Transportation Research Part F: Psychology and Behaviour 94 (2023) 254–267

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

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

Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving

Vishnu Radhakrishnan [a,*], Tyron Louw [a], Rafael Cirino Gonçalves [a], Guilhermina Torrao [a,1], Michael G. Lenné [b,c], Natasha Merat [a]

[a] Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, United Kingdom
[b] Seeing Machines Ltd., Canberra 2609, Australia
[c] Accident Research Centre, Monash University, Victoria 3800, Australia

A R T I C L E I N F O

Keywords:
Mental Workload
Car-following
Takeovers
Pupil diameter
Gaze
Visual attention

A B S T R A C T

This Horizon2020-funded driving simulator-based study on automated driving investigated the effect of different car-following scenarios, and takeover situations, on drivers' mental workload, as measured by eye tracking-based metrics of pupil diameter and self-reported workload ratings. This study incorporated a mixed design format, with 16 drivers recruited for the SAE Level 2 (L2; SAE International, 2021) automation group, who were asked to monitor the driving and road environment during automation, and 16 drivers in the Level 3 (L3) automation group, who engaged in a non-driving related task (NDRT; Arrows task) during automation. Drivers in each group undertook two experimental drives, lasting about 18 min each. To manipulate perceived workload, difficulty of the driving task was controlled by incorporating a lead vehicle which maintained either a Short (0.5 s) or Long (1.5 s) Time Headway (THW) condition during automated car-following (ACF). Each ACF session was followed by a subsequent request to takeover, which happened either in the presence or absence of a lead vehicle. Results from standard deviation of pupil diameter values indicated that drivers' mental workload levels fluctuated significantly more when monitoring the drive during L2 ACF, compared to manual car-following (MCF). Additionally, we found that drivers' mental workload, as indicated by their mean pupil diameter, increased steeply around takeovers, and was further exacerbated by the presence of a lead vehicle during the takeovers, especially in the Short THW condition, for both groups. Pupil diameter was found to be sensitive to subtle variations in mental workload, and closely resembled the trend seen in self-reported workload ratings. Further research is warranted to assess the feasibility of using eye-tracking-based metrics along with other physiological sensors, especially in real-world settings, to understand whether they can be used as real-time indicators of drivers' mental workload, in future driver state monitoring systems.

1. Introduction

With the recent push towards more automated control of vehicles, human-factors research into how this affects driver state has also

* Corresponding author.
E-mail addresses: V.Radhakrishnan@leeds.ac.uk (V. Radhakrishnan), T.L.Louw@leeds.ac.uk (T. Louw), tsrg@leeds.ac.uk (R. Cirino Gonçalves), guilhermina.torrao@vtt.fi (G. Torrao), mike.lenne@seeingmachines.com (M.G. Lenné), N.Merat@its.leeds.ac.uk (N. Merat).
[1] VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Finland.

https://doi.org/10.1016/j.trf.2023.02.015
Received 25 April 2022; Received in revised form 22 February 2023; Accepted 26 February 2023
Available online 3 March 2023

gained momentum. However, we are still a long way from achieving full autonomy of the driving task, where drivers are not required to intervene at all. Especially for automation levels that require some form of driver intervention, such as SAE Level 2 (L2) or Level 3 (L3; SAE International, 2021), drivers are required to have appropriate "readiness levels" (Gold et al., 2016; Zeeb et al., 2016) to safely resume control of the vehicle, when the system reaches limiting criteria, based on its Operational Design Domain, or when an unexpected event/fault causes the automated system to relinquish control to the driver (Mioch et al., 2017; Parasuraman et al., 2008).

Inappropriate levels of driver workload (underload or overload) can influence drivers' performance, safety, and readiness, especially when they have to resume manual control of an automated vehicle (Dogan et al., 2019; Parasuraman et al., 2008). Mental workload is described as the relation between the mental resources demanded by a task/activity, and an individual's information processing capacity, or the mental resources that are available to be supplied to the driver (de Waard, 1996; Parasuraman et al., 2008). Mental workload and task performance follow a complex and 'inverted U-shape' relationship (Bruggen, 2015; de Waard, 1996). Drivers have limited cognitive resources at their disposal, including a central resource pool that is used to perform all tasks (Kahneman, 1973), and additional multiple resources (such as visual or auditory), that are utilised, based on the task demand and modality (Wickens, 1984). The size of this resource pool can vary in capacity, based on the task demand (Young & Stanton, 2002). For example, exceptionally low workload levels (underload), resulting from low task demand, and/or the monotony imposed by automation, can reduce drivers' vigilance, or ability to maintain sustained attention (Young & Stanton, 2002). This also reduces drivers' ability to detect or perceive risk/hazards, resulting in deteriorating performance of the driving task, if and when drivers are required to resume control of the vehicle (Heikoop et al., 2019). Similarly, very high workload (overload) and increased demand from a competing task, such as engaging in a demanding non-driving related task (NDRT), can also result in a performance decrement, because drivers' physical and mental attention is taken away from the primary driving task (Gold et al., 2015; Zeeb et al., 2016). Therefore, in order to successfully resume control from automation and maintain appropriate readiness levels, drivers should maintain moderate workload and arousal levels, to reduce the likelihood of driving errors (Bruggen, 2015; de Waard, 1996). This paper focuses on understanding how drivers' mental workload is influenced by a series of car-following situations in manual and automated driving, and takeover scenarios after automation, in a simulator-based study on automated driving (SAE L2 and L3), investigating whether eye-based metrics can be used as an objective, non-invasive, indicator of drivers' mental workload.

Studies that have compared drivers' mental workload between monitoring the drive during automation and manual driving, suggest that drivers had similar levels of mental workload, and did not experience automation-induced underload during a monitoring phase of the drive, compared to manual driving (Lohani et al., 2020; Stapel et al., 2019). However, these studies measured average workload over a ~30-minute time window, and relied on subjective or heart-rate based measures. Observation studies within the aviation domain that have used subjective ratings, and a detection response task as performance indicators of workload (Wiener, 1989; Sarter, Woods, & Billings, 1997), have suggested that automation results in an uneven distribution of mental workload, rather than simply reducing or increasing it. In a recent study on workload in unmanned aerial vehicle (UAV) operators, Boehm et al. (2021) observed that mental workload can vary rapidly (in a matter of seconds) as a function of fluctuating task demands, and affects the performance of the operator. However, there are limited studies that have focused on understanding the moment-to-moment or phasic changes in drivers' mental workload (that is, rapid fluctuations in their workload levels) while monitoring the drive during automation, how this is different to fluctuations in mental workload during manual driving, and whether it affects driving performance during the resumption of control from automation.

Another factor influencing drivers' mental workload is car-following scenarios, where changes in the distance maintained from a lead vehicle can affect drivers' attentional demands (Liu et al., 2019; Siebert & Wallis, 2019). Car-following refers to the longitudinal following of a lead vehicle by a driver or automated system, and is a pre-cursor to rear-end collisions, which constitutes over 31 % of all collisions in the US (National Highway Traffic Safety Administration, 2009, p. 56). Time headway (THW), which is the elapsed time between the front of a lead vehicle passing the road, and the front of the ego vehicle passing the same point, has been used as a behavioural measure for understanding car-following (Vogel, 2003). Using a simulator study with manual driving, Liu et al. (2019), found that drivers reported significantly higher workload when the lead vehicle maintained shorter THWs of 1 s or less, compared to longer THWs of 1.5 s or more. Shorter THWs maintained by the ego vehicle have also been associated with higher risk perception (Louw et al., 2020) and discomfort (Siebert & Wallis, 2019), which can also lead to an increase in drivers' workload levels (Beggiato et al., 2019), during automated car-following scenarios. Therefore, to understand how drivers' mental workload was affected by the THW of a lead vehicle during both automated (ACF) and manual car-following (MCF), as well as the takeover period, this study included a series of car-following scenarios with different THWs.

Task switching within the automated driving context, for example, from monitoring the drive in a passive supervisory role, to resuming control of the vehicle in an active role as a driver, can increase driver workload (Monsell, 2003; Wickens et al., 2015). When a takeover request is issued, drivers are required to refocus their attention to the driving environment, and pay more attention to the road and any potential hazards. This sudden change in attentional demand from simply monitoring the drive, to the attention and effort required to perform a successful takeover, can result in a physical and mental overload during takeovers, where the task demand exceeds the resources available to the driver.

Accurate, objective measurement of drivers' mental workload levels can be challenging, due to the high inter-individual variability of how people perceive, and are affected by, different task demands. Nevertheless, from a performance and safety standpoint, it is crucial that we are able to objectively measure and understand variations in driver workload during HAD, as this can inform the automated system about the capabilities and limitations of the driver, especially if they have to resume control of the vehicle. Eventually, the vehicle can establish if the driver is indeed capable of taking back control of the vehicle, or if other measures, such as a Minimum Risk Manoeuvre (Thorn et al., 2018) should be considered to bring the car to a safe stop (Yu et al., 2021).

Skin conductance and heart-based sensors have been used extensively within the driving context, as objective indicators of mental workload (Foy & Chapman, 2018; Radhakrishnan et al., 2022; Reimer & Mehler, 2011), stress (Healey & Picard, 2005) or discomfort (Beggiato et al., 2019; Radhakrishnan et al., 2020). However, as a stand-alone measure, such skin conductance and heart-based measures are unable to identify the causal factor (such as stress, attention, workload, or arousal) that induce physiological changes, and the addition of other sensors, such as eye tracking, can aid our understanding of driver state. Also, skin conductance and heart-based sensors are affected by physical load, and in tasks such as manual driving, which involves both physical and mental load, it would be challenging to distinguish between physical and mental workload, using skin conductance or heart-based sensors alone. Moreover, some heart-based sensors can be intrusive in nature, and require extended preparation, prior to a study. Dash-based eye-trackers can be both non-invasive and unobtrusive (Marquart et al., 2015; Merat et al., 2012; Tsai et al., 2007), and eye tracking-based metrics, such as pupil diameter, blink frequency and blink duration, have been used in the past as indicators of mental workload, where shorter blink durations lead to high blink frequency, and vice versa. The relationship between blink frequency and duration, and the demand from different tasks is not always clear. For example, visual tasks can reduce blink frequency (Veltman & Gaillard, 1996), whereas a non-visual cognitive task (*n*-back) can increase blink frequency (Recarte et al., 2008). Given this ambiguity between blink duration and frequency and their relationship to workload in this paper, we only assessed pupil diameter, and its feasibility as an indicator of real-time driver workload.

Studies have shown that pupil diameter generally increases with an increase in drivers' mental load (Tsai et al., 2007). Mean pupil diameter has been used as an indicator of tonic dilation/constriction, with mean pupil diameter values increasing with an increase in mental load (Appel et al., 2018; Steinhauer et al., 2004), although these are generally 11 % or lower, compared to baseline values (Batmaz & Öztürk, 2008; Stern, 1997). However, one disadvantage of using mean pupil diameter is that it fails to capture the phasic fluctuations in pupil diameter, as the high and low fluctuations can cancel each other out, when taking the mean value of pupil diameter over the entire time series (Buettner, 2013). The standard deviation of pupil diameter has therefore been used as an indicator phasic fluctuations, with higher standard deviation of pupil diameter suggesting a higher fluctuation in drivers' workload levels (Beatty, 1982; Buettner, 2013). Pupil diameter is also known to be affected by individual and environmental differences, with variations in pupil diameter due to changes in light intensity being significantly higher than variations due to mental workload (Mathôt, 2018). For example, pupil diameter can increase in darker conditions to accommodate more light into the eyes (pupil light response, Mathôt, 2018; Spector, 1990). A shift in focus between near objects to those further away (pupil near response or accommodation reflux), can also result in pupil dilation (Mathôt, 2018), while a sudden change in the environment, for example, caused by sounds, movement or touch (orienting response), can also cause a small increase in magnitude or duration of pupil dilation, within 0.5 s to 1 s of stimulus onset (Mathôt, 2018).

1.1. *Current study*

This study was funded by the European Commission Horizon 2020 program under the L3Pilot project, grant agreement number 723051, and investigated how drivers' mental workload varied during different stages of manual and automated driving, and whether the changes in drivers' workload levels were reflected in their pupil diameter values. This study is also discussed in Radhakrishnan et al. (2022), where the authors investigated the effect driver workload on heart-rate and skin conductance-based physiological signals. However, the current paper differed from Radhakrishnan et al. (2022) in terms of the different segments and time windows used in analysis, as well as in the research questions, owing to the differences in the nature of metrics used. Using a mixed design, drivers were asked to monitor the drive during automation (L2) or engage in an NDRT (L3). Driver workload was also manipulated using two THW conditions, in a series of car-following scenarios (Short and Long), and two different types of takeovers, that happened either in the presence or absence of a lead vehicle (Lead/No Lead). To understand how drivers' mental workload varied during the monitoring phase, we compared drivers' mental workload levels during automated car-following (ACF) and manual car-following (MCF). To understand how drivers' mental workload varied at different points during the transition of control from automation to manual driving, we analysed drivers' mental workload levels before (Pre-Takeover), during (Takeover) and after resumption of control (Post-Takeover). We also investigated whether the presence of a lead vehicle, or shorter headway conditions, increased their mental workload during takeovers. Additionally, we investigated whether prior engagement in an NDRT (L3), as opposed to monitoring the drive during automation (L2), affected drivers' mental workload during the takeover. In addition to eye-tracking metrics such as pupil diameter, drivers were also asked to report their perceived workload using a subjective workload ratings scale (1–10, with 10 being extremely high workload). The research questions addressed by this paper were:

1. Does drivers' mental workload vary between automated car-following (ACF) and manual car-following (MCF), and is this reflected in drivers' pupil diameter values and subjective workload ratings, in the L2 group (ACF vs MCF),?
2. Is drivers' mental workload during the takeover different for the different takeover windows (Pre-Takeover, Takeover, and Post-Takeover), and whether this is affected by the presence of a lead vehicle (Lead vs No Lead), in the L2 group?
3. Is drivers' mental workload during the takeover affected by the THWs maintained by the automated controller (Short vs Long)?
4. Is drivers' mental workload during the takeover affected by the presence of a lead vehicle during the takeover (Lead vs No Lead)?
5. Is drivers' mental workload during the takeover affected by prior engagement in an NDRT during automation (L2 vs L3)?

2. Materials and methods

2.1. Participants

For this driving simulator study, a total of 32 participants were recruited, with 16 participants each in the L2 and L3 groups. However, the data from 3 participants from each group were discarded due to either missing data, or because they did not follow the instructions of the study. The remaining 13 participants (4 females, 9 males) in the L2 group had a mean age of 42 ± 17 years, with a mean driving experience of 22 ± 16 years. For the L3 group, the 13 participants (3 females, 10 males) had a mean age of 33 ± 8 years, with a mean driving experience of 14 ± 8 years. In accordance with the rules and regulations of the University of Leeds ethics committee (LTTRAN-054), all participants gave prior consent to take part in the study. Additionally, they were compensated with £25 upon completion of the experiment.

2.2. Apparatus

The University of Leeds Driving Simulator (UoLDS), which is a full motion-based driving simulator, was used to conduct the experiment. UoLDS consists of a 4 m diameter spherical projection dome, which has a 300° field of view projection system, and also includes a Jaguar S-type cab, housed within the dome. The electrical motion system for the simulator has 8 degrees of freedom, and consists of a 500 mm stroke-length hexapod motion platform, carrying the 2.5 T payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees of freedom of the Cartesian inertial frame. The electrical motion platform is mounted on a railed gantry, which further allows 5 m of effective motion in surge and sway. Drivers' eye-tracking data was recorded using Seeing Machines FOVIO eye-tracking hardware system and Seeing Machines PC-DMS software package.

When engaged, the Automated Driving System (ADS) used in this study was designed to control both lateral and longitudinal aspects of the driving task, driving at a constant speed of 40 mph. When the control of the vehicle was with the driver during manual driving, and the ADS was inactive, the HMI interface on the dashboard displayed a red steering icon (Fig. 1a), When the ADS was activated, the HMI interface displayed a green steering icon (Fig. 1b).

2.3. Study design

This study has already been presented in Radhakrishnan et al. (2022). A mixed design was incorporated in this study, with between-participant factor of Level of Automation (L2, L3) and a within-participant factor of Time Headway (Short, Long), Drive Mode (ACF, MCF) and Lead Vehicle (Lead, No Lead). With the exception of Drive Mode, all the other factors were counterbalanced.

Each participant undertook a ~ 10-minute practice drive, to become familiar with the simulator environment, and the driving controls. After the practice drive, each participant experienced two experimental drives, with a 5-minute break in between. All the experimental drives were performed in a single-carriageway urban environment, with a 40 mph speed limit, and low-density oncoming traffic. The first experimental drive consisted of free driving for ~ 2 min, following which a lead vehicle joined the driving lane. The participants were instructed to follow the lead vehicle as they normally would in such a scenario, without overtaking, for about 5 min. This was termed the manual baseline drive, and was used to collect drivers' baseline car-following behaviour, prior to any experimental manipulations. With the exception of the manual baseline drive, which was only present in the first experimental drive, the two experimental drives were similar. Each experimental drive consisted of 4 segments in the following order: Automated drive 1, Manual drive 1, Automated drive 2, and Manual drive 2 (Fig. 2).

At the end of the manual baseline drive, the lead vehicle exited the driving lane at an intersection, following which drivers experienced free driving for ~ 1 min. At the end of the free drive, participants were prompted with an auditory handover message, requesting them to engage the ADS: "Attention, engage automation". Drivers engaged the ADS using a button on the steering wheel.

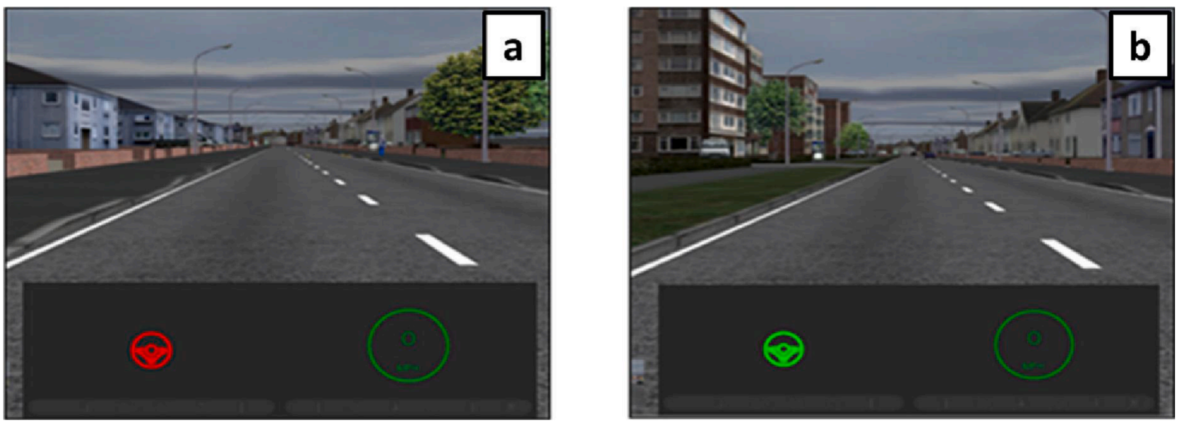

Fig. 1. HMI Interface on the dashboard: **(a)** when automation was disengaged; **(b)** Automation was engaged (from Radhakrishnan et al., 2022).

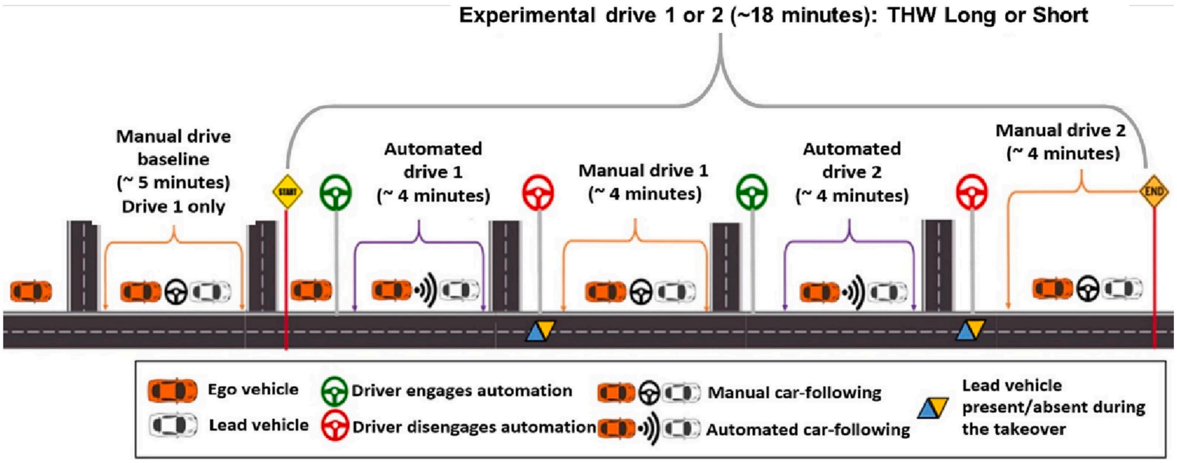

Fig. 2. Schematic representation of the experimental drives (from Radhakrishnan et al. 2022).

After about 1 min into the automated drive, a lead vehicle joined the driving lane, starting what we termed the automated car-following (ACF) segment. During the ACF segment, participants were exposed to one of two THW conditions (0.5 s for Short THW and 1.5 s for Long THW), in a counterbalanced order across the two experimental drives.

The end of an ACF segment was followed by an auditory takeover request: “Attention, get ready to takeover”. The takeover request was issued when the ADS reached a system limitation criterion, linked to faded lane markings on the road. The takeover request was followed by a short acoustic tone (1000 Hz), lasting 0.2 s, with increasing frequency (by a factor of 1.2 after each cycle), until the driver resumed manual control. To disengage the ADS, drivers pulled the left-hand indicator stalk, or rotated the steering wheel by more than 2°, or pressed the accelerator or brake pedals.

There were two types of takeover situations in this study, one with a lead vehicle present during the takeover (Lead), and the other without a lead vehicle when the takeover request was issued (No Lead). Each of the two experimental drives included the two lead vehicle conditions, with participants experiencing them after automation in both drives. The order in which participants experienced each of the two takeover situations was counterbalanced. In the Lead condition, the lead vehicle remained in the driving lane, when the takeover request was issued, and the participants took control of the vehicle. In the No Lead condition, the lead vehicle exited the driving lane at an intersection, about 5 s before the takeover request was given. About 10 s after the participants took control of the vehicle, a lead vehicle joined the driving lane, starting the manual car-following segment (MCF). Each experimental drive included an ACF, a Takeover and an MCF, repeated twice, in that order.

In this study, the Level of Automation determined whether the driver had to monitor the road during automated driving (L2), or could engage in an NDRT (L3). The Arrows task (Jamson & Merat, 2005) was used as the NDRT in this study. Participants were required to select an upward facing arrow, from a 4x4 grid of arrows, as shown in Fig. 3a. The arrows were displayed on a touchscreen, placed near the centre console and the gearshift, as shown in Fig. 3b. The screen also displayed participants’ cumulative score (each correct selection awarded them with one point) as well as a “score to beat”. To ensure that participants fully engaged in the Arrows Task, the participant briefing sheet offered an additional £5 reward to anyone who was able to beat the “score to beat”. However, for ethical reasons, every participant was paid this extra £5. Drivers in the L2 group were instructed to monitor the drive, when the ADS was engaged, with their hands off the steering wheel and foot off the pedals.

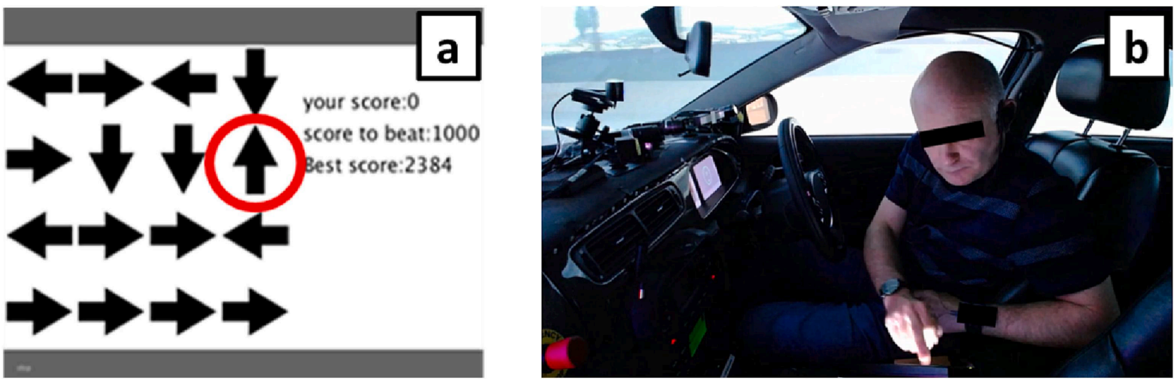

Fig. 3. **(a)** A representation of the arrows task with the upward facing arrow circled in red; **(b)** A participant engaging in arrows task in L3 group during automation.

2.4. Self-reported workload ratings

In addition to eye-tracking based metrics, we collected participants' subjective workload ratings, with each participant asked to rate their level of perceived workload, 13 times across the two drives (7 times in the first experimental drive, including once during the manual baseline drive, and 6 times in the second experimental drive). When prompted, participants rated their workload verbally, on a scale of 1–10, with 10 corresponding to the highest workload.

2.5. Procedure

Participants were briefed about the study upon arrival, after which they were requested to sign the consent form, and ask any questions they had about the study. After they signed the consent form, they entered the driving simulator for the practice drive. The practice drive included automated and manual driving, as well as a takeover situation. Participants were shown how to control the vehicle, engage and disengage the ADS, and provided information about the HMI. Participants in the L3 group were also given an opportunity to practise the Arrows Task. The practice drive was followed by the two experimental drives, lasting about 18 min each.

During the experimental drives, whenever participants were engaged in manual driving, they were advised to adhere to the speed limit of 40 mph, and asked to avoid overtaking the lead vehicle. A ~ 10-minute break followed each of the experimental drives, allowing participants to answer a set of questions about the experimental drive, including Arnett's Sensation Seeking Questionnaire (Arnett, 1994), traffic locus of control (Özkan & Lajunen, 2005) and the Driver Style Questionnaire (French et al., 1993), see Louw et al. (2020). The scope of the questionnaire data is beyond the objectives of this paper, and as such, were not included in our analysis.

2.6. Data analysis

The eye-tracking metrics were pre-processed using Seeing Machines' PC-DMS system, and all the data segmentation and plotting was done using MATLAB R2016a. Mean values were used to investigate drivers' overall mental workload levels as depicted by eye-based metrics, and standard deviations were used to show the variation or spread in mental workload, for each participant, across ~ 4-minute time windows of ACF and MCF. The time window for ACF was considered as the time from which the lead vehicle entered the driving lane after the drivers engaged automation, until the takeover request. The MCF time window started 10 s after taking over manual control of the vehicle, until ADS was re-engaged. As stated earlier, there were two ACF segments and two MCF segments, in each experimental drive (see Fig. 2). A set of t-tests on drivers' mean pupil diameter, and self-reported workload ratings, did not reveal any statistically significant differences between the two ACF and the two MCF segments, within each of the experimental drives. Therefore, the data for the two similar ACF and MCF segments were aggregated to a single representation, for both eye-metrics and self-reported ratings of workload, in the L2 group.

To analyse how drivers' mental workload varied around the takeover, we segmented the eye-metrics into 3 distinct time windows: Pre-Takeover, Takeover and Post-Takeover (Fig. 4). Only drivers in the L2 group were considered in this analysis, as drivers in the L3 group were engaged in the Arrows tasks, and their eyes were not captured by the eye tracker. The Pre-Takeover window was the time window from 10 s before the takeover request, until the takeover request was given. Takeover was the time window from when the takeover request was given, until the driver resumed manual control of the vehicle. This Takeover time window varied for each participant. Post-Takeover was the time window from when the driver resumed manual control of the vehicle, until 10 s after they resumed manual control. This time window was chosen because previous studies have shown that the peak in driving performance decrement was observed within 10 s after takeover of control (Merat et al., 2014).

To understand how drivers' mental workload levels during takeovers were affected by the presence of a lead vehicle (RQ 4), the THW maintained by this lead vehicle (RQ 3), and engagement in an NDRT during automation (RQ 5), we aggregated the eye-metrics

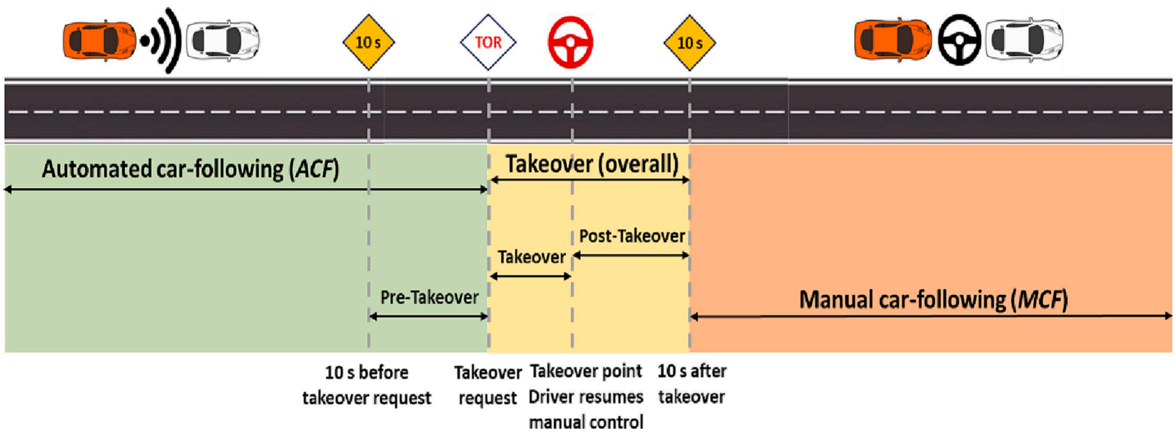

Fig. 4. Schematic depicting different takeover windows.

across the Takeover and Post-Takeover windows, into a single window labelled Takeover (overall) window. This analysis was done across drivers in the L2 and L3 groups, as drivers in the L3 group redirected their eyes onto the road environment, which was within the field of view of the eye-tracker, upon hearing the takeover request. The time from which a takeover request was given, until 10 s after they resumed manual control of the vehicle was considered as the overall takeover window (Fig. 4, yellow segment). Since pupil diameter has a rapid signal decay time, and is sensitive to instantaneous fluctuations in drivers' mental workload (Buettner, 2013; Mathôt, 2018), it is less likely to be influenced by previous workload inducing stimuli. Therefore, during the two No Lead takeover conditions in this study, presence of a lead vehicle in either Short or Long THW condition during automation, would not have influenced drivers' mean pupil diameter values at a later stage during the takeover. Additionally, paired-sample t-tests between the two No Lead takeover conditions did not reveal any significant differences, for either mean pupil diameter or self-reported workload ratings, across the L2 and the L3 groups. Hence, the data for the two No Lead conditions were combined to form a single representation, and labelled Infinite THW.

We computed both mean and standard deviation values of pupil diameter, during ACF and MCF segments. However, only mean values were used when comparing performance in the takeover segments. This was because the ACF and MCF segments were comparatively longer (~4 min) compared to the relatively short Takeover windows (maximum window width of ~ 15 s), and therefore, any fluctuations (high and low) in phasic pupil diameter values would have cancelled each other out, when taking only mean values. Standard deviation of pupil diameter is indicative of fluctuations in pupil diameter, and therefore, fluctuations in drivers' mental workload levels (Buettner, 2013). Table 1 below highlights the different research questions addressed in this paper and the corresponding data segments considered for analysis.

2.7. Statistical analysis

IBM SPSS Statistics 26 was used to conduct all the statistical tests in this study. Shapiro Wilk's test revealed that the majority (greater than90 %) of the group-level estimates were normally distributed for each of the dependent variables used, for every ANOVA test we conducted. For statistical significance, an α-value of 0.05 was used as a limiting criterion, and partial eta-squared was computed as an effect size statistic. Mauchly's test was used to test the sphericity of data. In cases where the data was not spherical, the degrees of freedom were corrected using Greenhouse-Geisser corrections. Levene's test was used to test whether the data was homogenous. Pair-wise comparisons were used to determine differences between the different factors.

Only 50 % of the data from self-reported workload ratings were normally distributed. However, the skewness and kurtosis of the non-normally distributed data were within the acceptable range +/- 2 (George & Mallery, 2010). The repeated measures ANOVA was robust enough to accommodate this violation of normality and homogeneity, with only a small effect on Type I error (Blanca et al., 2017; Pituch & Stevens, 2016).

3. Results

3.1. The effect of drive Mode on driver workload in the L2 group (RQ 1)

To understand how drivers' mental workload during monitoring the drive in automated car-following (ACF) compared to manual car-following (MCF), we performed a one-way repeated measures ANOVA, with within-participant factor of Drive Mode (ACF, MCF), on drivers' mean pupil diameter, standard deviation of pupil diameter, and self-reported workload ratings, for the L2 group. The use of Time Headway (THW) as a factor was excluded in these analyses, because drivers controlled their own THW in the MCF condition, as opposed to the fixed values imposed during ACF. Additionally, paired sample t-tests revealed no significant differences in pupil diameter values, or self-reported workload ratings, across the Short and Long THW conditions, during ACF. Therefore, we combined the values for the Short and Long THW conditions into a single representation, for both ACF and MCF. Since the ACF and MCF segments were about 4 min in length, we analysed the mean and standard deviation of pupil diameter, to understand if there were any overall differences and/or variations in mental workload within the ~ 4-minute segment.

Results showed no significant main effect of Drive Mode on the mean pupil diameter values (Table 2). This may be because the time window used for analysis of data was relatively long (~4 min), such that the phasic fluctuations in pupil diameter might have cancelled out when using the mean values for this metric. Previous research has shown that the standard deviation of pupil diameter provides a better indication of the phasic and dynamic aspect of pupil dilation, and is therefore a better measure for highlighting fluctuations in

Table 1
Research questions and corresponding data segments used in analysis.

Research question	Data segment (see Fig. 4)	Automation group	Remark
RQ 1	ACF and MCF	L2 group	Data for ACF and MCF was aggregated by averaging the pupil diameter values for two ACF and MCF conditions, and across Short and Long THW conditions.
RQ 2	Pre-Takeover, Takeover, Post-Takeover	L2 group	Data for Pre-Takeover, Takeover and Post-takeover windows for both Lead and No Lead condition, was analysed using separate ANOVAs for Short and Long THW conditions.
RQ 3, 4 and 5	Takeover (overall)	L2 and L3 group	Data for the two No Lead conditions, across Short and Long THW experimental drives, was aggregated as a single representation (Infinite THW), and compared across the two levels of automation.

Table 2
Effect of Drive Mode on drivers' pupil diameter, and self-reported workload ratings, in the L2 group.

Predictor	df1	df2	*F*	*p*	η_p^2
1. Pupil diameter (mean)	1	12	3.01	0.108	0.200
2. Pupil diameter (standard deviation)	1	12	59.09	<0.001	0.831
3. Self-reported workload ratings	1	9	0.421	0.533	0.045

mental workload (Buettner, 2013). As seen in Table 2, our results showed a significant main effect of Drive Mode on the standard deviation of drivers' pupil diameter, with drivers showing significantly higher standard deviation of pupil diameter during ACF, compared to MCF (Fig. 5, Fig. 6).

To further understand why pupil diameter values fluctuated more during ACF, compared to MCF, we visualised drivers' raw gaze data during ACF and MCF, across all participants and all drives, in the L2 group, using a 3D gaze contour plot. Results from the gaze contour plot suggest that there was a larger spread in gaze, across the driving scene, with drivers looking around more (drivers' gaze were likely directed towards the driving scene, HMI interface and right mirror), when monitoring the drive during ACF in the L2 group (Fig. 7a). However, when they were in control of the vehicle during MCF, their gaze was mostly concentrated around the centre of the driving scene, or approximately around the road centre area, suggesting they were more attentive to the driving task while being engaged in Manual driving (Fig. 7b).

Results from drivers' self-reported workload ratings did not reveal any significant effect of Drive mode, suggesting that drivers experienced similar levels of perceived workload, across both ACF and MCF, in the L2 group. However, there was only one workload rating for the entire duration of ACF or MCF, and hence, we could not compare the moment-to-moment changes in perceived workload, during the whole ACF or MCF period, using self-reported workload ratings.

3.2. *Changes in workload during different stages of a takeover in the L2 group (RQ 2)*

To understand how drivers' mental workload varied around the actual takeovers in the L2 group, and whether the presence of a lead vehicle affected this workload, we performed two 3x2 repeated measures ANOVAs with within-participant factors of Takeover Window (Pre-Takeover, Takeover, Post-Takeover) and Lead Vehicle (No Lead, Lead), on drivers' mean pupil diameter. A separate ANOVA was conducted for the Short and Long THW conditions, as adding THW as a factor would have been inaccurate because drivers did not experience any THW conditions, during the No Lead condition. Self-reported workload ratings were not captured separately during the three Takeover Windows considered in this analysis, and therefore, not included in the analysis.

There was a main effect of Takeover Window on drivers' mean pupil diameter in both the Short and Long THW conditions (Table 3). Given each of these takeover windows were under 10 s, standard deviation of pupil diameter was not analysed here, as mean pupil diameter was sufficient to reflect drivers' workload levels during such shorter time windows. Post-hoc tests revealed that drivers' mean pupil diameter increased significantly from the Pre-Takeover time window to the Takeover time window, as well as from the Pre-Takeover time window to the Post-Takeover time window, for both the Short and Long THW conditions (Fig. 8a and Fig. 8b). Additionally, the mean pupil diameter increased significantly from the Takeover time window to the Post-Takeover time window, for the Short THW condition. Taken together, our results suggest that drivers' mental workload increased during the takeover event, increasing from the Pre-Takeover time window to the Post-Takeover time window, across both the Short and Long THW conditions. There were no other main effects or interaction effects.

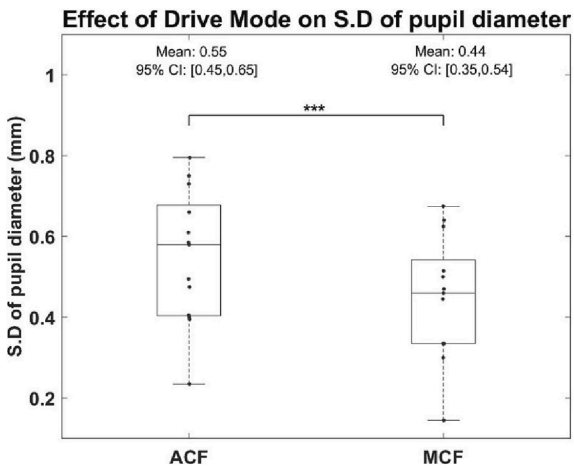

Fig. 5. Effect of Drive Mode (automated car-following or ACF, Manual car-following or MCF) on standard deviation of pupil diameter, in the L2 group. ***p ≤ 0.001.

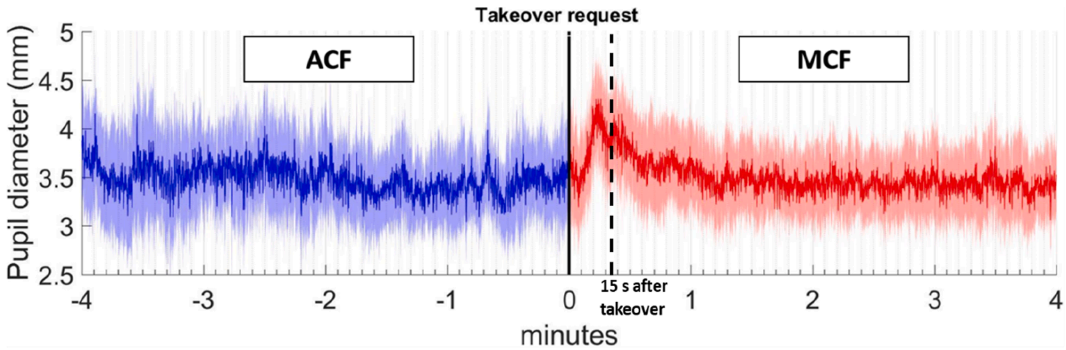

Fig. 6. Variation of drivers' pupil diameters, during automated car-following (ACF) and Manual car-following (MCF), in the L2 group. The darker blue and red lines denote mean values across all drivers in the L2 group, and the lighter blue and pink regions denote the 95 % confidence interval bands. Note that the MCF window used in the analysis only starts 10 s after drivers resumed manual control (~15 s after takeover request is issued), to filter out any variations due to the takeover.

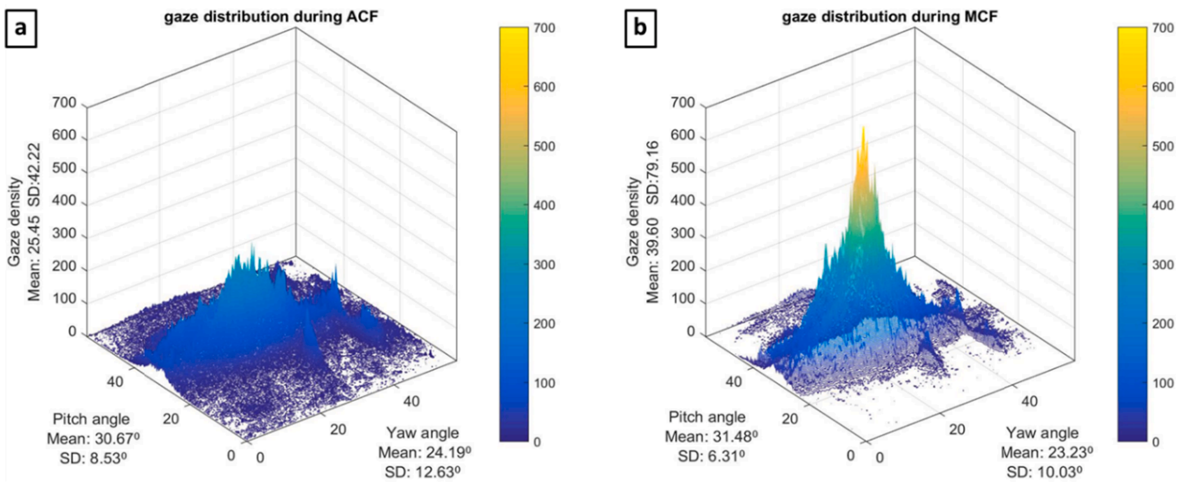

Fig. 7. 3D gaze contour plots across **(a)** Automated car-following (ACF) and **(b)** Manual car-following (MCF), in the L2 group. Colour-bar scale indicates number of gaze points in a particular bin, with the size of each square bin used to create the contour grid being 0.285°.

Table 3
Effect of Takeover Window on mean pupil diameter values, during Short and Long THW conditions, in the L2 group.

Predictor	df1	df2	*F*	*p*	η_p^2
Effect of Takeover Window					
1. Mean pupil diameter (Short THW)	2	24	26.95	<0.001	0.692
2. Mean pupil diameter (Long THW)	1	12	35.78	0.015	0.386
Effect of Lead Vehicle					
1. Mean pupil diameter (Short THW)	1	12	0.064	0.805	0.005
2. Mean pupil diameter (Long THW)	1	12	1.365	0.265	0.102

3.3. Effect of time Headway, lead vehicle and Level of automation on driver workload during takeovers (RQ 3, RQ4 and RQ 5)

To understand how drivers' mental workload was affected by the presence of a lead vehicle during the takeover, and its Time Headway, and whether their workload level during the takeover was affected by their prior engagement in an NDRT during automation, we performed a 3x2 mixed ANOVA on drivers' mean pupil diameter and self-reported workload ratings, using a within-participant factor of Time Headway (Short, Long, Infinite) and a between-participant factor of Level of Automation (L2, L3). As mentioned in **section 2.6**, the takeover window for this analysis is considered as the time from which the takeover request was given, to 10 s after they resumed manual control of the vehicle. For the Short and Long THW conditions, only the takeovers where a lead vehicle was present during the takeover, in either the Short or Long THW condition, was included. Additionally, the two No Lead conditions were consolidated to a single presentation labelled Infinite THW, as explained in **section 2.6**. We could not analyse the L3

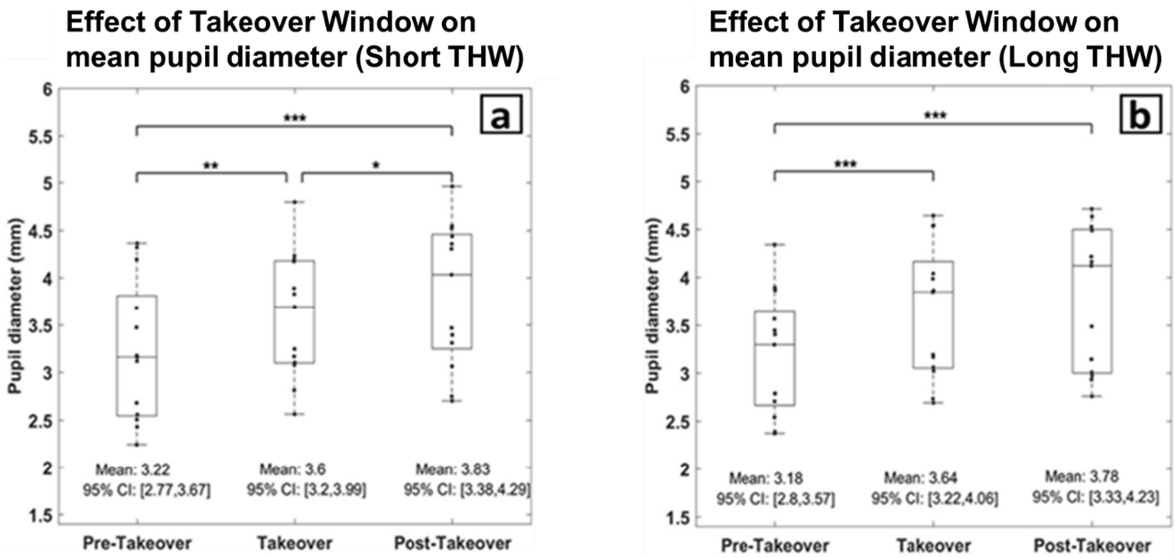

Fig. 8. Effect of Takeover Window on pupil diameter values during **(a)** Short and **(b)** Long THW conditions. *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001.

drivers' eye-metrics before the takeover request was issued (such as the Pre-Takeover window seen in **section 3.2**) as they were engaged in the Arrows task, and, hence their eyes were occluded from the field of view of the fixed-base eye tracker used in this study.

Results showed a main effect of the Time Headway condition on drivers' mean pupil diameter values, and self-reported workload ratings (Table 4, Fig. 9). Post-hoc tests revealed that drivers exhibited significantly higher mental workload levels in the Short THW condition, compared to when there was no Lead vehicle present during the takeover (Infinite THW), as revealed by both mean pupil diameter values ($p < .001$, Fig. 9a) and self-reported workload ratings ($p = .007$, Fig. 9b). Although statistically insignificant, drivers had higher mean pupil diameter values during the takeover in the Short THW condition, compared to the Long THW condition ($p = .098$). Drivers also reported higher workload levels during takeover in the Short THW condition, compared to the Long THW condition ($p = .033$, Fig. 9b). There were no other main effects or interaction effects on drivers' mean pupil diameter values, or self-reported workload ratings.

4. Discussion and conclusions

This study investigated changes in drivers' mental workload during a series of car-following situations in manual and automated driving, and examined how factors such as presence of a lead vehicle during the takeover, the Time Headway (THW) maintained by the lead vehicle, and prior engagement in an NDRT during automation affected drivers' mental workload. Drivers' self-reported levels of workload, and eye-tracking based metrics, such as mean pupil diameter and standard deviation of pupil diameter, were compared at different stages of automated and manual driving.

Paired-sampled t-tests revealed no significant differences in mean pupil diameter values or self-reported workload ratings, due to the different THW conditions, during the automated car-following (ACF). It is likely that the perceptual difference between the two THW conditions used in this study, was not prominent enough to affect drivers' mental workload levels, especially when they were not in control of the driving task. Additionally, drivers' mean pupil diameter values, and self-reported workload ratings, indicated that they experienced similar levels of overall mental workload across the entire automated drive, (which last around 4 min), when the L2 drivers were simply monitoring the driving environment during ACF, compared to manual car-following (MCF). This is consistent with results from past studies which report similar levels of mental workload, as indicated by physiological data (Lohani et al., 2021; Radhakrishnan et al., 2022), or subjective ratings (Stapel et al., 2019).

However, a comparison of the standard deviation of pupil diameter values in the L2 group suggests that drivers had a higher

Table 4
Effect of Time Headway (including No Lead condition represented as Infinite THW) and Level of Automation on drivers' mean pupil diameter values and self-reported workload ratings, during takeovers.

Predictor	df1	df2	F	p	η_p^2
Effect of Time Headway					
1. Mean pupil diameter	2	48	3.29	0.046	0.120
2. Self-reported workload ratings	2	36	4.02	0.011	0.220
Effect of Level of Automation					
1. Mean pupil diameter	1	24	1.92	0.179	0.074
2. Self-reported workload ratings	1	18	0.01	0.937	0.0004

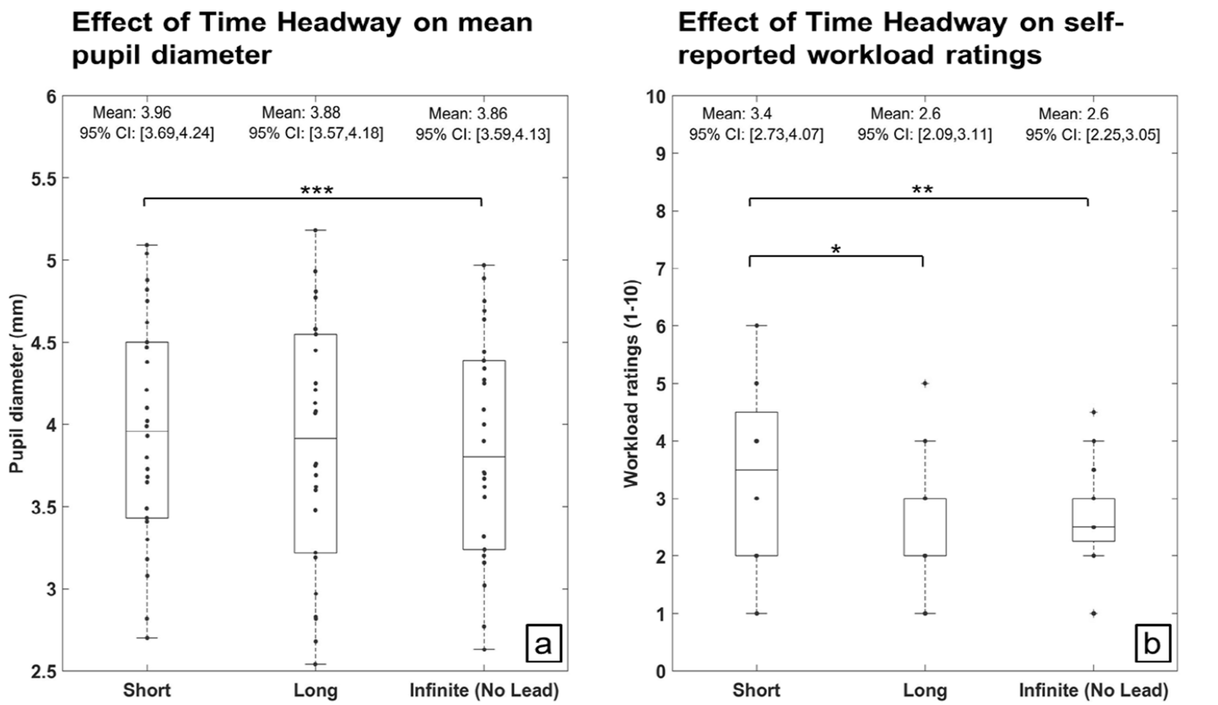

Fig. 9. Effect of Time Headway on driver workload, as reflected in **(a)** mean pupil diameter and **(b)** self-reported workload ratings, with drivers' self-reported workload increasing from 1 to 10. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

moment to moment variation in pupil diameter, when monitoring the drive during ACF, compared to manual driving during MCF. Such fluctuations in pupil diameter could be attributed to a host of factors, such as variations in light intensity (Ellis, 1981), pupil near response, that is, the pupillary response when the pupil constricts or dilates to focus on an object (Mathôt, 2018), a startle response (Mathôt, 2018) or variations in mental workload (Buettner, 2013; Kahneman & Beatty, 1966). To understand this result further, we compared drivers' gaze patterns during the monitoring task (in ACF) and manual driving (MCF). 3D Gaze contour plots revealed that drivers focused their attention more towards the road-centre area when they were in manual control of the vehicle, as would be expected. However, in agreement with previous studies in our lab (Louw, Madigan et al., 2017; Louw, Markula et al., 2017), and that of others (Noble et al., 2021; Schartmüller et al., 2021; Yang et al., 2022), a higher spread in gaze was observed during automation, with drivers looking away from the road centre. The higher spread in gaze during automation could have been caused by drivers looking around the driving scene more and shifting and refocusing their attention between objects near and further away in their field of view, resulting in pupil near response, also known as accommodation reflux, and explains the higher phasic fluctuations seen in our study (Fincham, 1951; Mathôt, 2018). Since the ACF and MCF driving environments in our study were quite similar in terms of lighting conditions, it is unlikely that variations in pupil diameter were due to startle response or variations in light intensity. At present, it is not possible to establish whether these higher fluctuations in pupil diameter during automation were due to variations in mental workload, or the movement of gaze to different regions, or both, as these causal factors are not mutually exclusive. Our results are in agreement with findings of Wiener (1989), who showed uneven distributions of attention for airline pilots during automation, as indicated by subjective ratings, and performance of the detection response task. This uneven distribution of drivers' mental workload during automation (e.g. extreme underload or overload) can be detrimental for a safe takeover (Merat et al., 2014), and highlights the value of accurate driver monitoring systems that can assist drivers with keeping the right level of attention for supervision of the automated system, as is requested by the SAE guidelines (SAE International, 2021).

We also analysed the L2 drivers' mental workload, as indicated by pupil diameter and subjective ratings, across the three takeover windows: Pre-Takeover, Takeover and Post-Takeover. Mean pupil diameter values steadily increased from the Pre-Takeover window to the Post-Takeover window. This clear increase in pupil diameter at each of the three stages, from monitoring the drive, to hearing the takeover request, to taking over shows how drivers' vigilance and attention levels, and likely their workload, is affected as they are asked to come back into the driving loop (Merat et al., 2018), and take on the responsibility of the driving task. Similar results have been observed for skin conductance responses, when workload was found to steadily increase upon the issuance of a takeover request (Du et al., 2020). In our study, drivers' workload levels, as reflected by their pupil diameter values, peaked at around 15 s after they resumed manual control of the vehicle (see Fig. 6). This is in line with driving simulator studies which show a peak in performance decrement around 15 s after a resumption of control from automated driving (Merat et al., 2014), after which the driver is able to stabilise the vehicle (Bueno et al., 2016).

In terms of differences in workload experienced between the two groups, engagement in the Arrows tasks did not seem to affect drivers' workload levels when they resumed control from automation, with similar pupil diameter values seen for the L2 and L3 group at the Pre-takeover stage. Since this study incorporated a non-critical takeover scenario, it is likely that drivers had adequate time to stop engaging in the Arrows task before resuming control of the vehicle, thereby eliminating the effect of any additional workload demands placed by the Arrows task on drivers in the L3 group, by the time they resumed manual control. Our results also indicated the presence of lead vehicle, especially at shorter THWs, significantly increased driver workload during the takeover, when compared to longer THWs, or takeovers without a lead vehicle present (Infinite THW). We did not observe any differences in drivers' mental workload, as reflected by mean pupil diameter values and self-reported workload ratings, between takeovers with a lead vehicle in the Long THW condition and takeovers without a lead vehicle, which suggests that the longer (1.5 s) THW conditions felt more comfortable for drivers during these non-critical takeover scenarios. Given the between subject nature of this design, factors such as age and driving experience could have influenced drivers' pupil diameter and workload levels. However, we found a similar pattern for drivers' skin conductance responses (SCR), which are known to be sensitive to stress and workload in driving (Du, Yang, & Zhou, 2020; Foy & Chapman, 2018), with higher SCRs observed during takeovers that were preceded by a lead vehicle maintaining a short THW (0.5 s), compared to those with longer THWs (1.5 s) (see Radhakrishnan et al., 2022). Overall, these results suggest that drivers' physiological response and self-reported workload ratings are sensitive to the demands of a takeover after automated driving, and especially prior to more (potentially) critical scenarios.

To conclude, our findings suggests that pupil diameter is sensitive to drivers' mental workload levels, and is sensitive to phasic variations in workload. One of the potential limitations of this study is that it was conducted in a controlled driving simulator environment, where the driving scene and its brightness levels were relatively similar throughout the drive. Further research is warranted to understand how pupil diameter is affected by different stages of automated driving in more real world settings. In this study, we were unable to use the eye tracker to objectively capture drivers' mental workload when they were engaged in the Arrows task during automation (L3 group), since drivers were looking away from the eye-tracking cameras (see Fig. 3b**)**. Since the chances of engaging in other activities will increase with higher levels of automation, it is important that drivers' workload levels and attention are monitored during automation, because, according to guidelines from SAE (SAE International, 2021), Level 3 driving still requires drivers to resume control from automation "when required". Therefore, combining eye-tracking measures with physiological sensors, can provide a more comprehensive, accurate and continuous prediction of drivers' mental workload and attention levels, even when drivers' eyes are occluded from the eye-tracking sensors. Additionally, combining eye-tracking data with physiological signals can also help in eliminating confounding factors (such as brightness) that can affect the eye tracking data. Further research is required to understand the value of these metrics, and their fusion, for a wider range of takeover scenarios, to help with the creation of more reliable driver monitoring systems.

Funding

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

CRediT authorship contribution statement

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

Declaration of Competing Interest

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

Data availability

The authors do not have permission to share data.

Acknowledgement

The authors would like to thank all partners within L3Pilot and the University of Leeds Driving Simulator for their cooperation and valuable contribution. Responsibility for the information and views set out in this publication lies entirely with the authors.

References

Appel, T., Scharinger, C., Gerjets, P., & Kasneci, E. (2018). Cross-subject workload classification using pupil-related measures. *Eye Tracking Research and Applications Symposium (ETRA), 18*. https://doi.org/10.1145/3204493.3204531

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

Batmaz, I., & Öztürk, M. (2008). Using pupil diameter changes for measuring mental workload under mental processing. *Journal of Applied Sciences, 8*(1), 68–76. https://doi.org/10.3923/jas.2008.68.76

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

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

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

Boehm, U., Matzke, D., Gretton, M., Castro, S., Cooper, J., Skinner, M., … Heathcote, A. (2021). Real-time prediction of short-timescale fluctuations in cognitive workload. *Cognitive Research: Principles and Implications, 6*(1), 30. https://doi.org/10.1186/s41235-021-00289-y

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

Bueno, M., Dogan, E., Selem, F. H., Monacelli, E., Boverie, S., & Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2040–2045*. https://doi.org/10.1109/ITSC.2016.7795886

Buettner, R. (2013). Cognitive workload of humans using artificial intelligence systems: Towards objective measurement applying eye-tracking technology. In I. J. Timm & M. Thimm (Eds.), *Advances in Artificial Intelligence. KI 2013. Lecture Notes in Computer Science* (Vol. 8077, pp. 37–48). Springer, Berlin, Heidelberg. Doi: 10.1007/978-3-642-40942-4_4.

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. s.n.

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

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

Ellis, C. J. K. (1981). The pupillary light reflex in normal subjects. *British Journal of Ophthalmology, 65*(11), 754–759. https://doi.org/10.1136/bjo.65.11.754

Fincham, E. F. (1951). The accommodation reflex and its stimulus. *British Journal of Ophthalmology, 35*(7), 381–393. https://doi.org/10.1136/bjo.35.7.381

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

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

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

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

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over Control from Highly Automated Vehicles in Complex Traffic Situations. *Human Factors, 58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*(2), 156–166. https://doi.org/10.1109/TITS.2005.848368

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

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

Kahneman, D. (1973). Attention and Effort. In *Prentice-Hall series in experimental psychology*. Prentice-Hall. Doi:10.2307/1421603.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Liu, K., Green, P., & Liu, Y. (2019). Traffic and Ratings of Driver Workload: The Effect of the Number of Vehicles and Their Distance Headways. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 2134–2138. https://doi.org/10.1177/1071181319631051

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., … Strayer, D. L. (2020). Driver Arousal and Workload Under Partial Vehicle Automation: A Pilot Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64*(1), 1955–1959. https://doi.org/10.1177/1071181320641471

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., … Strayer, D. L. (2021). No Difference in Arousal or Cognitive Demands Between Manual and Partially Automated Driving: A Multi-Method On-Road Study. *Frontiers in Neuroscience, 15*. https://doi.org/10.3389/fnins.2021.577418

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

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

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis and Prevention, 108*, 9–18. https://doi.org/10.1016/j.aap.2017.08.011

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing, 3*, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition, 1*(1), 1–23. https://doi.org/10.5334/joc.18

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

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

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., … Keinath, A. (2018). The "Out-of-the-Loop" concept in automated driving: Proposed definition, measures and implications. *Cognition, Technology & Work, 1–12*. https://doi.org/10.1007/s10111-018-0525-8

Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver Readiness Model for Regulating the Transfer from Automation to Human Control. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* -. https://doi.org/10.1145/3025171.3025199

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

National Highway Traffic Safety Administration. (2009). *Traffic Safety Facts 2009: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. Department of Transportation: U.S. http://www.nhtsa.gov/NCSA.

Noble, A. M., Miles, M., Perez, M. A., Guo, F., & Klauer, S. G. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention, 151*, Article 105959. https://doi.org/10.1016/J.AAP.2020.105959

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

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Pituch, K. A., & Stevens, J. P. (2016). Applied Multivariate Statistics for the Social Sciences. In *Applied Multivariate Statistics for the Social Sciences* (Sixth). Routledge, Taylor and Francis group. www.routledge.com/9780415836661.

Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Torrao, G., Lyu, W., … Lenné, M. G. (2022). Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 87*, 149–163. https://doi.org/10.1016/J.TRF.2022.04.002

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

Recarte, M.Á., Ṕerez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *Spanish Journal of Psychology, 11*(2), 374–385. https://doi.org/10.1017/s1138741600004406

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics*. https://doi.org/10.1080/00140139.2011.604431

SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://www.sae.org/standards/content/j3016_202104/.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). AUTOMATION SURPRISES. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* ((2nd ed.,, pp. 1926–1943). Wiley.

Schartmüller, C., Weigl, K., Löcken, A., Wintersberger, P., Steinhauser, M., & Riener, A. (2021). Displays for productive non-driving related tasks: Visual behavior and its impact in conditionally automated driving. *Multimodal Technologies and Interaction, 5*(4), 21. https://doi.org/10.3390/MTI5040021

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

Spector, R. H. (1990). The pupils. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths. https://www.ncbi.nlm.nih.gov/books/NBK381/.

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

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology, 52*(1), 77–86. https://doi.org/10.1016/J.IJPSYCHO.2003.12.005

Stern, J. A. (1997). *The pupil of the eye: What can it tell us about 'mental processes'?* Human Engineering for Quality Life (HQL). *Quarterly, 8*.

Thorn, E., Kimmel, S., & Chaka, M. (2018). *A Framework for Automated Driving System Testable Cases and Scenarios (Report No DOT HS 812 623)*. In National Highway Traffic Safety Administration: U.S. Department of Transportation. www.ntis.gov.

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation Space and Environmental Medicine*, 78(5 II).

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology, 42*(3), 323–342. https://doi.org/10.1016/0301-0511(95)05165-1

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

Wickens, C. D. (1984). Processing Resources in Attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Academic Press. Doi:R. Parasuraman, R. Davies.

Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2015). Discrete task switching in overload: A meta-analyses and a model. *International Journal of Human Computer Studies, 79*, 79–84. https://doi.org/10.1016/j.ijhcs.2015.01.002

Wiener, E. L. (1989). Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. *NASA Contractor Report*, Article 177528. https://human-factors.arc.nasa.gov/publications/HF_AdvTech_Aircraft.pdf.

Yang, S., Wilson, K., Roady, T., Kuo, J., & Lenné, M. G. (2022). Beyond gaze fixation: Modeling peripheral vision in relation to speed, Tesla Autopilot, cognitive load, and age in highway driving. *Accident Analysis & Prevention, 171*, Article 106670. https://doi.org/10.1016/j.aap.2022.106670

Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Yu, D., Park, C., Choi, H., Kim, D., & Hwang, S. H. (2021). Takeover safety analysis with driver monitoring systems and driver-vehicle interfaces in highly automated vehicles. *Applied Sciences (Switzerland), 11*(15), 6685. https://doi.org/10.3390/app11156685

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