RESEARCH ARTICLE

Pupillary dynamics during hands-off L2 driving and transitions of control under high cognitive load

Courtney M. Goodridge[1,2]*, Rafael C. Gonçalves[2], Ali Arabian[2], Anthony Horrobin[2], Albert Solernou[2], Yee Thung Lee[2], Audrey Bruneau[3], Jonny Kuo[4], Michael G. Lenné[4], Gaëtan Merlhiot[5], Yee Mun Lee[2], Natasha Merat[2]

1 School of Psychology, University of Leeds, Leeds, United Kingdom, **2** Institute for Transport Studies, University of Leeds, Leeds, United Kingdom, **3** Toyota Motor Europe, Brussels, Belgium, **4** Seeing Machines, Canberra, Australia, **5** VEDECOM Institute, Versailles, France

* c.m.goodridge@leeds.ac.uk

Check for updates

OPEN ACCESS

Citation: Goodridge CM, Gonçalves RC, Arabian A, Horrobin A, Solernou A, Lee YT, et al. (2026) Pupillary dynamics during hands-off L2 driving and transitions of control under high cognitive load. PLoS One 21(8): e0355165. https://doi.org/10.1371/journal.pone.0355165

Editor: Michael B. Steinborn, Julius-Maximilians-Universität Würzburg: Julius-Maximilians-Universitat Wurzburg, GERMANY

Received: October 23, 2025

Accepted: July 19, 2026

Published: August 11, 2026

Copyright: © 2026 Goodridge et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Data, analysis code, and models can be found in the following link: https://osf.io/vmxj3.

Funding: The authors would like to thank all partners within the Hi-Drive project for

Abstract

Arousal plays a vital role in facilitating the human ability to respond flexibly in a goal-directed manner, and pupillometry offers a non-invasive window into arousal-related neural processes given the close relationship between pupil size and Locus Coeruleus-Norepinephrine (LC-NE) activity. Whilst pupillometry has been used to detect cognitive load in manual and automated driving, the dynamic relationship between pre-stimulus (baseline) pupillary state and task-evoked pupillary responses (TEPRs) has not been investigated in driving contexts. This is important because variability in baseline pupil size – itself influenced by cognitive demands such as non-driving related task engagement – may contribute to variability in TEPRs independently of how drivers respond to critical events. This driving simulator experiment aimed to establish whether pupillometry during hands-off Level 2 (L2) driving was a reliable indicator of cognitive load, and to examine whether relationships between pupillary dynamics and behaviour established in controlled laboratory paradigms generalise to applied tasks. The size and reactivity of drivers' (N = 38) pupils were measured during hands-off L2 driving with and without a cognitive load task, followed by critical and non-critical transitions of control. Analysis revealed that mean, not standard deviation, of pupil diameter was a reliable indicator of cognitive load. Furthermore, higher baseline pupil diameter was associated with smaller TEPRs, and this relationship persisted after correcting for regression to the mean artefacts – suggesting that pre-stimulus pupillary state genuinely constrains subsequent TEPRs. Limited evidence was found that TEPRs or pre-stimulus pupillary variability predicted driver reaction times, potentially reflecting the motoric variability inherent in naturalistic driving responses. Finally, more critical events were associated with larger pupil diameter, indicating that drivers were exerting greater effort to manage the transition. These results indicate that pupillometry is a useful measure of cognitive load and that

their cooperation and valuable contribution. [This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101006664. The sole responsibility of this publication lies with the authors. Neither the European Commission nor CINEA – in its capacity of Granting Authority – can be made responsible for any use that may be made of the information this document contains]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

laboratory-established pupillometric relationships extend, at least partially, to applied driving contexts, with implications for the development of Driver Monitoring Systems (DMS).

1. Introduction

1.1 Arousal: From theory to application

Arousal is characterised by a combination of psychological and physiological responses to external stimuli (e.g., loud noises) or internal processes (e.g., solving a maths problem) [1], and influences both the available capacity for information processing and the selectivity of attentional allocation towards goal-relevant behaviour [2]. The relationship between arousal and performance is captured by the Yerkes-Dodson law: task performance increases linearly with arousal for simple tasks, but follows an inverted U-shape relationship for difficult tasks, such that too little and too much arousal are detrimental, with optimal performance occurring at intermediate arousal [3]. One proposed mechanism for this inverted- U relationship is that increasing arousal progressively narrows the range of cues an individual attends to; moderate arousal improves performance by filtering out task-irrelevant information; but excessive arousal causes task relevant cues to be excluded, paradoxically narrowing attentional focus whilst increasing distractibility [1,4].

(3), (4), and (2) bring together a shared focus on arousal as the primary governor of attentional focus, with some work extending the arousal system's role to regulating cognitive control [5,6]. The ubiquitous nature of the arousal system allows it to be applied to real-world tasks. For example, to maintain safe driving performance, a driver should act on the top of this hypothetical performance inverted-U curve [7,8]. However, the monotony of driving in simple conditions for experienced drivers can result in lowered states of awareness, resulting in failures to detect and react to changing road conditions [9,10]. This underload state corresponds to a deviation from the top of the inverted-U curve towards low arousal states and poor task performance. Conversely, overload can occur when physiological arousal is too high, also resulting in detrimental performance. The Cognitive Control Hypothesis (CCH) [11] proposes that driving tasks requiring cognitive control (e.g., negotiating intersections) or those that are non-practiced (e.g., novice drivers in new environments) are more likely to be negatively impacted by increased cognitive load.

Issues associated with underload and overload are likely to be exacerbated as vehicle automation increases; there are concerns that reducing driver mental demands may have detrimental effects on their vigilance and hazard perception [12,13]. Furthermore, drivers are more likely to engage in non-driving related tasks (NDRTs) with increased experience with automated driving functionalities [14–16]. Whilst there is evidence that NDRT engagement can alleviate underload by increasing physiological arousal [17], NDRTs may result in increasing driver cognitive load beyond optimal levels [18,19]. These issues are particularly problematic for

intermediate levels of automation (e.g., SAE Level 2 [L2] systems) where the driver is required to monitor the road environment and intervene when necessary.

Despite the implicit assumption that cognitive load negatively impacts transitions of control, research evidence does not always corroborate this. Steering and braking reaction times to obstacle avoidance events have consistently been unaffected by increased cognitive load [18,20–22]. This might be because reactions to strong looming signals (e.g., a decelerating lead vehicle) do not rely upon cognitive control; rather, they are automatised responses that are triggered by bottom-up saliency of visual looming – and thus are not negatively impacted [11,18]. The quality of transitions of control *have* been found to degrade with elevated cognitive load in some instances; the time between steering initiation and lane crossing increased with cognitive load [20]. Lane changes are responses that are likely to involve cognitive control [23]; drivers must engage top-down attentional control to shift their gaze between the wing mirrors, the road centre, an obstacle they want to overtake, and the destination lane [24,25]. Findings from [20] may, therefore, point towards evidence of a transition of control subtask that is detrimentally impacted by cognitive load.

The CCH suggests that manual driving responses that are novel or inherently uncertain are more likely to rely on cognitive control and, consequently, be negatively impacted under high cognitive load. Given that transitions of control are relatively infrequent, they are likely to remain less practiced than routine driving behaviours and may place greater demands on cognitive control processes. If these assumptions hold, fluctuations in arousal may directly influence the cognitive mechanisms required for successful transitions, highlighting the importance of understanding arousal's role in applied driving contexts. Hence the concept of Driver Readiness (DR) – defined as the probability of a successful take-over [26] – becomes central. Although efforts have been made to standardise DR metrics [27–30], it remains an abstract construct that is not directly observable and is therefore inferred from related measures. Cognitive load has emerged as a key predictor of successful transitions of control, positioning it as a critical component of DR. If the arousal system fundamentally governs cognitive performance, it is essential that future Driver Monitoring Systems (DMS) can reliably capture physiological indices of arousal. Pupillometry, given its close links to underlying neural dynamics, represents a promising candidate for real-time assessment in this domain.

1.2 Linking arousal to pupillary and neural dynamics

Aside from the pupil light response (PLR), pupil size varies as a function of a variety of cognitive phenomena including mental effort [31–33], decision making [34,35], surprise [36,37], and neural gain [38,39]. These Task-Evoked Pupillary Responses (TEPRs) are typically involuntary pupil dilations that peak any time from 500 ms [40] to 2 s post stimulus [41]. TEPRs follow systematic changes in neural activity throughout the cortex, regulated via neuromodulators [42]. Therefore, the pupillary response has been proposed as an easily measurable physiological reflection of arousal and attention networks [1,43].

Cognitively driven modulations of pupil size intersect through three main nuclei: (1) the Pretectal Olivary Nucleus (PON), which receives top-down information from brain areas implicated in higher order visual-saccadic processing [44,45]; (2) the Superior Colliculus (SC), which integrates sensory signals to facilitate the orientation of cross-modal attention [46]; and (3) the Locus Coeruleus (LC), which receives input from a variety of brain regions and is a primary source of norepinephrine (NE) (see [47] for a review). The LC-NE system is of particular interest because it has been hypothesised to regulate cortical arousal [48], alongside interacting with top-down cortical systems associated with attentional control and executive function [49,50]. The LC-NE system is therefore heavily involved in task-related shifts of arousal [51,52].

LC-NE-centred neural circuits have been heavily associated with pupillary responses [47,53]. Animal studies have revealed that pupillary dynamics in response to acoustic or effort-related stimuli positively correlate with LC activity [54,55]. Similarly, stimulation of the LC in monkeys and rats produced pupil dilations [54,56]. In humans, blood oxygenation level dependent (BOLD) signals in the LC have been positively correlated with pupil size when completing cognitive tasks [35,51]. Whilst pupil size is not a direct readout of LC activity, their close relationship provides a means for measuring changes in attentional state mediated by LC-NE activity [43,51].

1.3 Pupillary dynamics and applied task performance

Pupillometry offers a non-invasive window into brain processes associated with arousal and attention given the relationship between LC activity and pupillary responses. It is no wonder then, that researchers within the automotive domain have focused on pupillometry to understand cognitive resource engagement. In manual driving, conversational NDRTs [57], encountering unexpected pedestrians [58], and overtaking truck platoons [59] increase pupil size, indicating an increase in cognitive load during these interactions. During SAE Level 2 and Level 3, larger pupil size was also associated with smaller time headways and transitions of control due to increased cognitive demand [60].

However, pupillometry in driving contexts predominantly focus on tonic (e.g., sustained) pupil diameter as an index of cognitive load. Yet, sustained pupil diameter leaves largely unexamined whether the dynamic relationship between pre-stimulus (baseline) pupillary state and TEPRs is preserved in applied driving contexts. Research in controlled laboratory paradigms has established that pre-stimulus pupil size predicts the magnitude of subsequent TEPRs [61–63]. These studies propose that this relationship reflects the behaviour of task-evoked (phasic) LC responses that depend upon baseline (tonic) LC activity. Phasic responses to task-relevant events are thought to be maximised at intermediate tonic levels but attenuated when tonic activity is either too low or too high [53]. Adaptive Gain Theory (AGT) summarises this relationship between pupil size, LC activity, and goal directed behaviour. However, it is hard extrapolating pupillary data to LC-NE dynamics given the indirect relationship. Furthermore, an overlooked phenomenon is that large pre-stimulus pupils are likely to decrease in size afterwards, simply because larger pupils a more likely to become smaller [41]; thus requiring corrections of these regression to the mean artifacts [64].

It is important to outline that the baseline-TEPR relationships, and investigations of regression to the mean artefacts, have not been studied in driving. This is a key omission, because baseline pupil size varies considerably across individuals and situations in driving – including as a direct consequence of NDRT engagement – meaning that variability in pre-stimulus state will contribute to variability in TEPRs independently of how participants respond to critical events. Without understanding this relationship, it is not possible to determine whether a smaller TEPR in response to a request to intervene (RTI) reflects reduced responsiveness to the event itself or is simply a consequence of an elevated pre-stimulus baseline. Examining baseline-TEPR relationships in the context of transitions of control therefore has implications for the practical development of DMS that may use TEPRs to index driver state.

TEPRs have also been proposed as indicators of the degree to which attentional and cognitive resources are mobilised in response to a specific event [63]. In controlled laboratory paradigms, larger TEPRs have been associated with faster reaction times, suggesting that the magnitude of TEPRs reflect the efficiency of cognitive engagement with an incoming stimulus [65,66]. If this relationship generalises to driving, TEPRs in response to an RTI could provide an index of how effectively a driver is engaging with a transition of control. Examining whether TEPR magnitude predicts reaction time during transitions of control therefore represents an important test of whether laboratory-established pupillometric relationships extend to applied settings, with direct implications for whether TEPRs could serve as a useful real-time index of driver responsiveness in DMS.

Finally, a further dimension of pupillary dynamics that has received recent attention is short-term pupillary variability, indexed by the Micro-Pupillary Unrest Index (M-PUI). Rather than capturing the average level or evoked change in pupil size, M-PUI quantifies moment-to-moment fluctuations in pupil diameter over short time windows and has been proposed as an index of transient vigilance state [67]. In psychomotor vigilance tasks, higher pre-stimulus M-PUI has been associated with slower reaction times, suggesting that periods of greater pupillary instability reflect momentary lapses in alertness [68]. This is particularly relevant in the context of automated driving, where sustained monitoring of a system that rarely requires intervention may produce fluctuating vigilance states [12,69]. If M-PUI captures these transient fluctuations in vigilance, it may be sensitive to variability in DR that precedes a transition of control. However, this has not been tested in driving contexts, and it remains unclear whether the relationship between M-PUI and reaction time observed in simple laboratory tasks survives the increased complexity and response variability inherent in transitions of control.

1.4 Current work

Safe and effective transitions of control depend on the driver's readiness to intervene, which is strongly influenced by underlying arousal and cognitive control processes. However, DR cannot be measured directly and must be inferred from physiological and behavioural indicators. Pupillometry offers a promising, non-invasive index of these processes, yet it remains unclear whether theoretically grounded predictions of pupillary dynamics generalise to complex, real-world driving contexts.

The analysis in this manuscript is structured into three parts. First, we evaluate pupillometric indices as objective measures of cognitive load during hands-off L2 driving, focusing on the mean and standard deviation (SD) of pupil diameter as candidate metrics, given their relevance as indicators of arousal-related processing. In the second part, we examine three aspects of pupillary dynamics motivated by the broader pupillometry literature [65,66,68]. First, we investigate whether larger baseline pupil diameter is associated with smaller TEPRs during transitions of control, and whether this relationship persists after correcting for regression to the mean artefacts [64]. Second, we examine whether TEPR magnitude predicts driver reaction time during transitions of control. Third, we investigate whether pre-stimulus pupillary variability, indexed by M-PUI, is associated with driver reaction time. In the third part, we examine how task demands influence sustained pupil dynamics during transitions of control by manipulating cognitive load (N-back) and event criticality. Prior work has shown that mean pupil diameter increases with mental effort in both controlled tasks [70,71] and applied driving contexts [58–60,72]. We extend this by testing whether more critical driving events elicit greater sustained increases in pupil diameter, reflecting heightened arousal and task engagement during transition scenarios.

2. Method

2.1 Participants

41 participants originally took part in the experiment. However, three participants were removed because they failed to follow the experimental instructions or because eye tracking data was not captured correctly. The remaining 38 participants had a reasonable gender split between males (N = 22) and females (N = 16), an average age of 38.81 years (range = 22–65 years old, SD = 12.502), all had normal or corrected to normal vision, had a valid UK driving license for an average of 17.84 years (range = 4–43 years, SD = 12.119), and were regular drivers who drove an average of 9355.26 miles annually (range = 5000–20000 miles, SD = 4153.975).

2.2 Apparatus and materials

The experiment was conducted at the University of Leeds Driving Simulator, consisting of a Jaguar S-type cab encased within a 4 m spherical with a 300° field of view projection dome. The virtual environment created for this driving simulator experiment was kept at daytime luminance levels throughout. Therefore, the impact that changes in environmental lighting could have on pupil reactivity was experimentally controlled from the offset. Longitudinal and lateral movement was provided via a hexapod motion base and a 5 m x 5 m X-Y table. Pupillometry data were captured using the Seeing Machines' Driving Monitoring System (DMS) eye tracker, sampling at 60 Hz. The DMS – comprised of a driver-facing infrared camera – utilised proprietary algorithms to derive head position, facial feature detection, eyelid position, and pupil and gaze tracking to monitor the drivers' visual attention. The system monitored driver's gaze direction and identified gaze targets towards the forward roadway or towards other regions. The NASA Task Load Index (NASA-TLX) was used to measure subjective workload [73].

2.3 Design and procedure

Informed consent was obtained, both written (in terms of signing a consent form) and verbally (by asking the participant if they consented to take part in the research), and standardised instructions were delivered. All procedures were approved

by the University of Leeds Research Ethics Committee (Reference code: 2022-0353-206). Participants were recruited for this study between 27/09/2024–28/10/2024. The original design of the experiment was 2 x 2 Repeated Measures design. The two within-participant factors were event criticality (i.e., the time to collision [TTC] at the onset of a lead vehicle braking) and the manipulation of driver cognitive load. The event criticality was manipulated over two levels: TTC = 3 s corresponded to severe scenarios and TTC = 5 s corresponded to less severe scenarios [18,74–76]. Driver cognitive load was also manipulated over two levels; a no-load condition and a high load condition where participants completed an auditory version of the verbal response delayed digit recall task (N-back, where N = 2) (see [77] for details) during hands-off L2 driving. Digits were presented via the vehicle speaker system, whilst participant responses were verbal. Participants had the opportunity to practice the N-back task at the start of the experiment without any driving. Analysis of N-back performance and subjective workload ratings can be found in previous work [18,78].

Before the main experimental procedure, participants conducted a practice run to familiarise themselves with the simulator dynamics and the driving system. Participants were taught how to engage the L2 system and where to place their hands and feet when the L2 system was on. The experimental procedure consisted of two main drives – one where drivers had to complete N-back whilst monitoring their L2 system, and another where they just had to monitor the road and the driving system. Drive order was counterbalanced across the sample, and each drive lasted approximately 35 minutes. Each drive comprised of 10 discrete events. Following each drive, the NASA-TLX was used to measure subjective workload. Each event consisted of 30 s of manual driving, followed by approximately two minutes of hands-off L2 driving. After two minutes of hands-off L2 driving, a request to intervene (RTI) was delivered by a short auditory tone, and the simultaneous change in colour of a steering wheel icon positioned in the dash area from steady green (automation engaged) to flashing red (intervention required) (see [18,78] for more details on the procedure). Four of these events were critical, which involved a lead vehicle braking sharply, creating a rear-end scenario that drivers had to avoid. The remaining six were non-critical, which involved scenarios where there was no lead vehicle, or the lead vehicle did not brake. These were included to guard against learning effects such that the presence of a lead vehicle was not always associated with a critical event. Critical and non-critical events were presented in a pseudo-random order. Transitions of control could be initiated via 3 different modes: turning the steering wheel in either direction by more than 2°, pressing any of the operational pedals, or pressing a micro-switch button strapped to the steering wheel. For non-critical trials, a majority of transitions (77%) were initiated with the microswitch; for critical trials, a majority of transitions (84%) were initiated by using the operational pedals. Participants were instructed to respond as quickly and as safely as possible. Overall reaction time calculated as the fastest time between the RTI until the L2 system had been deactivated.

2.4 Data preparation and statistical modelling

2.4.1 Pupillometry pre-processing. Eye blinks were automatically identified and tagged via the Seeing Machines' Driving Monitoring System. To account for pupil size drop-offs, samples before and after blinks were also removed. The pupil time course was linearly interpolated across blinks. Mean pupil size was then computed by averaging the left and right pupil diameter. For investigating tonic pupil diameter during hands-off L2 driving and during the transition of control, and TEPRs in response to RTI, the pupil time series was bandpass filtered (third order butterworth, passband = 0.05–4 Hz). For the calculation of M-PUI, a Hanning filter was used to remove high frequency fluctuations. In line with previous research [68], we investigated four different time windows for the Hanning filter (50 ms, 100 ms, 200 ms, and 400 ms) to assess whether correlations between M-PUI and reaction times were consistent across filtering parameters.

For investigating pupil diameter during hands-off L2 driving, we calculated the mean and SD of pupil diameter during the 2-minute driving period. For investigating increases in pupil diameter associated with the transition of control, we calculated mean pupil diameter during the transition of control window. For calculating TEPRs, two time-based epochs were used. The first was the pre-RTI baseline epoch (−500-0 ms) and the second was the stimulus response epoch (300–1800 ms). For each trial, we subtracted these two epochs to give our primary dependent measure for assessing phasic pupillary

dilations: change in pupil response. The lower bound of the stimulus response epoch was chosen to eradicate implausibly fast pupillary dilations; the fastest a pupil reacts is within 200 ms [41]. The upper bound stimulus response epoch was chosen because – although pupillary dilations associated with orienting attention often peak 500 ms post-stimulus [40,79] – if there is a clearly defined triggering stimulus that requires the engagement of arousal, pupillary dilation can peak up to 1000 ms post-stimulus [41,80].

A key prediction from laboratory paradigms is that higher baseline pupil diameter should be associated with smaller TEPRs [41,61,81]. However, a correlation like this can also come about through regression to the mean given that pupils that are already large are more likely to constrict than become larger due to physiological constraints rather than cognitive related phenomena [41,64]. One way of correcting for this regression to the mean artefact is to analyse the spontaneous tendency of the pupil to change size on the time scale of an RTI-evoked pupil dilation, as a function of baseline pupil size. To implement this correction, we used a method proposed by [64]. We estimated the spontaneous tendency of the pupil to change as a function of baseline diameter using a pseudo-baseline and a pseudo-change window that was not time-locked to the transitions of control (i.e., a period where drivers were monitoring the L2 system). A cubic regression model was then fitted to predict the pseudo-change from baseline pupil size. This model was then used to predict the expected baseline-dependent change in pupil size for each trial based on its pre-stimulus baseline. The predicted change was subtracted from the observed TEPR, yielding a corrected measure of phasic pupil response. If higher baseline pupil diameter was still associated with smaller TEPRs following this baseline correction, this would suggest that the pre-stimulus pupillary state genuinely constrained the magnitude of subsequent TEPRs, rather than the correlation being a statistical consequence of regression to the mean.

Finally, we calculated M-PUI to estimate short-term vigilance levels. First, we applied a Hanning filter defining four different time windows prior to the RTI (50 ms, 100 ms, 200 ms, and 400 ms). For each of the smoothed time series datasets, M-PUI was calculated as the absolute value for the degree of change in pupillary fluctuations divided by the length of the pupillary time series.

2.4.2 Reaction time pre-processing. For investigating the relationship between reaction times and pupillary dynamics during transitions of control, implausibly small reaction times were filtered out of the dataset. We set a threshold of 300 ms based on previous research investigating human reactions to collision avoidance warning systems [82] (28 in total). Next, we calculated z scores and used these to remove long reaction times that did not represent valid behaviour; we set a threshold of 3 SDs above the mean. Overall, all reaction times in the final dataset were between 0.3–4.2 s.

2.4.3 Statistical modelling. A Bayesian approach was used to model the data. For each analysis, a Bayesian multilevel model was fitted. The distribution of each dependent variable was assessed prior to the analysis, and the appropriate distribution family was chosen for model building. Posterior distributions were estimated using the No-U-Turn sampler in the brms package in R [83] and weakly informative regularising priors were used. Weakly regularising priors attempt to "regularise" the posterior distribution by keeping it within reasonable bounds without trying to influence the data as much as possible. This helps rule out implausible effects; for example, predicting increases in pupil diameter that are physiologically implausible. Data, analysis code, and models can be found in the following link: https://osf.io/vmxj3.

Separate design matrices were specified for the three parts of the analysis, reflecting the different task demands associated with these phases. However, for all models, we used treatment coding for the categorical variables inputted into the model. Analysis of tonic pupil diameter during hands-off L2 driving focused on sustained cognitive demands preceding the RTI. The mean and SD of pupil diameter were calculated for each hands-off L2 driving segment and modelled as a function of an intercept (β_0), N-back (β_N), lead vehicle (β_L) and an interaction term between these variables ($\beta_{N:L}$). Both critical and non-critical trials were included; however, event criticality was not modelled, as no overt response to a developing hazard had occurred during the period where pupil size was measured.

TEPRs were modelled as a function of an intercept (β_0), baseline pupil diameter (β_B), N-back (β_N), event criticality (β_{TTC_3}, β_{TTC_5}), and all two way ($\beta_{B:N}$, $\beta_{B:TTC_3}$, $\beta_{B:TTC_5}$, $\beta_{N:TTC_3}$, $\beta_{N:TTC_5}$), and three way interactions ($\beta_{B:N:TTC_3}$, $\beta_{B:N:TTC_5}$)

between the predictors. The inclusion of these higher-level interactions was because we hypothesied that baseline pupil diameter may have interacted with the presence of N-back and event criticality when impacting TEPRs. Models were fitted before and after correction for regression to the mean effects to assess whether effects qualitatively changed following corrections. Reaction time was modelled as a function of TEPRs (β_{TEPR}), event criticality (β_{TTC}), two way interactions between TEPRs and event criticality ($\beta_{TEPR:TTC_3}$, $\beta_{TEPR:TTC_5}$), a two way interaction between TEPR and N-back ($\beta_{TEPR:N}$), and three way interactions between N-back, event criticality, and TEPRs ($\beta_{TEPR:TTC_3:N}$, $\beta_{TEPR:TTC_5:N}$). Main effects of N-back and the two way interaction between N-back and event criticality were omitted given that previous research analysed these data and found no effect on reaction times [18]. Reaction time was also modelled using the same parameters except the M-PUI metric replaced the TEPR metric. For the analysis of increases in pupil diameter during the transition of control, mean pupil diameter was modelled as a function of an intercept (β_0), N-back (β_N), event criticality (β_{TTC_3}, β_{TTC_5}), and two-way interactions between these predictors ($\beta_{N:TTC_3}$, $\beta_{N:TTC_5}$).

Each parameter was associated with probability distributions quantifying the level of uncertainty, conditioned on the data. Posterior distributions of parameters were described by their mean and a 95% Credible Interval (CI) [84]. The *probability of direction* (*pd*) was also reported for each fixed effect parameter, defined as the probability that an effect was positive or negative [85].

3. Results

3.1 Effect of N-back and lead vehicle presence on pupil diameter during hands-off L2 driving

First, we assessed the distributional properties of mean pupil diameter during hands-off L2 driving across N-back and lead vehicle conditions (see Fig 1).

A Bayesian model specifying a Gaussian distribution was fitted to assess changes in mean pupil diameter as a function of N-back and lead vehicle. The model (see Table 1) predicted a small increase in mean pupil diameter (0.032, [CI: −0.006, 0.050]) when lead vehicles were present (3.865 mm) relative to when they were not present (3.833 mm) during hands-off L2 driving. Despite this, the 95% credible intervals suggested that this effect could be close to 0. Conversely, the model predicted a strong effect of N-back. Mean pupil diameter increased (0.285, [CI: 0.192, 0.369]) when drivers completed the N-back task (4.118 mm) relative to the no N-back drives (3.833 mm). Further investigation of the heterogeneity of this average effect revealed that 90% of the population would be expected to demonstrate similar increases in

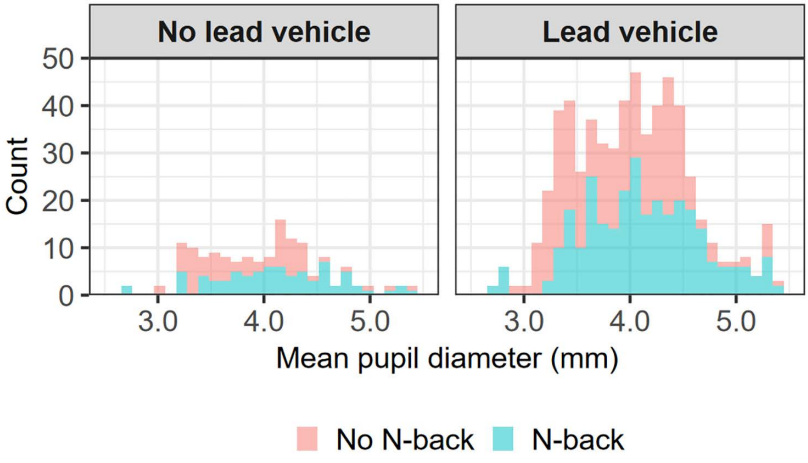


Fig 1. Distributions of mean pupil diameter during hands-off L2 driving with and without N-back and lead vehicle presence.

https://doi.org/10.1371/journal.pone.0355165.g001

Table 1. Posterior means and 95% CIs for parameters predicting mean pupil diameter.

	Coefficient (95% CI)	*pd*
β_0	**3.833 (3.650, 4.022)**	**100%**
β_N	**0.285 (0.194, 0.372)**	**100%**
β_L	**0.032 (−0.006, 0.070)**	**95.02%**
β_{NL}	−0.035 (−0.088, 0.020)	89.05%
Participants	38	
Observations	744	

https://doi.org/10.1371/journal.pone.0355165.t001

pupil diameter as a function of cognitive load during hands-off L2 driving. This is reflected in the current sample; only three participants demonstrated small *reductions* in pupil diameter whilst completing N-back (see Fig 2B).

Next, we assessed the distributional properties of the SD of pupil diameter across N-back and lead vehicle conditions (see Fig 3).

Modelling revealed no effect of lead vehicle; however, N-back was associated with increased SD of pupil diameter (0.039, [CI: 0.021, 0.057]), (0.236 mm) relative to no N-back conditions (0.196 mm) (see Table 2 and Fig 4). Further investigation of the heterogeneity of the average effect revealed that only 82% of the population would be predicted to have similar increases in the SD of pupil diameter as a function of N-back; almost a fifth of drivers demonstrated no differences or reversals of the average effect. Whilst these findings support previous work indicating that pupillometry can be a useful indicator of increased cognitive load [60], the current analysis reveals that mean, rather than the SD of pupil diameter, is likely to be a more reliable measure.

3.2 Correlations between baseline pupil diameter and TEPRs

Next, we examined the relationship between baseline pupil diameter and TEPRs. Mean pupil response was plotted for each condition, with the pre-stimulus baseline and stimulus response epochs shaded in grey (see Fig 5).

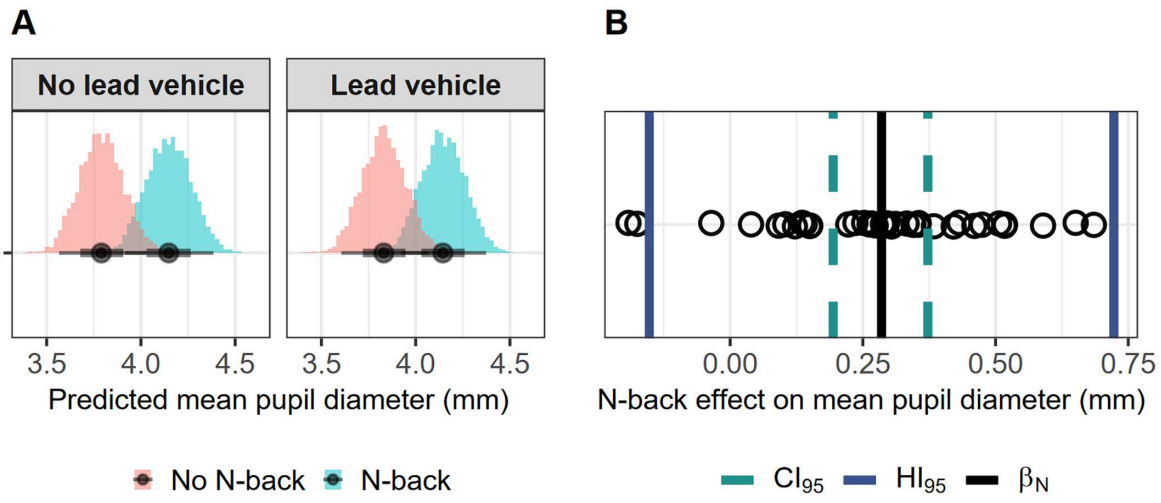


Fig 2. A) Posterior predictions of expected mean pupil diameter as a function of lead vehicle and N-back. Mean pupil diameter is predicted to increase during N-back relative to no N-back conditions B) Strip plot displaying a range of effects of N-back on mean pupil diameter. The black line denotes the average increase in mean pupil diameter, (fixed effect), the green dashed lines denote the heterogeneity of the average effect of N-back (95% Credible Intervals) and the blue solid lines denote the population heterogeneity of the effect of N-back.

https://doi.org/10.1371/journal.pone.0355165.g002

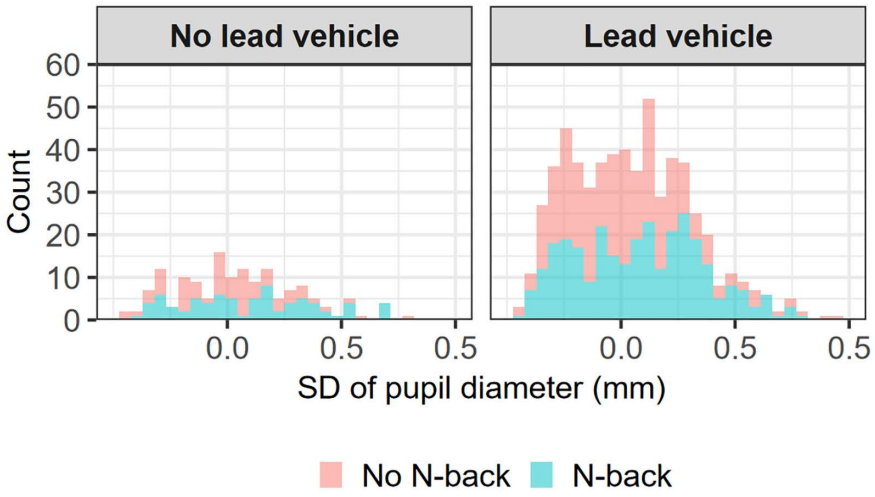


Fig 3. Distributions of the SD of pupil diameter during hands-off L2 driving with and without N-back and lead vehicle presence.

https://doi.org/10.1371/journal.pone.0355165.g003

Table 2. Posterior means and 95% CIs for parameters predicting SD of pupil diameter.

	Coefficient (95% CI)	*pd*
β_0	**0.196 (0.163, 0.229)**	**100%**
β_N	**0.039 (0.021, 0.057)**	**100%**
β_L	0.001 (−0.008, 0.009)	55.03%
β_{NL}	−0.006 (−0.019, 0.006)	84.38%
Participants	38	
Observations	744	

https://doi.org/10.1371/journal.pone.0355165.t002

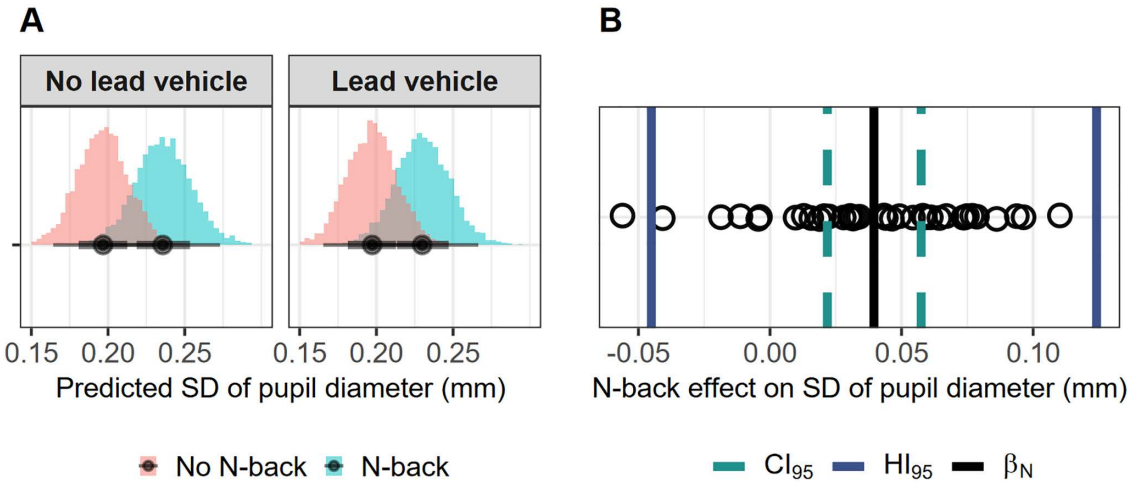


Fig 4. A) Posterior predictions of expected SD of pupil diameter as a function of lead vehicle and N-back. SD of pupil diameter is predicted to increase during N-back relative to no N-back conditions. However, this increase appears smaller and less certain than for mean pupil diameter B) Strip plot displaying a range of effects of N-back on SD of pupil diameter. The black line denotes the average increase in SD of pupil diameter, (fixed effect), the green dashed lines denote the heterogeneity of the average effect of N-back (95% Credible Intervals) and the blue solid lines denote the population heterogeneity of the effect of N-back.

https://doi.org/10.1371/journal.pone.0355165.g004

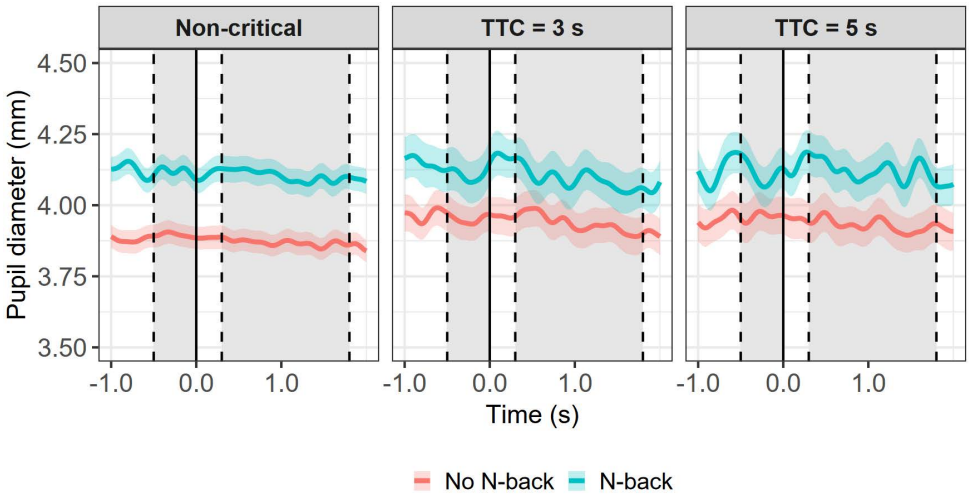


Fig 5. Mean pupil responses during N-back (in blue) and no N-back (in pink) conditions. The coloured shaded regions represent the standard error of the mean. Pre-stimulus baseline (−500-0) and stimulus response epochs (300-1800) are shaded in grey.

https://doi.org/10.1371/journal.pone.0355165.g005

The distribution of uncorrected (see Fig 6A) and corrected (see Fig 6B) TEPRs were plotted to assess distributional properties; a Gaussian model was then used for fitting.

The Bayesian model using the uncorrected TEPR revealed no effect of N-back or event criticality on TEPRs. However, there was evidence that higher baseline pupil diameter was associated with smaller TEPRs (−0.325 mm, [CI: −0.409, −0.249]) (see Fig 7A). Furthermore, the model revealed an interaction between baseline pupil diameter and N-back; the negative correlation between baseline pupil diameter and TEPR was stronger following no N-back trials (−0.341 mm, [CI: −0.427, −0.263]) relative to N-back trials (−0.309 mm, [CI: −0.394, −0.224]). These effects were maintained even after correcting for regression to the mean artefacts; whilst the relationship between baseline pupil diameter and TEPRs was smaller (−0.266 mm, [CI: −0.346, −0.191]) (see Fig 7B; Table 3), the interaction effect between no N-back (−0.285 mm, [CI: −0.367, −0.209]) and N-back trials (−0.246 mm, [CI: −0.336, −0.158]) remained similar in size.

3.3 Correlations between TEPRs and reaction time

Next, we analysed the relationship between driver reaction times and TEPRs. Distributions of reaction time data demonstrated characteristic negative skew (see Fig 8). As such, a log normal distribution was used to account for the extended right tail.

Modelling revealed no effect of corrected TEPR on reaction times (see Fig 9). However, we did find that for typical TEPRs, reaction times were faster for event criticalities of 3 s (−0.093 s, [CI: −0.137, −0.049]) and 5 s (−0.083 s, [CI: −0.129, −0.040]) relative to non-critical transitions (see Table 4).

3.4 Correlations between M-PUI and reaction time

Next, we analysed the relationship between driver reaction times and M-PUI. We fitted a separate model for each filter time window. Overall, we found limited evidence for the relationship between reaction time and M-PUI (see Table 5, see Fig 10). For small time windows (50 ms), there was some evidence that higher M-PUI was associated with longer reaction times. However, the *pd* indicated only a 94% probability that this correlation would be above zero. As time windows increased, the size and reliability of this effect reduced, as highlighted by the reduced size of the coefficient and *pd*, respectively.

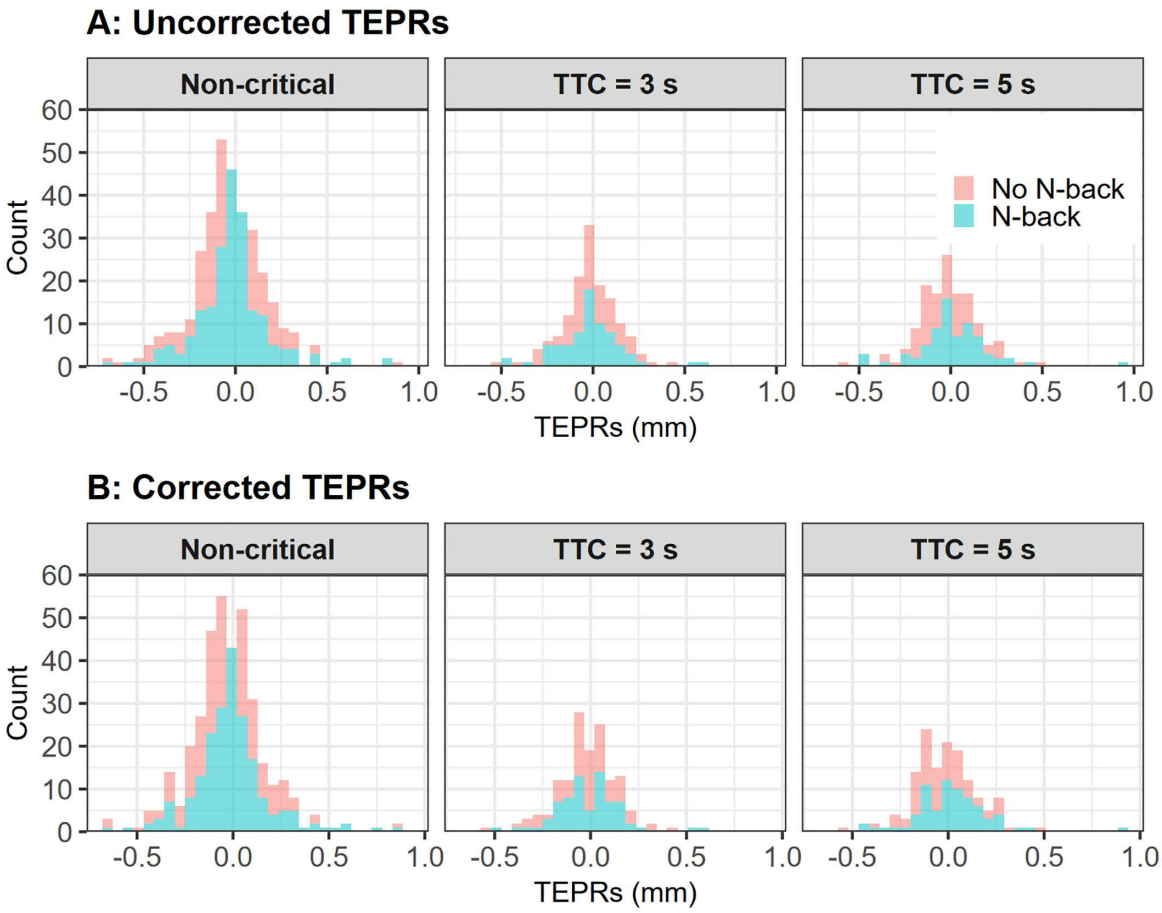


Fig 6. Distributions of corrected and uncorrected TEPRs within each event criticality and N-back condition.

https://doi.org/10.1371/journal.pone.0355165.g006

3.5 Increases in pupil diameter during transitions of control

Finally, we investigated tonic increases in pupil diameter associated with general increases in effort during the transitions of control. The transition of control time window was based on previous research [60]; 10 s from the RTI plus the average reaction time (~ 1 s). Mean pupil diameter during this window was used to investigate whether more critical events were associated with increased effort during the transition of control. Distributions of mean pupil diameter were plotted (see Fig 11) and as such, a Gaussian model was selected.

Fig 12 shows average pupil diameter divided into three sections: pre-RTI (0–10 s), during the transition of control (10–21 s), and post transition of control (21–30 s). For each condition, it was evident that there were large tonic increases in pupil diameter several seconds after the RTI (e.g., the first dashed vertical line in each panel of Fig 12) and after the initial disengagement of the automated system (approximately 1 s after the RTI).

Modelling mean pupil diameter during the transition of control time window revealed an interaction effect between N-back and event criticality conditions (see Table 6). For no N-back conditions, average pupil diameters for TTC = 3 s conditions (3.923 mm) were larger than for non-critical conditions (3.857 mm) (0.065, [CI: 0.022, 0.103]); a similar effect was observed between TTC = 5 s conditions (3.932 mm) and non-critical conditions (0.075, [CI: 0.037, 0.113]). Conversely, the model revealed no consistent changes between the event criticalities during N-back conditions: mean

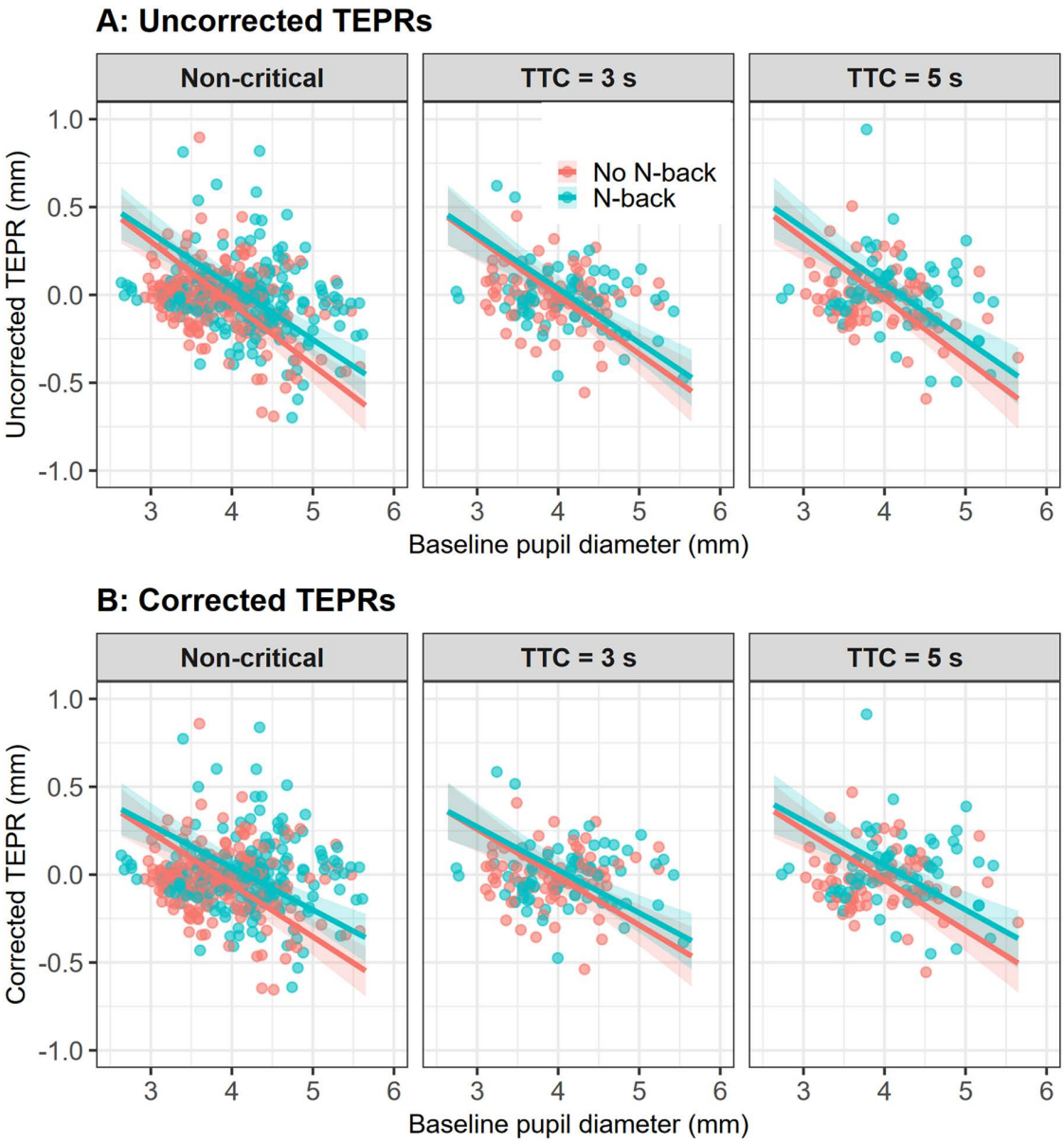


Fig 7. A negative correlation between baseline pupil diameter and uncorrected (A) and corrected (B) TEPRs. This indicates that pre-stimulus pupillary state genuinely constrains the magnitude of subsequent TEPRs, rather than the correlation being a statistical consequence of regression to the mean.

https://doi.org/10.1371/journal.pone.0355165.g007

pupil diameters in TTC = 3 s conditions (4.140 mm) were not reliably larger than non-critical conditions (4.141 mm) (0.001, [CI: −0.045, 0.039]); nor were TTC = 5 s conditions (4.146 mm) reliably larger than non-critical conditions (0.005, [CI: −0.031, 0.044]). Overall, these results are in-line with the hypothesis that more critical events are likely to increase physiological arousal as indicated by increases in pupil diameter. However, they also point towards ceiling effects in the N-back condition such that altering the criticality of an event did not further increase pupil diameter (and thus our measure of tonic arousal).

Table 3. Posterior means and 95% CIs for parameters predicting TEPRs.

Parameters	Uncorrected TEPR model		Corrected TEPR model	
	Coefficient (95% CI)	*pd*	Coefficient (95% CI)	*pd*
β_0	**1.359 (1.005, 1.005)**	**100%**	**1.136 (0.796, 1.470)**	**100%**
β_B	**−0.352 (−0.433, −0.266)**	**100%**	**−0.298 (−0.376, −0.216)**	**100%**
β_N	−0.095 (−0.307, 0.111)	81.13%	−0.128 (−0.337, 0.085)	88.90%
β_{TTC_3}	−0.050 (−0.336, 0.234)	63.57%	−0.060 (−0.337, 0.213)	66.83%
β_{TTC_5}	−0.005 (−0.270, 0.267)	51.83%	−0.021 (−0.290, 0.258)	57.32%
$\beta_{B:N}$	**0.048 (−0.002, 0.100)**	**96.98%**	**0.056 (0.004, 0.107)**	**98.25%**
$\beta_{B:TTC_3}$	0.023 (−0.048, 0.095)	73.42%	0.025 (−0.043, 0.095)	76.02%
$\beta_{B:TTC_5}$	0.007 (−0.061, 0.075)	58.93%	0.011 (−0.058, 0.079)	64.07%
$\beta_{N:TTC_3}$	0.051 (−0.324, 0.425)	60.62%	0.058 (−0.310, 0.427)	62.25%
$\beta_{N:TTC_5}$	0.076 (−0.306, 0.450)	65.33%	0.076 (−0.298, 0.449)	65.65%
$\beta_{B:N:TTC_3}$	−0.026 (−0.119, 0.066)	71.03%	−0.028 (−0.119, 0.062)	72.38%
$\beta_{B:N:TTC_5}$	−0.022 (−0.115, 0.072)	68.82%	−0.022 (−0.116, 0.070)	68.50%
Participants	38		38	
Observations	706		706	

https://doi.org/10.1371/journal.pone.0355165.t003

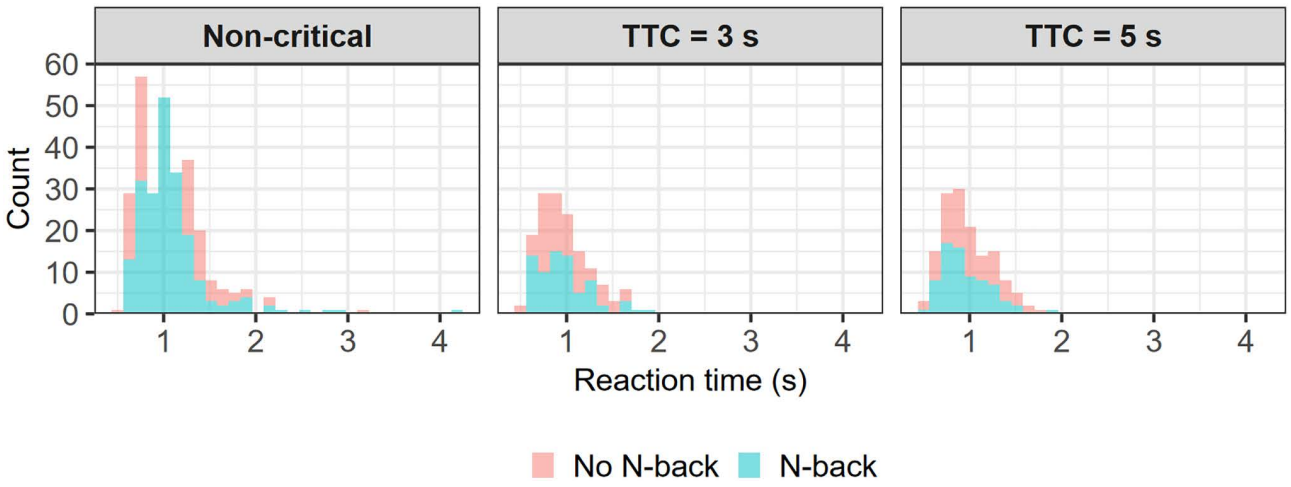


Fig 8. Distributions of raw reaction times with and without N-back for each event criticality.

https://doi.org/10.1371/journal.pone.0355165.g008

3.6 Post-hoc analysis: driver gaze during transitions of control

Finally, we conducted a post-hoc analysis of driver gaze distribution during transitions of control to understand where drivers were sampling visual information from. We calculated the SD of pitch angle during each of the three-time windows, and within each N-back and event criticality condition. Based on distributions of the raw data (see Fig 13), a Gaussian Bayesian model was fitted to investigate the effects. Once again, we used treatment coding for all categorical variables in the model. Thus we modelled SD of pitch angle as a function of an intercept (β_0), N-back (β_N), event criticality (β_{TTC_3}, β_{TTC_5}), time window ($\beta_{TW_{transition}}$, $\beta_{TW_{post-transition}}$), all two way interactions ($\beta_{N:transition}$, $\beta_{N:post-transntion}$, $\beta_{TTC_3:transition}$, $\beta_{TTC_5:transition}$, $\beta_{TTC_3:post-transition}$, $\beta_{TTC_5:post-transition}$ $\beta_{N:TTC_3}$, $\beta_{N:TTC_5}$), and three way interactions ($\beta_{N:TTC_3:transition}$, $\beta_{N:TTC_5:transition}$, $\beta_{N:TTC_3:post-transition}$, $\beta_{N:TTC_5:post-transition}$).

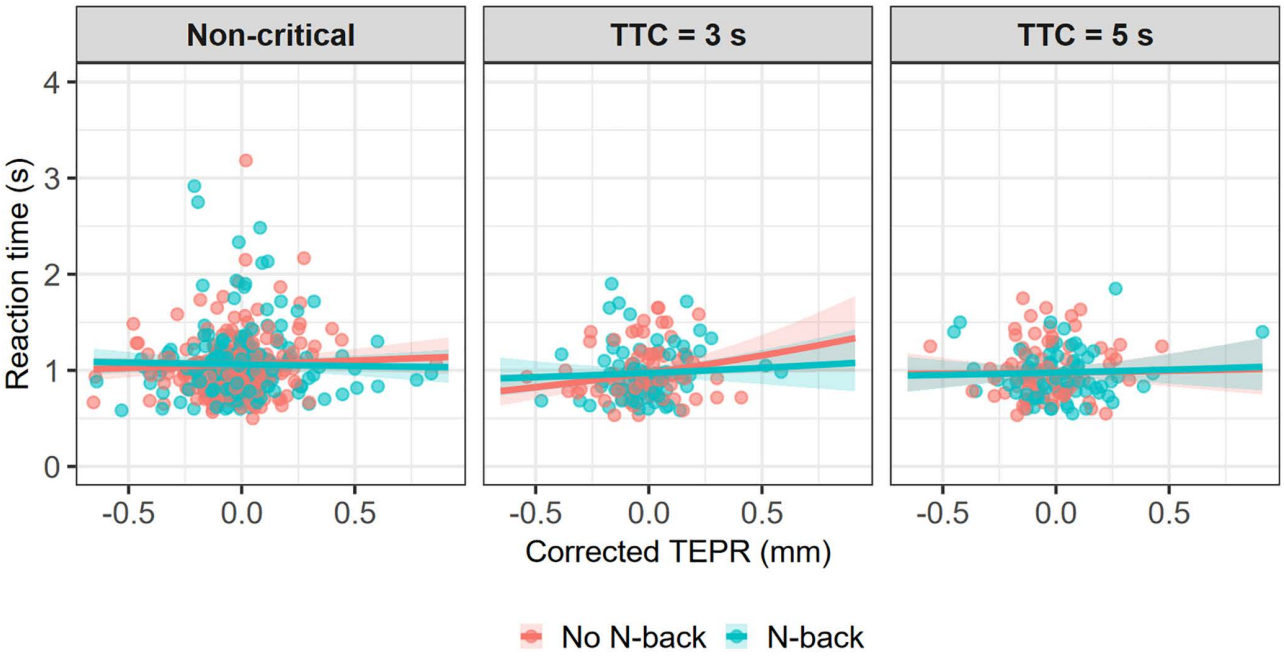


Fig 9. Relationship between TEPRs and reaction time within each N-back and event criticality condition.

https://doi.org/10.1371/journal.pone.0355165.g009

Table 4. Posterior means and 95% CIs for parameters predicting reaction time based on TEPRs.

	Coefficient (95% CI)	*pd*
β_0	0.032 (−0.023, 0.089)	88.63%
β_{TEPR}	0.073 (−0.090, 0.241)	80.35%
β_{TTC_3}	**−0.087 (−0.132, −0.042)**	**100%**
β_{TTC_5}	**−0.083 (−0.128, −0.038)**	**100%**
$\beta_{TEPR:N}$	−0.106 (−0.329, 0.115)	82.80%
$\beta_{TEPR:TTC_3}$	0.263 (−0.074, 0.602)	93.08%
$\beta_{TEPR:TTC_5}$	−0.043 (−0.379, 0.283)	59.60%
$\beta_{TEPR:TTC_3:N}$	−0.129 (−0.588, 0.318)	70.50%
$\beta_{TEPR:TTC_5:N}$	0.132 (−0.299, 0.558)	73.87%
Participants	38	
Observations	706	

https://doi.org/10.1371/journal.pone.0355165.t004

Modelling revealed two prominent interaction effects (see Table 7). Firstly, during the pre-RTI time window, SD of pitch angle was consistent for non-critical (3.397°), TTC = 3 s (3.333°), and TTC = 5 s (3.123°) conditions. However during the transition of control, SD of pitch angle increased to varying extents; the increase was largest for non-critical conditions (5.2289°) (2.191, [CI: 1.717, 2.656]) relative to TTC = 3 s conditions (4.258°) (0.927, [CI: 0.387, 1.542]) and TTC = 5 s conditions 3.854° (0.731, [CI: 0.107, 1.296]). Overall, this was indicative of drivers focusing more on the forward roadway during critical transitions control; conversely, drivers seemed to be looking down more during non-critical transitions, presumably looking for and obtaining vehicle status information (see Fig 14).

The second interaction effect was between N-back and time window. We found that during the pre-RTI phase, SD of pitch was lower during N-back conditions relative to no N-back conditions (−1.005, [CI: −1.303, −0.711]). This is consistent

Table 5. Posterior means and 95% CIs for parameters predicting reaction time based on M-PUI.

Parameters	50 ms time window		100 ms time window		200 ms time window		400 ms time window	
	Coefficient (95% CI)	***pd***	**Coefficient (95% CI)**	***pd***	**Coefficient (95% CI)**	***pd***	**Coefficient (95% CI)**	***pd***
β_0	0.010 (−0.049, 0.071)	63.42%	0.013 (−0.045, 0.072)	67.30%	0.018 (−0.039, 0.080)	71.97%	0.022 (−0.036, 0.079)	77.65%
β_{M-PUI}	0.569 (−0.161, 1.298)	94.05%	0.531 (−0.265, 1.306)	91.12%	0.483 (−0.313, 1.293)	87.87%	0.344 (−0.533, 1.226)	77.20%
$\boldsymbol{\beta_{TTC_3}}$	**−0.087 (−0.142, −0.034)**	**99.97%**	**−0.088 (−0.140, −0.037)**	**99.98%**	**−0.087 (−0.136, −0.036)**	**100.00%**	**−0.086 (−0.134, −0.040)**	**100.00%**
$\boldsymbol{\beta_{TTC_5}}$	**−0.069 (−0.126, −0.013)**	**99.27%**	**−0.071 (−0.124, −0.018)**	**99.57%**	**−0.074 (−0.124, −0.024)**	**99.75%**	**−0.077 (−0.125, −0.029)**	**99.88%**
$\beta_{M-PUI:N}$	0.007 (−0.594, 0.604)	51.23%	0.034 (−0.631, 0.701)	53.82%	0.055 (−0.685, 0.791)	56.17%	0.049 (−0.789, 0.881)	55.02%
$\beta_{M-PUI:TTC_3}$	0.065 (−0.746, 0.872)	56.12%	0.075 (−0.794, 0.949)	57.10%	0.031 (−0.868, 0.926	52.80%	0.011 (−0.941, 0.934)	51.03%
$\beta_{M-PUI:TTC_5}$	−0.023 (−0.892, 0.852)	52.90%	−0.023 (−0.894, 0.859)	52.03%	−0.049 (−0.936, 0.842)	53.87%	−0.012 (−0.979, 0.920)	50.97%
$\beta_{M-PUI:TTC_3:N}$	−0.078 (−0.892, 0.735)	57.20%	−0.043 (−0.911, 0.796)	53.62%	−0.048 (−0.950, 0.819)	53.73%	−0.038 (−1.000, 0.905)	52.33%
$\beta_{M-PUI:TTC_5:N}$	−0.467 (−1.297, 0.346)	87.15%	−0.443 (−1.295, 0.423)	84.42%	−0.367 (−1.269, 0.512)	79.10%	−0.203 (−1.133, 0.738)	66.32%

https://doi.org/10.1371/journal.pone.0355165.t005

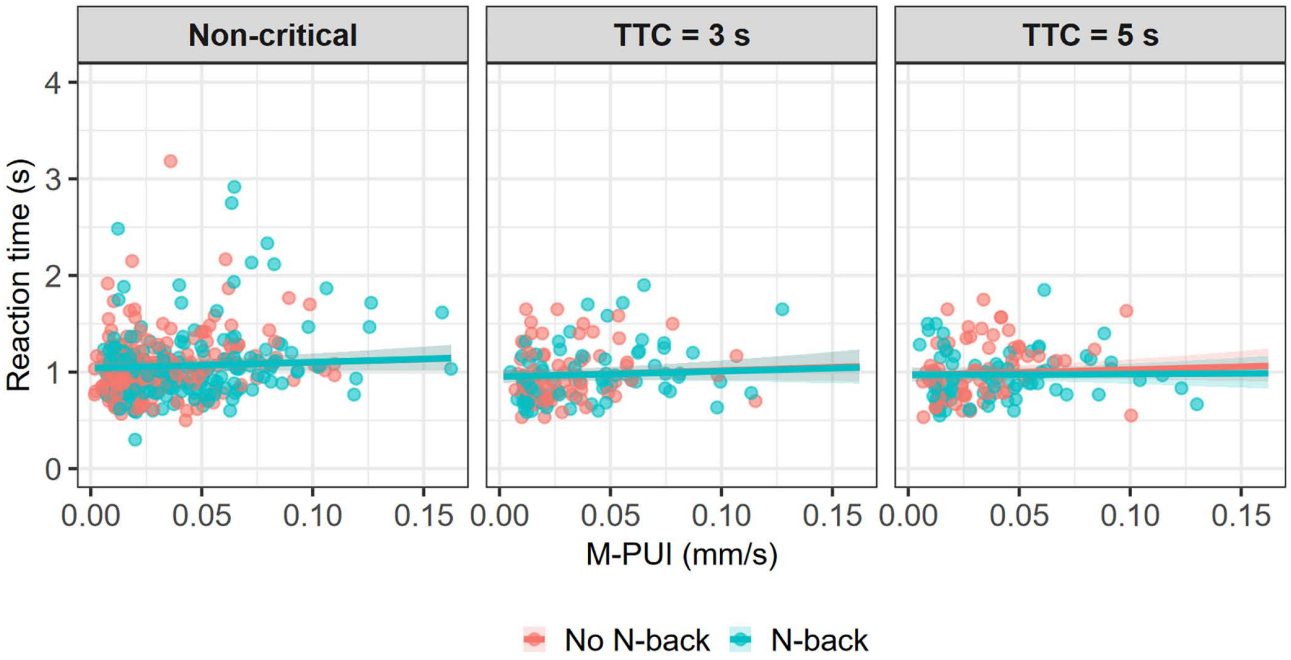


Fig 10. Relationship between M-PUI (50 ms time window) and reaction time within each N-back and event criticality condition.

https://doi.org/10.1371/journal.pone.0355165.g010

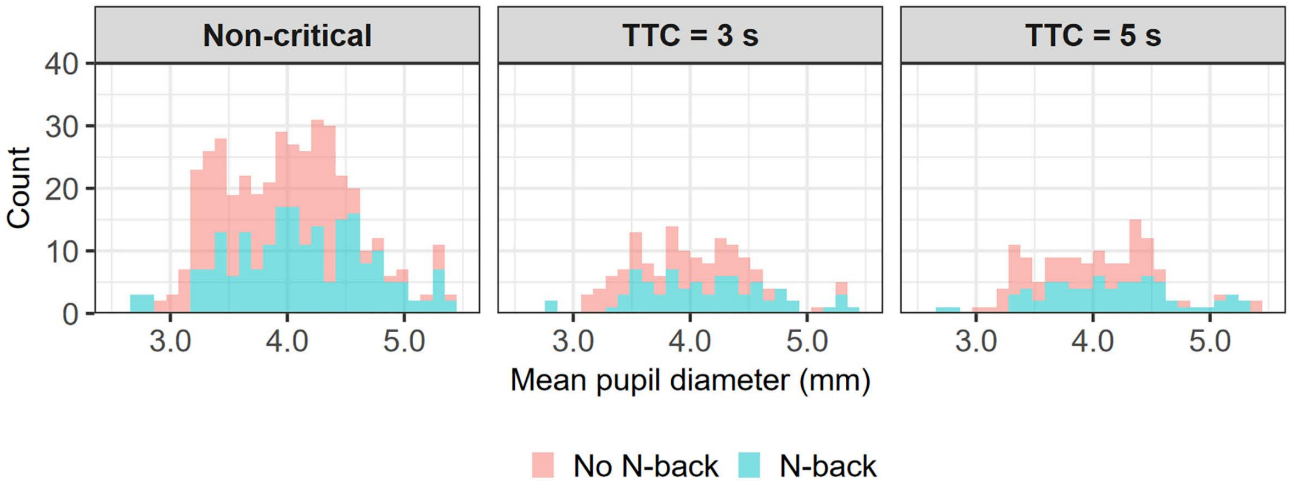


Fig 11. Distributions of mean pupil diameter during transitions of control with and without N-back for each event criticality.

https://doi.org/10.1371/journal.pone.0355165.g011

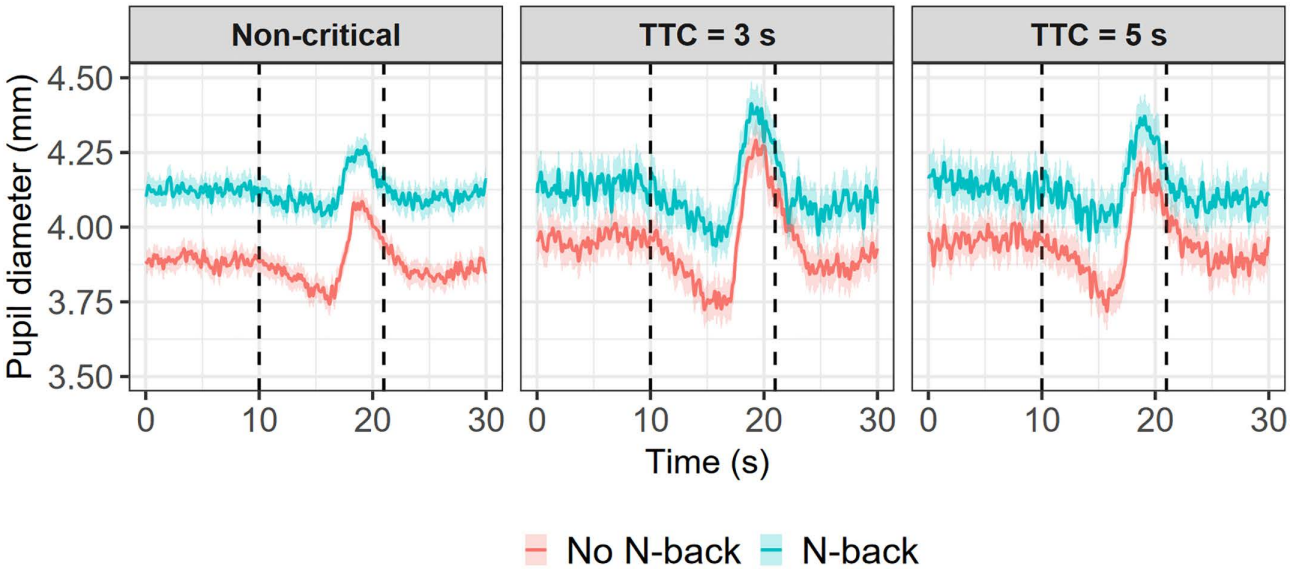


Fig 12. Average pupil diameter over time pre-RTI (0-10 s), during the transitions of control (10-21 s), and post transition of control (21-30 s).

https://doi.org/10.1371/journal.pone.0355165.g012

with evidence indicating that drivers were visually tunneling during N-back trials. However, during the transition of control (−0.535, [CI: −0.837, −0.226]), and post-transition (−0.105, [CI: −0.398, 0.198]) time windows, the difference between N-back and no N-back conditions were much smaller, once again indicating that drivers were looking down more during these phases regardless of their level of cognitive load.

4. Discussion

This investigation examined pupillary dynamics during hands-off L2 driving and transitions of control. Modelling revealed that mean pupil diameter was a more reliable indicator of cognitive load than the SD of pupil diameter; mean pupil

Table 6. Posterior means and 95% CIs for parameters predicting tonic pupil diameter increases.

	Coefficient (95% CI)	*pd*
β_0	3.857 (3.675, 4.036)	100%
β_N	0.282 (0.201, 0.366)	100%
β_{TTC_3}	0.065 (0.024, 0.105)	99.90%
β_{TTC_5}	0.075 (0.035, 0.112)	99.98%
$\beta_{N:TTC_3}$	−0.065 (−0.118, −0.014)	99.85%
$\beta_{N:TTC_5}$	−0.069 (−0.121, −0.017)	99.28%
Participants	38	
Observations	707	

https://doi.org/10.1371/journal.pone.0355165.t006

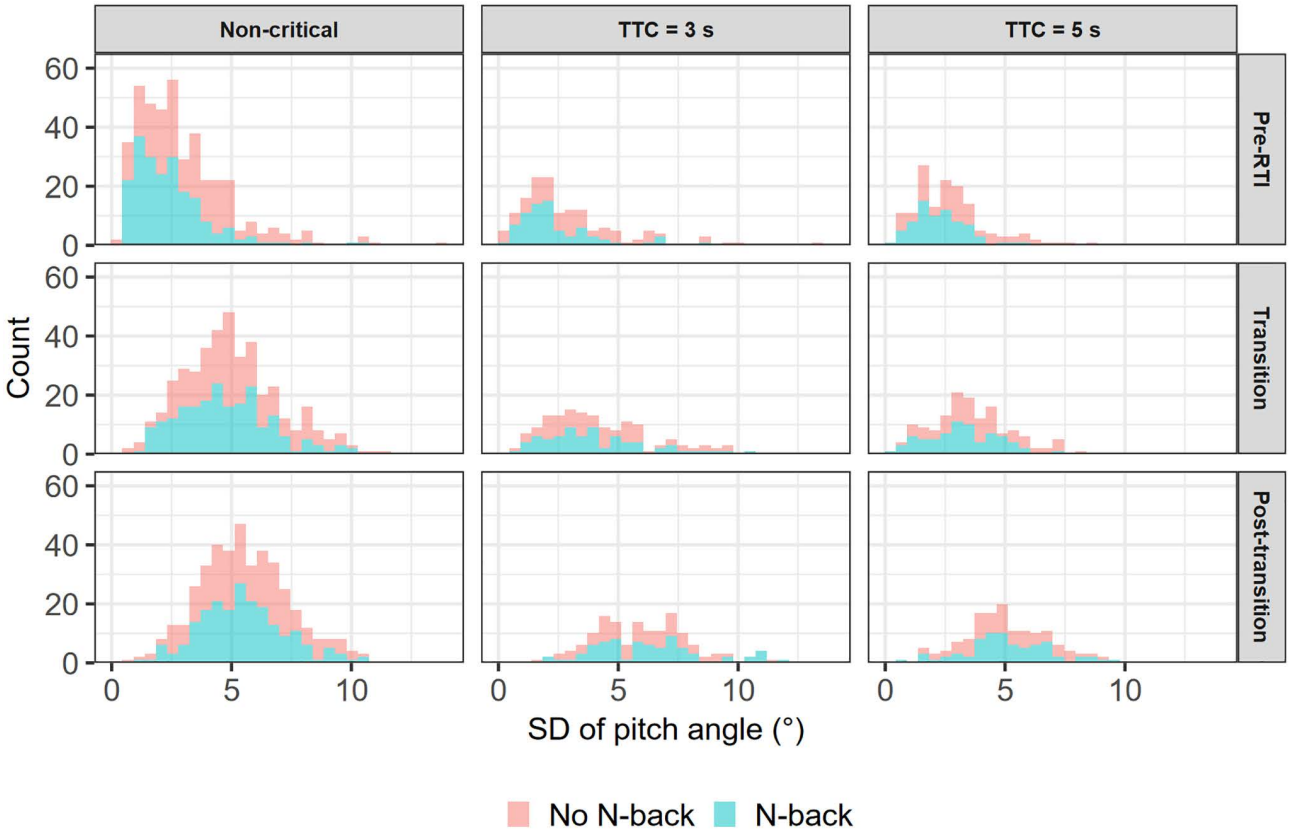


Fig 13. Distributions of the SD of pitch angle each time window with and without N-back for each event criticality.

https://doi.org/10.1371/journal.pone.0355165.g013

diameter increased during N-back trials, with 90% of the population predicted to show similar increases, compared to only 82% for the SD. We also examined three relationships between pupillary dynamics and behaviour established in controlled laboratory paradigms. A robust negative relationship between baseline pupil diameter and TEPRs was found, and crucially, this persisted after correcting for regression to the mean artefacts – suggesting that pre-stimulus pupillary state genuinely constrains subsequent evoked responsiveness rather than the correlation being a statistical consequence of physiological boundaries. This pattern is consistent with theoretical accounts linking tonic arousal state to the

Table 7. Posterior means and 95% CIs for parameters predicting SD of pitch angle.

Parameters	Coefficient (95% CI)	*pd*
β_0	**3.397 (3.027, 3.755)**	**100.00%**
β_N	**−1.005 (−1.306, −0.713)**	**100.00%**
β_{TTC_3}	−0.066 (−0.480, 0.347)	62.08%
β_{TTC_5}	−0.274 (−0.700, 0.152)	89.98%
$\beta_{TW_{transition}}$	**1.831 (1.413, 2.257)**	**100.00%**
$\beta_{TW_{post-transition}}$	**2.191 (1.723, 2.663)**	**100.00%**
$\beta_{N:TTC_3}$	0.075 (−0.510, 0.664)	60.23%
$\beta_{N:TTC_5}$	0.189 (−0.423, 0.781)	73.85%
$\beta_{N:transition}$	**0.469 (0.044, 0.896)**	**98.58%**
$\beta_{N:post-transition}$	**0.899 (0.493, 1.336)**	**100.00%**
$\beta_{TTC_3:transition}$	**−0.904 (−1.476, −0.321)**	**99.88%**
$\beta_{TTC_5:transition}$	**−1.099 (−1.697, −0.508)**	**100.00%**
$\beta_{TTC_3:post-transition}$	0.370 (−0.211, 0.967)	89.68%
$\beta_{TTC_5:post-transition}$	−0.274 (−0.862, 0.325)	81.78%
$\beta_{N:TTC_3:transition}$	0.235 (−0.563, 1.079)	70.78%
$\beta_{N:TTC_5:transition}$	−0.297 (−1.132, 0.527)	75.15%
$\beta_{N:TTC_3:post-transition}$	0.347 (−0.460, 1.173)	80.30%
$\beta_{N:TTC_5:post-transition}$	−0.045 (−0.885, 0.813)	54.70%
Participants	38	
Observations	2118	

https://doi.org/10.1371/journal.pone.0355165.t007

magnitude of stimulus-evoked neural responses, including accounts that implicate LC-NE dynamics in regulating the balance between sustained and evoked pupillary activity [35,54,86,87]. However, we found limited evidence that TEPRs or pre-stimulus pupillary variability predicted driver reaction times, likely reflecting the motoric variability introduced by naturalistic driving responses. Finally, more critical transitions of control were associated with larger increases in pupil diameter, consistent with heightened arousal and effortful task re-engagement following the RTI. A post-hoc analysis additionally revealed that drivers directed their gaze more towards the forward roadway during critical transitions, whereas non-critical transitions were associated with more time spent looking towards the dashboard area, presumably to obtain vehicle status information – a pattern consistent with prior research on driver gaze behaviour during transitions of control [76,88,89].

4.1 Pupillometry during hands-off L2 driving

Assessing whether an average effect is consistent across a population is important given that understanding variations in driver cognitive load is vital if DMS are to reliably and accurately detect driver state. For example, according to the European New Car Assessment Programme (Euro NCAP) 2026, as a minimum requirement, Forward Collision Warning (FCW) and/or Automated Emergency Braking (AEB) systems should differ by a least 200 ms between a distracted/impaired driver and an attentive driver in order to account for longer driver reaction times. The analysis reported here suggests that the SD of pupil diameter was not a reliable indicator of impairment in this context; and that despite an average increase, almost a fifth of the population would be expected to demonstrate null or opposite effects under conditions of high load. This is problematic because an algorithm detecting positive changes in the SD of pupil diameter may incorrectly conclude that no changes in cognitive load are present during periods of high cognitive load (i.e., false negatives).

Overall, these findings suggest that mean pupil diameter is a more robust candidate than short-term variability measures for monitoring changes associated with increased cognitive demand during hands-off L2 driving. Importantly, this

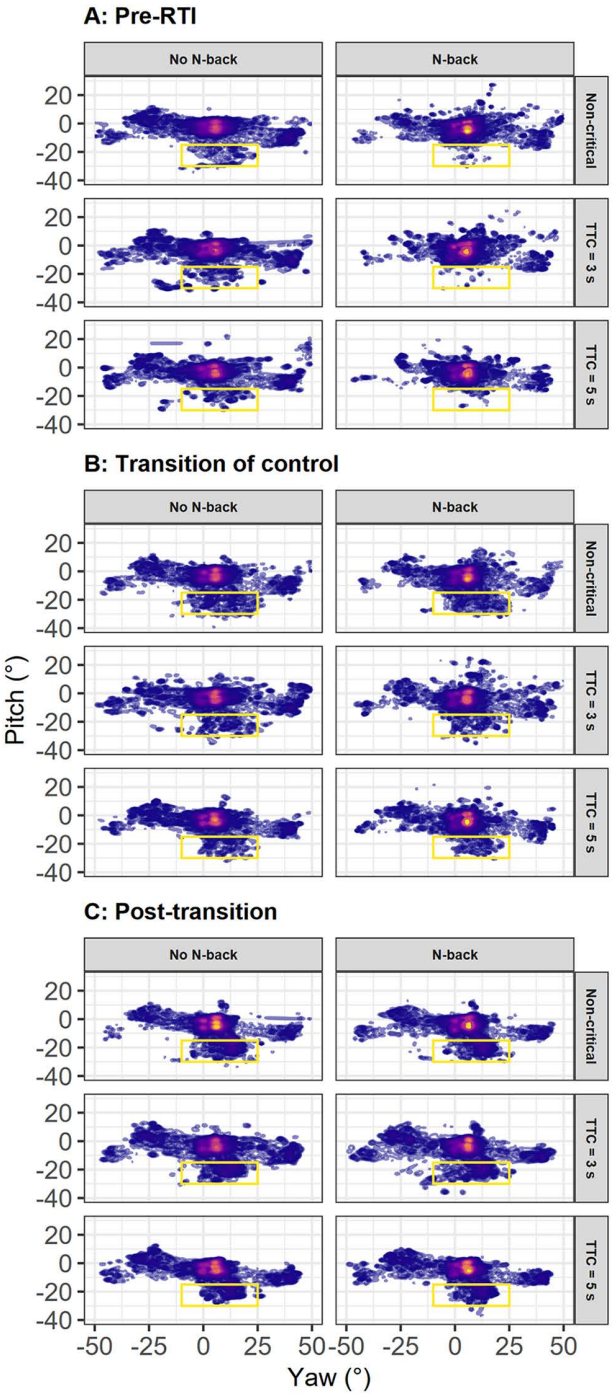


Fig 14. Contour plots show the average distribution of gaze during A) the 10 s prior to the RTI, B) the transition of control, and C) the 10 s following the transition of control. The yellow box highlights approximate location of the dashboard area. Drivers appeared to gaze more towards the dash area during the transition and post-transition phase relative to the pre-RTI phase.

https://doi.org/10.1371/journal.pone.0355165.g014

should not be interpreted as indicating that larger pupils necessarily reflect higher absolute arousal in all circumstances, as pupil diameter is influenced by multiple physiological and environmental factors [90]. Rather, under controlled luminance conditions, the results demonstrate that mean pupil diameter provides a more consistent marker of changes associated with increased cognitive demands than the SD of pupil diameter. For practical DMS applications, this suggests that algorithms should prioritise sustained changes in pupil diameter, while combining pupillometry with environmental sensing to account for non-cognitive influences on pupil size [91,92].

4.2 TEPRs in an applied context

Larger pre-stimulus pupil diameter was negatively associated with subsequent TEPRs, and this association survived correction for regression to the mean artefacts. This finding is consistent with AGT – a unifying account linking arousal, pupil size, and behavioural performance via LC-NE dynamics [53]. AGT proposes that the strength of task-evoked (phasic) LC responses depends on baseline (tonic) LC activity. When tonic LC activity is low, task-related events elicit relatively weak phasic responses, corresponding to reduced engagement and poorer performance typically observed in low arousal states. At intermediate tonic levels, phasic LC responses are maximised, enhancing neural gain and supporting optimal task performance. In contrast, elevated tonic LC activity attenuates phasic responses while promoting more indiscriminate increases in neural gain, resulting in reduced selectivity and greater distractibility. Controlled laboratory studies have found similar correlations between pre-stimulus (tonic) pupil size and TEPRs [61,63,81]. We found similar effects in a more applied task setting, a pattern *consistent with* the tonic-phasic trade-off predicted by AGT, though the present data cannot confirm LC-NE involvement directly.

Previous research had also reported negative correlations between TEPRs and response times [53,61,63,65,93], suggesting that larger TEPRs were associated with faster responses. This relationship is thought to reflect NE release from LC projections, increasing the gain of neural interactions and facilitating more efficient information processing [53,93], which would be expected to reduce response latency. However, the present study found no relationship between TEPRs and reaction time. One possible explanation is the presence of multiple potential takeover actions. These responses (e.g., braking, steering, or pressing a button) differ in their minimum motor execution latencies, as reflected in the slightly slower reaction times observed for non-critical transitions of control (which were predominantly initiated via a button press rather than operational pedals). Such variability in motor demands may have obscured any underlying relationship between pupillary dynamics and reaction time. In contrast, controlled laboratory studies typically require participants to respond in highly controlled and uniform ways. Although allowing drivers to respond naturally increases ecological validity, it also introduces motoric variability that may reduce sensitivity to detect associations between TEPRs and behavioural responses.

Evidence for an association between M-PUI and reaction times was limited. A small relationship was observed when M-PUI was computed over short time windows, but this effect attenuated as window length increased. This pattern is broadly consistent with prior findings linking elevated M-PUI to slower responses in psychomotor vigilance tasks [68] and suggests that M-PUI may be most sensitive to transient fluctuations in vigilance. One interpretation is that shorter windows better capture moment-to-moment variability in arousal-related processes, whereas longer windows may average over these fluctuations, reducing sensitivity to their behavioural consequences. Similarly to the relationships between reaction times and TEPRs, the relatively weak effects observed in the present driving context may also reflect increased variability in behavioural responses. Unlike psychomotor vigilance tasks, which typically involve simple button presses, driving responses involve more complex and variable motor actions, which may reduce the detectability of subtle vigilance-related influences on reaction time.

More broadly, the present findings demonstrate that relationships between baseline pupil diameter and subsequent TEPRs reported in tightly controlled laboratory paradigms also emerge during transitions of control in automated driving. However, these data should not be interpreted as identifying specific arousal states or confirming the operation of LC-NE

mechanisms. Rather, they indicate that pre-stimulus pupillary state constrains subsequent pupillary responsiveness in a predictable manner. For DMS development, this suggests that interpreting TEPRs without accounting for baseline pupil diameter may underestimate or overestimate responsiveness to RTIs. Incorporating baseline correction into TEPR-based algorithms may therefore improve the interpretation of task-evoked pupillary responses.

4.3 Pupillometry during transitions of control

Consistent with prior work demonstrating increased pupil diameter during transitions of control [60], we found that more critical transitions – manipulated via TTC – were associated with larger increases in pupil diameter. Interestingly, this effect was observed only in the no N-back condition. One possible explanation for this is a ceiling effect: the N-back manipulation increased overall pupil size, potentially limiting further dilation during critical transitions due to physiological constraints. Crucially, the present design does not allow us to determine whether increases associated with event criticality reflect elevated cognitive load specifically, or more general physiological arousal associated with critical events. Previous research suggests that responses to critical rear-end scenarios may be highly automatised [18], implying that the earliest behavioural reactions are unlikely to require substantial cognitive control. In this context, increased pupil diameter could plausibly reflect sympathetic activation – a "fight-or-flight" response triggered by imminent collision threat. Supporting this interpretation, pupil diameter has been shown to predict accident severity in real-world contexts [94], suggesting sensitivity to acute threat. Notably, however, the pupil increases observed in the present study emerged several seconds after the transition of control. This delayed elevation is less consistent with a purely reflexive sympathetic response and may instead reflect the sustained effort required to re-engage with the dynamic driving task after monitoring a hands-off L2 system. From this perspective, critical transitions may demand increased cognitive control, error monitoring, and vigilance, resulting in sustained arousal adjustments rather than brief phasic responses.

Taken together, these findings suggest that sustained pupil diameter during the transition itself may provide a useful physiological indicator of the overall demands imposed by a takeover event. However, these increases should not be interpreted as direct measures of arousal, but rather as reflecting the combined cognitive and behavioural demands associated with managing the transition. For practical applications, this suggests that DMS could benefit from monitoring changes in tonic pupil diameter across the takeover period, rather than relying solely on pre-transition measurements, to identify situations in which drivers are experiencing elevated task demands.

4.4 Future work and conclusions

Although there is a close relationship between pupil diameter and LC activity, the mechanism is not direct and involves many interconnected pathways [47]. Future research may want to confirm this relationship through functional neuroimaging techniques such as electroencephalography (EEG) or functional Near-Infrared Spectroscopy (fNIRs). Whilst fNIRs does not have the spatial resolution to image the LC, the LC does extensively innervate the cerebral cortex, including the dorsolateral and dorsomedial prefrontal cortex (PFC), which fNIRs has access to [53,95–97]. Hence the combined use of fNIRs and pupillometry could elucidate the interactions between the LC and the PFC in applied tasks. The P300 event related potential (ERP) has been proposed as an electrophysiological reflection of the phasic activity of the LC-NE system [98]. There is evidence to suggest that pupil diameter and the P300 amplitude are sensitive to similar experimental parameters [99,100]. Large P300 amplitudes are associated with large TEPRs during oddball tasks [101,102]. This is particularly interesting given that P300 amplitudes are seen to be *dampened* during auditory oddballs presented during states of high cognitive load in manual and automated driving [103–106]. Taken in context with the current results, there appears to be circumstantial evidence that phasic pupillary responses and the P300 component could be associated in a driving context; albeit more robust evidence would be needed to confirm this. This association may have safety related implications given that high cognitive load has also been found to reduce P300 amplitudes towards driving related sensory information

(e.g., lead vehicle brake light onset) [103]. If TEPRs demonstrate similar behaviour, this could be a useful non-invasive indication as to whether a driver has attended to, and encoded, task-relevant information in the road environment during non-critical and critical situations.

Future work may also want to further investigate the impact that drowsiness has on pupillometry. The Pupillary Unrest Index (PUI) is a common measure of sleepiness that uses pupillometry, however it requires a standardised experimental procedure for calculation [107]. It might be interesting to understand whether M-PUI – which is calculated over much smaller timescales – is associated with drowsiness. This could be operationalised in a number of ways. For example, studies may want to investigate whether M-PUI is impacted by the length of time participants are using an automated system given that extended use may result in reduced alertness [108,109]. Similarly, the relationship between M-PUI and reliable subjective (the Karolinska Sleepiness Scale [KSS]) and objective (percentage of eye closure [PERCLOS], and amplitude velocity ratio [AVR]) measures of alertness could be assessed [110–113].

In conclusion, this study demonstrates that pupillometry can provide a sensitive physiological measure of changes associated with cognitive demands during hands-off L2 driving and transitions of control. Mean pupil diameter reliably distinguished periods of increased cognitive load, while baseline pupil diameter consistently constrained the magnitude of subsequent TEPRs, even after accounting for regression to the mean artefacts. Importantly, these findings should not be interpreted as direct evidence of specific arousal states or LC-NE activity. Rather, they show that well-established relationships between different pupillary measures observed in controlled laboratory paradigms also generalise, at least in part, to a complex applied driving task. From a practical perspective, the results suggest that DMS could benefit from combining sustained pupil diameter with baseline-adjusted TEPRs when interpreting driver state, while integrating these measures with other behavioural and environmental information. More broadly, this work helps define what pupillometry can – and cannot – tell us about DR, providing a more principled foundation for the development of physiologically informed monitoring systems for road vehicles.

Author contributions

Conceptualization: Courtney M. Goodridge, Audrey Bruneau, Jonny Kuo, Michael G. Lenné, Yee Mun Lee, Natasha Merat.

Data curation: Courtney M. Goodridge, Rafael C. Gonçalves, Ali Arabian.

Formal analysis: Courtney M. Goodridge.

Funding acquisition: Natasha Merat.

Investigation: Courtney M. Goodridge, Yee Thung Lee.

Methodology: Courtney M. Goodridge, Rafael C. Gonçalves, Audrey Bruneau, Jonny Kuo, Michael G. Lenné, Yee Mun Lee.

Project administration: Courtney M. Goodridge, Natasha Merat.

Resources: Albert Solernou, Natasha Merat.

Software: Courtney M. Goodridge, Anthony Horrobin, Albert Solernou, Jonny Kuo, Michael G. Lenné.

Supervision: Albert Solernou, Natasha Merat.

Validation: Courtney M. Goodridge.

Visualization: Courtney M. Goodridge.

Writing – original draft: Courtney M. Goodridge.

Writing – review & editing: Courtney M. Goodridge, Rafael C. Gonçalves, Audrey Bruneau, Gaëtan Merlhiot, Yee Mun Lee, Natasha Merat.

References

1. Huang R, Clewett D. The locus coeruleus: where cognitive and emotional processing meet the eye. In: Papesh MH, Goldinger SD, editors. Modern pupillometry [Internet]. Cham: Springer International Publishing; 2024 [cited 2026 Apr 21]. p. 3–75. Available from: https://link.springer.com/10.1007/978-3-031-54896-3_1
2. Kahneman D. Attention and effort [Internet]; 1973 [cited 2026 Apr 21]. Available from: https://philpapers.org/rec/kahaae
3. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation [Internet]; 1908 [cited 2026 Apr 21]. Available from: http://psychclassics.yorku.ca/Yerkes/Law/
4. Easterbrook JA. The effect of emotion on cue utilization and the organization of behavior. Psychol Rev. 1959;66(3):183–201. https://doi.org/10.1037/h0047707 PMID: 13658305
5. Grueschow M, Kleim B, Ruff CC. Role of the locus coeruleus arousal system in cognitive control. J Neuroendocrinol. 2020;32(12):e12890. https://doi.org/10.1111/jne.12890 PMID: 32820571
6. Verguts T, Notebaert W. Adaptation by binding: a learning account of cognitive control. Trends Cogn Sci. 2009;13(6):252–7. https://doi.org/10.1016/j.tics.2009.02.007 PMID: 19428288
7. de Waard D, Brookhuis KA. Assessing driver status: a demonstration experiment on the road. Accid Anal Prev. 1991;23(4):297–307. https://doi.org/10.1016/0001-4575(91)90007-r PMID: 1883469
8. Wiener EL, Curry RE, Faustina ML. Vigilance and task load: in search of the inverted U. Hum Factors J Hum Factors Ergon Soc. 1984;26(2):215–22. https://doi.org/10.1177/001872088402600208 PMID: 6479984
9. Matthews G, Desmond PA. Underload and performance impairment: evidence from studies of stress and simulated driving. In: Engineering psychology and cognitive ergonomics [Internet]. Routledge; 2017 [cited 2026 Apr 21]. p. 355–61. Available from: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315094496-39/underload-performance-impairment-evidence-studies-stress-simulated-driving-gerald-matthews-paula-desmond
10. Murata A, Doi T, Karwowski W. The effects of countermeasures for preventing vigilance decrement during manual and automated driving. IEEE Access. 2025;13:88648–61. https://doi.org/10.1109/access.2025.3570073
11. Engström J, Markkula G, Victor T, Merat N. Effects of cognitive load on driving performance: the cognitive control hypothesis. Hum Factors. 2017;59(5):734–64. https://doi.org/10.1177/0018720817690639 PMID: 28186421
12. McWilliams T, Ward N. Underload on the road: measuring vigilance decrements during partially automated driving. Front Psychol. 2021;12:631364. https://doi.org/10.3389/fpsyg.2021.631364 PMID: 33935882
13. Young MS, Stanton NA. Attention and automation: new perspectives on mental underload and performance. Theor Issues Ergon Sci. 2002;3(2):178–94. https://doi.org/10.1080/14639220210123789
14. Dunn NJ, Dingus TA, Soccolich S, Horrey WJ. Investigating the impact of driving automation systems on distracted driving behaviors. Accid Anal Prev. 2021;156:106152. https://doi.org/10.1016/j.aap.2021.106152 PMID: 33932819
15. Noble AM, Miles M, Perez MA, Guo F, Klauer SG. Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. Accid Anal Prev. 2021;151:105959. https://doi.org/10.1016/j.aap.2020.105959 PMID: 33385962
16. Reagan IJ, Teoh ER, Cicchino JB, Gershon P, Reimer B, Mehler B, et al. Disengagement from driving when using automation during a 4-week field trial. Transp Res Part F Traffic Psychol Behav. 2021;82:400–11. https://doi.org/10.1016/j.trf.2021.09.010
17. Zhang W, Carsten O, Goodridge CM, Zhang W. Effects of adopting voice assistants countermeasures on alerting drowsy drivers across age. Appl Ergon. 2025;129:104616. https://doi.org/10.1016/j.apergo.2025.104616 PMID: 40780011
18. Goodridge CM, Gonçalves RC, Arabian A, Horrobin A, Solernou A, Lee YT, et al. The impact of N-back-induced mental workload and time budget on takeover performance. Accid Anal Prev. 2026;225:108327. https://doi.org/10.1016/j.aap.2025.108327 PMID: 41297089
19. Mueller AS, Cicchino JB, Calvanelli JV Jr. Habits, attitudes, and expectations of regular users of partial driving automation systems. J Safety Res. 2024;88:125–34.
20. Choi D, Sato T, Ando T, Abe T, Akamatsu M, Kitazaki S. Effects of cognitive and visual loads on driving performance after take-over request (TOR) in automated driving. Appl Ergon. 2020;85:103074. https://doi.org/10.1016/j.apergo.2020.103074 PMID: 32174362
21. Louw T, Markkula G, Boer E, Madigan R, Carsten O, Merat N. Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. Accid Anal Prev. 2017;108:9–18. https://doi.org/10.1016/j.aap.2017.08.011 PMID: 28837837
22. Merlhiot G, Bueno M. How drowsiness and distraction can interfere with take-over performance: a systematic and meta-analysis review. Accid Anal Prev. 2022;170:106536. https://doi.org/10.1016/j.aap.2021.106536 PMID: 34969510
23. Gonçalves RC, Louw TL, Madigan R, Quaresma M, Romano R, Merat N. The effect of information from dash-based human-machine interfaces on drivers' gaze patterns and lane-change manoeuvres after conditionally automated driving. Accid Anal Prev. 2022;174:106726. https://doi.org/10.1016/j.aap.2022.106726 PMID: 35716544
24. Salvucci DD, Liu A, Boer ER. Control and monitoring during lane changes. Vis Veh [Internet]; 2001 [cited 2026 Apr 21];9. Available from: http://web.mit.edu/amliu/www/Papers/VIV9-SalLiuBoer.pdf
25. Tijerina L, Garrott WR, Stoltzfus D, Parmer E. Eye glance behavior of van and passenger car drivers during lane change decision phase. Transp Res Rec J Transp Res Board. 2005;1937(1):37–43. https://doi.org/10.1177/0361198105193700106

26. Georg JM, Hecht T, Lienkamp M, Wang C. Development of a human driver model during highly automated driving for the ASIL controllability classification. 8. Tagung Fahrerassistenz [Internet]; 2017 [cited 2026 Apr 21]. Available from: https://mediatum.ub.tum.de/1421389
27. Gonçalves RC, Goodridge CM, Kuo J, Lenné MG, Merat N. Using driver monitoring to estimate readiness in automation: a conceptual model based on simulator experimental data. Cogn Tech Work. 2024;26(4):569–84. https://doi.org/10.1007/s10111-024-00777-3
28. Kim J, Kim W, Kim H-S, Lee S-J, Kwon O-C, Yoon D. A novel study on subjective driver readiness in terms of non-driving related tasks and take-over performance. ICT Express. 2022;8(1):91–6. https://doi.org/10.1016/j.icte.2021.04.008
29. Mioch T, Kroon L, Neerincx MA. Driver readiness model for regulating the transfer from automation to human control. Proceedings of the 22nd International Conference on Intelligent User Interfaces [Internet]. Limassol Cyprus: ACM; 2017 [cited 2026 Apr 21]. p. 205–13. Available from: https://dl.acm.org/doi/10.1145/3025171.3025199
30. Mariajoseph M, Gallina B, Carli M, Bibbo D. A physiology-based driver readiness estimation model for tuning ISO 26262 controllability. 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring) [Internet]. IEEE; 2020 [cited 2026 Apr 21]. p. 1–5. Available from: https://ieeexplore.ieee.org/abstract/document/9129132/
31. Hess EH, Polt JM. Pupil size as related to interest value of visual stimuli. Science. 1960;132(3423):349–50. https://doi.org/10.1126/science.132.3423.349
32. Hess EH, Polt JM. Pupil size in relation to mental activity during simple problem-solving. Science. 1964;143(3611):1190–2. https://doi.org/10.1126/science.143.3611.1190 PMID: 17833905
33. van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: a review. Psychon Bull Rev. 2018;25(6):2005–15. https://doi.org/10.3758/s13423-018-1432-y PMID: 29435963
34. Cavanagh JF, Wiecki TV, Kochar A, Frank MJ. Eye tracking and pupillometry are indicators of dissociable latent decision processes. J Exp Psychol Gen. 2014;143(4):1476–88. https://doi.org/10.1037/a0035813 PMID: 24548281
35. Murphy PR, O'Connell RG, O'Sullivan M, Robertson IH, Balsters JH. Pupil diameter covaries with BOLD activity in human locus coeruleus. Hum Brain Mapp. 2014;35(8):4140–54. https://doi.org/10.1002/hbm.22466 PMID: 24510607
36. Braem S, Coenen E, Bombeke K, van Bochove ME, Notebaert W. Open your eyes for prediction errors. Cogn Affect Behav Neurosci. 2015;15(2):374–80. https://doi.org/10.3758/s13415-014-0333-4 PMID: 25588818
37. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci. 2012;15(7):1040–6. https://doi.org/10.1038/nn.3130 PMID: 22660479
38. Cheadle S, Wyart V, Tsetsos K, Myers N, de Gardelle V, Herce Castañón S, et al. Adaptive gain control during human perceptual choice. Neuron. 2014;81(6):1429–41. https://doi.org/10.1016/j.neuron.2014.01.020 PMID: 24656259
39. van den Brink RL, Murphy PR, Nieuwenhuis S. Pupil diameter tracks lapses of attention. PLoS One. 2016;11(10):e0165274. https://doi.org/10.1371/journal.pone.0165274 PMID: 27768778
40. Mathôt S, Dalmaijer E, Grainger J, Van der Stigchel S. The pupillary light response reflects exogenous attention and inhibition of return. J Vis. 2014;14(14):7. https://doi.org/10.1167/14.14.7 PMID: 25761284
41. Mathôt S, Vilotijević A. Methods in cognitive pupillometry: design, preprocessing, and statistical analysis. Behav Res Methods. 2023;55(6):3055–77. https://doi.org/10.3758/s13428-022-01957-7 PMID: 36028608
42. Zénon A. Eye pupil signals information gain. Proc R Soc B Biol Sci [Internet]. 2019 [cited 2026 Apr 21];286(1911). Available from: https://royalsocietypublishing.org/rspb/article/286/1911/20191593/85281
43. Strauch C, Wang C-A, Einhäuser W, Van der Stigchel S, Naber M. Pupillometry as an integrated readout of distinct attentional networks. Trends Neurosci. 2022;45(8):635–47. https://doi.org/10.1016/j.tins.2022.05.003 PMID: 35662511
44. Huerta MF, Krubitzer LA, Kaas JH. Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: I. Subcortical connections. J Comp Neurol. 1986;253(4):415–39. https://doi.org/10.1002/cne.902530402 PMID: 3793998
45. Leichnetz GR. Preoccipital cortex receives a differential input from the frontal eye field and projects to the pretectal olivary nucleus and other visuomotor-related structures in the rhesus monkey. Vis Neurosci. 1990;5(2):123–33. https://doi.org/10.1017/s095252380000016x PMID: 2177637
46. Wurtz RH, Albano JE. Visual-motor function of the primate superior colliculus. Annu Rev Neurosci. 1980;3:189–226. https://doi.org/10.1146/annurev.ne.03.030180.001201 PMID: 6774653
47. Joshi S, Gold JI. Pupil size as a window on neural substrates of cognition. Trends Cogn Sci. 2020;24(6):466–80. https://doi.org/10.1016/j.tics.2020.03.005 PMID: 32331857
48. Berridge CW. Noradrenergic modulation of arousal. Brain Res Rev. 2008;58(1):1–17. https://doi.org/10.1016/j.brainresrev.2007.10.013 PMID: 18199483
49. Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from environment to theory of mind. Neuron. 2008;58(3):306–24. https://doi.org/10.1016/j.neuron.2008.04.017 PMID: 18466742
50. Sara SJ, Bouret S. Orienting and reorienting: the locus coeruleus mediates cognition through arousal. Neuron. 2012;76(1):130–41. https://doi.org/10.1016/j.neuron.2012.09.011 PMID: 23040811

51. Alnæs D, Sneve MH, Espeseth T, Endestad T, van de Pavert SHP, Laeng B. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. J Vis. 2014;14(4):1. https://doi.org/10.1167/14.4.1 PMID: 24692319

52. Samuels ER, Szabadi E. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. Curr Neuropharmacol. 2008;6(3):235–53. https://doi.org/10.2174/157015908785777229 PMID: 19506723

53. Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci. 2005;28(1):403–50. https://doi.org/10.1146/annurev.neuro.28.061604.135709

54. Joshi S, Li Y, Kalwani RM, Gold JI. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. Neuron. 2016;89(1):221–34. https://doi.org/10.1016/j.neuron.2015.11.028 PMID: 26711118

55. Varazzani C, San-Galli A, Gilardeau S, Bouret S. Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. J Neurosci. 2015;35(20):7866–77. https://doi.org/10.1523/JNEUROSCI.0454-15.2015 PMID: 25995472

56. Liu Y, Rodenkirch C, Moskowitz N, Schriver B, Wang Q. Dynamic lateralization of pupil dilation evoked by locus coeruleus activation results from sympathetic, not parasympathetic, contributions. Cell Rep. 2017;20(13):3099–112. https://doi.org/10.1016/j.celrep.2017.08.094 PMID: 28954227

57. Palinko O, Kun AL, Shyrokov A, Heeman P. Estimating cognitive load using remote eye tracking in a driving simulator. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA'10 [Internet]. Austin (TX): ACM Press; 2010 [cited 2026 Apr 21]. 141 p. Available from: http://portal.acm.org/citation.cfm?doid=1743666.1743701

58. Bozkir E, Geisler D, Kasneci E. Assessment of driver attention during a safety critical situation in VR to generate VR-based training. ACM Symposium on Applied Perception 2019 [Internet]. Barcelona, Spain: ACM; 2019 [cited 2026 Jun 30]. p. 1–5. Available from: https://dl.acm.org/doi/10.1145/3343036.3343138

59. Lin Z, Chen F, Öztürk İ, Merat N. Using psychophysiological metrics for understanding drivers' mental workload and visual attention when overtaking automated truck platoons. Accid Anal Prev. 2026;233:108577. https://doi.org/10.1016/j.aap.2026.108577 PMID: 42070493

60. Radhakrishnan V, Louw T, Cirino Gonçalves R, Torrao G, Lenné MG, Merat N. Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. Transp Res Part F Traffic Psychol Behav. 2023;94:254–67. https://doi.org/10.1016/j.trf.2023.02.015

61. De Gee JW, Knapen T, Donner TH. Decision-related pupil dilation reflects upcoming choice and individual bias. Proc Natl Acad Sci. 2014;111(5). https://doi.org/10.1073/pnas.1317557111

62. Knapen T, de Gee JW, Brascamp J, Nuiten S, Hoppenbrouwers S, Theeuwes J. Cognitive and ocular factors jointly determine pupil responses under equiluminance. PLoS One. 2016;11(5):e0155574. https://doi.org/10.1371/journal.pone.0155574 PMID: 27191166

63. van Kempen J, Loughnane GM, Newman DP, Kelly SP, Thiele A, O'Connell RG, et al. Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. Elife. 2019;8:e42541. https://doi.org/10.7554/eLife.42541 PMID: 30882347

64. Mridha Z, de Gee JW, Shi Y, Alkashgari R, Williams J, Suminski A, et al. Graded recruitment of pupil-linked neuromodulation by parametric stimulation of the vagus nerve. Nat Commun. 2021;12(1):1539. https://doi.org/10.1038/s41467-021-21730-2 PMID: 33750784

65. de Gee JW, Tsetsos K, Schwabe L, Urai AE, McCormick D, McGinley MJ, et al. Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. Elife. 2020;9:e54014. https://doi.org/10.7554/eLife.54014 PMID: 32543372

66. Murphy PR, Boonstra E, Nieuwenhuis S. Global gain modulation generates time-dependent urgency during perceptual choice in humans. Nat Commun. 2016;7:13526. https://doi.org/10.1038/ncomms13526 PMID: 27882927

67. Yamashita J, Terashima H, Yoneya M, Maruya K, Oishi H, Kumada T. Pupillary fluctuation amplitude preceding target presentation is linked to the variable foreperiod effect on reaction time in Psychomotor Vigilance Tasks. PLoS One. 2022;17(10):e0276205. https://doi.org/10.1371/journal.pone.0276205 PMID: 36264952

68. Yamashita J, Terashima H, Yoneya M, Maruya K, Koya H, Oishi H, et al. Pupillary fluctuation amplitude before target presentation reflects short-term vigilance level in Psychomotor Vigilance Tasks. PLoS One. 2021;16(9):e0256953. https://doi.org/10.1371/journal.pone.0256953 PMID: 34534237

69. Biondi FN, McDonnell AS, Mahmoodzadeh M, Jajo N, Balasingam B, Strayer DL. Vigilance decrement during on-road partially automated driving across four systems. Hum Factors J Hum Factors Ergon Soc. 2024;66(9):2179–90. https://doi.org/10.1177/00187208231189658

70. Pavlov YG, Kasanov D, Kosachenko AI, Kotyusov AI, Busch NA. Pupillometry and electroencephalography in the digit span task. Sci Data. 2022;9(1):325. https://doi.org/10.1038/s41597-022-01414-2 PMID: 35715429

71. Kosachenko AI, Kasanov D, Kotyusov AI, Pavlov YG. EEG and pupillometric signatures of working memory overload. Psychophysiology. 2023;60(6):e14275. https://doi.org/10.1111/psyp.14275 PMID: 36808118

72. Goodridge CM, Gonçalves RC, Reher A, Kuo J, Lenné MG, Merat N. Assessing data imbalance correction methods and gaze entropy for collision prediction. PLoS One. 2025;20(11):e0336777. https://doi.org/10.1371/journal.pone.0336777 PMID: 41259399

73. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Advances in psychology [Internet]. Elsevier; 1988 [cited 2026 Apr 21]. p. 139–83. Available from: https://www.sciencedirect.com/science/article/pii/S0166411508623869

74. Gold C, Damböck D, Lorenz L, Bengler K. "Take over!" How long does it take to get the driver back into the loop? Proc Hum Factors Ergon Soc Annu Meet. 2013;57(1):1938–42. https://doi.org/10.1177/1541931213571433

75. Mok BKJ, Johns M, Lee KJ, Ive HP, Miller D, Ju W. Timing of unstructured transitions of control in automated driving. 2015 IEEE Intelligent Vehicles Symposium (IV) [Internet]. IEEE; 2015 [cited 2026 Apr 21]. p. 1167–72. Available from: https://ieeexplore.ieee.org/abstract/document/7225841/
76. Louw T, Merat N. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. Transp Res Part C Emerg Technol. 2017;76:35–50. https://doi.org/10.1016/j.trc.2017.01.001
77. Mehler B, Reimer B, Dusek J. MIT AgeLab delayed digit recall task (n-back). MIT Cent Transp Logist Res Pap [Internet]; 2011 [cited 2026 Apr 21];(2011/018). Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6206338
78. Goodridge CM, Gonçalves RC, Arabian A, Horrobin A, Solernou A, Lee YT, et al. Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. Accid Anal Prev. 2024;202:107560. https://doi.org/10.1016/j.aap.2024.107560 PMID: 38677239
79. Wang C-A, Munoz DP. Modulation of stimulus contrast on the human pupil orienting response. Eur J Neurosci. 2014;40(5):2822–32. https://doi.org/10.1111/ejn.12641 PMID: 24911340
80. Hoeks B, Levelt WJM. Pupillary dilation as a measure of attention: a quantitative system analysis. Behav Res Methods Instrum Comput. 1993;25(1):16–26. https://doi.org/10.3758/bf03204445
81. Gilzenrat MS, Nieuwenhuis S, Jepma M, Cohen JD. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci. 2010;10(2):252–69. https://doi.org/10.3758/CABN.10.2.252 PMID: 20498349
82. Sugimoto Y, Sauer C. Effectiveness estimation method for advanced driver assistance system and its application to collision mitigation brake system. Proceedings: International Technical Conference on the Enhanced Safety of Vehicles [Internet]. National Highway Traffic Safety Administration; 2005 [cited 2026 Apr 21]. p. 8p–8p. Available from: https://www.safetylit.org/citations/index.php?fuseaction=citations.viewdetails&citationIds[]=citjournalarticle_245170_38
83. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. J Stat Softw. 2017;80:1–28.
84. Kruschke J. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan [Internet]; 2014 [cited 2026 Apr 21]. Available from: https://books.google.com/books?hl=en&lr=&id=FzvLAwAAQBAJ&oi=fnd&pg=PP1&dq=Kruschke,+J.+(2014).+Doing+Bayesian+data+analysis:+A+tutorial+with+R,+JAGS,+and+Stan.&ots=ChwmUYti1I&sig=x61LeUY1IK9RSMzF597noVS9Ox0
85. Makowski D, Ben-Shachar M, Lüdecke D. bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. JOSS. 2019;4(40):1541. https://doi.org/10.21105/joss.01541
86. Breton-Provencher V, Sur M. Active control of arousal by a locus coeruleus GABAergic circuit. Nat Neurosci. 2019;22(2):218–28. https://doi.org/10.1038/s41593-018-0305-z PMID: 30643295
87. Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, et al. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nat Commun. 2016;7:13289. https://doi.org/10.1038/ncomms13289 PMID: 27824036
88. Chen C, Lin Z, Zhang S, Chen F, Chen P, Zhang L. The compatibility between the takeover process in conditional automated driving and the current geometric design of the deceleration lane in highway. Sustainability. 2021;13(23):13403. https://doi.org/10.3390/su132313403
89. Pipkorn L, Dozza M, Tivesten E. Driver visual attention before and after take-over requests during automated driving on public roads. Hum Factors. 2024;66(2):336–47. https://doi.org/10.1177/00187208221093863
90. Mathôt S. Pupillometry: psychology, physiology, and function. J Cogn. 2018;1(1):16.
91. Cai Y, Strauch C, Van der Stigchel S, Naber M. Open-DPSM: an open-source toolkit for modeling pupil size changes to dynamic visual inputs. Behav Res Methods. 2024;56(6):5605–21. https://doi.org/10.3758/s13428-023-02292-1 PMID: 38082113
92. Palinko O, Kun A. Exploring the influence of light and cognitive load on pupil diameter in driving simulator studies. Driving Assessment Conference [Internet]. University of Iowa; 2011 [cited 2026 Apr 21]. Available from: https://pubs.lib.uiowa.edu/driving/article/id/28666/
93. Mather M, Clewett D, Sakaki M, Harley CW. Norepinephrine ignites local hotspots of neuronal excitation: how arousal amplifies selectivity in perception and memory. Behav Brain Sci. 2016;39:e200. https://doi.org/10.1017/S0140525X15000667 PMID: 26126507
94. Tang R, Kim JH, Parker R, Jeong YJ. Indicating severity of vehicle accidents using pupil diameter in a driving simulator environment. In: Duffy VG, editor. Digital human modeling. Applications in health, safety, ergonomics, and risk management [Internet]. Cham: Springer International Publishing; 2018 [cited 2026 Apr 21]. p. 647–56. (Lecture Notes in Computer Science). Available from: https://link.springer.com/10.1007/978-3-319-91397-1_53
95. Arnsten AF, Goldman-Rakic PS. Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey. Brain Res. 1984;306(1–2):9–18. https://doi.org/10.1016/0006-8993(84)90351-2 PMID: 6466989
96. Geissler CF, Schneider J, Frings C. Shedding light on the prefrontal correlates of mental workload in simulated driving: a functional near-infrared spectroscopy study. Sci Rep. 2021;11(1):705. https://doi.org/10.1038/s41598-020-80477-w PMID: 33436950
97. Yeung MK, Lee TL, Han YMY, Chan AS. Prefrontal activation and pupil dilation during n-back task performance: a combined fNIRS and pupillometry study. Neuropsychologia. 2021;159:107954. https://doi.org/10.1016/j.neuropsychologia.2021.107954 PMID: 34252415
98. Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the P3, and the locus coeruleus-norepinephrine system. Psychol Bull. 2005;131(4):510–32. https://doi.org/10.1037/0033-2909.131.4.510 PMID: 16060800
99. Steinhauer SR, Hakerem GAD. The pupillary response in cognitive psychophysiology and schizophrenia [Internet]; 1992 [cited 2026 Apr 21]. Available from: https://psycnet.apa.org/record/1992-98867-009

100. Nieuwenhuis S. Learning, the P3, and the locus coeruleus-norepinephrine system. In: Neural basis of motivational and cognitive control. The MIT Press; 2011. p. 208–22. https://doi.org/10.7551/mitpress/9780262016438.003.0012

101. Murphy PR, Robertson IH, Balsters JH, O'connell RG. Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. Psychophysiology. 2011;48(11):1532–43. https://doi.org/10.1111/j.1469-8986.2011.01226.x PMID: 21762458

102. Strayer DL. Do pupillometry and the P3 event-related potential reflect locus coeruleus-norepinephrine system activity in humans? 2020.

103. Strayer DL, Drews FA. Cell-phone–induced driver distraction. Curr Dir Psychol Sci. 2007;16(3):128–31. https://doi.org/10.1111/j.1467-8721.2007.00489.x

104. Strayer DL, Turrill J, Cooper JM, Coleman JR, Medeiros-Ward N, Biondi F. Assessing cognitive distraction in the automobile. Hum Factors J Hum Factors Ergon Soc. 2015;57(8):1300–24. https://doi.org/10.1177/0018720815575149 PMID: 26534847

105. van der Heiden RMA, Janssen CP, Donker SF, Hardeman LES, Mans K, Kenemans JL. Susceptibility to audio signals during autonomous driving. PLoS One. 2018;13(8):e0201963. https://doi.org/10.1371/journal.pone.0201963 PMID: 30102723

106. Van Der Heiden RMA, Kenemans JL, Donker SF, Janssen CP. The effect of cognitive load on auditory susceptibility during automated driving. Hum Factors. 2022;64(7):1195–209. https://doi.org/10.1177/0018720821998850

107. Wilhelm B, Wilhelm H, Lüdtke H, Streicher P, Adler M. Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. Sleep. 1998;21(3):258–65. https://doi.org/10.1093/sleep/21.3.258 PMID: 9595604

108. Neubauer CE. Alertness maintaining tasks: a fatigue countermeasure during vehicle automation? [Internet]. University of Cincinnati; 2014 [cited 2026 Apr 21]. Available from: https://search.proquest.com/openview/4bc402e0ff73e083050953bffc4bae36/1?pq-origsite=gscholar&cbl=18750

109. Warm JS, Dember WN, Hancock PA. Vigilance and workload in automated systems. In: Automation and human performance [Internet]. CRC Press; 2018 [cited 2026 Apr 21]. p. 183–200. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9781315137957-9/vigilance-workload-automated-systems-joel-warm-william-dember-peter-hancock

110. Abe T. PERCLOS-based technologies for detecting drowsiness: current evidence and future directions. Sleep Adv. 2023;4(1):zpad006. https://doi.org/10.1093/sleepadvances/zpad006 PMID: 37193281

111. Johns MW. The amplitude-velocity ratio of blinks: a new method for monitoring drowsiness. Sleep [Internet]. 2003 [cited 2026 Apr 21];26(SUPPL). Available from: https://www.mwjohns.com/wp-content/uploads/2017/05/apss_2003_06_03_the_amplitude_velocity_ratio_of_blinks_a_new_method_of_monitoring_drowsiness_poster_69.pdf

112. Johns MW, Tucker AJ. The amplitude-velocity ratios of eyelid movements during blinks: changes with drowsiness. Sleep. 2005;28:A122.

113. Kundinger T, Riener A, Sofra N, Weigl K. Drowsiness detection and warning in manual and automated driving: results from subjective evaluation. Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications [Internet]. Toronto, ON, Canada: ACM; 2018 [cited 2026 Apr 21]. p. 229–36. Available from: https://dl.acm.org/doi/10.1145/3239060.3239073