Impact of Multi-Dimensional Cognitive Demands on Takeover Performance, Physiological and Eye-tracking Measures in Conditionally Automated Vehicles

Ange Wang[a,b], Wenxin Shi[c], Dengbo He[a,b,c*], Hai Yang[a,b]

[a]*Thrust of Intelligent Transportation, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China*

[b]*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China*

[c]*Thrust of Robotics and Autonomous Systems, Systems Hub, The Hong Kong University of Science and Technology (Guangzhou), China*

Abstract

Before fully autonomous vehicles come true, drivers are still required to be responsible for driving safety and take over control of the vehicle when prompted by takeover requests in conditionally automated vehicles. Thus, drivers' takeover performance can affect the safety of conditionally automated driving. However, though cognitive distraction can impair takeover performance in general, the influence of multi-dimensionality in the cognitive resources was under-investigated. At the same time, it is unknown how physiological and eye-tracking metrics are associated with different modalities of cognitive tasks in conditionally automated vehicles. Thus, through a driving simulation study with 42 participants, we investigated the effects of multidimensional cognitive demands, as imposed by multiple types of non-driving-related tasks, on drivers' takeover performance, physiological responses, and eye-tracking metrics in conditionally automated vehicles. Results showed that certain takeover performance (i.e., vehicle speed and lateral acceleration), and physiological and eye-tracking metrics (i.e., differences between consecutive R-peaks, skin conductance level, variation in respiratory intervals, number of fixations, number of saccades and saccade angle) are still responsive to cognitive load in the context of driving automation. Further, the modality of the cognitive tasks can moderate the takeover performance (i.e., takeover time) and certain physiological (i.e., ratio of spectral power in the low- and high-frequency range and respiration depth) metrics. These findings suggest that, in future conditionally automated vehicles, in-vehicle task designs should consider the modality of the cognitive demands for driving safety and for the performance of the driver monitoring systems.

Keywords

Automated driving, Multi-dimensional cognitive load, Takeover performance, Physiological response, Eye-tracking metrics

1. Introduction

Though driving automation is claimed to enhance safety (Jeong et al., 2017) by replacing drivers in some tasks, fully autonomous vehicles are still not yet ready. According to the Society of Automotive Engineers (SAE) (*J3016 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, 2018), the highest level of driving commercially available automation is Level 3, which requires drivers to take over the control of vehicles in case of emergencies. However, one of the benefits of driving automation is that drivers can be relieved from driving tasks and make use of commute time, which has been found to negatively impact takeover performance (Schartmüller et al., 2021). Thus, evaluating the impact of non-driving-related tasks (NDRTs) on takeover performance and monitoring the drivers' states are crucial to the driving safety of SAE Level-3 vehicles.

Previous research has found that both visual-manual NDRT (Guo et al., 2023; Zeeb et al., 2016) and cognitive NDRT (Choi et al., 2020) can impair take-over performance in vehicles with driving automation and cognitive load is related to human error (Dingus et al., 2019; Seyfzadehdarabad et al., 2023). Such negative cognitive load effects can last up to 5 minutes after takeover events (Xu et al., 2024). The negative impacts of the visual-manual tasks are expected as driving mostly relies on visual information (Tivesten & Dozza, 2014). However, how cognitive load affects takeover performance deserves further investigation, given multiple cognitive resources can be claimed in multiple tasks or even a single task (Wickens, 1991). For example, recalling a message in a conversation requires memory resources, while following navigation would require spatial resources. At the same time, the cognitive resources required for the transfer of control can also be multi-dimensional (e.g., spatial resources to rebuild traffic scenarios and memory resources to be aware of system boundaries). Thus, treating the cognitive task as unidimensional may not fully reveal the association between cognitive tasks and take-over performance. Given the widespread high cognitive load inducers for drivers (e.g., in-vehicle tasks, complex traffic, and bad weather), we need to understand the impact of different types of cognitive tasks on takeover performance to better inform driving automation designs and in-vehicle task designs in SAE Level-3 vehicles.

At the same time, driver-state monitoring has been acknowledged as an effective way to ensure driving safety, and such a system will become mandatory in some regions (Aisin Mobility, 2022). Thus, previous research has explored models to estimate drivers' states in SAE Level-3 vehicles (Shi et al., 2024; Wang et al., 2024), mostly using physiological measures and eye-tracking measures (Das & Maiti, 2024; Rahman et al., 2020; Kumar et al., 2022; Tjolleng et al., 2017). However, to the best of our knowledge, no research has differentiated the influence of cognitive task types on drivers' physiological and eye-tracking metrics. Thus, we also aim to explore how the type of cognitive tasks can affect drivers' physiological and eye-tracking responses in takeover events, which may support more accurate and targeted drivers' cognitive load estimation in SAE Level 3 vehicles.

In this study, a driving simulator experiment was conducted. Three commonly used standardized cognitive tasks were used to impose different types of high cognitive load among drivers, i.e., the n-back task (Jaeggi et al., 2010) that requires memory resources, the mathematical task (Meteier et al., 2021; Wiediartini et al., 2023) that requires numerical processing resources, and the spatial reasoning task (Liang & Lee, 2010) that requires spatial cognitive resources. These three tasks have been extensively used as NDRTs to impose high cognitive loads among drivers (He et al., 2022; Liang & Lee, 2010; Meteier et al., 2021) in non-automated driving and driving automation contexts.

2. Related Work

2.1 Commonly Used Cognitive Tasks in Driving Tasks

To manipulate cognitive load among drivers, a variety of cognitive tasks were adopted in previous driving research. Memory tasks are commonly used to manipulate cognitive load in drivers. For instance, the n-back task has been widely adopted as it simulates working memory processes, i.e., maintaining, retrieving, and updating information (Wang et al., 2016). Initially proposed by Mehler et al. (Mehler et al., 2012a), the n-back has several variations, including matching (Deng et al., 2024; Meteier et al., 2022; Nilsson et al., 2022; Öztürk et al., 2023; Rahman et al., 2020), repeating (Cegovnik et al., 2018; Gable et al., 2015; Hajek et al., 2013; Mehler et al., 2009, 2009, 2012b; Tjolleng et al., 2017), and counting (He et al., 2019). In

addition to n-back tasks, mathematical tasks have also been widely adopted, which require one's computational resources. Common mathematical tasks include addition (Brookhuis et al., 1991; Harbluk et al., 2007; Zhou et al., 2022), subtraction (Meteier et al., 2021; Shi et al., 2024; Wang et al., 2024), and multiplication operations (Mathissen et al., 2022). The mathematical task may be implemented as interactive question-answer sessions (Harbluk et al., 2007; Mathissen et al., 2022; Zhou et al., 2022) or continuous computation by participants (Meteier et al., 2021; Shi et al., 2024; Wang et al., 2024). Finally, spatial tasks were also adopted, including virtual reality navigation tasks (Liang & Lee, 2010), object-seeking tasks (Radhakrishnan et al., 2023), and spatial image tasks (Engström et al., 2005; Liang & Lee, 2010; Recarte & Nunes, 2000), which require participants to process spatial information while driving.

2.2 The Effect of NDRTs on Takeover Performance

Engaging in NDRTs has been found to delay drivers' responses in takeover events. For example, in a driving simulator study, Liu et al. (2024) found that different NDRTs led to varying takeover performances, from lowest to highest, the performance ranked as error detection, monitoring, texting, and conversing. Lin et al. (2020) found that visual NDRTs (i.e., reading the news and watching videos) can delay drivers' responses to TOR and impair the takeover quality in safety-critical events. Through a meta-analysis study, Zhang (2024) found that both visual-cognitive-motor tasks and visual-cognitive tasks have a significant negative impact on takeover time, but the former has a greater effect than the latter. At the same time, cognitive NDRT can also impair one's takeover performance. For example, Lee et al. (2021) found that engaging in cognitive NDRTs (e.g., taking with passengers) was negatively associated with both longitudinal and lateral control metrics. Roche et al. (2019) further confirmed that the NDRT type can affect the takeover performance. More specifically, the standard deviation (SD) of lane position deviation after the takeover is larger with auditory than visual-auditory cognitive NDRT.

2.3 The association between cognitive NDRTs, eye-tracking, and physiological measures

The cognitive load has been found to be associated with drivers' physiological and eye-tracking responses in both non-automated vehicles and vehicles with driving automation. For

example, in a simulator study, Liang and Lee (2010) found that the increased cognitive load led to an increase in blink rate, but a decrease in the vertical and horizontal SD of gaze positions in vehicles without driving automation. Similarly, in vehicles with driving automation, through driving simulation, Chen et al. (2022) suggested that the pupil diameter, the frequency and duration of saccades, the length of fixations, and 3D gaze entropy are reliable measures of a driver's mental strain in vehicles with the advanced driving assistant system (ADAS). Similarly, Radhakrishnan et al. (2023) found that the size of the pupil is responsive to changes in visual task-induced mental effort when driving with ADAS in a driving simulation experiment, which mostly requires spatial processing resources. The association between high cognitive load and drivers' physiological and eye-tracking metrics has also been observed in on-road studies. For example, in on-road studies, the high cognitive load was found to be associated with a declined frequency of checking peripheries, mirrors, and instruments (Harbluk et al., 2007), elevated heart rate (HR), and diminished heart rate variability (HRV) (Brookhuis et al., 1991), and larger pupil diameter (PD) (Recarte & Nunes, 2000, 2003) in vehicles without ADAS. However, as mentioned earlier, no research has differentiated the influence of different cognitive resources on drivers' physiological and eye-tracking metrics, especially in vehicles with driving automation.

2.4 This Study

Existing research has shown that cognitive tasks during automated driving affect takeover performance, but most studies treat "cognitive load" as a unidimensional construct. To address this research gap, through a driving simulator experiment, we aim to investigate how drivers' takeover performance varies across different types of cognitive loads (memory recall, numerical processing, spatial reasoning) and whether eye-tracking patterns and physiological responses vary across cognitive task types and cognitive load levels in vehicles with driving automation.

3. Method

3.1 Participants

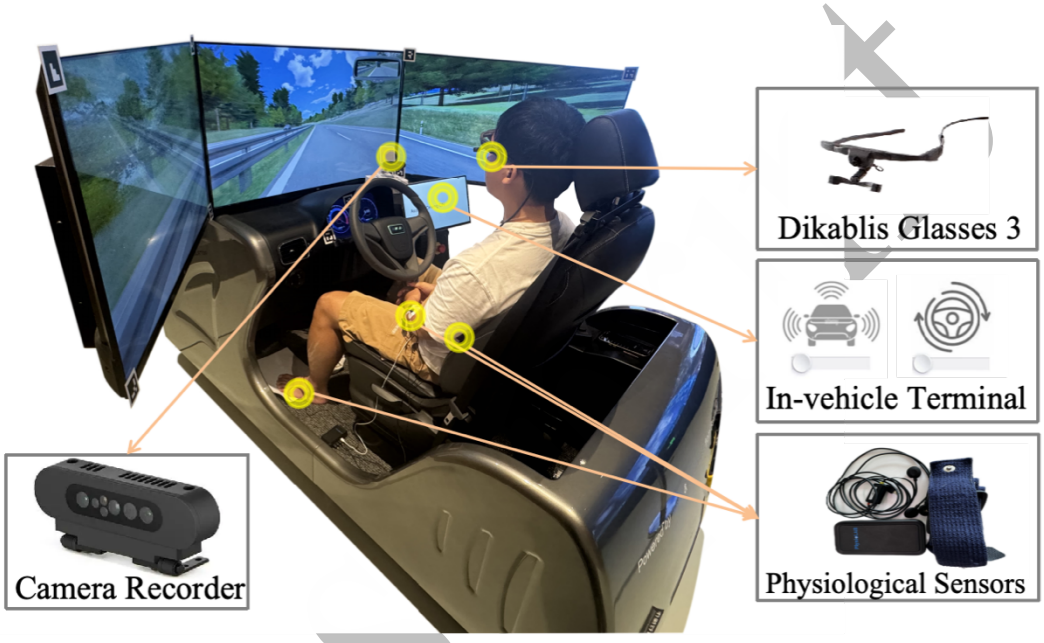
A total of 42 drivers, consisting of 25 males and 17 females with an average age of 35.3

years (SD = 9.10, ranging from 23 to 53 years), participated in our study. In this study, we did not specifically control for a balanced gender distribution among participants. Instead, we aimed to reflect the current gender distribution of licensed drivers in China, where the experiment was conducted. As of the latest statistics, female drivers account for 33.7% of the total licensed drivers in China in 2022 (approximately 162 million out of 481 million licensed drivers) (Car Circle, 2022). In our study, 40.5% of the participants were female (17 out of 42), which is slightly higher than the national gender ratio but still reflects the trend in which the proportion of female drivers has been increasing in recent years. The female driver population has been growing at a rate approximately double that of male drivers (Car Circle, 2022), making it relevant to simulate a future scenario in which the percentage of female drivers may rise further. Eligibility criteria included holding a valid driver's license for at least one year, having no prior experience with ADAS features (including Lane Centering Control, Adaptive Cruise Control, or any other more advanced ADAS functions such as automated lane-changing), and having driven a minimum of 5,000 kilometers in the past year. We did not control the age distribution of the drivers, but our participant sample was generally young, which matches the age distribution of major buyers of ADAS-equipped vehicles in China (Souhu Auto, 2024). We also targeted drivers without ADAS experience, as most drivers nowadays still have limited experience with ADAS, and they have diverse attitudes toward ADAS (Huang et al., 2023). Thus, technologies supporting this group of drivers may better enhance driving safety with ADAS. Participants were informed that they would receive compensation of 70 RMB per hour, with the possibility of earning an additional performance-based bonus of up to 30 RMB for completing cognitive tasks during the experiment. The study was approved by the Hong Kong University of Science and Technology (HREP-2023-0199).

3.2 Equipment

The experiment utilized a fixed-base driving simulator (Figure 2a) equipped with three 43-inch screens, providing a 150° horizontal and 47° vertical viewing angle. Participants could engage the driving automation system by pressing virtual buttons on a 15-inch screen positioned next to the steering wheel. They could disengage the automation by pressing another

virtual button, turning the steering wheel, or using the brake pedal. Driving data was logged at a rate of 60 Hz using the Silab 7.1 simulation software from WIVW GmbH. Physiological measurements, including ECG, RESP, and EDA, were collected with sensors from Ergoneers GmbH at a sampling rate of 100 Hz (see Figure 2b). Eye-tracking data was recorded using the Dikablis and D-Lab software, also at a 60 Hz sampling rate. All data streams were synchronized using the Human Research Tool (HRT) software by Info Instrument.



(a) (b)

Figure 1. Driving simulator, eye-tracker and physiological sensor placements.

3.3 Driving Task

Each participant undertook a single drive on a six-lane dual carriageway composed of 21 straight segments, each 7 km in length. The speed limit was set at 110 km/h, with traffic flowing freely at a density of 6 cars per kilometer per lane. Participants were informed that the vehicle was equipped with an SAE Level 3 Advanced Driving Assistance System (ADAS), requiring them to engage the automation whenever possible and only take manual control when necessary or in response to a Takeover Request (TOR). As shown in Figure 2, the drive included three types of takeover scenarios: 1) exiting a ramp, where drivers needed to shift to the right lane,

merge onto the exit ramp, and reduce speed to below 60 km/h; 2) navigating around a traffic accident, where drivers had to change lanes to avoid the accident area; and 3) proceeding through a foggy area, where drivers were required to reduce speed. Each takeover event was preceded by a TOR, with a lead time of 10 seconds before automation disengagement, following the methodology of previous research (Chen et al., 2023).

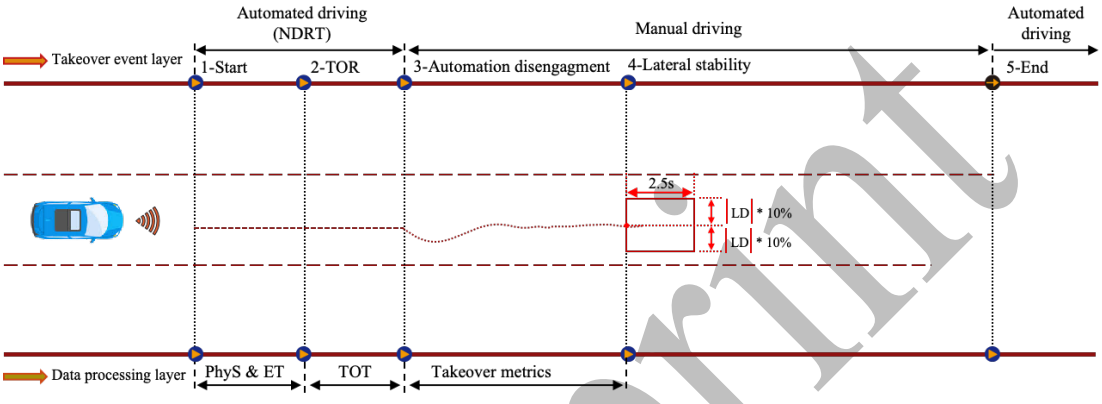


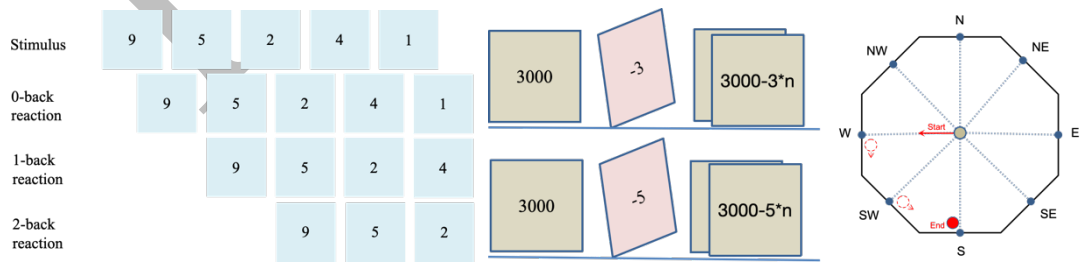
Figure 2. Design of takeover scenarios. In the figure, NDRT: Non-Driving Related Task; TOR: Takeover Request; TOT: Takeover Time; LD: Lateral Deviation; PhyS: Physiological Signals; ET: Eye-Tracking.

3.4 Non-Driving-Related Tasks (NDRTs)

Cognitive load is multidimensional, engaging various types of cognitive resources (Wang et al., 2024). To evaluate the impact of different cognitive dimensions on takeover performance, three types of cognitive tasks were used, as shown in Table 1 and Figure 3: calculation, memory, and spatial processing, which simulates the cognitive demands when estimating an electric vehicle's remaining range, processing and recalling traffic scenarios; and navigation activities, respectively. Additionally, the complexity of each NDRT was adjusted to introduce varying levels of cognitive load, simulating diverse traffic conditions.

Table 1. Summary of NDRTs.

Task	Descriptions	Levels	Cognitive Resource	Abbreviation
N-back Task (Jaeggi et al., 2010)	A series of stimuli (numbers/letters) are presented with a pause between each. Participants recall the stimulus that is *n* positions earlier.	Three levels: 0-back, 1-back and 2-back tasks.	Memory	MT
Calculation Task (Meteier et al., 2021)	Oral backward counting from 3,000 by increments of 3 (non-integer) or 5 (integer).	Two levels: Counting backward by 3 or 5 from 3,000.	Calculation	CT
Spatial Task (Liang & Lee, 2010)	Participants listen to an audio clip describing a route and identify the main direction faced at the end, simulating high cognitive demands similar to those in navigation systems.	One level: "What direction is this person facing when she goes to the north station and moves two stations clockwise?" (Answer: East)	Spatial process	ST



(a) (b) (c)

Figure 3. Cognitive tasks: (a) n-back task, (b) math task, and (c) spatial task.

3.5 Experiment Design

A 3 (Takeover Scenarios) by 7 (NDRTs) within-subject design was used. To reduce the potential effect of experiment order on the results, the three takeover scenarios were counterbalanced using a Latin squared design. Within each scenario, the 7 NDRT types were also counterbalanced using a Latin squared design, leading to 7 possible orders. This combination of scenario and NDRT resulted in a total of 21 unique experimental orders. Each combination of experimental conditions was experienced by two participants. Specifically, after completing all NDRTs within one Takeover Scenario, participants took a 10-minute break before proceeding to the next seven takeover scenarios in order to reduce the potential fatigue effects.

3.6 Procedures

Participants were required to follow regular sleep routines, avoid alcohol, and abstain from caffeine consumption for 24 hours before the experiment. As shown in Figure 4, upon arrival, they provided written consent and attended a 30-minute orientation covering the experimental procedures, vehicle operation, and cognitive tasks. This was followed by a practice driving session where each participant encountered one TOR on the highway. Afterward, physiological sensors and eye-tracking devices were put on and calibrated before starting the formal experiment. During the experiment, each participant encountered 21 TORs, with each TOR occurring on one of the straight sections of the road. Following each TOR, participants stopped to complete two questionnaires: the NASA Task Load Index (NASA-TLX) for perceived workload (Hart & Staveland, 1988). The entire session took around three hours to complete.

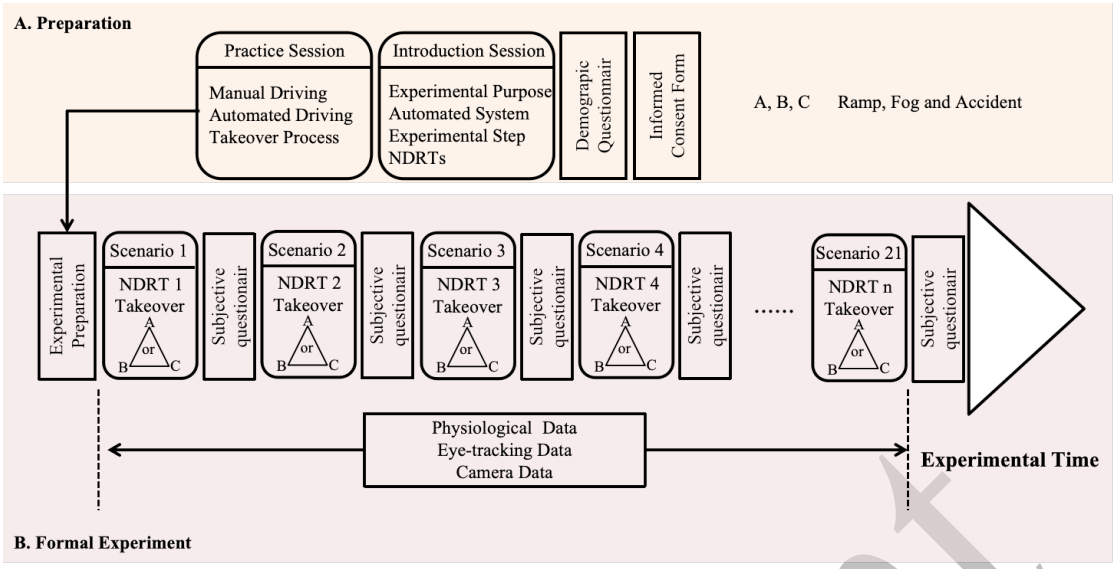


Figure 4. Experimental process.

3.7 Variable Extraction and Signal Processing

As shown in Table 2, the variables of interest include takeover performance, physiological responses, and eye-tracking measures. For the physiological measures, band-pass filters were applied to reduce the noise in the data before metric extraction. Specifically, a fourth-order Butterworth low-pass filter was applied to mitigate high-frequency disturbances for EDA. ECG signals were processed through a band-pass filter between 3 Hz to 45 Hz, followed by R-wave detection using an enhanced Pan-Tompkins algorithm (Sathyapriya et al., 2014) to calculate the R-R intervals. For RESP, band-pass filters were applied within the frequency range between 0.1 Hz to 0.35 Hz. All physiological metrics and eye-tracking metrics were extracted 120 seconds before TOR to the TOR onset. The data extraction periods for the variables are provided in Figure 2, where the vehicle achieving lateral stability was defined as the moment when the lateral deviation in the following 2.5 seconds was within 10% of the maximum lateral deviation after the automation disengagement (Chen et al., 2023).

1 Table 2. Description of dependent variables (DVs)

DV	Unit	Category	Description (Note)	Calculation Details	Trend with Increased Cognitive Load
TOT	s	Takeover performance	Takeover time, i.e., from the moment the takeover request is initiated to the deactivation of the autonomous driving	The average/max value of the metrics during the period from the automation disengagement to the vehicle achieving lateral stability. (Cao et al., 2021; Yao et al., 2021).	↑ (Guo et al., 2023; Zeeb et al., 2017; Zhang, 2024)
V	m/s		Vehicle speed		Average↓ (Chen et al., 2023; Mehler et al., 2012b; Zeeb et al., 2016); ↑ (Mehler et al., 2009)
LD	m		The lateral distance from the centerline of the lane, including mean and max LD		↑ (Zeeb et al., 2017; Zhang et al., 2023)
SW	rad		Steering wheel angel		↑ (Mehler et al., 2012b; Zhang et al., 2023)
AX	m/s^2		Lateral acceleration		↑ (Louw et al., n.d.)
AY	m/s^2		Longitudinal acceleration		No traffic ↑ (Du et al., 2020); Heavy oncoming traffic density ↓(Du et al., 2020)
PA	px	Eye-tracking metrics	Pupil area	Average of the metrics from 120 seconds prior to the TOR to the moment of	↑ (Guo et al., 2023; Nilsson et al., 2022; Zénon, 2019)
NF	Times		Number of fixations		↓ (Broadbent et al., 2023); ↑ (Chen et al., 2022)

FT	%		Fixations time (The percentage of fixation time per minute)	the TOR initiation.	↑ (Broadbent et al., 2023; Chen et al., 2022; Harris et al., 2023)
NS	Times		Number of saccades		↑ (Chen et al., 2022)
ST	%		Saccade time (The percentage of saccade time per minute)		
SA	Degrees		Saccade angle (The angle between two consecutive saccades)		↓ (Harris et al., 2023)
SCL	μS	Electrodermal activity (EDA)	Skin conductance level (The average skin conductance over time)		↑ (He et al., 2019; Mehler et al., 2009, 2012b; Nilsson et al., 2022)
SCR	μS		Skin conductance response (The difference between peak skin conductance and the baseline level)		
HR	Beats/minute	Electrocardiogram (ECG)	Heart Rate		↑ (Gable et al., 2015; He et al., 2019; Liu et al., 2017; Mehler et
RMSSD	0.01s		Magnitude of the differences between consecutive R-R		↓(Zhang et al., 2023)
SDNN	s		Variability of the time intervals between consecutive normal heartbeats		↓ (Zhang et al., 2023)
LF	ms^2		Spectral power in the low-frequency range (usually 0.04 to 0.15 Hz)		↓ (Zheng et al., 2021)
HF	ms^2		Spectral power in the high-frequency range (usually 0.15 to 0.4 Hz)		↓ (Zheng et al., 2021)
LF/HF	%		Ratio of LF and HF		↑ (Zhang et al., 2023)

RR	Respirations / minute	Respiration (RESP)	Respiration rate	↑ (Grassmann et al., 2016; Mehler et al., 2009)
RD	mm		Respiration depth (the amount of air that is inhaled and exhaled during each breath, as measured by the movement of the chest)	↓ (He et al., 2019)
RV	%		Variation in respiratory intervals	No significant change

Note: in the table, speed - V, braking pedal depth - BP, steering wheel angle - SW, maximum lateral deviation - Max LD, mean lateral deviation - Mean LD, longitudinal acceleration – AX, lateral acceleration - AY, pupil area - PA, number of fixations - NF, fixations time - FT, number of saccades - NS, saccade time - ST, saccade angle - SA, skin conductance level – SCL, skin conductance response - SCR, heart rate - HR, respiratory rate - RR, respiratory depth - RD, respiratory variation – RV. The ↑ and ↓ mean the metrics increased or decreased with the increase of the cognitive load level.

3.8 Statistical Models

For continuous variables, mixed linear models were built using the PROC MIXED procedure. For count and rate data (i.e., NF, NS, and RR), the negative binomial model was built using the PROC GENMOD procedure, given the over-dispersion of the samples. All models were built in "SAS OnDemand for Academics." The independent variables (IVs) include NDRT type (see Table 1) and perceived cognitive load as quantified by NASA-TLX (Hart & Staveland, 1988), ranging from 1 (low) to 21 (high). The perceived cognitive load instead of the task load levels (e.g., 1-back versus 2-back) was selected, given different participants may have different capabilities in conducting even the same type of NDRT. Thus, the perceived cognitive can better quantify drivers' spare cognitive resources and facilitate more a fair comparison of the effects of NDRT type on the variables of interest. The collinearity test results indicate a correlation coefficient of 0.4 and a variance inflation factor (VIF) of 3.9, which were acceptable (Kalnins, 2018). The generalized estimating equation model accounted for the repeated measures (Zeger et al., 1988). Tukey-Kramer post-hoc tests (Kramer, 1956) were conducted for significant ($p < .05$) variables.

4. Results

4.1 Takeover Performance

4.1.1 Takeover Time

As shown in Table 4, the interaction between the cognitive load level and NDRT type was a significant predictor of TOT (takeover time). Overall, the TOT increased with the cognitive load level. The changes in TOT in response to the increase in cognitive load level can be found in Figure 5a. Additionally, the pairwise comparisons show that the changes in cognitive load caused by the memory task led to significantly smaller changes in TOT compared to the calculation task, with a difference (Δ) of -0.07 and a 95% confidence interval (95% CI) between -0.12 and -0.02 (p=.003).

4.1.2 Takeover Quality

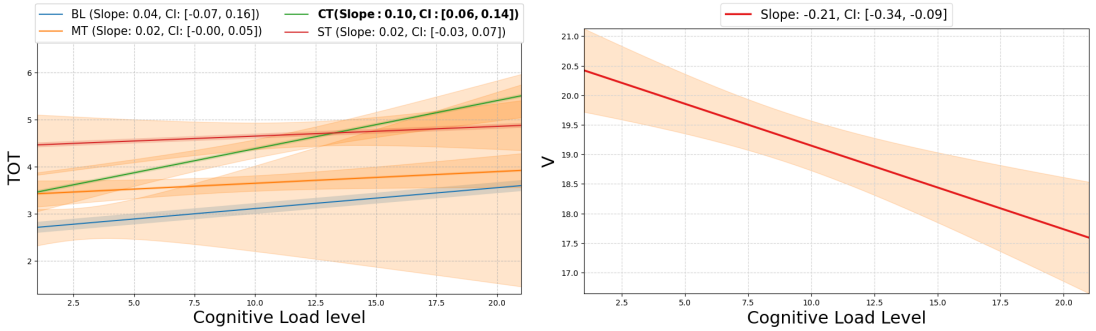
Cognitive load level significantly affected V (speed) and AX (lateral acceleration). Specifically, as shown in Figures 5(b) and 5(c), for every one-unit increase in cognitive load,

speed decreased by 0.21 m/s (95% CI: [-0.34, -0.09], t(806) = -3.49, $p < .001$), and lateral acceleration decreased by 0.02 m/s² (95% CI: [-0.03, -0.005], t(806) = -2.68, $p = .008$).

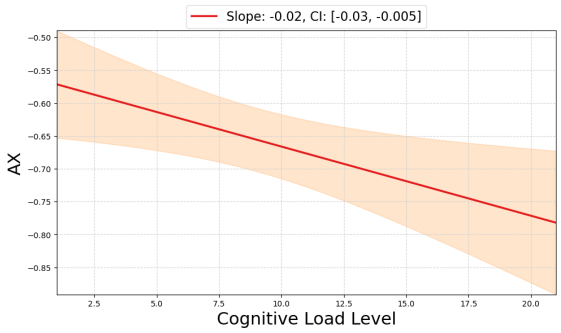
Table 4. Statistic results of multi-dimensional cognitive resources on takeover performance.

Dependent variables		Independent variables	F value	*p*-value
Takeover time	TOT	Cognitive load level	F(1, 806) = 9.15	**.003**
		NDRT type	F(3, 806) =7.71	**<.001**
		Cognitive load level * NDRT type	F(3, 806) = 3.47	**.015**
Takeover quality	V	Cognitive load level	F(1, 806) = 12.2	**<.001**
		NDRT type	F(3, 806) = 0.23	.873
		Cognitive load level* NDRT type	F(3, 806) = 0.5	.685
	SW	Cognitive load level	F(1, 806) = 0.78	.378
		NDRT type	F(3, 806) = 0.92	.429
		Cognitive load level* NDRT type	F(3, 806) = 0.5	.681
	AX	Cognitive load level	F(1, 806) = 7.19	**.008**
		NDRT type	F(3, 806) = 0.98	.404
		Cognitive load level* NDRT type	F(3, 806) = 1.05	.370
	AY	Cognitive load level	F(1, 806) = 1.04	.308
		NDRT type	F(3, 806) = 2.5	.059
		Cognitive load level* NDRT type	F(3, 806) = 1.38	.247
	Mean LD	Cognitive load level	F(1, 806) = 0.00	.970
		NDRT type	F(3, 806) = 1.78	.149
		Cognitive load level* NDRT type	F(3, 806) = 1.3	.272
	Max LD	Cognitive load level	F(1, 806) = 0.12	.724
		NDRT type	F(3, 806) = 0.27	.845
		Cognitive load level* NDRT type	F(3, 806) = 0.32	.807

Notes: * means interaction effect. The bold font indicates significant results ($p < .05$).



(a) (b)



(c)

Figure 5. The effects of NDRT type and cognitive load level on takeover performance. The shaded areas are the 95% CI. The bold fonts mean the *p*-value < .05.

4.2 Physiological Responses

Table 5 summarizes the influence of NDRT type and cognitive load level on drivers' physiological responses.

4.2.1 ECG metrics

The cognitive load level was significantly associated with RMSSD. Specifically, as shown in Figure 6(a), for each unit increase in cognitive load, the RMSSD decreased by 0.001 seconds (95% CI: [-0.002, 0], t(806) = -2.19, p = .029). At the same time, the interaction between NDRT type and cognitive load level significantly affected LF/HF. As shown in Figure 6(b), in baseline without NDRT, cognitive load led to a larger increase in LF/HF than the memory task (Δ = 0.29, 95% CI: [0.10, 0.47], p = .002), the calculation task (Δ = 0.24, 95% CI: [0.04, 0.43], p = .022), and the spatial task (Δ = 0.27, 95% CI: [0.06, 0.47], p = .013).

4.2.2 EDA

Cognitive load significantly affected SCL. Specifically, as shown in Figure 6(c), for each 1-point increase in cognitive load level, SCL increased by 0.07 μs (95% CI: [0.002, 0.13], t(806) = 2.04, p = .042).

4.2.3 RESP

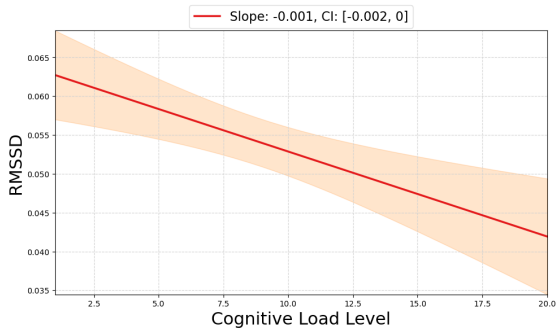
The cognitive load level was associated with RV. Specifically, as shown in Figure 6(d), for every 1-unit increase in cognitive load level, RV increased by 0.03% (95% CI: [0.01, 0.05], t(806) = 3.2, p = .001). At the same time, the interaction between cognitive load level and NDRT type significantly affected RD, as shown in Figure 6(e). Without NDRT, the cognitive load led to larger increase in RD than that when the drivers were provided with the memory task (Δ = 1.09, 95% CI: [0.38, 1.80], p = .003), the calculation task (Δ = 0.82, 95% CI: [0.08, 1.56], p = .031), and the spatial task (Δ = 0.84, 95% CI: [0.04, 1.63], p = .038).

Table 5. Statistic results of multi-dimensional cognitive resources on physiological responses.

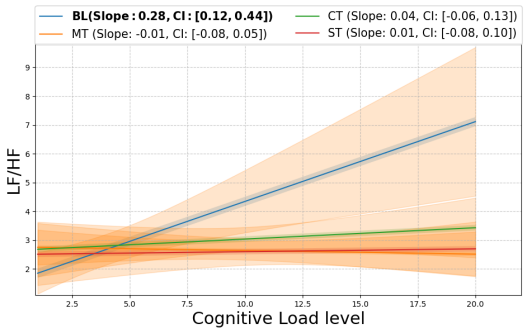
DV		IV	F/χ^2 value	p-value
ECG	HR	Cognitive load level	F(1, 806) = 0.46	.450
		NDRT type	F(3, 806) = 2.17	.090
		Cognitive load level * NDRT type	F(3, 806) = 0.27	.844
	RMSSD	Cognitive load level	F(1, 806) = 4.82	**.029**
		NDRT type	F(3, 806) = 0.90	.440
		Cognitive load level* NDRT type	F(3, 806) = 0.34	.793
	SDNN	Cognitive load level	F(1, 806) = 0.39	.532
		NDRT type	F(3, 806) = 0.11	.957
		Cognitive load level * NDRT type	F(3, 806) = 0.67	.569
	LF	Cognitive load level	F(1, 806) = 3.12	.078
		NDRT type	F(3, 806) = 1.28	.280
		Cognitive load level * NDRT type	F(3, 806) = 1.07	.423
	HF	Cognitive load level	F(1, 806) = 0.05	.824
		NDRT type	F(3, 806) = 0.65	.585
		Cognitive load level * NDRT type	F(3, 806) = 0.55	.646
		Cognitive load level	F(1, 806) = 6.96	**.009**

	LF/HF	NDRT type	F(3, 806) = 1.53	.205
		Cognitive load level * NDRT type	F(3, 806) = 3.20	**.023**
EDA	SCL	Cognitive load level	F(1, 806) = 4.17	**.042**
		NDRT type	F(3, 806) = 0.38	.767
		Cognitive load level * NDRT type	F(3, 806) = 0.88	.452
	SCR	Cognitive load level	F(1, 806) = 2.87	.091
		NDRT type	F(3, 806) = 0.19	.901
		Cognitive load level * NDRT type	F(3, 806) = 0.04	.988
RESP	RD	Cognitive load level	F(1, 806) = 9.07	**.003**
		NDRT type	F(3, 806) = 0.64	.591
		Cognitive load level * NDRT type	F(3, 806) = 3.29	**.020**
	RV	Cognitive load level	F(1, 806) = 10.27	**.001**
		NDRT type	F(3, 806) = 0.23	.877
		Cognitive load level * NDRT type	F(3, 806) = 0.46	.712
	RR	Cognitive load level	$\chi^2(1) = 0$	.952
		NDRT type	$\chi^2(3) = 3.26$	.354
		Cognitive load level * NDRT type	$\chi^2(3) = 7.54$	.056

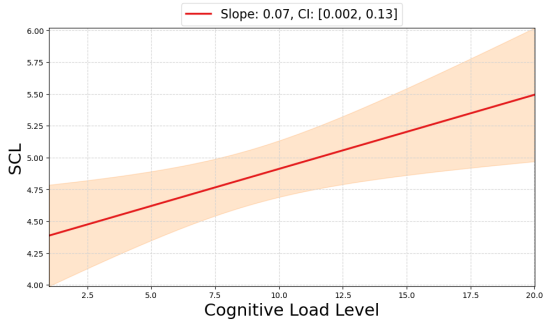
Notes: * means interaction effect. The bold font indicates significant results ($p < .05$).



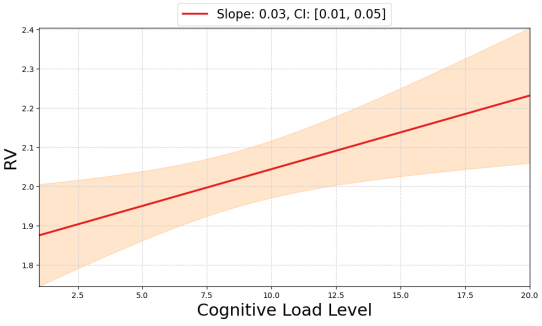
(a)



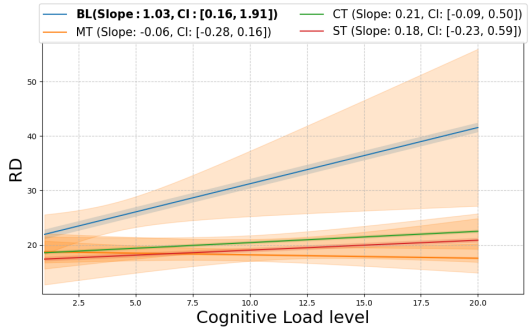
(b)



(c)



(d)



(e)

Figure 6. The effects of NDRT type and cognitive load level on physiological response. The shaded areas are the 95% CI. The bold fonts mean the *p*-value < .05.

4.3 Eye-tracking metrics

Table 6 shows the influence of NDRT type and cognitive load level on eye-tracking metrics. In general, cognitive load level significantly affected NF, NS, and SA; and NDRT types significantly affected PA. Specifically, for each 1-unit increase in cognitive load level, NF increased by 1.02% (95% CI: [1.01%, 1.03%], χ^2 (1) = 13.78, $p < .001$), the expected value of NS increases by 1.03% (95% CI: [1.01%, 1.04%], χ^2 (1) = 18.03, $p < .001$), and SA decreased by 0.11 degrees (95% CI: [-0.19, -0.02], t(806) = -2.39, $p = .017$). At the same time, when the cognitive load level is controlled, the PA in baseline condition (without NDRT) was lower than that with spatial task (Δ = -155.14, 95% CI: [-295.37, -14.92], $p = .029$).

Table 6. Statistic results of multi-dimensional cognitive resources on eye-tracking behavior.

DV		IV	F/χ^2 value	*p*-value
Eye-tracking	PA	Cognitive load level	F(1, 806) = 2.86	.993
		NDRT type	F(3, 806) = 0.00	**.036**
		Cognitive load level * NDRT type	F(3, 806) = 0.99	.397
	FT	Cognitive load level	F(1, 806) = 0.6	.438
		NDRT type	F(3, 806) = 0.69	.558
		Cognitive load level * NDRT type	F(3, 806) = 1.13	.336
	NF	Cognitive load level	χ^2(1) = 5.68	**.017**
		NDRT type	χ^2(3) = 4.71	.194

		Cognitive load level * NDRT type	$\chi^2(3) = 3.2$	.361
	ST	Cognitive load level	$F(1, 806) = 0.02$	.902
		NDRT type	$F(3, 806) = 0.68$	.562
		Cognitive load level * NDRT type	$F(3, 806) = 0.03$	.992
	NS	Cognitive load level	$\chi^2(1) = 5.52$	**.019**
		NDRT type	$\chi^2(3) = 4.27$	.234
		Cognitive load level * NDRT type	$\chi^2(3) = 2.41$	.492
	SA	Cognitive load level	$F(1, 806) = 5.52$	**.017**
		NDRT type	$F(3, 806) = 1.52$	.208
		Cognitive load level * NDRT type	$F(3, 806) = 1.4$	.242

Notes: * means interaction effect. The bold font indicates significant results ($p < .05$).

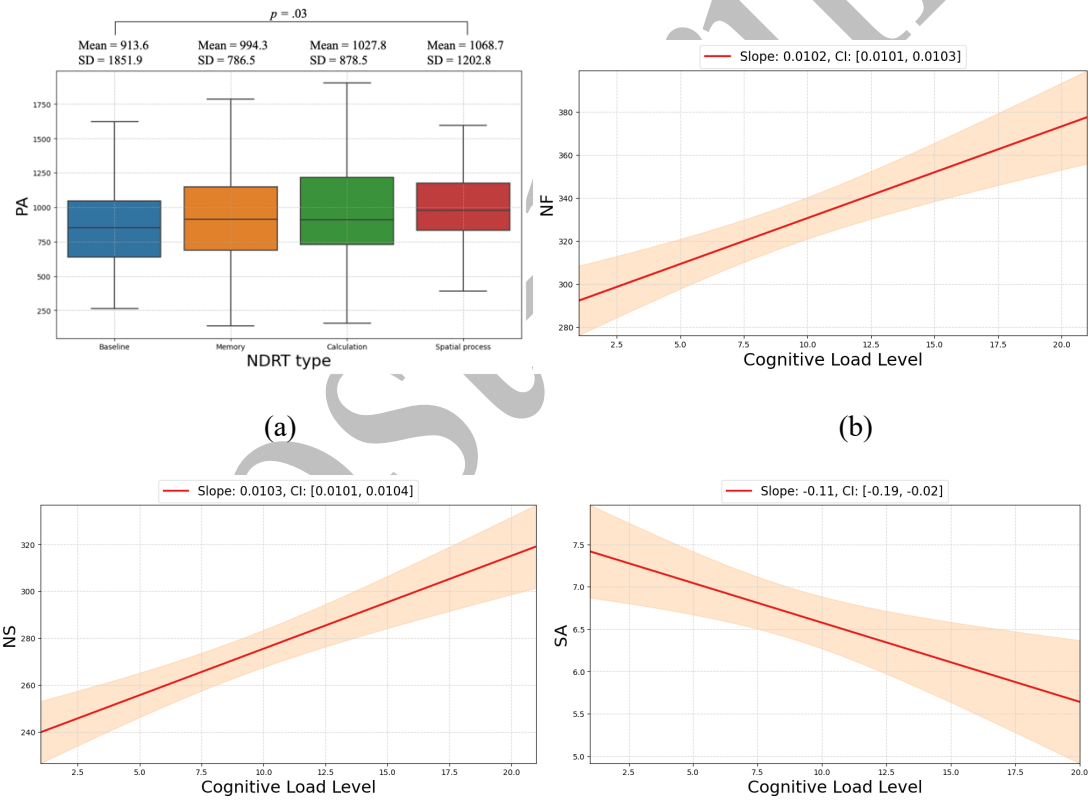


Figure 7. The effects of NDRT type and cognitive load level on eye-tracking metrics. The shaded areas are the 95% CI. The bold fonts mean the *p*-value < .05.

5. Discussions

5.1 Cognitive Tasks and Takeover Performance

In general, in line with previous driving-related research (e.g., (Guo et al., 2023, 2023; Zhang, 2024)), a higher level of perceived cognitive load would lead to a longer takeover time. This is expected, as the increased cognitive load would occupy the cognitive resources needed to perceive the environment and select appropriate actions (Wickens, 1991). In addition, as mentioned, cognitive resources are not unidimensional, and our research has provided a higher resolution regarding the influence of the task modality on takeover performance. We found that given a cognitive load level, the calculation task, memory task, and spatial task all led to longer TOT than the baseline condition without NDRTs. However, even with the same level of perceived cognitive load, different types of cognitive load influenced the takeover performance differently. Especially, the cognitive load levels due to the calculation task had a stronger influence on the TOT, as compared to memory tasks. Thus, though all high cognitive tasks should be minimized in the cabin, designers should be more cautious when implementing in-vehicle tasks involving complex cognitive calculations (e.g., calculating the remaining distance in navigation) compared to tasks involving only memory tasks (e.g., recalling the name of a destination), as for the same level of perceived cognitive load, cognitive calculation tasks are expected to delay the takeover response time by 0.07s, as compared to memory tasks. These findings indicate that the cognitive load level alone may not fully explain the performance differences when the drivers are cognitively distracted and we need to consider cognitive load and task type when designing systems for automated driving.

At the same time, also in line with previous research (Chen et al., 2023; Zeeb et al., 2016), the perceived cognitive load before TOR can affect the post-takeover performance, with a decreased speed and smaller lateral acceleration being observed with increased cognitive load. This phenomenon may be attributed to the self-modulation phenomenon (Kim et al., 2023), where participants tended to adjust their driving behaviors in response to the changes in their perceived risk (Fuller, 2005) (as modulated by the spare cognitive recourses in our scenarios). However, the NDRT type had little effect on takeover quality. It is likely that after the takeover,

drivers' attention has shifted from NDRT to driving tasks. Thus, the original task modality has little influence on the takeover quality.

5.2 The Effect of Cognitive Tasks on Physiological Responses and Eye-Tracking Metrics

We found that RMSSD decreased and SCL increased as cognitive load increased, likely due to increased sympathetic activity (Heine et al., 2017). These results are in line with previous studies (Nilsson et al., 2022; Tjolleng et al., 2017). For example, in a driving simulation experiment with the n-back task as the secondary task, Tjolleng et al. (2017) found that the time-domain HRV metrics decreased with increasing mental workload. Previous research has also found that cognitive load levels can affect SCL. For example, Ding et al. (2020), and Nourbakhsh et al. (2012) both reported increased skin conductance when drivers are engaging in mental arithmetic tasks. This is because when the sympathetic nervous system is highly activated (as a result of high cognitive load), sweat gland activity increases, leading to a higher skin conductance. In addition, selected eye-tracking metrics were also found to be associated with cognitive load levels. We found that the number of fixations (NF) and the number of saccades (NS) increased with the increase of drivers' perceived cognitive load, while the saccade angle (SA) decreased. These results are in line with previous research (Chen et al., 2022; Harris et al., 2023; Zagermann et al., 2018) and confirmed that the NDRTs we introduced have successfully imposed high cognitive load among drivers. These findings also indicate that certain eye-tracking and physiological measures may still be used for cognitive load estimation in SAE Level-3 vehicles.

However, the responses of some metrics were unexpected. For example, we observed no association between HR and cognitive load level; while most previous research observed a positive relationship between HR and cognitive load (Gable et al., 2015; He et al., 2019; Liu et al., 2017; Mehler et al., 2009, 2012b; Wang et al., 2024), though a few other studies also failed to observe the relationships between the cognitive load and HR (Iani et al., 2004). It is worth noting that these studies were based on manual driving, while in L3 automated driving mode, the vehicle is controlled by the automated system, and most of the time, the driver does not

need to actively operate the vehicle, leading to a lower overall workload. This may explain the more pronounced HR variations in manual driving mode, given HR is more sensitive to stress and task complexity (Chalmers et al., 2022).

In addition, again, we found that the drivers' physiological and eye-tracking metrics were not solely decided by cognitive load level. For example, LF/HF and respiration depth (RD) were affected by both the perceived cognitive load level and the NDRT type. Specifically, only in baseline, the LF/HF and RD increase with the increase of the cognitive load. In other words, the change in LF/HF was significantly higher under no NDRT condition compared to that when memory, calculation, and spatial tasks were provided. At the same time, we found that the PA (i.e., the area of pupil diameter) was solely associated with the NDRT type but not with the increase of the perceived cognitive load. These findings suggest that cognitive load alone affected certain physiological and eye-tracking metrics, but this effect may be shadowed by other modalities of the NDRTs. In other words, the influence of cognitive resources needed in other NDRT modalities may have shadowed the influence of the cognitive load in the context of SAE conditionally automated driving. Thus, future driver state monitoring algorithms should consider the modality of the NDRT task to improve the model performance.

5.3 Practical Applications for Driver Monitoring Systems (DMS)

For the first time, we found that the cognitive task types can affect the physiological, eye-tracking metrics and takeover speed in vehicles with driving automation. Thus, future DMS should not treat the high cognitive load as a single state and may need to take the task-type effects into consideration when predicting the takeover performance or estimating drives' cognitive load. At the same time, as not all physiological and eye-tracking metrics are responsive to all types of cognitive tasks, an algorithm relying on single metrics may fail to capture the high cognitive load induced by specific types of cognitive tasks. Thus, future cognitive load estimation algorithms may need to take multiple metrics into consideration if the model aims to evaluate the general cognitive load of drivers.

6. Limitations

It should be noted that, although the experiment included 42 participants, which is comparable to previous studies (Chen et al., 2023; Du et al., 2020; He et al., 2019), the sample size may still not be large enough to draw solid conclusions regarding the relationships among some variables. Further, the inherent individual variability may also influence the results (Goodridge et al., 2024). However, we opted not to take the individual differences into consideration due to the relatively small sample size. Instead, to reduce the uncontrolled variation within participants, we focused on a relatively young population that had limited experience with ADAS. Future research may address this limitation by having a more diverse participant population, for example, experienced ADAS users and older drivers, as they may exhibit different physiological and eye-tracking responses to high cognitive loads. At the same time, another limitation of this study is the gender imbalance in the sample, as we did not control for a balanced female-male ratio. This was done intentionally to reflect the current gender distribution of licensed drivers in the region where the data was collected (Car Circle, 2022). However, we acknowledge that a more balanced gender distribution might better mitigate any potential gender-related effects on cognitive load or performance in automated driving tasks. Future studies may benefit from a more controlled sample to investigate the potential impact of gender differences on these factors. Next, we only evaluated the physiological measures and eye-tracking measures for practical considerations (both signals have the potential to be collected non-intrusively). However, EEG signals also have the potential to be used for cognitive load estimation (Ma et al., 2024), which should be evaluated in future studies. Finally, participants experienced multiple takeover scenarios within a relatively short period, with

actual road risks being eliminated in a driving simulator. Consequently, drivers' behaviors and physiological responses may have been skewed from what they are in the real world as a result of the potential fatigue and learning effects, as well as the simulated traffic environment, though the counterbalancing technique was used in the experiment design. Future studies should address these limitations by conducting on-road experiments with larger sample sizes in the field or naturalistic driving studies.

7. Conclusions

This study investigated the effects of multidimensional cognitive demands on drivers' takeover performance, physiological responses, and eye-tracking metrics in conditionally automated driving in a driving simulator study. We found that:

- Cognitive load can affect takeover performance, with a higher load leading to longer takeover time. However, the effects of the cognitive load can be moderated by the cognitive task type, with the cognitive load imposed by the calculation task leading to the largest variation in takeover time. At the same time, increased cognitive load led to more conservative driving behavior, as reflected by reduced speed and lateral acceleration after the takeover. These findings suggest that future in-vehicle task design should consider the modality of the cognitive demands for the driving safety of conditionally automated vehicles.

- Selected physiological responses (RMSSD, RV, SCL) and eye-tracking metrics (NF, NS, and SA) are still sensitive to the variations of cognitive load during conditionally automated driving. However, the relationships between certain physiological features (LF/HF, RD) and cognitive load can be moderated by the types of cognitive tasks, indicating the need to consider NDRT types when estimating drivers' cognitive load in the context of conditionally automated driving.

These findings can help develop tailored in-vehicle tasks and improve the performance of driver monitoring systems in conditionally automated driving systems. Future work should explore how these effects translate to real-world driving performance and safety interventions.

CRediT authorship contribution statement

Ange WANG: Data curation, Methodology, Software, Validation, Formal analysis, Writing – original draft. Wenxin SHI: Data curation, Investigation, Writing – review & editing. Dengbo HE: Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision. Hai YANG: Supervision.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the Natural Science Foundation of Guangdong Province of China (2024A1515010392), and partially by the National Natural Science Foundation of China (52202425), Guangzhou Municipal Science and Technology Project (2023A03J0011) and the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (2023B1212010007).

References

Aisin Mobility. (2022, February 3). *Deadlines for driver monitoring systems are approaching fast*. Aisin Mobility. https://www.aisin-mobility.eu/whatsnew/get-out-there-hxgeh

Broadbent, D. P., D'Innocenzo, G., Ellmers, T. J., Parsler, J., Szameitat, A. J., & Bishop, D. T. (2023). Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*, 121–132. Scopus. https://doi.org/10.1016/j.trf.2022.11.013

Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, *23*(4), 309–316. https://doi.org/10.1016/0001-4575(91)90008-S

Cao, Y., Zhou, F., Pulver, E. M., Molnar, L. J., Robert, L. P., Tilbury, D. M., & Yang, X. J. (2021). Towards standardized metrics for measuring takeover performance in conditionally automated driving: A systematic review. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 1065–1069. https://doi.org/10.1177/1071181321651213

Car Circle. (2022, March). *In the past five years, the growth rate of female drivers was more than twice that of men*. https://chejiahao.autohome.com.cn/info/10331585/

Cegovnik, T., Stojmenova, K., Jakus, G., & Sodnik, J. (2018). An analysis of the

suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *APPLIED ERGONOMICS*, *68*, 1–11. Scopus. https://doi.org/10.1016/j.apergo.2017.10.011

Chalmers, T., Hickey, B. A., Newton, P., Lin, C.-T., Sibbritt, D., McLachlan, C. S., Clifton-Bligh, R., Morley, J., & Lal, S. (2022). Stress Watch: The Use of Heart Rate and Heart Rate Variability to Detect Stress: A Pilot Study Using Smart Watch Wearables. *Sensors*, *22*(1), Article 1. https://doi.org/10.3390/s22010151

Chen, H., Zhao, X., Li, Z., Fu, Q., Wang, Q., & Zhao, L. (2023). Construction and Analysis of Driver Takeover Behavior Modes Based on Situation Awareness Theory. *IEEE Transactions on Intelligent Vehicles*, 1–15. https://doi.org/10.1109/TIV.2023.3280077

Chen, H., Zhao, X., Li, Z., Li, H., Gong, J., & Wang, Q. (2023). Study on the influence factors of takeover behavior in automated driving based on survival analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, *95*, 281–296. https://doi.org/10.1016/j.trf.2023.04.012

Chen, W., Sawaragi, T., & Hiraoka, T. (2022). Comparing eye-tracking metrics of mental workload caused by NDRTs in semi-autonomous driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *89*, 109–128. Scopus. https://doi.org/10.1016/j.trf.2022.05.004

Choi, D., Sato, T., Ando, T., Abe, T., Akamatsu, M., & Kitazaki, S. (2020). Effects of cognitive and visual loads on driving performance after take-over request (TOR) in automated driving. *Applied Ergonomics*, *85*, 103074. https://doi.org/10.1016/j.apergo.2020.103074

Das, S., & Maiti, J. (2024). Assessment of cognitive workload based on information theory enabled eye metrics. *Safety Science*, *176*, 106567. https://doi.org/10.1016/j.ssci.2024.106567

Deng, M., Gluck, A., Zhao, Y., Li, D., Menassa, C. C., Kamat, V. R., & Brinkley, J. (2024). An analysis of physiological responses as indicators of driver takeover readiness in conditionally automated driving. *Accident Analysis & Prevention*, *195*, 107372. https://doi.org/10.1016/j.aap.2023.107372

Ding, Y., Cao, Y., & Wang, Y. (2020). Physiological Indicators of Mental Workload in Visual Display Terminal Work. In R. S. Goonetilleke & W. Karwowski (Eds.), *Advances in Physical Ergonomics and Human Factors* (pp. 86–94). Springer International Publishing. https://doi.org/10.1007/978-3-030-20142-5_9

Dingus, T. A., Owens, J. M., Guo, F., Fang, Y., Perez, M., McClafferty, J., Buchanan-King, M., & Fitch, G. M. (2019). The prevalence of and crash risk associated with primarily cognitive secondary tasks. *Safety Science*, *119*, 98–105. https://doi.org/10.1016/j.ssci.2019.01.005

Du, N., Kim, J., Zhou, F., Pulver, E., Tilbury, D. M., Robert, L. P., Pradhan, A. K., & Yang, X. J. (2020). Evaluating Effects of Cognitive Load, Takeover Request Lead Time, and Traffic Density on Drivers' Takeover Performance in Conditionally Automated Driving. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 66–73. https://doi.org/10.1145/3409120.3410666

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 97–120. https://doi.org/10.1016/j.trf.2005.04.012

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention*, *37*(3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003

Gable, T., Walker, B., Kun, A., Winton, R., & ACM. (2015). *Comparing Heart Rate and Pupil Size as Objective Measures of Workload in the Driving Context: Initial Look* (WOS:000382169000004). 20–25. Scopus. https://doi.org/10.1145/2809730.2809745

Goodridge, C. M., Gonçalves, R. C., Arabian, A., Horrobin, A., Solernou, A., Lee, Y. T., Lee, Y. M., Madigan, R., & Merat, N. (2024). Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. *Accident Analysis & Prevention*, *202*, 107560. https://doi.org/10.1016/j.aap.2024.107560

Grassmann, M., Vlemincx, E., von Leupoldt, A., Mittelstädt, J. M., & Van den Bergh, O. (2016). Respiratory Changes in Response to Cognitive Load: A Systematic Review. *Neural Plasticity*, *2016*, e8146809. https://doi.org/10.1155/2016/8146809

Guo, L., Xu, L., Ge, P., & Wang, X. (2023). How Resource Demands of Nondriving-Related Tasks and Engagement Time Affect Drivers' Physiological Response and Takeover Performance in Conditional Automated Driving. *IEEE Transactions on Human-Machine Systems*, *53*(3), 600–609. https://doi.org/10.1109/THMS.2023.3268095

Hajek, W., Gaponova, I., Fleischer, K. H., & Krems, J. (2013). Workload-adaptive cruise control—A new generation of advanced driver assistance systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, *20*, 108–120. Scopus. https://doi.org/10.1016/j.trf.2013.06.001

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, *39*(2), 372–379. https://doi.org/10.1016/j.aap.2006.08.013

Harris, A. M., Eayrs, J. O., & Lavie, N. (2023). Establishing gaze markers of perceptual load during multi-target visual search. *Cognitive Research: Principles and Implications*, *8*(1), 56. https://doi.org/10.1186/s41235-023-00498-7

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

He, D., Donmez, B., Liu, C. C., & Plataniotis, K. N. (2019). High Cognitive Load Assessment in Drivers Through Wireless Electroencephalography and the Validation of a Modified N-Back Task. *IEEE Transactions on Human-Machine Systems*, *49*(4), 362–371. IEEE Transactions on Human-Machine Systems. https://doi.org/10.1109/THMS.2019.2917194

He, D., Wang, Z., Khalil, E. B., Donmez, B., Qiao, G., & Kumar, S. (2022). Classification of Driver Cognitive Load: Exploring the Benefits of Fusing Eye-

Tracking and Physiological Measures. *Transportation Research Record: Journal of the Transportation Research Board*, *2676*(10), 670–681. https://doi.org/10.1177/03611981221090937

Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., & Deml, B. (2017). Electrocardiographic features for the measurement of drivers' mental workload. *Applied Ergonomics*, *61*, 31–43. https://doi.org/10.1016/j.apergo.2016.12.015

Huang, C., He, D., Wen, X., & Yan, S. (2023). Beyond adaptive cruise control and lane centering control: Drivers' mental model of and trust in emerging ADAS technologies. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1236062

Iani, C., Gopher, D., & Lavie, P. (2004). Effects of task difficulty and invested mental effort on peripheral vasoconstriction. *Psychophysiology*, *41*(5), 789–798. https://doi.org/10.1111/j.1469-8986.2004.00200.x

J3016—Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. (2018). SAE International J3016_201806 (2018).

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, *18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Jeong, E., Oh, C., & Lee, S. (2017). Is vehicle automation enough to prevent crashes? Role of traffic operations in automated driving environments for traffic safety. *Accident Analysis & Prevention*, *104*, 115–124. https://doi.org/10.1016/j.aap.2017.05.002

Kalnins, A. (2018). Multicollinearity: How common factors cause Type 1 errors in multivariate regression. *Strategic Management Journal*, *39*(8), 2362–2385. https://doi.org/10.1002/smj.2783

Kim, K., Akbas, T., Lee, R., Manella, K., & Sulzer, J. (2023). Self-modulation of rectus femoris reflex excitability in humans. *Scientific Reports*, *13*(1), 8134. https://doi.org/10.1038/s41598-023-34709-4

Kramer, C. Y. (1956). Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications. *Biometrics*, *12*(3), 307–310. https://doi.org/10.2307/3001469

Lee, S. C., Yoon, S. H., & Ji, Y. G. (2021). Effects of Non-Driving-Related Task Attributes on Takeover Quality in Automated Vehicles. *International Journal of Human–Computer Interaction*, *37*(3), 211–219. https://doi.org/10.1080/10447318.2020.1815361

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, *42*(3), 881–890. https://doi.org/10.1016/j.aap.2009.05.001

Lin, Q., Li, S., Ma, X., & Lu, G. (2020). Understanding take-over performance of high crash risk drivers during conditionally automated driving. *Accident Analysis & Prevention*, *143*, 105543. https://doi.org/10.1016/j.aap.2020.105543

Liu, Y., Ayaz, H., & Shewokis, P. A. (2017). Multisubject "Learning" for Mental

Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00389

Liu, Y., Wu, C., Zhang, H., Ding, N., Xiao, Y., Zhang, Q., & Tian, K. (2024). Safety evaluation and prediction of takeover performance in automated driving considering drivers' cognitive load: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *103*, 35–52. https://doi.org/10.1016/j.trf.2024.03.021

Louw, T., Kountouriotis, G., & Carsten, O. (n.d.). DRIVER DISTRACTION DURING VEHICLE AUTOMATION: HOW DOES DRIVER ENGAGEMENT AFFECT RESUMPTION OF CONTROL? *New South Wales*.

Ma, S., Yan, X., Billington, J., Merat, N., & Markkula, G. (2024). Cognitive load during driving: EEG microstate metrics are sensitive to task difficulty and predict safety outcomes. *Accident Analysis & Prevention*, *207*, 107769. https://doi.org/10.1016/j.aap.2024.107769

Mathissen, M., Hennes, N., Faller, F., Leonhardt, S., & Teichmann, D. (2022). Investigation of Three Potential Stress Inducement Tasks During On-Road Driving. *IEEE Transactions on Intelligent Transportation Systems*, *23*(5), 4823–4832. Scopus. https://doi.org/10.1109/TITS.2021.3112811

Mehler, B., Reimer, B., & Coughlin, J. (2012a). *MIT AgeLab Delayed Digit Recall Task (n-back)(Paper 2011-3B)(2011-06-28)* [Dataset].

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *TRANSPORTATION RESEARCH RECORD*, *2138*, 6–12. Scopus. https://doi.org/10.3141/2138-02

Mehler, B., Reimer, B., & Coughlin, J. F. (2012b). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task: An On-Road Study Across Three Age Groups. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(3), 396–412. https://doi.org/10.1177/0018720812442086

Meteier, Q., Capallera, M., Ruffieux, S., Angelini, L., Abou Khaled, O., Mugellini, E., Widmer, M., & Sonderegger, A. (2021). Classification of Drivers' Workload Using Physiological Signals in Conditional Automation. *Frontiers in Psychology*, *12*. Scopus. https://doi.org/10.3389/fpsyg.2021.596038

Meteier, Q., De Salis, E., Capallera, M., Widmer, M., Angelini, L., Abou Khaled, O., Sonderegger, A., & Mugellini, E. (2022). Relevant Physiological Indicators for Assessing Workload in Conditionally Automated Driving, Through Three-Class Classification and Regression. *Frontiers in Computer Science*, *3*. Scopus. https://doi.org/10.3389/fcomp.2021.775282

Nilsson, E. J., Bärgman, J., Ljung Aust, M., Matthews, G., & Svanberg, B. (2022). Let Complexity Bring Clarity: A Multidimensional Assessment of Cognitive Load Using Physiological Measures. *Frontiers in Neuroergonomics*, *3*, 787295. https://doi.org/10.3389/fnrgo.2022.787295

Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. *Proceedings*

of the 24th Australian Computer-Human Interaction Conference, 420–423. https://doi.org/10.1145/2414536.2414602

Öztürk, İ., Merat, N., Rowe, R., & Fotios, S. (2023). The effect of cognitive load on Detection-Response Task (DRT) performance during day- and night-time driving: A driving simulator study with young and older drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *97*, 155–169. https://doi.org/10.1016/j.trf.2023.07.002

Radhakrishnan, V., Louw, T., Cirino Gonçalves, R., Torrao, G., Lenné, M. G., & Merat, N. (2023). Using pupillometry and gaze-based metrics for understanding drivers' mental workload during automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *94*, 254–267. Scopus. https://doi.org/10.1016/j.trf.2023.02.015

Rahman, H., Ahmed, M. U., Barua, S., & Begum, S. (2020). Non-contact-based driver's cognitive load classification using physiological and vehicular parameters. *Biomedical Signal Processing and Control*, *55*, 101634. https://doi.org/10.1016/j.bspc.2019.101634

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied*, *6*(1), 31–43. https://doi.org/10.1037/1076-898X.6.1.31

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*(2), 119–137. https://doi.org/10.1037/1076-898X.9.2.119

Roche, F., Somieski, A., & Brandenburg, S. (2019). Behavioral Changes to Repeated Takeovers in Highly Automated Driving: Effects of the Takeover-Request Design and the Nondriving-Related Task Modality. *Human Factors*, *61*(5), 839–849. Scopus. https://doi.org/10.1177/0018720818814963

S. Kumar, D. He, G. Qiao, & B. Donmez. (2022). Classification of Driver Cognitive Load based on Physiological Data: Exploring Recurrent Neural Networks. *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*, 19–24. https://doi.org/10.1109/ICARM54641.2022.9959588

Sathyapriya, L., Murali, L., & Manigandan, T. (2014). Analysis and detection R-peak detection using Modified Pan-Tompkins algorithm. *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 483–487. https://doi.org/10.1109/ICACCCT.2014.7019490

Schartmüller, C., Weigl, K., Löcken, A., Wintersberger, P., Steinhauser, M., & Riener, A. (2021). Displays for Productive Non-Driving Related Tasks: Visual Behavior and Its Impact in Conditionally Automated Driving. *Multimodal Technologies and Interaction*, *5*(4), Article 4. https://doi.org/10.3390/mti5040021

Seyfzadehdarabad, F., Sadeghi-Firoozabadi, V., Shokri, O., Bagheri, M., & Sadeghi Firoozabadi, A. (2023). Cognitive correlates of maritime pilots' human errors. *Safety Science*, *165*, 106196. https://doi.org/10.1016/j.ssci.2023.106196

Shi, W., Wang, Z., Wang, A., & He, D. (2024). Classification of Driver Cognitive Load in Conditionally Automated Driving: Utilizing Electrocardiogram-Based Spectrogram with Lightweight Neural Network. *Transportation Research*

Record. https://doi.org/10.1177/03611981241252797

Tivesten, E., & Dozza, M. (2014). Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 258–272. https://doi.org/10.1016/j.trf.2014.08.004

Tjolleng, A., Jung, K., Hong, W., Lee, W., Lee, B., You, H., Son, J., & Park, S. (2017). Classification of a Driver's cognitive workload levels using artificial neural network on ECG signals. *Applied Ergonomics*, *59*, 326–332. https://doi.org/10.1016/j.apergo.2016.09.013

Wang, A., Huang, C., Wang, J., & He, D. (2024). The association between physiological and eye-tracking metrics and cognitive load in drivers: A *meta*-analysis. *Transportation Research Part F: Traffic Psychology and Behaviour*, *104*, 474–487. https://doi.org/10.1016/j.trf.2024.06.014

Wang, A., Wang, J., Shi, W., & He, D. (2024). Cognitive Workload Estimation in Conditionally Automated Vehicles Using Transformer Networks Based on Physiological Signals. *Transportation Research Record*. https://doi.org/10.1177/03611981241250023

Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2016). Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task. *IEEE Transactions on Human-Machine Systems*, *46*(3), 424–435. IEEE Transactions on Human-Machine Systems. https://doi.org/10.1109/THMS.2015.2476818

Wickens, C. D. (1991). Processing resources and attention. *Multiple-Task Performance*, *1991*, 3–34.

Wiediartini, Ciptomulyono, U., & Dewi, R. S. (2023). Evaluation of physiological responses to mental workload in n-back and arithmetic tasks. *Ergonomics*, 1–13. https://doi.org/10.1080/00140139.2023.2284677

Xiashuo. (2024, June 28). *2024 intelligent driving interest group research report*. Sohu Auto. https://roll.sohu.com/a/789116026_407401

Xu, J., Fard, M., Zhang, N., Davy, J. L., & Robinson, S. R. (2024). Cognitive load and task switching in drivers: Implications for road safety in semi-autonomous vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *107*, 1175–1197. https://doi.org/10.1016/j.trf.2024.11.005

Yao, H., An, S., Zhou, H., & Itoh, M. (2021). Driver takeover performance in conditionally automated driving: Sudden system failure situation versus ODD exit situation. *SICE Journal of Control, Measurement, and System Integration*, *14*(1), 89–96. https://doi.org/10.1080/18824889.2021.1913861

Zagermann, J., Pfeil, U., & Reiterer, H. (2018). Studying Eye Movements as a Basis for Measuring Cognitive Load. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3188628

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

Zeeb, K., Härtel, M., Buchner, A., & Schrauf, M. (2017). Why is steering not the same as braking? The impact of non-driving related tasks on lateral and longitudinal driver interventions during conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *50*, 65–79. https://doi.org/10.1016/j.trf.2017.07.008

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics*, *44*(4), 1049–1060. https://doi.org/10.2307/2531734

Zénon, A. (2019). Eye pupil signals information gain. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1911), 20191593. https://doi.org/10.1098/rspb.2019.1593

Zhang, N., Fard, M., Xu, J., Davy, J. L., & Robinson, S. R. (2023). Influence of non-driving related tasks on driving performance after takeover transition in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *96*, 248–264. https://doi.org/10.1016/j.trf.2023.05.009

Zhang, Y. (2024). *Non-Driving-Related Tasks Influencing Drivers' Takeover Time: A Meta-Analysis* (arXiv:2405.18667). arXiv. https://doi.org/10.48550/arXiv.2405.18667

Zheng, L., Qiao, X.-Q., Ni, T., Yang, W., & Li, Y.-N. (2021). Driver Cognitive Loads Based on Multi-dimensional Information Feature Analysis. *Zhongguo Gonglu Xuebao/China Journal of Highway and Transport*, *34*(4), 240–250. Scopus. https://doi.org/10.19721/j.cnki.1001-7372.2021.04.021

Zhou, Y., Xu, Z., Niu, Y., Wang, P., Wen, X., Wu, X., & Zhang, D. (2022). Cross-Task Cognitive Workload Recognition Based on EEG and Domain Adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 50–60. https://doi.org/10.1109/TNSRE.2022.3140456