.12	–0.01	–0.03	0.02	0.21	0.25	–0.01	–					
12. TLX – Physical demand (%)	0.24	–0.43	0.46	0.24	0.14	–0.14	0.23	0.18	0.35	–0.26	0.53	–				
13. TLX – Temporal demand (%)	–0.17	–0.31	0.05	0.26	0.03	0.04	0.10	0.07	0.27	–0.18	0.69	0.41	–			
14. TLX – Performance (%)	–0.25	0.01	–0.22	–0.33	–0.43	0.00	0.06	–0.02	0.01	–0.13	–0.09	0.01	–0.10	–		
15. TLX – Effort (%)	–0.02	–0.31	0.24	0.32	0.16	–0.02	0.12	0.38	0.42	–0.11	0.72	0.68	0.48	0.08	–	
16. TLX – Frustration (%)	–0.16	0.07	0.17	0.39	0.40	–0.20	0.47	0.27	0.48	–0.25	0.27	0.46	0.33	–0.05	0.39	–
17. Hit rate DT condition (%)	–0.01	–0.24	0.24	0.20	–0.09	0.37	–0.14	0.51	0.05	0.12	0.55	0.46	0.47	0.03	0.31	0.24

Note. Correlations of magnitude 0.43 or greater are statistically significant from 0 ($p < 0.05$).

4.1. Signal-detection performance

Participants adhered to the task instructions: Not one participant pressed the response button in the NT condition, whereas all 22 participants detected at least 61 out of 70 target cars in the DT condition, with a high mean hit rate of 94.7% and a low mean false alarm rate of 0.8%. These results suggest that the saying ‘man is a poor monitor’ does not apply to our automated platooning environment. Although our miss rate was low in comparison to some of the classical vigilance research (e.g., Mackworth, 1948, showing that the miss rate rose from 16% to 28% over a 2-h experiment), missing 5.3% of targets may still pose a high risk in real life if the targets (i.e., the situations that require manual intervention) occur frequently and/or if misses have severe consequences. To illustrate, if each miss of the DT condition resulted in a collision, then there would have been an unacceptably high number of 82 (5.3% of 22 participants \times 70 red cars) collisions in 15 h of driving (22 participants \times 40 min). The low miss rate, which is in disagreement with our hypothesis, may be caused by the fact that the environment of the open road is more complex and dynamic than classic vigilance task environments, for which operators may be more likely to remain engaged (for further discussion on task complexity and vigilance see Cummings et al., 2016; Montague et al., 1965; Molloy and Parasuraman, 1996). The detection task itself might have been considered engaging as well; Neubauer et al. (2014) showed that interacting with media devices is considered to be engaging. Alternatively, it could be that our detection task was easy due to the high saliency of the red cars (i.e., a high signal-to-noise ratio); see Körber et al. (2015) for a similar interpretation regarding the results of their auditory detection task used during partially automated driving.

4.2. Self-reported stress and workload

Our results indicate that automated platooning in general (i.e., each of the three task instructions) resulted in a loss of task engagement with respect to the pre-task score (i.e., standardized change scores below zero). These results resemble those of Saxby et al. (2007), who found that during a passive (fully automated) condition drivers reported a significant loss of task engagement.

Furthermore, participants found being occupied with a (either voluntary [VT] or mandatory [DT]) task significantly less worrisome than having no task to do (NT). This may be due to the fact that when having no task to do, participants have more opportunity to

focus attention on themselves and worry about personal matters. Additionally, the voluntary task yielded lower distress scores than the other two task conditions, which is in agreement with our hypothesis.

Performing the detection task was considered to be somewhat mentally demanding, with a mean TLX mental demand rating of 39% for the DT condition versus 26% for the VT condition, on a scale from *very low* (0%) to *very high* (100%) (Fig. 3). Thus, consistent with our hypothesis, overall, participants found the VT condition the least stressful, and the DT task the most workload-inducing.

By means of personal observations and post-experiment interviews it was found that in the VT condition participants performed actions ranging from eating a sandwich and interacting with one's mobile phone, to performing the detection task. Indeed, half of participants in the VT condition decided to detect red cars even though this was not required (Fig. 2). Some of them explained afterwards that they preferred doing the detection task in the VT condition to ensure they remained attentive to the road (and see Miller et al., 2015 for a study showing that engaging in a non-driving task can prevent drowsiness).

4.3. Heart rate and heart rate variability

Our hypotheses regarding workload and stress were not accepted regarding the mean heart rate and SDNN: there were no statistically significant difference between the three conditions. However, heart rate variability in terms of the LF/HF ratio differentiated the VT condition from the other two conditions, supporting the hypothesis that the VT condition was the least workload-inducing. Moreover, we found that the heart rate reduced and SDNN increased substantially during the experiment from Run 1 to Run 3. This suggests that, regardless of the driver's task in an automated platoon, the driver will become less stressed as well as underloaded over time. These results resemble those of Körber et al. (2015), who found significant time-on-task effects for several eye measures such as blink frequency. Collectively, our results indicate that not only the type of task, but also time-on-task has substantial effects on the state of the driver, suggesting that both these effects should be taken into consideration when designing automated driving systems.

4.4. Limitations

Certain limitations of this experiment are acknowledged. First, it

is impossible to uniquely attribute a specific psychological state (e.g., workload) to a specific overt physiological recording (Cacioppo and Tassinari, 1990). Because psychological states are often substantially correlated (e.g., Desmond and Matthews, 2009), we recognize the risk of ‘construct proliferation’ (for a review see Heikoop et al., 2016a). The interpretation of heart rate variability is particularly difficult, because the time-domain measure (SDNN) was sensitive to time-on-task (Run 1 vs. Run 3) but not to the type of task, whereas the frequency-domain measure (LF/HF ratio) was sensitive to the type of task (VT vs. DT) but not to time-on-task. It is known that physical activity and posture (e.g., sitting vs. supine or upright position) have substantial effects on heart rate variability (e.g., Bernardi et al., 1996; Castiglioni et al., 2009; Pomeranz et al., 1985) and so do individual characteristics such as age (Voss et al., 2015) and gender (Koenig and Thayer, 2016) (see also the correlation analysis in Table 1). It cannot be ruled out that these variables have interacted with the experimental conditions, although our within-subject design in which each participant serves as his/her own control ought to be robust against individual differences. The correlation analysis indicated that the heart rate variability measures were not meaningfully related to self-reported workload, but were substantially associated with basic heart rate, raising questions about the discriminant validity of these two measures (see Mehler et al., 2011, for a similar observation). The interpretation of heart rate variability as an index of workload or sympathetic versus parasympathetic activity is the topic of ongoing research and debate (e.g., Billman, 2013; Reyes del Paso et al., 2013; Thayer et al., 2012), and we concur with Vollmer (2015) that “the average heart rate and heart rate changes can act as confounding variables” (p. 610). Thus, it appears that heart rate variability is able to discriminate between task conditions, yet the causal pathways and neurophysiological mechanisms remain to be elucidated. This situation is similar to other physiological signals, such as pupil movements, which are known to “empirically reflect variations in central processing load” but for which the physiological causes are indirect and complex (Beatty and Lucero-Wagoner, 2000).

Second, although previous studies have found substantial correlations between driving behaviour in a STISIM simulator and driving behaviour on the road (Bédard et al., 2010; Lee et al., 2003; Mayhew et al., 2011; Wang et al., 2010), participants’ behaviours and physiological states during a simulator experiment are not necessarily representative of real-world platooning. Participants knew that a potential crash would not cause them physical harm and were therefore probably not as stressed as in a real life scenario. Moreover, the fixed-base simulator did not provide vestibular motion feedback, whereas in on-road automated driving, issues of visual-vestibular conflict may influence driver comfort and the uptake of secondary tasks (Diels and Bos, 2016). Thus, a replication of this experiment in a real world-driving scenario is advised.

Third, this research did not contain a control condition in which participants were to drive manually (e.g., Barnard and Lai, 2010; Saxby et al., 2013), nor did we vary the degree of automation reliability and availability (cf. Neubauer et al., 2012). For example, it is likely that drivers will become considerably frustrated if automation requires regular manual intervention (De Winter et al., 2016). Conversely, if automation is guaranteed to be safe and no manual intervention is ever to be expected (i.e., Fully Automated Driving or level 4 automation, per BAST and NHTSA definitions, respectively), like traveling in a train, participants are likely to be less stressed.

Fourth, although hazard perception is a critical component of (automated) car driving (Underwood et al., 2011), it is unknown whether detecting red cars is a realistic representation of such task. It remains to be investigated how our results generalize to the anticipation, detection, and response to diverse and realistic

hazards that require manual intervention, such as vehicles deviating from their paths, automation malfunction, sudden decelerations of lead cars, or stationary objects.

Fifth, the significant drop of heart rate and rise of heart rate variability might reflect acclimatization to the laboratory setting rather than workload and stress per se. To rule out this effect, multi-day sessions of the same participants to the driving simulator are necessary (cf. Beggiato and Krebs, 2013; Beggiato et al., 2015; Kazi et al., 2007; Pereira et al., 2015).

Sixth, our sample was relatively small ($N = 22$), and therefore the statistical power to detect small effects is low. Specifically, in order to detect effects small effects ($d_z = 0.2$) with 80% power and a significance level of 0.05/3, 265 participants would be needed (Faul et al., 2007). Only by means of fundamentally different types of research (e.g., large-sample cross-institutional research), it is possible to detect subtle effects of stress and workload that may exist in the population.

Lastly, the eye tracker used in this experiment was prone to missing data, a problem that is common in naturalistic driving tasks (Ahlstrom et al., 2012). However, in our case, the amount of missing data itself discriminated more strongly between the task conditions than what was actually measured with the eye-tracker. For future research, we recommend the use of a head-mounted eye-tracker if the goal is to measure eye-closure.

5. Conclusion

In conclusion, this research shows that the idea that automated driving puts humans in an extremely stressful monitoring role for which they are “magnificently disqualified” (Hancock, 2015, p. 138) does not generalize to a simulator-based platooning task. Although the VT condition was the least workload-inducing and least stressful, participants in the DT condition remained attentive to the road for 40 min, with an average hit rate of 94.7% on the detection task. Furthermore, our results indicate clear time-on-task effects regarding heart rate. Similar to Szalma et al. (2004) we recommend that in order to increase performance and reduce stress, both type of task and time-on-task should be considered in system design.

Acknowledgment

The authors are involved in the Marie Curie Initial Training Network (ITN): HFAuto (PITN-GA-2013-605817).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.apergo.2016.10.016>.

References

- Ahlstrom, C., Victor, T., Wege, C., Steinmetz, E., 2012. Processing of eye/head-tracking data in large-scale naturalistic driving data sets. *IEEE Trans. Intell. Transp. Syst.* 13, 553–564. <http://dx.doi.org/10.1109/TITS.2011.2174786>.
- Banks, V.A., Stanton, N.A., Harvey, C., 2014. Sub-systems on the road to vehicle automation: hands and feet free but not ‘mind’ free driving. *Saf. Sci.* 62, 505–514. <http://dx.doi.org/10.1016/j.ssci.2013.10.014>.
- Barnard, Y., Lai, F., 2010. Spotting sheep in Yorkshire: using eye-tracking for studying situation awareness in a driving simulator. In: de Waard, D., Axelsson, A., Berglund, M., Peters, B., Weikert, C. (Eds.), *Human Factors: A System View of Human, Technology, and Organisation*. Shaker Publishing, Maastricht, The Netherlands, pp. 249–261.
- Barutcu, I., Esen, A.M., Kaya, D., Turkmen, M., Karakaya, O., Melek, M., Esen, O.B., Basaran, Y., 2005. Cigarette smoking and heart rate variability: dynamic influence of parasympathetic and sympathetic maneuvers. *Ann. Noninvasive Electrocardiol.* 10 (3), 324–329. <http://dx.doi.org/10.1111/j.1542-474X.2005.00636.x>.
- Beatty, J., Lucero-Wagoner, B., 2000. The pupillary system. In: Cacioppo, J., Tassinari, L., Berntson, G. (Eds.), *Handbook of Psychophysiology*. Cambridge University Press, Cambridge, MA, pp. 142–162.

- Bédard, M., Parkkari, M., Weaver, B., Riendeau, J., Dahlquist, M., 2010. Assessment of driving performance using a simulator protocol: validity and reproducibility. *Am. J. Occup. Ther.* 64, 336–340. <http://dx.doi.org/10.5014/ajot.64.2.336>.
- Beggiato, M., Krems, J.F., 2013. The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transp. Res. F Traffic Psychol. Behav.* 18, 47–57. <http://dx.doi.org/10.1016/j.trf.2012.12.006>.
- Beggiato, M., Pereira, M., Petzoldt, T., Krems, J., 2015. Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transp. Res. F Traffic Psychol. Behav.* 35, 75–84. <http://dx.doi.org/10.1016/j.trf.2015.10.005>.
- Bergenheim, C., Shladover, S., Coelingh, E., 2012. Overview of platooning systems. In: *Proceedings of the 19th ITS World Congress*, October 22–26, Vienna, Austria.
- Bernardi, L., Valle, F., Coco, M., Calciati, A., Sleight, P., 1996. Physical activity influences heart rate variability and very-low-frequency components in Holter electrocardiograms. *Cardiovasc. Res.* 32, 234–237. [http://dx.doi.org/10.1016/0008-6363\(96\)00081-8](http://dx.doi.org/10.1016/0008-6363(96)00081-8).
- Billman, G.E., 2013. The LF/HF ratio does not accurately measure cardiac sympathovagal balance. *Front. Physiol.* 4, 54–58. <http://dx.doi.org/10.3389/fphys.2013.00026>.
- Brookhuis, K.A., De Waard, D., 2010. Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. <http://dx.doi.org/10.1016/j.aap.2009.06.001>.
- Cabrall, C.D., Happee, R., De Winter, J.C.F., 2016. From Mackworth's clock to the open road: a literature review on driver vigilance task operationalization. *Transp. Res. F Traffic Psychol. Behav.* 40, 169–189. <http://dx.doi.org/10.1016/j.trf.2016.04.001>.
- Cacioppo, J.T., Tassinary, L.G., 1990. Inferring psychological significance from physiological signals. *Am. Psychol.* 45, 16–28. <http://dx.doi.org/10.1037/0003-066X.45.1.16>.
- Casner, S.M., Hutchins, E.L., Norman, D., 2016. The challenges of partially automated driving. *Commun. ACM* 59, 70–77. <http://dx.doi.org/10.1145/2830565>.
- Castiglioni, P., Parati, G., Civijian, A., Quintin, L., Di Rienzo, M., 2009. Local scale exponents of blood pressure and heart rate variability by detrended fluctuation analysis: effects of posture, exercise, and aging. *IEEE Trans. Biomed. Eng.* 56, 675–684. <http://dx.doi.org/10.1109/TBME.2008.2005949>.
- Cha, D., 2003. Driver Workload Comparison Among Road Sections of Automated Highway Systems (No. 2003-01-0119). SAE Technical Paper. <http://dx.doi.org/10.4271/2003-01-0119>.
- Cinaz, B., Arnrich, B., La Marca, R., Tröster, G., 2013. Monitoring of mental workload levels during an everyday life effort-work scenario. *Pers. Ubiquitous Comput.* 17, 229–239. <http://dx.doi.org/10.1007/s00779-011-0466-1>.
- Corrales, M., Torres, B., Esquivel, A., Salazar, M., Orellana, J., 2012. Normal values of heart rate variability at rest in a young, healthy and active Mexican population. *Health* 4, 21183. <http://dx.doi.org/10.4236/health.2012.47060>.
- Cummings, M.L., Gao, F., Thornburg, K.M., 2016. Boredom in the workplace: a new look at an old problem. *Hum. Factor J. Hum. Factor Ergon. Soc.* 58, 279–300. <http://dx.doi.org/10.1177/0018720815609503>.
- Davis, J., Animashaun, A., Schoenherr, E., McDowell, K., 2008. Evaluation of semi-autonomous convoy driving. *J. Field Robot.* 25, 880–897. <http://dx.doi.org/10.1002/rob.20263>.
- De Waard, D., Van der Hulst, M., Hoedemaeker, M., Brookhuis, K.A., 1999. Driver behaviour in an emergency situation in the automated highway system. *Transp. Hum. Factor* 1, 67–82. http://dx.doi.org/10.1207/s10101_7.
- De Winter, J.C.F., 2014. Controversy in human factors constructs and the explosive use of the NASA TLX: a measurement perspective. *Cogn. Technol. Work* 16, 289–297. <http://dx.doi.org/10.1007/s10111-014-0275-1>.
- De Winter, J.C.F., Happee, R., Martens, M., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transp. Res. F Traffic Psychol. Behav.* 27, 196–217. <http://dx.doi.org/10.1016/j.trf.2014.06.016>.
- De Winter, J.C.F., Stanton, N.A., Price, J.S., Mistry, H., 2016. The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *Int. J. Veh. Des.* 70, 297–324. <http://dx.doi.org/10.1504/IJVD.2016.076736>.
- Desmond, P.A., Matthews, G., 2009. Individual differences in stress and fatigue in two field studies of driving. *Transp. Res. F Traffic Psychol. Behav.* 12, 265–276. <http://dx.doi.org/10.1016/j.trf.2008.12.006>.
- Diels, C., Bos, J.E., 2016. Self-driving carsickness. *Appl. Ergon.* 53, 374–382. <http://dx.doi.org/10.1016/j.apergo.2015.09.009>.
- Fairclough, S.H., Venables, L., 2005. Prediction of subjective states from psychophysiology: a multivariate approach. *Biol. Psychol.* 71, 100–110. <http://dx.doi.org/10.1016/j.biopsycho.2005.03.007>.
- Fallah, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A.R., Miyake, S., 2016. Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. *Appl. Ergon.* 52, 95–103. <http://dx.doi.org/10.1016/j.apergo.2015.07.009>.
- Farber, E.L., 1999. Comments on "Driving behaviour in an emergency situation in the automated highway system". *Transp. Hum. Factor* 1, 83–85. http://dx.doi.org/10.1207/s10101_8.
- Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Method* 39, 175–191. <http://dx.doi.org/10.3758/BF03193146>.
- Fenton, R.E., Cosgriff, R.L., Olson, K.W., Blackwell, L.M., 1968. One approach to highway automation. *Proc. IEEE Spec. Syst. Compon.* 56, 556–566. <http://dx.doi.org/10.1109/PROC.1968.6344>.
- Funke, G., Matthews, G., Warm, J.S., Emo, A.K., 2007. Vehicle automation: a remedy for driver stress? *Ergonomics* 50, 1302–1323. <http://dx.doi.org/10.1080/00140130701318830>.
- Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., Reed, N., 2014. Driving next to automated vehicle platoons: how do short time headways influence non-platoon drivers' longitudinal control? *Transp. Res. F Traffic Psychol. Behav.* 27, 264–273. <http://dx.doi.org/10.1016/j.trf.2014.03.003>.
- Hancock, P.A., 2015. Automobility: the coming use of fully-automated on-road vehicles. In: *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, March 9–12, Orlando, FL, pp. 137–139. <http://dx.doi.org/10.1109/COGSIMA.2015.7108188>.
- Hancock, P.A., Parasuraman, R., 1992. Human factors and safety in the design of Intelligent Vehicle-Highway Systems (IVHS). *J. Saf. Res.* 23, 181–198. [http://dx.doi.org/10.1016/0022-4375\(92\)90001-P](http://dx.doi.org/10.1016/0022-4375(92)90001-P).
- Harris, D.H., 2002. How to really improve airport security. *Ergon. Des. Q. Hum. Factor Appl.* 10, 17–22. <http://dx.doi.org/10.1177/106480460201000104>.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. [http://dx.doi.org/10.1016/S0166-4115\(08\)62386-9](http://dx.doi.org/10.1016/S0166-4115(08)62386-9).
- Healey, J.A., Picard, R.W., 2004. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* 6, 156–166. <http://dx.doi.org/10.1109/ITITS.2005.848368>.
- Heikoop, D.D., De Winter, J.C.F., Van Are, B., Stanton, N.A., 2016a. Psychological constructs in driving automation: a consensus model and critical comment on construct proliferation. *Theor. Issue Ergon. Sci.* 17, 284–303. <http://dx.doi.org/10.1080/1463922X.2015.1101507>.
- Heikoop, D.D., De Winter, J.C.F., Van Are, B., Stanton, N.A., 2016b. Acclimatizing to automation: driver workload and stress during automated car following in real traffic. (Manuscript submitted for publication).
- Helton, W.S., Warm, J.S., Matthews, G., Corcoran, K.J., Dember, W.N., 2002. Further tests of an abbreviated vigilance task: effect of signal salience and jet aircraft noise on performance and stress. *Proc. Hum. Factor Ergon. Soc. Annu. Meet.* 46, 1546–1550. <http://dx.doi.org/10.1177/154193120204601704>.
- Hochstädter, A., Cremer, M., 1997. Investigating the potential of convoy driving for congestion dispersion. In: *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, November 9–12, Boston, MA, pp. 735–740. <http://dx.doi.org/10.1109/ITSC.1997.660565>.
- Jorna, P.G.A.M., 1992. Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. *Biol. Psychol.* 34, 237–257. [http://dx.doi.org/10.1016/0301-0511\(92\)90017-O](http://dx.doi.org/10.1016/0301-0511(92)90017-O).
- Karaaslan, U., Varaiya, P., Walrand, J., 1991. Two proposals to improve freeway traffic flow. In: *American Control Conference*, June 26–28, Boston, MA, pp. 2539–2544.
- Karakaya, O., Barutcu, I., Kaya, D., Esen, A.M., Saglam, M., Melek, M., Kaymaz, C., 2007. Acute effect of cigarette smoking on heart rate variability. *Angiology* 58, 620–624.
- Kazi, T.A., Stanton, N.A., Walker, G.H., Young, M.S., 2007. Designer driving: drivers' conceptual models and level of trust in adaptive cruise control. *Int. J. Veh. Des.* 45, 339–360. <http://dx.doi.org/10.1504/IJVD.2007.014909>.
- Kibler, A.W., 1965. The relevance of vigilance research to aerospace monitoring tasks. *Hum. Factor J. Hum. Factor Ergon. Soc.* 7, 93–99. <http://dx.doi.org/10.1177/001872086500700202>.
- Kim, A., Perlman, D., Bogard, D., Harrington, R., 2016. Review of Federal Motor Vehicle Safety Standards (FMVSS) for Automated Vehicles: Identifying Potential Barriers and Challenges for the Certification of Automated Vehicles Using Existing FMVSS. Preliminary Report No. DOT-VNTSC-OSTR-16–03. National Highway Traffic Safety Administration, Washington, DC.
- Knoop, V., Hoogendoorn, S., Van Zuylen, H., 2008. Capacity reduction at incidents: empirical data collected from a helicopter. *Transp. Res. Rec. J. Transp. Res. Board* 2071, 19–25. <http://dx.doi.org/10.3141/2071-03>.
- Koenig, J., Thayer, J.F., 2016. Sex differences in healthy human heart rate variability: a meta-analysis. *Neurosci. Biobehav. Rev.* 64, 288–310. <http://dx.doi.org/10.1016/j.neubiorev.2016.03.007>.
- Körber, M., Cingel, A., Zimmermann, M., Bengler, K., 2015. Vigilance decrement and passive fatigue caused by monotony in automated driving. In: *Proceedings of the 6th International Conference on Applied Human Factors and Ergonomics*, July 26–30, Las Vegas, NV, pp. 365–371. <http://dx.doi.org/10.1016/j.promfg.2015.07.499>.
- Kunze, R., Haberstroh, M., Ramakers, R., Henning, K., Jeschke, S., 2011. Automated truck platoons on motorways—a contribution to the safety on roads. In: *Jeschke, S., Isenhardt, I., Henning, K. (Eds.), Automation, Communication and Cybernetics in Science and Engineering 2009/2010*. Springer, Berlin, Heidelberg, pp. 415–426. http://dx.doi.org/10.1007/978-3-642-16208-4_38.
- Lal, S.K.L., Craig, A., 2002. Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology* 39, 313–321. <http://dx.doi.org/10.1017/S004857720139095>.
- Larson, J., Liang, K.Y., Johansson, K.H., 2015. A distributed framework for coordinated heavy-duty vehicle platooning. *IEEE Trans. Intell. Transp. Syst.* 16, 419–429. <http://dx.doi.org/10.1109/ITITS.2014.2320133>.
- Lee, H.C., Cameron, D., Lee, A.H., 2003. Assessing the driving performance of older adult drivers: on-road versus simulated driving. *Accid. Anal. Prev.* 35, 797–803. [http://dx.doi.org/10.1016/S0001-4575\(02\)00083-0](http://dx.doi.org/10.1016/S0001-4575(02)00083-0).
- Leviton, L., Golembiewski, G., Bloomfield, J.R., 1998. Human factors issues in automated highway systems. *ITS J. Technol. Plann. Oper.* 4, 21–47. <http://dx.doi.org/10.1080/10248079808903735>.
- Llaneras, R.E., Salinger, J., Green, C.A., 2013. Human Factors issues associated with

- limited ability autonomous driving systems: drivers' allocation of visual attention to the forward road. In: Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, June 17–20, Bolton Landing, NY, pp. 92–98.
- Luque-Casado, A., Zabala, M., Morales, E., Mateo-March, M., Sanabria, D., 2013. Cognitive performance and heart rate variability: the influence of fitness level. *PLoS One* 8, e56935. <http://dx.doi.org/10.1371/journal.pone.0056935>.
- Luque-Casado, A., Perales, J.C., Cárdenas, D., Sanabria, D., 2016. Heart rate variability and cognitive processing: the autonomic response to task demands. *Biol. Psychol.* 113, 83–90. <http://dx.doi.org/10.1016/j.biopsycho.2015.11.013>.
- Mackie, R.R. (Ed.), 1977. *Vigilance: Theory, Operational Performance, and Physiological Correlates*. Plenum, New York, NY.
- Mackworth, N.H., 1948. The breakdown of vigilance during prolonged visual search. *Q. J. Exp. Psychol.* 1, 6–21. <http://dx.doi.org/10.1080/17470214808416738>.
- Manzano, B.M., Vanderlei, L.C.M., Ramos, E.M., Ramos, D., 2011. Acute effects of smoking on autonomic modulation: analysis by Poincaré plot. *Arq. Bras. Cardiol.* 96, 154–160. <http://dx.doi.org/10.1590/S0066-782X2011005000013>.
- Matthews, G.S., 2014. *Task-induced Fatigue, Operator Engagement and the Control of Attention* [PowerPoint slides].
- Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., Huggins, J., 1999. Validation of a comprehensive stress state questionnaire: towards a state 'Big Three'?. In: Mervielde, I., Deary, I.J., de Fruyt, F., Ostendorf, F. (Eds.), *Personality Psychology in Europe*, vol. 7. Tilburg University Press, Tilburg, pp. 335–350.
- Matthews, G., Campbell, S.E., Falconer, S., July 2000. Personality, energetics and individual differences in motivation. In: Proceedings of the Tenth European Conference on Personality, Krakow, Poland.
- Matthews, G., Campbell, S.E., Falconer, S., Joyner, L.A., Huggins, J., Gilliland, K., Grier, R., Warm, J.S., 2002. Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. *Emotion* 2, 315–340. <http://dx.doi.org/10.1037/1528-3542.2.4.315>.
- Mayhew, D.R., Simpson, H.M., Wood, K.M., Lonero, L., Clinton, K.M., Johnson, A.G., 2011. On-road and simulated driving: concurrent and discriminant validation. *J. Saf. Res.* 42, 267–275. <http://dx.doi.org/10.1016/j.jsr.2011.06.004>.
- McKenzie, B., Rapino, M., 2011. *Commuting in the United States: 2009*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.
- Mehler, B., Reimer, B., Wang, Y., 2011. A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In: Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 590–597.
- Mets, M.A., Kuipers, E., Senerpont Domis, L.M., Leenders, M., Olivier, B., Verster, J.C., 2011. Effects of alcohol on highway driving in the STISIM driving simulator. *Hum. Psychopharmacol. Clin. Exp.* 26, 434–439.
- Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., Ju, W., 2015. Distraction becomes engagement in automated driving. *Proc. Hum. Factor Ergon. Soc. Annu. Meet.* 59, 1676–1680. <http://dx.doi.org/10.1177/1541931215591362>.
- Molloy, R., Parasuraman, R., 1996. Monitoring an automated system for a single failure: vigilance and task complexity effects. *Hum. Factor J. Hum. Factor Ergon. Soc.* 38, 311–322. <http://dx.doi.org/10.1177/001872089606380211>.
- Montague, W.E., Webber, C.E., Adams, J.A., 1965. The effects of signal and response complexity on eighteen hours of visual monitoring. *Hum. Factor J. Hum. Factor Ergon. Soc.* 7, 163–172. <http://dx.doi.org/10.1177/001872086500700208>.
- Mueller, S.T., Piper, B.J., 2014. The psychology experiment building language (PEBL) and PEBL test battery. *J. Neurosci. Method* 222, 250–259. <http://dx.doi.org/10.1016/j.jneumeth.2013.10.024>.
- Neubauer, C., Matthews, G.S., Langheim, L., Saxby, D.J., 2012. Fatigue and voluntary utilization of automation in simulated driving. *Hum. Factor J. Hum. Factor Ergon. Soc.* 54, 734–746. <http://dx.doi.org/10.1177/0018720811423261>.
- Neubauer, C., Matthews, G.S., Saxby, D.J., 2014. Fatigue in the automated vehicle: do games and conversation distract or energize the driver? *Proc. Hum. Factor Ergon. Soc. Annu. Meet.* 58, 2053–2057. <http://dx.doi.org/10.1177/1541931214581432>.
- Norman, D.A., 2015. The human side of automation. In: Meyer, G., Beiker, S. (Eds.), *Road Vehicle Automation*, vol. 2. Springer International Publishing, Switzerland, pp. 73–79.
- Omae, M., Hashimoto, N., Sugamoto, T., Shimizu, H., 2005. Measurement of driver's reaction time to failure of steering controller during automatic driving. *Rev. Automot. Eng.* 26, 213–215.
- Pasaoglu, G., Fiorello, D., Martino, A., Zani, L., Zubaryeva, A., Thiel, C., 2014. Travel patterns and the potential use of electric cars—Results from a direct survey in six European countries. *Technol. Forecast. Soc. Change* 87, 51–59. <http://dx.doi.org/10.1016/j.techfore.2013.10.018>.
- Pereira, M., Beggiato, M., Petzoldt, T., 2015. Use of adaptive cruise control functions on motorways and urban roads: changes over time in an on-road study. *Appl. Ergon.* 50, 105–112. <http://dx.doi.org/10.1016/j.apergo.2015.03.002>.
- Ploeg, J., Van de Wouw, N., Nijmeijer, H., 2014. *Lp* string stability of cascaded systems: application to vehicle platooning. *IEEE Trans. Contr. Syst. Technol.* 22, 786–793. <http://dx.doi.org/10.1109/TCST.2013.2258346>.
- Pomeranz, B., Macaulay, R.J.B., Caudill, M.A., Kutz, I., Adam, D., Gordon, D., Benson, H., 1985. Assessment of autonomic function in humans by heart rate spectral analysis. *Am. J. Physiol. Heart Circ. Physiol.* 248, H151–H153.
- Pritchett, A.R., Lewis, D.S., 2010. Cognitive engineering and aviation safety. *Encycl. Aeronaut. Eng.* <http://dx.doi.org/10.1002/9780470686652.eae508>.
- Ren, W., Green, D., 1994. Continuous platooning: a new evolutionary and operating concept for automated highway systems. In: Proceedings of the American Control Conference, June 29–July 1, Baltimore, MD. <http://dx.doi.org/10.1109/ACC.1994.751685>.
- Reyes del Paso, G.A., Langewitz, W., Mulder, L.J., Roon, A., Duschek, S., 2013. The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies. *Psychophysiology* 50, 477–487. <http://dx.doi.org/10.1111/psyp.12027>.
- Saffarian, M., De Winter, J.C.F., Happee, R., 2012. Automated driving: Human-factors issues and design solutions. *Proc. Hum. Factor Ergon. Soc. Annu. Meet.* 56, 2296–2301. <http://dx.doi.org/10.1177/1071181312561483>.
- Saxby, D.J., Matthews, G.S., Hitchcock, E.M., Warm, J.S., 2007. Development of active and passive fatigue manipulations using a driving simulator. *Proc. Hum. Factor Ergon. Soc. Annu. Meet.* 51, 1237–1241. <http://dx.doi.org/10.1177/154193120705101839>.
- Saxby, D.J., Matthews, G.S., Warm, J.S., Hitchcock, E.M., Neubauer, C., 2013. Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *J. Exp. Psychol. Appl.* 19, 287–300. <http://dx.doi.org/10.1037/a0034386>.
- Scerbo, M.W., 2001. Stress, workload, and boredom in vigilance: a problem and an answer. In: Hanco, P.A., Desmond, P.A. (Eds.), *Stress, Workload, and Fatigue*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 267–278.
- Shaffer, F., Combatalade, D.C., 2013. Don't add or miss a beat: a guide to cleaner Heart Rate Variability recordings. *Biofeedback* 41, 121–130. <http://dx.doi.org/10.5298/1081-5937-41.3.04>.
- Sheridan, T.B., 1996. Allocating functions among humans and machines. In: Beevis, D., Essens, P., Schuffel, H. (Eds.), *Improving Function Allocation for Integrated Systems Design*. Wright Patterson AFB: Crew Systems Ergonomics Information Analysis Center, pp. 7–27. Report No. CSERIAC SOAR 96-01.
- Sheridan, T.B., 2002. *Humans and Automation: Systems Design and Research Issues*. John Wiley & Sons, Inc, New York, NY.
- Sheridan, T.B., 2011. Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: distinctions and modes of adaptation. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 41, 662–667. <http://dx.doi.org/10.1109/TSMCA.2010.2093888>.
- Skottke, E.M., Debus, G., Wang, L., Huestegge, L., 2014. Carryover effects of highly automated convoy driving on subsequent manual driving performance. *Hum. Factor J. Hum. Factor Ergon. Soc.* 56, 1272–1283. <http://dx.doi.org/10.1177/0018720814524594>.
- Stanislaw, H., Todorov, N., 1999. Calculation of signal detection theory measures. *Behav. Res. Method Instrum. Comput.* 31, 137–149. <http://dx.doi.org/10.3758/BF03207704>.
- Stearman, E.J., Durso, F.T., 2016. Vigilance in a dynamic environment. *J. Exp. Psychol. Appl.* 22, 107–123. <http://dx.doi.org/10.1037/xap0000075>.
- Society of Automotive Engineers [SAE], 2014. *Taxonomy and Definitions for Terms Related to On-road Motor Vehicle Automated Driving Systems* (Standard No. J3016). SAE International. Retrieved from: http://standards.sae.org/j3016_201401/.
- Suriya-Prakash, M., John-Preetham, G., Sharma, R., 2015. Is heart rate variability related to cognitive performance in visuospatial working memory? *PeerJ Prepr.* 3, e1711.
- Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A., 2004. Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. *Hum. Factor J. Hum. Factor Ergon. Soc.* 46, 219–233. <http://dx.doi.org/10.1518/hfes.46.2.219.37334>.
- Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation* 93, 1043–1065. <http://dx.doi.org/10.1161/01.CIR.93.5.1043>.
- Teller, A., 2015. Moonshots and reality. In: Keynote Presentation Given at South by Southwest (SXSW), March 17, Austin TX.
- Thayer, J.F., Åhs, F., Fredrikson, M., Sollers, J.J., Wager, T.D., 2012. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* 36, 747–756. <http://dx.doi.org/10.1016/j.neubiorev.2011.11.009>.
- Thorpe, C., Jochem, T., Pomerleau, D., 1998. The 1997 automated highway free agent demonstration. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems, pp. 496–501. <http://dx.doi.org/10.1109/ITSC.1997.660524>.
- Tsugawa, S., Kato, S., Aoki, K., 2011. An automated truck platoon for energy saving. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4109–4114. <http://dx.doi.org/10.1109/IROS.2011.6094549>.
- Underwood, G., Crundall, D., Chapman, P., 2011. Driving simulator validation with hazard perception. *Transp. Res. F Traffic Psychol. Behav.* 14, 435–446. <http://dx.doi.org/10.1016/j.trf.2011.04.008>.
- United Nations, November 8, 1968. *Convention on Road Traffic*. Vienna, Austria.
- Vollmer, M., 2015. A robust, simple and reliable measure of heart rate variability using relative RR intervals. In: IEEE Conference on Computing in Cardiology Conference (CinC), pp. 609–612. <http://dx.doi.org/10.1109/CIC.2015.7410984>.
- Voss, A., Schroeder, R., Heitmann, A., Peters, A., Perz, S., 2015. Short-term heart rate variability—Influence of gender and age in healthy subjects. *PLoS ONE* 10, e0118308. <http://dx.doi.org/10.1371/journal.pone.0118308>.
- Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L.A., Coughlin, J.F., 2010. The validity of driving simulation for assessing differences between in-vehicle informational interfaces: a comparison with field testing. *Ergonomics* 53, 404–420. <http://dx.doi.org/10.1080/00140130903464358>.
- Warm, J.S., Parasuraman, R., Matthews, G., 2008. Vigilance requires hard mental

- work and is stressful. *Hum. Factor* 50, 433–441. <http://dx.doi.org/10.1518/001872008X312152>.
- Wiener, E.L., 1985. Beyond the sterile cockpit. *Hum. Factor* 27, 75–90. <http://dx.doi.org/10.1177/001872088502700107>.
- Wiener, E.L., Curry, R.E., 1980. Flight-deck automation: promises and problems. *Ergonomics* 23, 995–1011. <http://dx.doi.org/10.1080/00140138008924809>.
- Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J., Kirn, C.L., 1994. *Research on Vehicle-based Driver Status/Performance Monitoring: Development, Validation, and Refinement of Algorithms for Detection of Driver Drowsiness. Final Report (DOT HS 808 247)*. National Highway Traffic Safety Administration, Washington, DC.
- Young, M.S., Stanton, N.A., 2007. What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics* 50, 1324–1339. <http://dx.doi.org/10.1080/00140130701318855>.