Delft University of Technology

Acclimatizing to automation

Driver workload and stress during partially automated car following in real traffic

Heikoop, Daniël D.; de Winter, Joost C.F.; van Arem, Bart; Stanton, Neville A.

DOI
10.1016/j.trf.2019.07.024

Publication date
2019

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Citation (APA)
Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2019). Acclimatizing to automation: Driver workload and stress during partially automated car following in real traffic. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 503-517. https://doi.org/10.1016/j.trf.2019.07.024

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.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

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.

Transportation Research Part F 65 (2019) 503–517

Contents lists available at ScienceDirect

Transportation Research Part F

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

Acclimatizing to automation: Driver workload and stress during partially automated car following in real traffic

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, United Kingdom
[b] Department of Cognitive Robotics, Delft University of Technology, the Netherlands
[c] Department of Transport & Planning, Delft University of Technology, the Netherlands

A R T I C L E I N F O

Article history:
Received 23 July 2018
Received in revised form 18 July 2019
Accepted 30 July 2019

Keywords:
Automated driving
Workload
Stress
Object detection
On-road

A B S T R A C T

Automated driving systems are increasingly prevalent on public roads, but there is currently little knowledge on the level of workload and stress of drivers operating an automated vehicle in a real environment. The present study aimed to measure driver workload and stress during partially automated driving in real traffic. We recorded heart rate, heart rate variability, respiratory rate, and subjective responses of nine test drivers in the Tesla Model S with Autopilot. The participants, who were experienced with driver assistance systems but naïve to the Tesla, drove a 32 min motorway route back and forth while following a lead car in regular traffic. In one of the two drives, participants performed a heads-up detection task of bridges they went underneath. Averaged across the two drives, the participants' mean self-reported overall workload score on the NASA Task Load Index was 19%. Moreover, the participants showed a reduction in heart rate and self-reported workload over time, suggesting that the participants became accustomed to the experiment and technology. The mean hit (i.e., pressing the button near a bridge) rate in the detection task was 88%. In conclusion, driving with the Tesla Autopilot on a motorway involved a low level of workload that decreased with time on task.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Workload and stress in automated driving

Cars that provide combined longitudinal and lateral automated control support have recently been introduced on the market. Automated driving may be expected to reduce workload and stress as compared to manual driving because the driver does not have to control the vehicle. However, unless the driving task is fully automated, automated driving may cause high workload and stress, because the driver needs to supervise both the human-machine interface and the state of the car in relation to the outside environment (for an illustration, see Fig. 1). More specifically, the driver of an automated car has to remain attentive to reclaim manual control if required (Casner, Hutchins, & Norman, 2016; Stanton, Young, & McCaulder, 1997) a task that may be demanding and stressful (Hancock, 2015). Furthermore, the type of supervisory control shown in Fig. 1 may cause out-of-the-loop problems, such as loss of situation awareness and mode errors, which resemble those

* Corresponding author.
E-mail address: d.d.heikoop@tudelft.nl (D.D. Heikoop).

https://doi.org/10.1016/j.trf.2019.07.024
1369-8478/© 2019 Elsevier Ltd. All rights reserved.

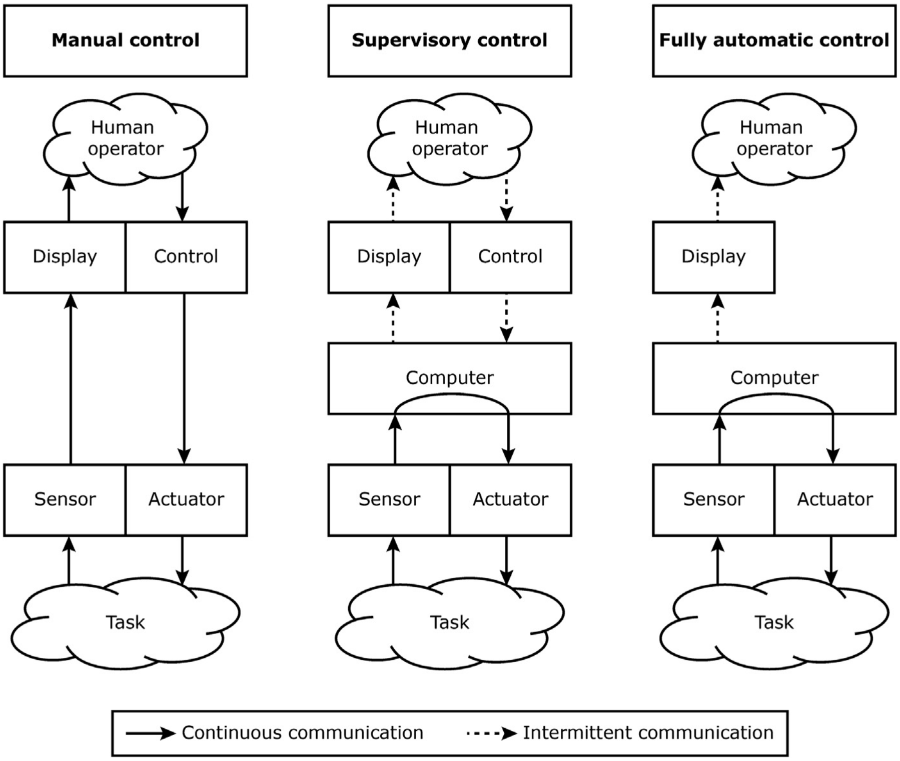


Fig. 1. Manual control (left), supervisory control (middle), and fully automatic control (right). In manual control, the driver controls the car via manipulators (steering wheel & pedals) and continuously receives information from the car in the environment (i.e., task). In fully automatic driving, the human has no contribution to the driving task other than to set a destination (or to press an emergency stop button). Hence, the driver is out of the control loop completely. In supervisory control, the driver interacts with a computer that closes the control loop via sensors and actuators, while the driver intermittently (1) provides instructions to the computer, (2) receives information via displays, and (3) receives information from the car in the environment. Adapted from Sheridan (1992).

observed in aviation and process control (Haslbeck & Hoermann, 2016; Kaber & Endsley, 1997; Metzger & Parasuraman, 2001; see also Stanton & Marsden, 1996). A survey by Dikmen and Burns (2016) among 121 Tesla owners found that automation failures (e.g., failure to detect lanes) were frequent but not perceived as risky. Furthermore, the majority of respondents indicated that it is important to remain alert and to be aware of the automation's limitations.

1.2. Prior research on workload and stress in automated driving

The majority of Human Factors research on driver workload in automated vehicles has been conducted in driving simulators (see De Winter, Happee, Martens, & Stanton, 2014, for a review). Overall, the results indicate that the self-reported workload as assessed with the NASA Task Load Index is substantially lower in automated driving than in manual driving, and below 50% on a scale from 0 to 100% (see Table 1 for an overview).

A small number of on-road studies are available. Recently, Endsley (2017) conducted a single-subject naturalistic driving study using a Tesla Model S over a six-month period. She reported that her situation awareness increased when using automation because less focus was needed on controlling the vehicle, and more attention could be devoted to looking at traffic and road signage. However, Endsley also experienced various issues of mode confusion and unexpected automation transitions, as well as a loss of attention. Endsley further found that ratings of satisfaction, usefulness, and trust gradually increased from months 1–2 towards months 5–6, which is in line with the results of a longitudinal naturalistic driving study on adaptive cruise control with 15 participants (Beggiato, Pereira, Petzoldt, & Krems, 2015). Additionally, overall self-reported workload was towards the lower end of the scale (Endsley, 2017).

Eriksson, Banks, and Stanton (2017) let 12 test drivers use the Tesla autopilot for about 20 minutes per participant. Participants each experienced approximately 12 automation-to-manual control transitions and completed the NASA Task Load Index after the ride. The mean overall workload was 19%. Stapel, Mullakkal-Babu, and Happee (2019) conducted an on-road highway driving study in which 15 participants used the Tesla Autopilot for about 20 min. The authors found overall low levels of workload among participants (between 10% and 43%), with the type of road (busy city ring versus relatively empty highway) and prior experience with the Tesla Model S being moderator variables (Table 1). In another on-road study, Banks and Stanton (2016) tested a prototype version of automated longitudinal and lateral control in addition to a driver-initiated

Table 1
Overview of workload measurements on the NASA Task Load Index in automated driving studies (determined in 2017).

Reference	Simulator/road	Sample size	Mean workload
Banks and Stanton (2016)	On-road	32	42% (median)
Borojeni, Chuang, Heuten, and Boll (2016)	Simulator	21	30%
Damböck, Weißgerber, Kienle, and Bengler (2013)	Simulator	24	33%
De Winter, Stanton, Price, and Mistry (2016)	Simulator	24	31% (exp. 1)
	Simulator	27	31% (exp. 2)
Eriksson et al. (2017)	On-road (Tesla)	12	19%
Eriksson and Stanton (2017)	Simulator	26	21%
Heikoop et al. (2017)	Simulator	22	28%
Large, Banks, Burnett, Baverstock, and Skrypchuk (2017)	Simulator	30	36% (partial automation), 21% (high automation)
Manawadu et al. (2015)	Simulator	6 (novices)	36%
	Simulator	6 (experienced)	30%
McDowell et al. (2008)	On-road (military)	11	40%
Petermeijer, Bazilinskyy, and De Winter (2017)	Simulator	24	28% (with auditory and vibrotactile feedback)
Petermeijer, Cieler, and De Winter (2017)	Simulator	18	22%, 36% (with N-Back task)
Saxby, Matthews, Warm, Hitchcock, and Neubauer (2013)	Simulator	36	34% (exp. 1)
	Simulator	56	27% (exp. 2)
Schwalk, Kalogerakis, and Maier (2015)	Simulator	24	21%
Stapel et al. (2019)	On-road (Tesla)	8 (no experience with Tesla)	25% (empty highway), 43% (city ring)
	On-road (Tesla)	7 (experienced with Tesla)	10% (empty highway), 24% (city ring)
Young (2000)	Simulator	18	23%
Young and Stanton (2004)	Simulator	12	12% (exp. 1)
	Simulator	12	12% (exp. 2)
Young and Stanton (2007)	Simulator	24 (novice drivers)	11%
	Simulator	30 (learner drivers)	13%
	Simulator	30 (expert drivers)	20%
	Simulator	30 (advanced drivers)	24%

auto-overtaking system. These authors found a relatively high workload on the NASA Task Load Index (median of 42%) during 9 min of automated driving per participant.

The discrepancy between the results of Banks and Stanton (2016) and the findings of Eriksson et al. (2017) and Stapel et al. (2019) may be caused by the fact that the prototype system tested by Banks and Stanton, which included a heads-up display and offered overtake suggestions, was difficult to use or that participants were still learning how to use it. Because the participants in Banks and Stanton (2016) drove only 9 min with the automation system, the high workload levels "may be a simple reflection of the fact that these ratings were collected during first-time use of the automated system", (p. 393).

McDowell, Nunez, Hutchins, and Metcalfe (2008) and Davis, Animashaun, Schoenherr, and McDowell (2008) performed on-road trials with automated military convoys. In these studies, where there was no other traffic and, because they were military experiments, object detection was of primary importance. The results showed that automated driving reduced workload and improved performance in object detection in comparison to manual driving.

On-road studies may be expected to yield higher workload than simulator research because the latter involves no physical risk of accidents. However, in some cases, on-road studies yielded lower workload than simulator-based studies. For example, the reported workload in Eriksson et al. (2017) was 19%, compared to the 33% in the simulator study of Manawadu, Ishikawa, Kamezaki, and Sugano (2015). This difference might be attributable to the participant pools, added events, or secondary tasks. Specifically, the study of Eriksson et al. (2017) involved experienced test drivers and did not include a secondary task; participants were merely required to take over and relinquish control of the vehicle throughout the experiment. In Manawadu et al. (2015), critical events were triggered, to which the participants had to respond.

1.3. Aim of the present study

The present study aimed to assess whether on-road automated driving with the Tesla Model S alleviates driver workload *over time*. The on-road studies of Eriksson et al. (2017) and Stapel et al. (2019) consisted of approximately 20 min of highway driving with the automation engaged (excluding a familiarization drive) and did not report on temporal effects. Our study consisted of 64 min of automated highway driving per participant. In a driving simulator study by Heikoop, De Winter, Van Arem, and Stanton (2017), use was made of cardiovascular equipment to measure stress and workload; that study found that participants' heart rate dropped during the experiment. Our aim was to replicate these findings on-road with the same equipment as used in the simulator study.

Additionally, in our study, a simple detection task was used to add task demands on top of the regular monitoring demands during automated driving. More specifically, in one of the two drives, participants were instructed to press a handheld button when driving underneath a bridge. The task of detecting bridges is practically convenient in an on-road experiment because bridges are irregularly spaced, and the locations of bridges are retrievable from Google Maps. This detection task is conceptually similar to the approach taken in a previous platooning experiment in a driving simulator (Heikoop et al., 2017). In Heikoop et al. (2017), it was found that the detection task (i.e., to detect red cars on the road) increased self-reported mental demands compared to not performing a detection task. We expected to find a similar effect in this study.

2. Methods

2.1. Participants

Nine participants (seven males, two females) aged between 25 and 47 years ($M = 35.44$; $SD = 8.26$) with 6–30 years of self-reported driving experience ($M = 17.56$; $SD = 8.46$) took part in this experiment. The participants were employees of a large automotive company. Eight participants indicated that they drove every day and one participant indicated driving 4–6 days a week. Two participants indicated they drove up to 10,000 miles, five up to 20,000, one up to 30,000, and one up to 50,000 miles in the past year. All participants had completed level-2 driver training, an extended driver training specifically designed for people who drive as part of their job and which serves as a legal requirement for insurance purposes. All participants had driven various supercars before, and had experience with advanced driver assistance systems (e.g., adaptive cruise control, lane keeping assist), but had no experience with the Tesla Autopilot. We refrained from recruiting participants who had experience with automated driving systems, such as the Tesla's Autopilot, because participants in our previous experiment were not experienced with automated driving systems either (Heikoop et al., 2017) (and see Stapel et al., 2019, for a comparison of self-reported workload between automation-experienced drivers and automation-inexperienced drivers).

No incentive was provided to the participants, and all participants gave written informed consent. The study was approved by the Ethics Research Governance Office of the University of Southampton under submission ERGO number 19091.

2.2. Apparatus

The experiment was performed with a Tesla Model S 90D with Autopilot as the participants' vehicle and a Jaguar XF as a lead vehicle. The lead vehicle was used for safety reasons and for creating a persistent car following task. With a forward-looking radar, forward-facing camera, and ultrasonic sensors, the Autopilot can steer, adjust speed, detect obstacles, and apply brakes automatically ("Full self-driving hardware on all cars", 2015). The Tesla Autopilot can be characterised as level 2 automation (i.e., partial automation) because both steering and speed control are automated, and the driver is still expected to monitor the driving environment (NHTSA, 2017).

The Traffic-Aware Cruise Control of the participant's vehicle was set to 1, which was the closest following distance and which translates to a time headway of about 1 s. This headway corresponds to common headways in highway traffic (Brackstone & McDonald, 2007; Neubert, Santen, Schadschneider, & Schreckenberg, 1999; Song & Wang, 2010; Treiber, Kesting, & Helbing, 2006), and was sufficiently short to have a low likelihood of other cars merging in between the participant's vehicle and the lead vehicle.

Participants wore a Dikablis eye-tracker (but data were deemed unusable) and electrocardiography (ECG) equipment linked to LabChart 8 that captured their cardiovascular and respiratory activity. This ECG equipment consisted of the AD Instruments PowerLab 26 T Teaching Series, three MLA2505 biopotential electrodes and lead wires with disposable ECG electrode patches, and the MLT1132 respiratory belt transducer. The electrodes were placed in a triangular configuration. For male participants, one electrode was placed over the xiphoid process, and two electrodes below the far ends of the collar. For female participants, one electrode was placed at the top of the sternum and two electrodes below the ribs on both sides. This gender-based distinction was mainly made for comfort purposes (see e.g., Shaffer & Combatalade, 2013). The respiratory belt was placed over the clothes around the chest.

2.3. Environment

The experiment took place on March 14–18, 2016. Participants drove on the left (slow) lanes of the British dual three-lane motorways M40, M42, and M5, for which the speed limit is 70 mph (112 km/h). Participants completed two drives during daytime outside of rush hours. The first drive was completed between entry point 14 of the M40 northbound and M5 northbound exit point 3 (Fig. 2). In the second drive, the participants drove back to the starting point. Specifically, the second drive was completed between the motorway entry point at the service stations after entry point 3 of the M5 southbound and the M40 southbound until exit point 14. In one of the two drives, participants completed a bridge detection task (see Section 2.5).

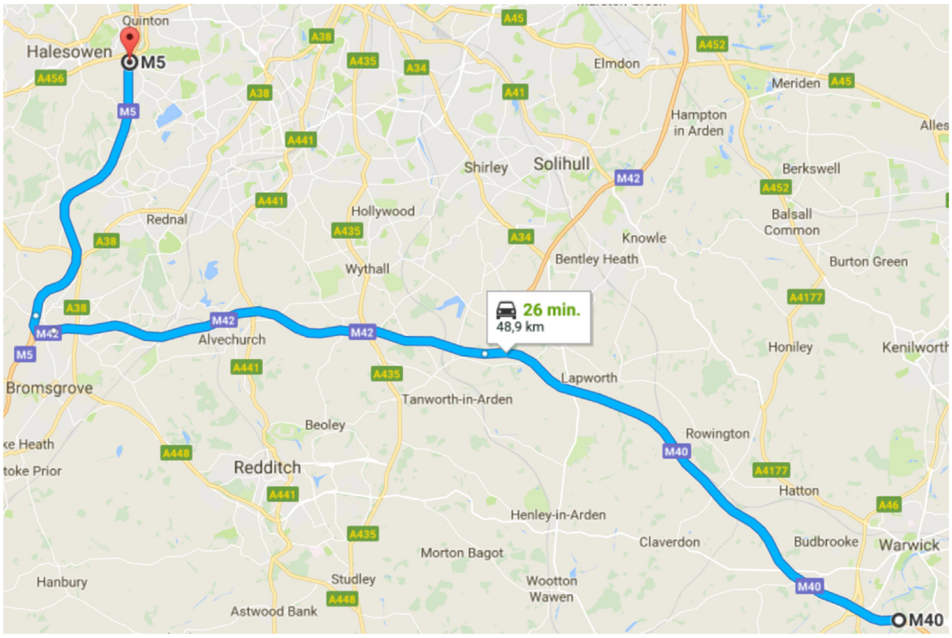


Fig. 2. Map (from Google Maps) displaying the northbound route, starting at entry point 14 of the M40, and ending at exit point 3 of the M5. The southbound route went in the opposite direction, starting at entry point 3 of the M5, and ending at exit point 14 of the M40.

2.4. Procedure

All participants received a training trial and completed two drives of approximately 32 min each. Vigilance research has shown that detection performance exhibits a decay function with time on task (Mackworth, 1964). Furthermore, it has been found that after 15 min the most substantial deterioration of detection performance has taken place (see a review by Teichner (1974), reporting that "at least half of the final loss is completed within the first 15 min" (p. 348)). Because the average driving trip in Europe and the U.S. is between 20 and 30 min (McKenzie & Rapino, 2011; Pasaoglu et al., 2014), it may be assumed that the present study is representative of the first exposures to a new automated driving system on public roads.

Before the experiment, the participant performed a test drive on a test track. Upon arrival at the test track site, the participant received paper instructions explaining that he/she would be driving within a highly automated platoon. Furthermore, a consent form, a demographics questionnaire, and the pre-task Dundee Stress State Questionnaire (DSSQ) were provided. After having completed these questionnaires, the participant was taken to the passenger seat of the participant's vehicle and introduced to the safety driver, who was a professional driver, trained to intervene in emergencies. The safety driver sat in the passenger seat throughout the experiment for legal and safety reasons. The safety driver performed a lap on the test track and showcased the Autopilot, as well as several details of the car. After that lap, the participant and safety driver changed seats, and the participant drove the car until they were comfortable driving manually and with the Autopilot feature. Then the ECG electrodes were attached, after which the participant drove to the selected motorway entry point, following the lead vehicle. After entering the motorway, the Autopilot was engaged by the participant, and the experiment started. The safety driver sat in the seat next to the participant and verbally intervened if the participant did not act appropriately or safely. A verbal intervention could be a situation where the participant did not override the automation when he/she should, for example, when the Autopilot had difficulty keeping the lane in case of diverging lane markers upon approaching an exit ramp. The experimenter sat in the rear seat, monitoring the equipment and making notes of events during the experiment. Before the first drive, the participant was discouraged from interacting with the safety driver or the experimenter for the duration of the experiment. Thus, the interaction between the safety driver and the experimenter was kept to a minimum.

In the occasions where another vehicle merged in between the participant's vehicle and the lead vehicle, the participants were instructed by the safety driver to remain in automated mode and follow this other vehicle. However, if the gap with the lead vehicle became large, then the participants were calmly instructed to follow the lead vehicle again by overtaking the outside traffic, while it was emphasised to try to remain in automated mode. An automated lane change could be performed by using the indicator stalk while holding the steering wheel. All events such as lane changes, merges, and Autopilot (dis) engagements were recorded by the experimenter using paper and pencil. Summed across the nine participants, a total of 30 and 37 lane changes (of which 16 and 21 automated) occurred for Drive 1 and Drive 2, respectively. A manual override occurred 9 and 4 times during Drive 1 and Drive 2, respectively.

At the end of the first drive, the participants exited the motorway and stopped at a nearby parking lot. They were then provided with the post-task DSSQ and the NASA Task Load Index. Once completed, the participants followed the lead vehicle to the motorway again and performed the second drive. At the end of the second drive, the participants were again provided with the post-task DSSQ and the NASA Task Load Index.

2.5. Independent variable – Bridge detection task

The experiment consisted of two drives, either with or without a detection task, in counterbalanced order. Specifically, five participants completed the second 'southbound' drive with the detection task, and four participants completed the first 'northbound' drive with the detection task. Without the detection task, participants had to follow the lead vehicle as their only objective. With the detection task, they also had to detect the bridges they went underneath by pressing a handheld button (Fig. 3). In the first drive, participants drove underneath 50 bridges, and during the second drive, participants drove underneath 47 bridges. Photos of the bridges are available as supplementary material. Learning/acclimatization effects were assessed by comparing the results of the first drive with the results of the second drive.

2.6. Dependent measures

The following measures were recorded per participant, for each of the two drives:

- Time of day (hh:mm).
- Duration of the drive (s).
- Mean speed (km/h), recorded with a GPS application on a smartphone.
- Manual overrides (number of times the participant took over control from the Autopilot to drive manually).
- Manual lane changes (number of manually performed lane changes). Lane changes that were initiated in Autopilot mode and taken over by the participant to be performed manually were also regarded as manual lane changes.
- Automated lane changes (number of lane changes performed by the participant while remaining in Autopilot mode during the lane change).
- Hit rate of the bridges (% of bridges detected). The hit rate was calculated by linking the known locations of the bridges (as retrieved from Google Maps) with the locations at the moments of button presses (recorded with a GPS application on a smartphone).

  An algorithm was written that matched bridges with the nearest button press in terms of radial distance until all bridges were assigned to a button press or no button presses were left (each button press could be assigned to one bridge only). Button presses which followed each other within $\frac{2}{3}$ s (i.e., accidental double pressing of the button) were discarded.

Fig. 3. Photo taken during the experiment: the participant on the right presses the handheld button when detecting a bridge. The safety driver is sitting in the left (passenger) seat. The vehicle in front is the lead vehicle the participant had to follow during the experiment.

Furthermore, if the nearest button press was more than 1250 m from the bridge, then this bridge was marked as a miss. The liberal threshold of 1250 was used because there were several sources of inaccuracy in the locations of the button presses. Specifically, (1) The GPS signal had a limited temporal resolution (0.2 Hz, which at an average speed of 86 km/h amounts to a travelled distance of about 120 m), (2) Some participants pressed the button late (i.e., when being beneath a bridge) while others pressed the button early (i.e., when the bridge could first be seen), and (3) The GPS recording had limited accuracy (the 50th, 95th, and 99th percentile of the estimated accuracy were 24 m, 249 m, and 965 m, respectively). By definition, the miss rate equals 100% minus the hit rate (Tanner, Wilson, & Swets, 1954).

- False alarm rate (% of false alarms relative to the number of bridges). A button press was considered a false alarm when after determining the hits, there were still button presses unaccounted for. Fig. 4 provides an illustration of the hits and false alarms for Participant #1.
- Heart rate (bpm). The heart rate was regarded as a measure of stress and mental workload (Healey & Picard, 2004; Mehler, Reimer, & Coughlin, 2012).
- Standard deviation of the normal-to-normal interval (SDNN) (ms), a time-domain measure of mental workload (e.g., De Waard, 1996; Jorna, 1992). The SDNN was defined as the mean of the standard deviation (*SD*) of all Normal to Normal peak intervals (NN) in the ECG signal per 5-min segment along the drive (SDNN index, see Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996). A low SDNN value is interpreted as high workload (Fallahi, Motamedzade, Heidarimoghadam, Soltanian, & Miyake, 2016; see also Heikoop et al., 2017). See Fig. 5 for an illustration of the calculation process.
- Low frequency/high frequency (LF/HF) ratio, a frequency-domain measure of mental workload. This spectral analysis of the NN interval calculates the power in the low-frequency (LF) 0.04–0.15 Hz range relative to the power in the high-frequency (HF) 0.15–0.40 Hz range. A high LF/HF ratio is indicative of high workload (Cinaz, Arnrich, La Marca, & Tröster, 2013; Suriya-Prakash, John-Preetham, & Sharma, 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 (Vollmer, 2015).
- Respiratory rate (bpm). Because the respiratory belt transducer produced a noisy signal (presumably because of in-vehicle vibrations) and may contain drifts and other artefacts, the signal was filtered with a second-order Butterworth 0.1–1.0 Hz bandpass filter. This frequency range incorporates a typical human respiratory rate of 0.25 Hz. Next, the data were rank transformed to remove outliers, and subsequently, a discrete Fourier transformation was applied to retrieve the frequency with maximum amplitude (see Fig. 6 for illustration).
- The Dundee Stress State Questionnaire (DSSQ), a self-report measure of stress and fatigue (Matthews, Szalma, Panganiban, Neubauer, & Warm, 2013). In this experiment, version 1.3 of the DSSQ was used (Matthews, Campbell, & Falconer, 2000). Standardized change scores for each scale of the DSSQ were calculated as follows: (post-score – pre-score)/(standard deviation of the pre-score) (Helton, Warm, Matthews, Corcoran, & Dember, 2002). The scores for the three scales (Engagement, Distress, and Worry) were calculated by averaging four subscales and averaging them to result in one score for each element (based on Fairclough & Venables, 2005; Matthews, 2014; Matthews et al., 2002). Task

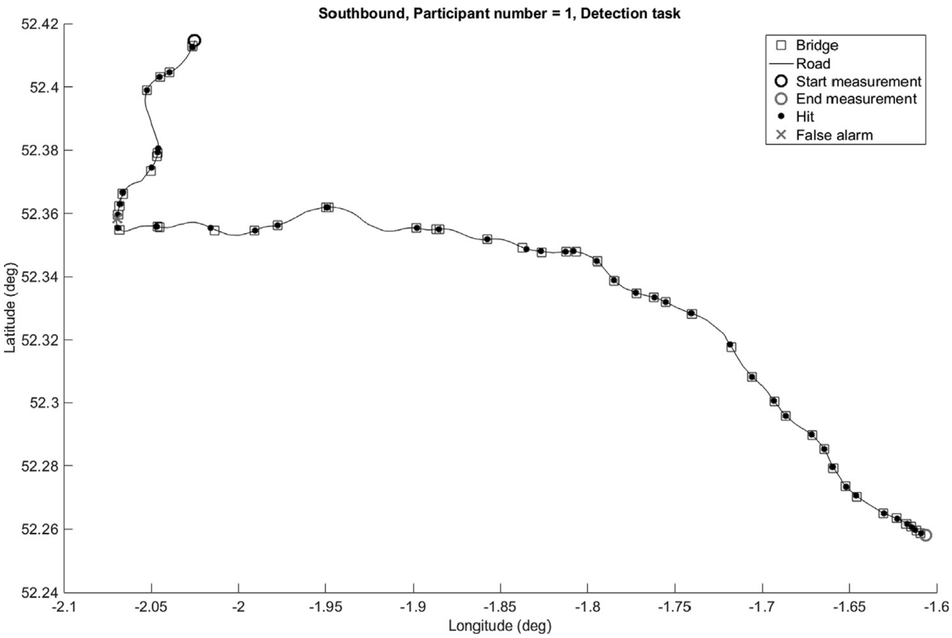
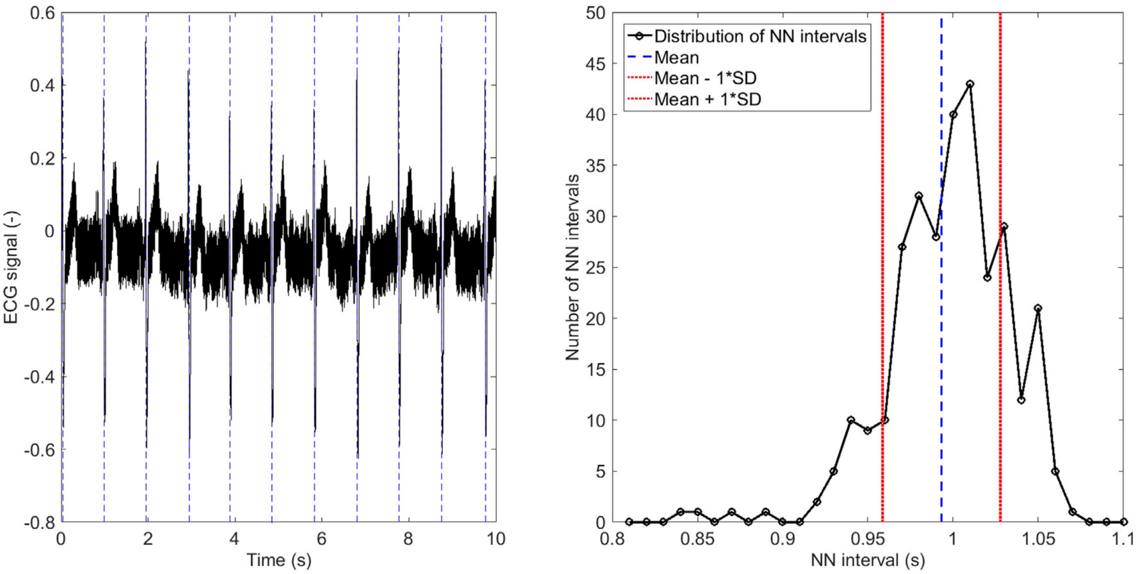
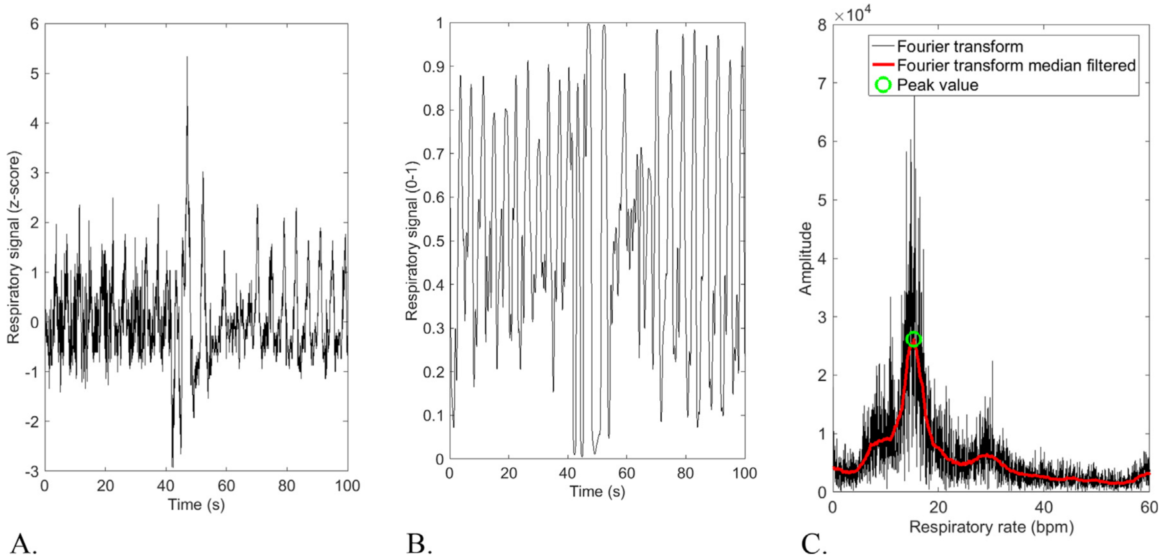


Fig. 4. Illustration of GPS data and the detection task. In this case, the participant detected all 47 bridges (hit rate = 100%) and had 1 false alarm (false alarm rate = 2.1%). Where a single dot is visible for two bridges, two button presses appeared in the same GPS sample (the GPS recorded the position every 5 s).



A. B.

Fig. 5. Illustration of the calculation of SDNN. (A) ECG signal with extracted NN intervals (first 10 s of Drive 1 of Participant #1). (B) Distribution of NN intervals with the mean and standard deviation of the NN intervals (SDNN = 34.6 ms; based on the first 300 s of Drive 1 of Participant #1).



A. B. C.

Fig. 6. Illustration of data processing of the respiratory signal. (A) *z*-transformed raw signal (first 100 s of Drive 1 of Participant #1). (B) Filtered signal, rank-transformed and scaled from 0 to 1 (first 100 s of Drive 1 of Participant #1). (C) Amplitude (non-normalized) of a discrete Fourier transform (calculated using a fast Fourier transform) with identified peak value (based on the entire Drive 1 of Participant #1).

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 & Confidence, and (8) Anger/Frustration. Finally, Worry consists of (9) Self-Focused Attention, (10) Self-Esteem, (11) Task-Relevant Interference, and (12) Task-Irrelevant Interference. We imputed missing answers (4% of the total) using the nearest-neighbour method.

- NASA Task Load Index, a self-report measure to assess workload (Hart & Staveland, 1988). The 'raw' approach was used, which does not apply weights to the scales (Hart, 2006).

The mean speed, duration, heart rate, SDNN, LF/HF ratio, and respiratory rate were calculated from the moment that the participant was 200 m in front of the first bridge until 200 m after the participant passed the last bridge. SDNN was calculated as the average across six available 5-min segments.

3. Results

Table 2 presents the results per participant. Due to the small number of participants in this study, statistical tests are not reported, as these were deemed unreliable.

Participants drove on average about 32 km per drive, at a mean speed of 86 km/h (Table 2), which is well below the speed limit of 112 km/h (the speed limit on British motorways is 70 mph, which equals 112 km/h). The difference in duration between Drive 1 and Drive 2 is caused by the fact that Drive 1 was about 4 km longer than Drive 2. This was due to the respective entry- and exit points being in different locations.

The 9 participants together took control from the automated driving system 13 times, of which 6 were due to the Autopilot not anticipating the traffic merging between the lead vehicle and participant's vehicle, 2 due to the Autopilot following the undesired line at an exit point, 2 due to the participant not trusting the Autopilot to perform correctly, 1 to an unexpected disengagement of the Autopilot, and 1 to the participant disengaging the Autopilot without apparent reason. The remaining Autopilot disengagement occurred for unknown reasons. Furthermore, lane changes were performed 27 and 40 times during the no task and detection task condition, respectively, of which 14 and 16 were manual. The safety driver verbally intervened to take over manual control at two instances, both at the point where the M5 and M42 merge. However, no dangerous situations occurred during any of the drives.

Table 3 shows the results of the self-report questionnaires. Averaged across the two drives, the mean self-reported overall workload was 19% (21% in Drive 1, 16% in Drive 2; 18% for no task drives, 19% for detection task drives). The mean (*SD*) per item of the NASA Task Load Index was 26% (19%) for Physical Demand, 8% (6%) for Mental Demand, 12% (9%) for Temporal Demand, 27% (29%) for Performance, 17% (13%) for Effort, and 21% (23%) for Frustration. The DSSQ results showed that participants exhibited an overall disengagement from the task, and a worriless attitude towards the task as compared to the pre-task DSSQ (i.e., the standardized change scores are smaller than 0).

Table 2
Descriptive results (time of day, duration, mean speed, manual overrides, manual and automated lane changes) per participant and drive number.

Participant	Time of day Drive		Duration (s) Drive		Mean speed (km/h) Drive		Manual overrides Drive		Manual lane changes Drive		Automated lane changes Drive	
	1	2	1	2	1	2	1	2	1	2	1	2
1 (NT, DT)	10:52	11:55	2158	1925	80.5	82.9	1	0	2	1	0	2
2 (DT, NT)	14:29	15:18	2032	1892	85.2	84.1	0	1	1	4	5	1
3 (NT, DT)	10:43	11:46	2057	1653	85.4	96.5	2	1	1	4	1	4
4 (DT, NT)	14:40	15:20	1440	1405	85.6	86.1	1	0	2	0	2	2
5 (NT, DT)	10:33	11:28	2073	1871	83.6	85.1	2	0	0	1	3	2
6 (DT, NT)	14:41	15:33	2013	1834	86.1	87.0	0	0	2	2	1	1
7 (NT, DT)	10:25	11:18	1991	1813	87.1	87.9	0	1	1	2	2	3
8 (DT, NT)	14:34	15:34	1928	1800	88.9	88.2	2	0	2	1	1	2
9 (NT, DT)	10:40	11:42	2008	1886	86.1	84.6	1	1	3	1	1	4
Average			2033	1834	85.4	86.9	1.00	0.44	1.56	1.78	1.78	2.33

Note. NT = No Task, DT = Detection Task. Participant #4 did not complete the entire route because the batteries of the car were emptying and the car needed to be charged. This participant was excluded from the calculation of the average duration.

Table 3
Self-reported overall workload on the NASA Task Load Index and standardized change scores of self-reported stress (DSSQ) per participant and drive number.

Participant	Workload score (%) Drive		DSSQ engagement Drive		DSSQ distress Drive		DSSQ worry Drive	
	1	2	1	2	1	2	1	2
1 (NT, DT)	12	8	−0.10	−0.21	−0.43	−0.49	−0.35	−0.57
2 (DT, NT)	18	14	−0.12	−0.18	−0.97	−1.13	0.17	−0.16
3 (NT, DT)	45	46	−1.10	−0.19	0.67	0.23	−0.78	0.14
4 (DT, NT)	17	3	−0.01	0.26	−0.82	−1.41	−2.34	−1.93
5 (NT, DT)	31	20	−0.37	−0.61	0.40	−0.16	−0.51	−0.72
6 (DT, NT)	14	14	0.18	0.46	−0.30	−0.63	−0.34	−0.32
7 (NT, DT)	9	6	−1.48	−0.52	0.46	0.26	0.49	−0.68
8 (DT, NT)	40	25	−2.60	−1.88	2.67	−0.90	0.81	1.59
9 (NT, DT)	5	7	−1.43	−0.02	1.69	1.13	−1.13	−1.69
Average	21	16	−0.78	−0.32	0.38	−0.34	−0.44	−0.48

Note. NT = No Task, DT = Detection Task.

Table 4
Cardiovascular and respiratory results per participant and drive number.

Participant	Heart rate (bpm)		SDNN (ms)		LF/HF ratio		Respiratory rate (bpm)	
	Drive		Drive		Drive		Drive	
	1	2	1	2	1	2	1	2
1 (NT, DT)	61.3	61.9	44.7	46.2	0.79	0.72	15.3	15.7
2 (DT, NT)	99.4	90.4	20.3	33.6	1.52	2.00	N.A.	N.A.
3 (NT, DT)	59.7	56.6	50.2	53.9	0.97	0.73	14.1	15.1
4 (DT, NT)	72.2	67.5	48.3	59.4	1.02	1.14	18.5	17.8
5 (NT, DT)	69.0	66.4	55.7	47.7	1.26	1.26	15.6	14.9
6 (DT, NT)	82.1	73.0	43.6	42.4	0.91	0.91	18.4	17.2
7 (NT, DT)	59.2	58.9	44.9	51.5	0.89	0.94	19.2	19.7
8 (DT, NT)	80.4	70.7	35.3	39.5	0.99	0.96	20.2	20.3
9 (NT, DT)	56.6	52.2	93.2	101.3	1.18	1.02	18.4	18.7
Average	71.1	66.4	48.5	52.8	1.06	1.08	17.0	16.5

Note. NT = No Task, DT = Detection Task. The respiratory rate of Participant #2 is not provided, as there was no clear peak value to be identified from the Fourier transformation.

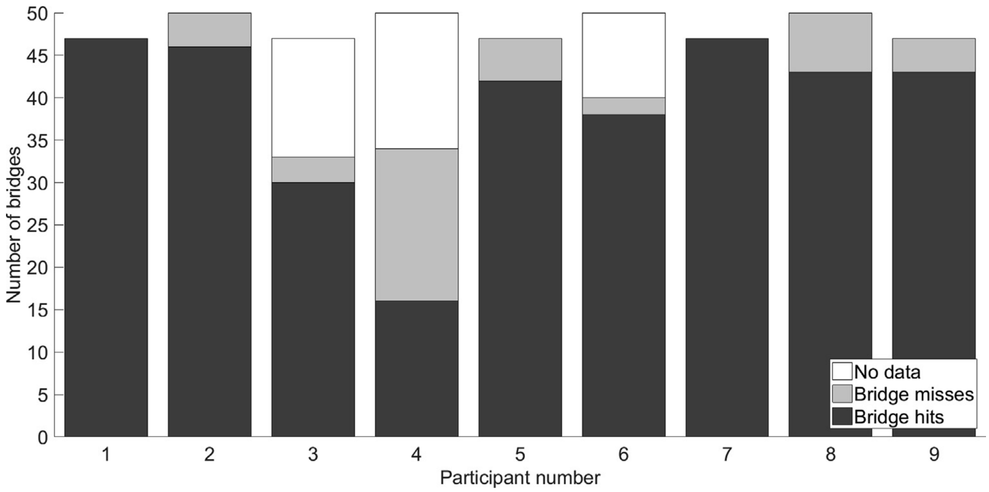


Fig. 7. Number of hits and misses per participant. The number of false alarms was 1, 0, 0, 0, 1, 0, 0, 0, 2 for Participants 1–9.

Table 4 shows the results of the physiological measures. An acclimatization effect can be seen, with the heart rate being lower in Drive 2 than in Drive 1 for 8 out of 9 participants. The SDNN exhibits a negative correlation with the heart rate (Spearman rank-order correlation = −0.77) (see also Heikoop et al., 2017). Here, for 7 of 9 participants, SDNN was higher in Drive 2 than in Drive 1. The LF/HF ratio and the respiratory rate remained relatively constant throughout the two drives.

Due to technical issues, participants #3, 4, and 6 had no button press data for 14, 16, and 10 bridges respectively. Hit rates and false alarm rates were calculated for the remaining number of bridges for these participants. The mean (*SD*) hit and false alarm rates of the bridges were 88.0% (16.0%) and 0.9% (1.5%), respectively. The lowest hit rate was 47.1%, whereas two participants had hit rates of 100% (Fig. 7). The mean hit rate and mean false alarm rate correspond to a perceptual sensitivity (d') of 3.52 and response bias (β) of 7.89 (Stanislaw & Todorov, 1999). These results indicate that participants were well able to distinguish the bridges from the non-bridges with a conservative response strategy.

4. Discussion

This study aimed to measure levels of workload and stress during automated driving with the Tesla Autopilot. The literature has shown that automated driving yields low ratings of self-reported overall workload (averaging at 23%, see De Winter et al., 2014). A previous study using the Tesla Autopilot has found overall workload scores ranging from 10% for experienced Tesla drivers on an empty highway to 43% on a city ring with drivers who had not driven in the Tesla before (Table 1; Stapel et al., 2019). Another on-road experiment found high workload for automated driving compared to manual driving, with overall workload scores for automated driving being 42% (Banks & Stanton, 2016). It was unclear whether the novelty of the automation in Banks and Stanton (2016) created elevated levels of workload, so they proposed extended exposure to automation, which was the purpose of the current study.

Our results of a 2 × 32 min of automated driving found that the mean overall workload as measured with the NASA Task Load Index was low, with a score of 21% in Drive 1 and 16% in Drive 2. In other words, our results provided an indication that automated driving involves a level of self-reported workload that is within the range of the workload observed in driving simulators (see Table 1, showing a minimum overall workload of 11% and a maximum of 36%). The fact that the participants had to follow a lead vehicle may have contributed to the low overall workload by limiting their decision-making requirements. The workload was particularly low for the Physical Demand item, which may be because the participants did not have to move the pedals or steering wheel for large portions of the time. Similarly, the average heart rate and respiratory rate were close to the resting rates of a typical person (American Heart Association, 2015; Lindh, Pooler, Tamparo, & Dahl, 2009). Thus, automated driving in the present experiment could not be considered demanding or stressful. Because of the small number of participants, we refrained from reporting the results of null hypothesis significance tests and recommend that the raw data in Tables 2–4 be used in future meta-analysis with larger statistical power. However, several results are robust and may not require larger samples. For example, the overall NASA Task Load Index workload score averaged across the two drives per participant ranged between 6% and 45% with a 95% confidence interval of 8% to 29%. Hence, it appears to be legitimate to conclude that overall workload is 'low' and in line with prior research, such as a similar study by Stapel et al. (2019).

Compared to the detection task in Heikoop et al. (2017), participants in the current study performed somewhat worse (hit rate of 95% in Heikoop et al., 2017; 88% in the present study). Furthermore, self-reported workload in the present study was not substantially different between the detection task and no task conditions. It is possible that, for safety reasons or due to the presence of the safety driver sitting next to them, participants in the present study were trying harder to stay alert to the primary driving task as compared to the participants in the simulator study, thereby having a lower incentive to detect the bridges. Indeed, with the Tesla Autopilot, participants do have to remain alert due to the potential need to intervene and take manual control of the vehicle. In a survey study by Dikmen and Burns (2016), 62% of Autopilot users reported that they had experienced at least one unexpected or unusual behaviour when driving in automated mode. In our experiment, the 9 participants combined took over manual control 13 times, for a variety of reasons, and performed 67 lane changes, of which 45% were manual. The relatively large percentage of manual lane changes may be due to the somewhat cumbersome technique required to perform an automated lane change. For an automated lane change to succeed, the driver pressed the indicator stalk while having his/her hands on the steering wheel with enough weight for the Autopilot to recognize their presence. In our experiment, this often resulted in a slight turn of the steering wheel, disengaging the Autopilot, after which the lane change had to be performed manually.

Fig. 8 provides a comparison between the results of the current on-road study and two previous driving simulator studies (Heikoop et al., 2017; Heikoop, De Winter, Van Arem, & Stanton, 2018). The two simulator studies are similar to the current study as all three studies involved car following in an automated car, drives of 30–40 min duration, identical ECG equipment,

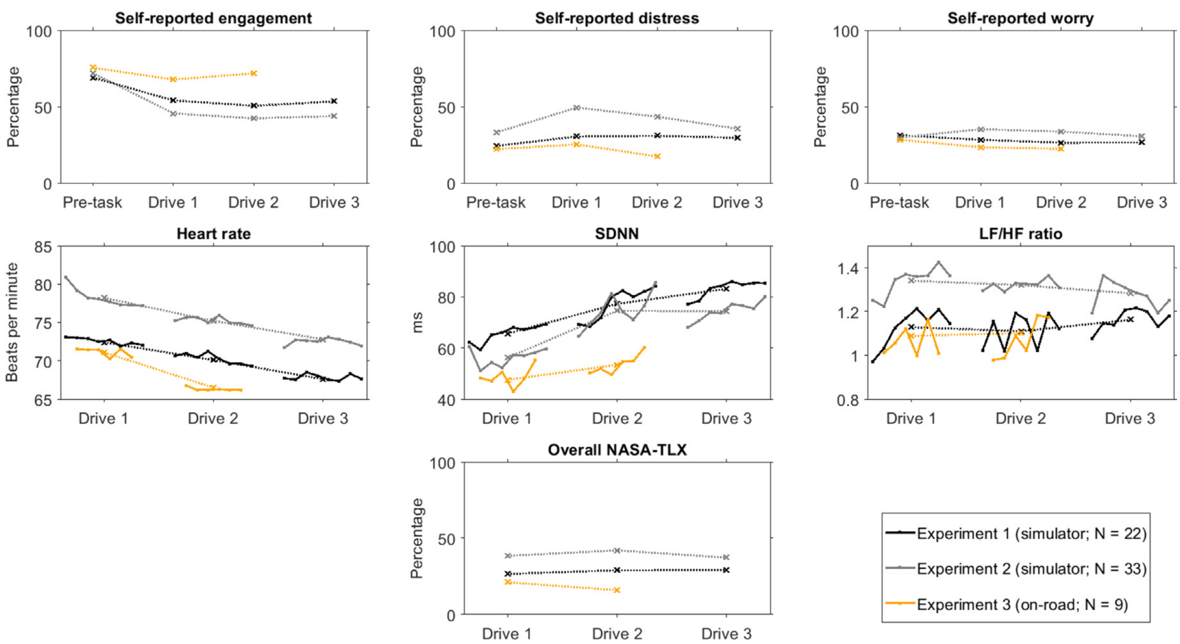


Fig. 8. Comparisons of three independent experiments. Top: The Dundee Stress State Questionnaire (DSSQ). Middle: Cardiovascular measures per 5-min interval. Bottom: Self-reported overall workload. Experiment 1: Heikoop et al. (2017) ($N = 22$; 3 × 40 min of simulated driving). Experiment 2: Heikoop et al. (2018) ($N = 33$, but $N = 29$ for heart rate measures; 3 × 40 min of simulated driving). Experiment 3: present experiment ($N = 9$; 2 × 32 min of on-road driving). Note: The DSSQ used in Experiment 2 is the short version of the DSSQ. The DSSQ and workload scores were scaled from 0% (minimum possible) to 100% (maximum possible).

and a secondary button-pressing task (detecting bridges in the present on-road study, detecting red cars in the simulator studies). Of course, the comparison between the three experiments cannot be made one-on-one, as the nature of the experiment and the participant pool varied between the studies, with expert drivers in the on-road study and participants from the university community in the simulator-based studies. Furthermore, the second simulator study was more demanding than the first because a think-aloud method and memory task were applied during driving. Despite several differences between the three studies, interesting observations can be made from Fig. 8. First, it can be seen that participants felt relatively engaged during the on-road study. The relatively high level of engagement may be because the participants in the present study prioritized safety and tried to stay alert, as noted above. Participants of this on-road study reported relatively low levels of distress and worry compared to the participants in the simulator studies. This may be because the participants in the on-road study were expert drivers who were accustomed to driving with advanced driver assistance systems. Furthermore, the overall self-reported workload was low (a mean score of 19% on a scale from 0% to 100%). Both the on-road and simulator-based studies found that self-reported workload remained approximately constant, heart rate decreased, and SDNN increased as a function of time. Previous on-road studies into manual driving (Lisper, Laurell, & Stening, 1973; Schmidt et al., 2009) and partially automated driving using the Tesla Autopilot (Stapel et al., 2019) also found that the heart rate decreases with time on task. Finally, the LF/HF ratios in the simulator and on the road were within the same range (Fig. 8). Summarizing, the results from the current on-road experiment provide trends that are similar to results from simulator studies, with a decrease in heart rate over time and self-reported workload at the low-end of the scale.

5. Conclusions and recommendations

This experiment complements existing research on stress and workload during automated driving (Table 1), and may form a basis for more extensive research on this topic. The results point to an effect of acclimatization as demonstrated by a drop in perceived workload over time, and a decrease in heart rate. Our sample was small ($N = 9$); in order to acquire higher statistical power, replication studies with more participants are advised. Nonetheless, our study produced insights into the effects of workload and acclimatization to automation and could serve as a foundation for future research into this phenomenon.

This experiment used expert drivers, and it remains to be investigated how the results translate to less trained drivers. It is likely that the participants in this study, who were experienced with various supercars and had completed advanced driver training, are more adept to novel technology and less likely to be stressed than the general population. Stapel et al. (2019) found in their on-road study that people who have experience with driving in a Tesla reported a lower workload than people who had not driven a Tesla before, both during manual and automated driving. Besides prior experience with automated driving systems, other types of individual differences may also be of relevance for future research. These include age-related psychometric abilities (e.g., speed of information processing) and factors like boredom proneness and propensity to daydream, as well as general trust in automation (Körber & Bengler, 2014). Future research should take into account these effects in order to reach more generalizable conclusions about how people interact with partially automated driving systems.

The methods used for psychophysiological measurement need consideration. Although the LF/HF ratio is regarded as a valid measure of workload (Cinaz et al., 2013; Suriya-Prakash et al., 2015), the results of the LF/HF ratio in this experiment did not yield the same time-on-task effects as the self-reported overall workload. The lack of sensitivity of the LF/HF ratio could be due to various confounding effects such as driver posture, vibrations in the vehicle, or a dependency on the heart rate itself (as also discussed by Heikoop et al., 2017). Also, the use of a smartphone-based GPS app did not allow for high accuracy data (see Section 2.6). It is recommended that future researchers use vehicle state data combined with differential GPS.

The focus of this experiment was on how drivers are affected by automated driving over time. Future on-road research could include a control condition in which people drive manually (cf. Stapel et al., 2019) or include different levels of automation such as driving with adaptive cruise control, or adaptive cruise control combined with steer assist (see Naujoks, Purucker, & Neukum, 2016). In our experiment, control transitions were not wanted, yet occurred several times per participant. A closed-track on-road experiment (cf. Albert, Lange, Schmidt, Wimmer, & Bengler, 2015; Omae, Hashimoto, Sugamoto, & Shimizu, 2005) could investigate the psychological effects of transition of control to and from automated driving in a controlled manner. Furthermore, it could be investigated whether the present findings generalize to driving tasks situations without a safety driver, and with actual hazards for which manual intervention is necessary. It is possible that drivers would score better in an environment in which the target stimuli represent actual safety-critical events. Additionally, it is noted that physiological measures and the NASA Task Load Index capture only certain facets of stress and workload. Recent on-road research indicated that using the Tesla Autopilot is associated with higher response times on a secondary task as compared to manual driving, possibly due to monitoring demands (e.g., monitoring of the roadway or automation status) or difficulty with multitasking (Biondi, Lohani, Hopman, Mills, Cooper, & Strayer, 2018; Stapel et al., 2019). Also, it remains to be studied how our findings generalize to other driving scenarios, such as driving during rush hours. Our experiment could be extended by investigating, for example, the effects on behavioural adaptation and trust, complementing recent on-road studies with partially automated driving systems by, amongst others, Banks, Eriksson, O'Donoghue, and Stanton (2018), Fridman, Brown, Kindelsberger, Angell, Mehler, and Reimer (2019), Koskinen, Lyyra, Mallat, and Tuunainen (2019), Lin, Ma, and Zhang (2018), and Walker, Boelhouwer, Alkim, Verwey, and Martens (2018).

Finally, we note that other types of automated driving systems, such as Volvo's Pilot Assist, may not be comparable to our present findings for the Tesla Autopilot, as the former requires the driver to touch the wheel permanently (Solís-Marcos, Ahlström, & Kircher, 2018). Whether a hands-on or hands-off approach is optimal in terms of workload and safety remains a topic for further investigation (see e.g., Wandtner, Schömig, & Schmidt, 2018).

Acknowledgements

This experiment was performed in collaboration with Jaguar Land Rover. The authors are involved in the Marie Curie Initial Training Network: HFAuto (PITN-GA-2013-605817).

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.4121/uuid:cbe0f5c6-0439-4533-86c2-faeebedf5662.

References

Albert, M., Lange, A., Schmidt, A., Wimmer, M., & Bengler, K. (2015). Automated driving – Assessment of interaction concepts under real driving conditions. In *Proceedings of the 6th international conference on applied human factors and ergonomics (AHFE2015), and the affiliated conferences* (pp. 2832–2839). https://doi.org/10.1016/j.promfg.2015.07.767.

American Heart Association. (2015). Target heart rates. Retrieved from http://www.heart.org/HEARTORG/HealthyLiving/PhysicalActivity/FitnessBasics/Target-Heart-Rates_UCM_434341_Article.jsp#.V8Q9T_krKWg.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study. *Applied Ergonomics, 68*, 138–145. https://doi.org/10.1016/j.apergo.2017.11.010.

Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control: Automating automobiles of the future. *Applied Ergonomics, 53*, 389–395. https://doi.org/10.1016/j.apergo.2015.06.020.

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. *Transportation Research Part F, 35*, 75–84. https://doi.org/10.1016/j.trf.2015.10.005.

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. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 62, No. 1, pp. 1878–1882). Sage CA: Los Angeles, CA: SAGE Publications. http://dx.doi.org/10.1177/1541931218621427.

Borojeni, S. S., Chuang, L., Heuten, W., & Boll, S. (2016). Assisting drivers with ambient take-over requests in Highly Automated Driving. In Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications (AutomotiveUI '16), 237–244, October 24–26, Ann Arbor, MI. http://dx.doi.org/10.1145/3003715.3005409.

Brackstone, M., & McDonald, M. (2007). Driver headway: How close is too close on a motorway? *Ergonomics, 50*, 1183–1195. https://doi.org/10.1080/00140130701318665.

Casner, S. M., Hutchins, E. L., & Norman, D. (2016). The challenges of partially automated driving. *Communications of the ACM, 59*, 70–77. https://doi.org/10.1145/2830565.

Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and Ubiquitous Computing, 17*, 229–239. https://doi.org/10.1007/s00779-011-0466-1.

Damböck, D., Weißgerber, T., Kienle, M., & Bengler, K. (2013). Requirements for cooperative vehicle guidance. In Proceedings of the 16th international IEEE annual conference on intelligent transportation Systems (pp. 1656–1661). The Hague, The Netherlands. http://dx.doi.org/10.1109/itsc.2013.6728467.

Davis, J., Animashaun, A., Schoenherr, E., & McDowell, K. (2008). Evaluation of semi-autonomous convoy driving. *Journal of Field Robotics, 25*, 880–897. https://doi.org/10.1002/rob.20263.

De Waard, D. (1996). *The measurement of drivers' mental workload* Doctoral dissertation. University of Groningen.

De Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 196–217. https://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. *International Journal of Vehicle Design, 70*, 297–324. https://doi.org/10.1504/IJVD.2016.076736.

Dikmen, M., & Burns, C. M. (2016). Autonomous driving in the real world: Experiences with Tesla Autopilot and Summon. In Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications (AutomotiveUI '16), 225–228, October 24–26, Ann Arbor, MI. http://dx.doi.org/10.1145/3003715.3005465.

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making, 11*, 225–238. https://doi.org/10.1177/1555343417695197.

Eriksson, A., Banks, V. A., & Stanton, N. A. (2017). Transition to manual: Comparing simulator with on-road control transitions. *Accident Analysis & Prevention, 102*, 227–234. https://doi.org/10.1016/j.aap.2017.03.011.

Eriksson, A., & Stanton, N. A. (2017). Take-over time in highly automated vehicles: Non-critical transitions to and from manual control. *Human Factors, 59*, 689–705. https://doi.org/10.1177/0018720816685832.

Fairclough, S. H., & Venables, L. (2005). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology, 71*, 100–110. https://doi.org/10.1016/biopsycho.2005.03.007.

Fallahi, 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. *Applied Ergonomics, 52*, 95–103. https://doi.org/10.1016/j.apergo.2015.07.009.

Fridman, L., Brown, D. E., Kindelsberger, J., Angell, L., Mehler, B., & Reimer, B. (2019). Human side of Tesla Autopilot: Exploration of functional vigilance in real-world human-machine collaboration. Retrieved from https://www.researchgate.net/publication/332462137_Human_Side_of_Tesla_Autopilot_Exploration_of_Functional_Vigilance_in_Real-World_Human-Machine_Collaboration.

Full self-driving hardware on all cars. (2015). Retrieved from: https://www.tesla.com/en_GB/autopilot.

Hancock, P. A. (2015). Automobility: The coming use of fully-automated on-road vehicles. In *Proceedings of the IEEE international multi-disciplinary conference on cognitive methods in situation awareness and decision support (CogSIMA)* (pp. 137–139). https://doi.org/10.1109/cogsima.2015.7108188.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society 50th annual meeting, (Vol. 50, pp. 904–908). http://dx.doi.org/10.1177/154193120605000909.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183. https://doi.org/10.1016/s0166-4115(08)62386-9.

Haslbeck, A., & Hoermann, H.-J. (2016). Flying the needles: Flight deck automation erodes fine-motor flying skills among airline pilots. *Human Factors, 58*, 533–545. https://doi.org/10.1177/0018720816640394.
Healey, J. A., & Picard, R. W. (2004). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*, 156–166. https://doi.org/10.1109/TITS.2005.848368.
Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, stress, and fatigue: A driving simulator study. *Applied Ergonomics, 60*, 116–127. https://doi.org/10.1016/j.apergo.2016.10.016.
Heikoop, D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2018). Effects of mental demands on situation awareness during platooning: A driving simulator study. *Transportation Research Part F, 58*, 193–209. https://doi.org/10.1016/j.trf.2018.04.015.
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. In Proceedings of the human factors and ergonomics society annual meeting, (Vol. 46, 1546–1550). https://doi.org/10.1177/154193120204601704..
Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology, 34*, 237–257. https://doi.org/10.1016/0301-0511(92)90017-O.
Kaber, D. K., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress, 16*, 126–131. https://doi.org/10.1002/prs.680160304.
Koskinen, K., Lyyra, A., Mallat, N., & Tuunainen, V. K. (2019). Trust and risky technologies: Aligning and coping with Tesla Autopilot. In Proceedings of the 52nd Hawaii International Conference on System Sciences (pp. 5777–5786). http://dx.doi.org/10.24251/hicss.2019.697.
Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. Interacción 2014, Puerto de la Cruz, Tenerife, Spain. http://dx.doi.org/10.1145/2662253.2662275.
Large, D. R., Banks, V. A., Burnett, G. E., Baverstock, S., & Skrypchuk, L. (2017). Exploring the behaviour of distracted drivers during different levels of automation in driving. In Proceedings of the 5th international conference on driver distraction and inattention (DDI2017), March 20–22, Paris, France.
Lin, R., Ma, L., & Zhang, W. (2018). An interview study exploring Tesla drivers' behavioural adaptation. *Applied Ergonomics, 72*, 37–47. https://doi.org/10.1016/j.apergo.2018.04.006.
Lindh, W. Q., Pooler, M. S., Tamparo, C. D., & Dahl, B. M. (2009). Vital signs and measurements. In W. Q. Lindh, M. S. Pooler, C. D. Tamparo, & B. M. Dahl (Eds.), *Delmar's comprehensive medical assisting: administrative and clinical competencies, Chapter 24, 573*. Clifton Park, NY: Cengage Learning Inc.
Lisper, H.-O., Laurell, H., & Stening, G. (1973). Effects of experience of the driver on heart-rate, respiration-rate, and subsidiary reaction time in a three hours continuous driving task. *Ergonomics, 16*, 501–506. https://doi.org/10.1080/00140137308924539.
Mackworth, J. F. (1964). Performance decrement in vigilance, threshold, and high-speed perceptual motor tasks. *Canadian Journal of Psychology, 18*, 209–223. https://doi.org/10.1037/h0083302.
Manawadu, U., Ishikawa, M., Kamezaki, M., & Sugano, S. (2015). Analysis of individual driving experience in autonomous and human-driven vehicles using a driving simulator. In Proceedings of the IEEE international conference on advanced intelligent mechatronics (AIM), 299–304, July 7–11, Busan, Korea. http://dx.doi.org/10.1109/AIM.2015.7222548.
Matthews, G. S. (2014). Task-induced fatigue, operator engagement and the control of attention. [PowerPoint slides]. Retrieved from https://www.sintef.no/globalassets/project/hfc/documents/hfc-forum-bergen-fatigue.pdf.
Matthews, G. S., Campbell, S. E., & Falconer, S. (2000). Personality, energetics and individual differences in motivation. In Proceedings of the tenth European conference on personality, Krakow, Poland, July 2000.
Matthews, G. S., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., ... Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress and worry. *Emotion, 2*, 315–340. https://doi.org/10.1037//1528-3542.2.4.315.
Matthews, G. S., Szalma, J., Panganiban, A. R., Neubauer, C., & Warm, J. S. (2013). Profiling task stress with the Dundee Stress State Questionnaire. In L. Cavalcanti & S. Azevedo (Eds.), *Psychology of Stress* (pp. 49–91). Nova Science Publishers.
McDowell, K., Nunez, P., Hutchins, S., & Metcalfe, J. S. (2008). Secure mobility and the autonomous driver. *IEEE Transactions on Robotics, 24*, 688–697. https://doi.org/10.1109/TRO.2008.924261.
McKenzie, B., & Rapino, M. (2011). Commuting in the United States: 2009. American Community Survey Reports, US Census Bureau. September, 2011.
Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups. *Human Factors, 54*, 396–412. https://doi.org/10.1177/0018720812442086.
Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors, 43*, 519–528. https://doi.org/10.1518/001872001775870421.
Naujoks, F., Purucker, C., & Neukum, A. (2016). Secondary task engagement and vehicle automation – Comparing the effects of different automation levels in an on-road experiment. *Transportation Research Part F: Traffic Psychology and Behaviour, 38*, 67–82. https://doi.org/10.1016/j.trf.2016.01.011.
Neubert, L., Santen, L., Schadschneider, A., & Schreckenberg, M. (1999). Single-vehicle data of highway traffic – a statistical analysis. *Physical Review E, 60*, 6480–6490. https://doi.org/10.1103/PhysRevE.60.6480.
NHTSA (2017). Automatic vehicle control systems. Investigation PE 16-007. Retrieved from https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.pdf.
Omae, M., Hashimoto, N., Sugamoto, T., & Shimizu, H. (2005). Measurement of driver's reaction time to failure of steering controller during automated driving. *Review of Automotive Engineering, 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. *Technological Forecasting and Social Change, 87*, 51–59. https://doi.org/10.1016/j.techfore.2013.10.018.
Petermeijer, S. M., Cieler, S., & De Winter, J. C. F. (2017). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis and Prevention, 99*, 218–227. https://doi.org/10.1016/j.aap.2016.12.001.
Petermeijer, S., Bazilinskyy, P., & De Winter, J. C. F. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204–215. https://doi.org/10.1016/j.apergo.2017.02.023.
Saxby, D. J., Matthews, G., 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. *Journal of Experimental Psychology: Applied, 19*, 287–300. https://doi.org/10.1037/a0034386.
Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis and Prevention, 41*, 1087–1093. https://doi.org/10.1016/j.aap.2009.06.007.
Schwalk, M., Kalogerakis, N., & Maier, T. (2015). Driver support by a vibrotactile seat matrix – Recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving. In Proceedings of the 6th international conference on applied human factors and ergonomics (AHFE2015), and the affiliated conferences (pp. 2466–2473). Las Vegas, NV. http://dx.doi.org/10.1016/j.promfg.2015.07.507.
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. https://doi.org/10.5298/1081-5937-41.3.04.
Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT Press.
Solís-Marcos, I., Ahlström, C., & Kircher, K. (2018). Performance of an additional task during Level 2 automated driving: An on-road study comparing drivers with and without experience with partial automation. *Human Factors, 60*, 778–792. https://doi.org/10.1177/0018720818773636.
Song, M., & Wang, J.-H. (2010). Studying the tailgating issues in Rhode Island and its treatment. Journal of the Transportation Research Forum, 49, Arlington, VA. http://dx.doi.org/10.5399/osu/jtrf.49.3.2581.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*, 137–149. https://doi.org/10.3758/BF03207704.
Stanton, N. A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles. *Safety Science, 24*, 35–49. https://doi.org/10.1016/s0925-7535(96)00067-7.

Stanton, N. A., Young, M., & McCaulder, B. (1997). Drive-by-wire: The case of mental workload and the ability of the driver to reclaim control. *Safety Science, 27*, 149–159. https://doi.org/10.1016/s0925-7535(97)00054-4.

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.

Suriya-Prakash, M., John-Preetham, G., & Sharma, R. (2015). Is heart rate variability related to cognitive performance in visuospatial working memory? *PeerJ PrePrints, 3*, e1711.

Tanner, W. P., Jr., Wilson, P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401–409. https://doi.org/10.1037/h0058700.

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. https://doi.org/10.1161/01.CIR.93.5.1043.

Teichner, W. H. (1974). The detection of a simple visual signal as a function of time of watch. *Human Factors, 16*, 339–353. https://doi.org/10.1177/001872087401600402.

Treiber, M., Kesting, A., & Helbing, D. (2006). Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps. *Physical Review E, 74*, 016123. https://doi.org/10.1103/PhysRevE.74.016123.

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. 609612) https://doi.org/10.1109/CIC.2015.7410984.

Walker, F., Boelhouwer, A., Alkim, T., Verwey, W. B., & Martens, M. H. (2018). Changes in trust after driving level 2 automated cars. *Journal of Advanced Transportation*. https://doi.org/10.1155/2018/1045186.

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human Factors, 66*, 870–881. https://doi.org/10.1177/0018720818768199.

Young, M. S. (2000). *Attention, automaticity, and automation: New perspectives on mental underload and performance* Doctoral dissertation. UK: University of Southampton.

Young, M. S., & Stanton, N. A. (2004). Taking the load off: Investigations of how adaptive cruise control affects mental workload. *Ergonomics, 47*, 1014–1035. https://doi.org/10.1080/00140130410001686348.

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. https://doi.org/10.1080/00140130701318855.