

Evaluating the Impacts of Situational Awareness and Mental Stress on Takeover Performance Under Conditional Automation

Shubham Agrawal¹, Srinivas Peeta^{2,*}

¹ Department of Sociology, Anthropology and Criminal Justice, Clemson University, Clemson, SC

² School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA

Abstract

Several safety concerns emerge for the transition of control from the automated driving system to a human driver after the vehicle issues a takeover warning under conditional vehicle automation (SAE Level 3). In this context, recent advances in in-vehicle driver monitoring systems enable tracking drivers' physiological indicators (e.g., eye-tracking and heart rate (HR) measures) to assess their real-time situational awareness (SA) and mental stress. This study seeks to analyze differences in driver's SA and mental stress over time (i.e., successive experiment runs) using these physiological indicators to assess their impacts on takeover performance. We use eye-tracking measures (i.e., on-road glance rate and road attention ratio) as indicators of driver's SA during automated driving. Further, we use the pre-warning normalized HR (NHR) and HR variability (HRV) as well as the change in NHR and HRV after the takeover warning as indicators of mental stress immediately before and the change in mental stress after the takeover warning, respectively. To analyze the effects of driver state (in terms of SA and mental stress) on the overall takeover performance, this study uses a comprehensive metric, Takeover Performance Index (TOPI), proposed in our previous work (Agrawal and Peeta, 2020). The TOPI combines multiple driving performance indicators while partly accounting for their interdependencies. Results from statistical analyses of data from 134 participants using driving simulator experiments illustrate significant differences in driver state over successive experiment runs, except for the change in mental stress after the takeover warning. Some significant correlations were found between the physiological indicators of SA and mental stress used in this study. Takeover performance model results illustrate a significant negative effect of change in NHR after the takeover warning on the TOPI. However, none of the other physiological indicators show significant impacts on takeover performance. The study findings provide valuable insights to auto manufacturers for designing integrated in-vehicle driver monitoring and warning systems that enhance road safety and user experience.

Keywords: Automated vehicle, Takeover performance, Eye-tracking, Electrocardiography, Driving simulator

* Corresponding author. Address: School of Civil and Environmental Engineering, and H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332. Tel.: +1 (404) 894-2243. Email address: peeta@gatech.edu (S. Peeta)

1. Introduction

SAE Level 2 or partially automated vehicles have been on public roads for some time. These vehicles assist drivers with certain dynamic driving tasks such as lane-keeping and headway maintenance to reduce their driving effort. Unlike partially automated vehicles that expect continuous supervision by drivers, SAE Level 3 or conditionally automated vehicles can control all aspects of the dynamic driving task and allow drivers to engage in non-driving related tasks (NDRTs) (SAE J3016, 2018). Even though drivers are not expected to monitor the vehicle under conditional automation, they are required to immediately take over vehicle control if it issues a takeover warning for venturing out of its operational bounds or for experiencing a system failure. This leads to several concerns regarding the driver's ability to safely resume vehicle control due to NDRT-induced distraction and the resulting loss in situational awareness (SA) (de Winter et al., 2014). Further, like in partially automated vehicles, drivers may also exhibit mental underload, vigilance decrement, and passive fatigue in conditionally automated vehicles after a sustained period of automated driving, which can negatively impact their takeover performance (i.e., resuming manual control after a takeover request) (Körber et al., 2015; Young and Stanton, 2002). Therefore, it is imperative to investigate drivers' mental state (hereafter referred to as "driver state") using human factors constructs such as SA and mental stress that may affect their takeover performance in conditionally automated vehicles. A comprehensive understanding of the impacts of driver state on takeover performance can help auto manufacturers to devise mechanisms (e.g., takeover warning system design) to manage driver state in automated vehicles, and thereby mitigate its adverse effects on road safety and user experience.

Previous studies have evaluated driver state in automated vehicles using self-reported subjective surveys (offline measures); for example, Situation Awareness Rating Technique and Situation Awareness Global Assessment Technique to assess SA (Franz et al., 2015; Large et al., 2018), NASA Task Load Index and Rating Scale Mental Effort to assess workload (Dogan et al., 2019; Yoon and Ji, 2019), and Dundee Stress State Questionnaire to assess mental stress (Heikoop et al., 2018). The drawbacks of these methods, however, are that they are either intrusive (if conducted during the experiment), impractical for real-world applications, or subjected to several memory biases (Nguyen et al., 2019; Spector, 1994).

To avoid the aforementioned limitations of subjective measures, several studies have evaluated driver state using objective physiological indicators. Physiological indicators used to assess SA mainly consist of eye-tracking measures such as horizontal gaze dispersion (Louw and Merat, 2017), blink frequency (Radlmayr et al., 2018a), road attention ratio (Petersen et al., 2019), and on-road glance frequency and duration (Zeeb et al., 2015). Workload and mental stress have been analyzed using a variety of physiological data, including eye-tracking (e.g., pupil diameter, percentage of time with eyes closed, and blink frequency) (de Winter et al., 2014; Merat et al., 2012), cardiovascular activity (e.g., heart rate (HR) and its variability) (Alrefaie et al., 2019; Heikoop et al., 2018), skin conductance (Choi et al., 2018), and electroencephalography (Solís-Marcos and Kircher, 2018). Some studies have also used indirect and less precise methods, including NDRT performance (Petersen et al., 2019) and verbal protocol analysis (where drivers say out loud whatever they are thinking at the moment) (Heikoop et al., 2018), to assess SA in automated driving. Recent advances in in-vehicle driver monitoring systems are providing opportunities to utilize physiological data for estimating driver state in real-time. In this context, our study investigates the efficacy of driver physiological indicators that can be obtained from in-vehicle driver monitoring systems in estimating driver state, and analyzes their impacts on takeover performance under conditional automation. Specifically, we use eye-tracking measures, including on-road glance rate (OGR; i.e., number of eye glances at the road per time unit) and road attention ratio (RAR; i.e., percentage of time spent glancing at the road), as physiological indicators of driver's SA during the automated driving and before the takeover warning. OGR is used because it measures how frequently the driver is monitoring the road

and, thereby, how frequently they obtain updates on the dynamic road environment. Similarly, RAR is used because it measures the ratio of the time duration the driver is monitoring the road compared to looking away from the road and, hence, informs on the level (assuming it is proportional to time) at which they monitor the road environment. Moore and Gugerty (2010) reported that the percentage of time fixating on an area-of-interest (similar to RAR in our context) is a good predictor of SA measured using a questionnaire in an air traffic control task, with a higher fixation time indicating a higher SA. However, it should be noted that SA is a complex construct involving not only the perception of elements in the current situation, as captured by eye-tracking indicators, but also the comprehension of elements and ability to project the future situation (Endsley, 1988). We use pre-warning normalized HR (pre-NHR) and the change in normalized HR (Δ NHR) after the takeover warning as indicators of mental stress immediately before and the change in mental stress after the takeover warning, respectively. HR was normalized to alleviate the individual differences in the base HR. In addition to the normalized HR measures, we also use pre-warning HR variability (HRV) and the change in HRV after the takeover warning to assess mental stress. Heart rate and its variability are used as they have been shown to have a significant correlation with mental stress in the literature, with higher HR and lower HRV associated with higher mental stress (Heikoop et al., 2018; Taelman et al., 2008).

Most existing studies analyze the effects of driver state on takeover performance by manipulating experimental conditions such as NDRT (Yoon and Ji, 2019), traffic density (Gold et al., 2016), and driver impairment (e.g., fatigue and alcohol) (Kreuzmair et al., 2017; Wiedemann et al., 2018). However, these studies do not explicitly analyze the effects of driver’s pre-warning state on takeover performance. In our previous work (Agrawal and Peeta, 2020), we used brain electrical activity patterns recorded using an electroencephalogram (EEG) to create a detailed profile of driver’s pre-warning cognitive state, which cannot be done by using eye-tracking and HR measures, and analyzed its impacts on takeover performance. However, this study focuses on eye-tracking and HR measures because, unlike EEG-based measures, they can be collected via in-vehicle driving monitoring systems, and thereby, have real-world applications. Since drivers’ physiological indicators (e.g., HR and its variability) have been reported to change over time in automated vehicles (de Winter et al., 2014; Heikoop et al., 2018, 2017), we also investigate the differences in driver’s SA and mental stress over time (i.e., successive takeover situation experiences) using their corresponding physiological indicators. In addition, we analyze the correlations between these physiological indicators to investigate the relationship between driver’s SA and mental stress.

Takeover performance is typically modeled using driver reaction time (e.g., takeover time) (Zhang et al., 2019), driving performance indicators (e.g., minimum time-to-collision and maximum deceleration rate) (Lu et al., 2019; Radlmayr et al., 2014), physical performance (e.g., hands-on-wheel time) (Yoon and Ji, 2019), and visual performance (e.g., time-to-first-gaze) (Kim et al., 2018). However, these indicators are often modeled independently, and therefore do not provide a comprehensive understanding of takeover performance. A few recent efforts have been made to propose comprehensive takeover performance metrics (Agrawal and Peeta, 2020; Radlmayr et al., 2018b). This study uses the Takeover Performance Index (TOPI) proposed in our previous work (Agrawal and Peeta, 2020), as it provides a single metric to analyze the effects of driver state on the overall takeover performance (see Section 2.5 for details).

Although several studies have evaluated driver state using physiological indicators measured using electrocardiogram (ECG) and eye-tracking instruments as discussed before, few have analyzed their impacts on takeover performance in conditionally automated vehicles. Zeeb et al. (2015) used on-road glance frequency, on-road glance duration, and maximum eyes-off-road time to classify drivers into low-, medium-, and high-risk groups, and analyzed the differences in takeover performance using brake reaction times and percentage of collisions among these groups. Alrefaie et al. (2019) analyzed the impacts of normalized HR and normalized pupil diameter on driver’s reaction time during the takeover event. They

also correlated the physiological indicators with the mean percentage change in the vehicle's speed and heading angle before the takeover event to evaluate the driver's preparedness to respond to the event. A few other studies investigated the differences in physiological indicators under different experimental conditions (e.g., type of takeover warning) in the context of takeover performance, but did not directly analyze their impacts on takeover performance (Lu et al., 2019; Petersen et al., 2019). A key drawback of the aforementioned studies is that they do not capture the impacts of physiological indicators on the overall takeover performance, as they model multiple takeover performance measures independently. Our study analyzes the impacts of ECG and eye-tracking-based measures on the overall takeover performance using the TOPI.

This study designs driving simulator experiments to evaluate takeover performance in conditionally automated vehicles. Using these experiments, it estimates driver state (i.e., SA and mental stress) from physiological data (i.e., eye-tracking and HR measures) that in the real world can be measured using in-vehicle driver monitoring systems. The study also analyzes the differences in driver state over time (i.e., successive takeover situation experiences through multiple experiment runs). Further, it explores the correlations between the physiological indicators of SA and mental stress, and analyzes their overall impacts on takeover performance. Hereafter, we use the term "physiological indicators" to refer to the six eye-tracking and HR measures used in this study: OGR and RAR during the automated driving before the takeover warning, pre-NHR and pre-HRV immediately before the takeover warning, and the Δ NHR and Δ HRV immediately before and after the takeover warning.

2. Methodology

2.1. Apparatus

2.1.1. Driving simulator

The study experiments were conducted using a driving simulator (AVSimulation, 2020), as shown in Figure 1. The simulator features a full-scale driving cockpit and a steering wheel with force feedback. The driving view was projected on three LCD screens with a 120 degrees field-of-view. Side-view mirrors, rear-view mirror, and speedometer were presented on the screens. A four-lane divided highway with a medium level of ambient traffic (about 6-7 vehicles in a 500-meter radius) and a speed limit of 105 kph (about 65 mph) was created using the SCANeRStudio® 1.7 software. Since a straight road could artificially reduce the driver's mental stress to monitor the driving environment and potentially enhance SA, the highway was deliberately chosen to replicate the road curvatures of sections of the U.S. Interstate 65 between West Lafayette, IN and Chicago, IL. Driving data such as the ego vehicle's speed, acceleration, lateral position, steering wheel angle, curvilinear (i.e., along the curved road) inter-vehicular distance from the immediate downstream vehicle in the same lane (leader vehicle; if any), and the speed of the leader vehicle were collected at 20 Hz frequency.



Figure 1 Driving simulator and non-driving related task

2.1.2. Automated driving system

The automated driving system (ADS) implemented in this study mimicked a conditionally automated vehicle that could perform both longitudinal and lateral control on the highway. It was set to observe the speed limit and stay in the right-most lane unless it had to pass a slower-moving vehicle in the lane. Drivers could activate the ADS by pulling the indicator stalk on the left of the steering wheel while driving in the middle of a lane, and disengage it by either braking or steering at any time. A green icon with the alphabet “A” on the rear-view mirror indicated that the ADS was active.

The ADS provided two types of takeover warnings: an uncertainty alert and a takeover request (TOR). An uncertainty alert (presented as a single auditory chime) denoted that the system is unsure of its ability to drive safely and prompted for driver’s attention to the road without the need for manual intervention (non-mandatory takeover). In contrast, a TOR (presented as three auditory chimes in quick succession) denoted a system failure and required immediate driver intervention (mandatory takeover). In this study, a TOR was always preceded by an uncertainty alert. Thus, if active, the ADS was capable of maintaining road safety even after an uncertainty alert unless a TOR was issued, at which point the ADS was automatically deactivated. The warnings were delivered via two multimedia speakers positioned on each side of the driving cockpit below the screens.

2.1.3. Non-driving related task (NDRT)

During the automated drive, participants were asked to engage in an NDRT that consisted of a repetitive number transcription task. The NDRT was delivered on a touchscreen tablet that was positioned near the center console (see Figure 1). The participants completed a single task by inputting two six-digit numbers and a mathematical operator shown on the left side of the tablet screen into a calculator interface and pressing “submit” when done. This simulated visual-motor distraction for the participants. Numeric characters were chosen over alphanumeric characters to avoid potential language-specific biases.

2.1.4. Electrocardiogram (ECG)

Participants’ heart electrical activities were recorded during the simulated runs using the B-Alert X-24 EEG and ECG systems (Advanced Brain Monitoring, 2017). The ECG recorded electrical signals at a sampling rate of 256 Hz using four electrodes: one at the right collar bone, one at the lowest left rib, and two at the mastoids (reference for measuring the power of the signal). We used the B-Alert software to extract heart rate (HR) using the electrical signals at 1-second intervals (B-Alert, 2009).

2.1.5. Eye-tracking glasses

Wearable eye-tracking glasses, Dikablis Glasses 3 (Ergoneers, 2018), were used to collect participants' eye gaze patterns. Its design allows it to be equipped on top of participants' corrective glasses, if needed, without affecting measurements. We defined a road area-of-interest (AOI) in the DLAB software (Ergoneers, 2018) by attaching QR code tags on the corners of the three simulator screens to compute two eye-tracking statistics related to the road monitoring behavior: on-road glance rate (OGR; the number of glances at the road AOI per unit time) and road attention ratio (RAR; percentage of time spent glancing at the road AOI).

2.2. Experiment design

The study experiments were designed as a part of a larger study that investigates the effects of introductory information on driver's trust in automation and takeover performance. The study participants were divided into two groups based on the content of the introduction video that they viewed. However, our study disregards this grouping as no statistically significant differences were found in the takeover performance between the two groups (see Souders et al., 2021 for details).

Table 1 Safety-critical event details

Run	Time budget for takeover warning		Safety-critical event description
	Uncertainty alert	Takeover request	
1	6 seconds	-	Lateral automation late to change lane (2-second time budget), resulting in a close pass of a slower-moving bus (35 kph)
2	8 seconds	4 seconds	A broken-down vehicle (0 kph) occluded by a slower-moving bus (35 kph) blocking the lane
3	7 seconds	-	Longitudinal automation late to apply brakes (3-second time budget) for a slower-moving bus (55 kph)

Participants completed three experiment runs (approximately 10, 10, and 7 minutes, respectively) with moderate ambient traffic density. Each run culminated in a different safety-critical event that was preceded by a takeover warning. The first and third runs presented a non-mandatory takeover event (i.e., driver intervention is not required to maintain road safety), while the second run presented a mandatory takeover event (i.e., driver intervention is required to avoid a crash). The safety-critical event in the first run starts with the ego vehicle approaching a slower-moving bus in the right lane. The ADS, if activated, issues an uncertainty alert with a 6-second time budget (i.e., the time-to-collision between the ego vehicle and the bus) and shifts the ego vehicle to the left lane with a 2-second time budget if the driver does not intervene, resulting in a close pass. In the second run, the safety-critical event starts with the ego vehicle following a slower-moving bus in the right lane, which completely occludes the driver's view of a broken-down vehicle downstream. The ADS issues an uncertainty alert as the bus moves to the left lane revealing the broken-down vehicle on the screen with an 8-second time budget. If the driver does not intervene until the 4-second time budget mark, the ADS issues a TOR and disengages. This makes it mandatory for the driver to either apply brakes (where applying the maximum deceleration of 12 m/s^2 within 2.78 seconds of the TOR would just prevent a crash), switch to the left lane, or perform a combination of both to avoid crashing with the broken-down vehicle. Similar to the first run, the safety-critical event in the third run starts with the ego vehicle following a slower-moving bus in the right lane. The ADS issues an uncertainty warning with a 7-second time budget and applies brakes to slow down if the driver does not intervene until the time budget is 3 seconds. It then follows the bus until the end of the scenario or intervention from the driver. The safety-

critical events are summarized in Table 1. Further, to avoid unwanted interactions of the ego vehicle with ambient traffic (other than the event vehicles), the ambient traffic was smoothly dissipated by changing the speeds of the ambient vehicles about 2000 meters before the takeover event. Since the larger study focuses on driver's trust in automation that can change with prior experiences, the order of the runs was not randomized to control the order of the takeover experiences for the participants.

2.3. Procedure

Before arriving at the lab for the driving simulator experiments, participants completed an online survey designed to check their participation eligibility (see Section 2.6 for the eligibility criteria) and to gather their sociodemographic and other behavioral/attitudinal details (see Souders et al., 2021 for details). Eligible participants were directed to a website for scheduling a driving simulator session. During the lab visit, participants signed an informed consent form and were introduced to the complete experiment procedure. Then, they completed a practice run in the simulator that involved manually driving the simulated vehicle in an urban environment. This run was designed to acclimatize them with the simulator and to serve as a screening procedure for participants with simulator sickness, based on verbal confirmation from participants and the experimenter's assessment of their physical condition. Next, participants viewed the introductory video for their randomly assigned group. A double-blind manipulation, where neither the experimenter nor the participant knew about the assigned group, was implemented. Then, participants were equipped with the eye-tracker and the EEG and ECG instruments. This study does not analyze EEG data; please see Agrawal and Peeta (2020) for the detailed EEG-related analysis. Next, participants completed another practice run with the ADS and NDRT concurrently on a four-lane divided highway to get acquainted with the ADS's activation/deactivation procedures and takeover warnings. They were also informed about the expected actions for the two takeover warnings (i.e., focusing attention on the road after an uncertainty alert without the need for intervention, and immediate intervention required after a TOR). There was no ambient traffic in the practice run, except for the three slower-moving vehicles that were used initially to illustrate the capabilities of the ADS to pass them. Traffic cones were used as obstacles to simulate potential takeover events that trigger the issuance of a takeover warning. In addition, each takeover warning was immediately preceded by a voice message (i.e., "pay attention to the road" for an uncertainty alert and "take over manual control" for a TOR) to prepare the participants for the event. Participants experienced both types of takeover warnings multiple times (at least three times each) during the practice run. They were asked to activate the ADS about 30 seconds before each warning during the practice. The minimum duration of this practice run was about six minutes; about two minutes for introducing the NDRT and ADS and its capabilities, and about four minutes of minimum takeover practice with the warnings. However, the practice run continued until the participants verbally confirmed that they are comfortable with handling the ADS. Note that the practice run was not designed to provide a realistic takeover experience to participants (relaxed time budget, voice message, and use of traffic cones instead of vehicles), but rather to make them feel comfortable with handling the ADS. Following the practice run, participants completed three experiment runs. They were instructed to keep the ADS activated throughout the run unless they are concerned about road safety. In conditionally automated driving, it is expected that drivers would disengage from driving and engage in other NDRTs. Thus, to promote such behavior, participants were informed that their final compensation for participation would depend on the number of correct responses to the NDRT as well as road safety across all three runs. Regardless, all participants received full compensation of \$45 at the end of the experiment. Participants also completed a few surveys between runs and after the three runs (see Souders et al., 2021 for details). Our study does not use data collected from these surveys. Finally, they were debriefed and compensated for their participation. The experiment protocol was approved by the Purdue University's Institutional Review Board (protocol #1811021326).

2.4. Data preprocessing

Since the base HR differs across individuals, HR was normalized for each participant to alleviate the individual differences. The normalized heart rate (NHR) for each individual is computed using Eq. (1), where HR_t is the heart rate at time step t , and $HR_{\min}$ and $HR_{\max}$ are the minimum and maximum heart rates, respectively, over all three runs. Therefore, the value of NHR is between 0 and 1.

$$NHR = \frac{HR_t - HR_{\min}}{HR_{\max} - HR_{\min}} \quad (1)$$

Pre-warning NHR (pre-NHR) and post-warning NHR (post-NHR) are computed by averaging the NHR for 15-second time windows before and after the issuance of the uncertainty alert, respectively. Similarly, pre-warning heart rate variability (HRV) and post-warning HRV are computed by taking the standard deviation of HR within their respective 15-second time windows. A 15-second time window was chosen to match the time window for the driving data used to compute the TOPI (see Section 2.5). The change in NHR (ΔNHR) and the change in HRV (ΔHRV) due to the takeover warning were calculated as the difference between post- and pre-warning values of NHR and HRV, respectively. We use pre-NHR and pre-HRV to assess driver's pre-warning mental stress, and ΔNHR and ΔHRV to assess the change in mental stress after the takeover warning.

The eye-tracking measures (OGR and RAR) for each run are computed from the beginning of each run until 10 seconds before the issuance of the uncertainty alert. A 10-second time window is chosen to have as little gap as possible before the safety-critical events' vehicle(s) visibly respond to execute the event. Time intervals during which the participant disengaged the ADS (i.e., manual driving) were excluded while computing these measures. We use these measures to assess driver's SA during the automated driving before the takeover event.

2.5. Takeover performance metric

This study uses the comprehensive takeover performance metric, Takeover Performance Index (TOPI), proposed in our previous work (Agrawal and Peeta, 2020). Unlike other takeover performance metrics (e.g., Radlmayr et al., 2018b), the TOPI captures the interdependencies between different driving indicators and provides a single metric to evaluate the overall takeover performance; thereby enabling us to analyze the effects of driver state (i.e., SA and mental stress) and other factors (i.e., novelty in takeover experience and type of takeover warning) on takeover performance.

The TOPI (see Agrawal and Peeta, 2020 for more details) is calculated by combining multiple driving performance indicators representing different aspects of takeover performance: risk of collision, intensity of driver's response, and trajectory quality. For each of the driving performance indicators, a takeover performance sub-index is calculated on a unit interval (i.e., a value between 0 and 1) using a sigmoid (non-linear) function, as illustrated in Eq. (2).

$$T_i = \frac{1}{1 + e^{-W_i(x_i - c_i)}} \quad (2)$$

In Eq. (2), T_i denotes the takeover performance sub-index for the driving performance indicator i and x_i is the value of the indicator. W_i and c_i are the spreading and centering parameters of the sigmoid function that are determined by using insights from the literature and/or by making appropriate assumptions for the given takeover situation. W_i is back-calculated by substituting x_i in Eq. (2) with a reference value of driving performance indicator $\tilde{x}_i$ (spreading parameter reference) at the 75th percentile of T_i (i.e., T_i equal to 0.75).

The TOPI is calculated by taking a geometric mean of all sub-indices, which results in the TOPI value ranging from 0 to 1. A lower value of the TOPI indicates a poor takeover performance. Unlike arithmetic mean, geometric mean inherently assigns more weight to lower values, which results in poor sub-indices (i.e., values closer to 0) having a larger effect on the TOPI value. This makes the TOPI particularly useful to assess safety-critical situations.

Since the current study and our previous study (Agrawal and Peeta, 2020) use data from the same experiments, we use the same four driving performance indicators (i.e., minimum time-to-collision, maximum longitudinal deceleration rate, maximum steering wheel angle, and maximum lateral road deviation) and their corresponding parameter values to compute the TOPI value. Table 2 shows the parameter values used to calculate the TOPI in this study. We use driving data collected within a 15-second time window after the issuance of the uncertainty alert to compute driving performance indicators for the TOPI. Further, the TOPI is not calculated for runs in which participants took over vehicle control before an uncertainty alert was issued.

Table 2 Parameter values used in this study for computing the TOPI

Driving performance indicator	Centering parameter (c_i)	Spreading parameter reference ($\tilde{x}_i$)
Minimum time-to-collision	2.5 seconds	1 second
Maximum deceleration rate	4.5 m/s ²	-2 m/s ²
Maximum steering wheel angle	40 degrees	-10 degrees
Maximum lateral road deviation	0.915 m	-0.457 m

2.6. Participants

Participants were recruited from the Greater Lafayette community in Indiana, through advertisements in the Purdue University's weekly email newsletter, paper fliers at community events, and word of mouth. Participant eligibility criteria included: (i) being 18 years of age or older, (ii) having a valid driver's license, (iii) having no predisposition to motion sickness, (iv) self-reporting the absence of any mental impairments or physical impairments that would make it difficult for them to get in and out of the simulator unassisted, and (v) passing an over-the-phone memory impairment test (Wechsler Logical Memory Scale; Wechsler, 1997) if over the age of 64 years. In addition, participants were asked to avoid consuming any medication or caffeine for at least 8 hours prior to the driving simulator experiment session (see Agrawal and Peeta, 2020 for details).

Overall, 134 drivers participated in this study, yielding a total of 402 experiment runs. Of these, participants took manual control of the vehicle after a takeover warning was issued in 340 runs. Nineteen runs were excluded from the analysis due to an unexpected error (a rogue vehicle appearing in front of the ego vehicle) during the takeover event, resulting in a total of 321 runs for 131 participants with valid TOPI values. Figure 2 shows the age (28.9 ± 13.8 years) and gender (68 females and 63 males) distribution of the participants for these runs. Further, data was filtered to exclude runs with invalid eye-tracking and HR measures from their corresponding analysis (see Section 4). The invalid physiological data was due to two possible reasons: (i) eye-tracker or ECG equipment malfunctioned, resulting in their corresponding missing or poor-quality data, and (ii) problems in synchronizing the physiological data with the driving data due to missing timestamps. Thus, we partly or fully lost physiological data for some experiment runs, including data loss for all three runs for some participants. Table 3 presents the valid data counts for the TOPI, physiological indicators, and their combinations. Thus, the driving data used in this study and a previous study (Agrawal and Peeta, 2020) are drawn from the same overall sample, but use different, non-exclusive

subsamples based on the data exclusion due to physiological data quality issues. The two studies use different physiological data; this study uses eye-tracking and ECG data while the previous study used EEG data.

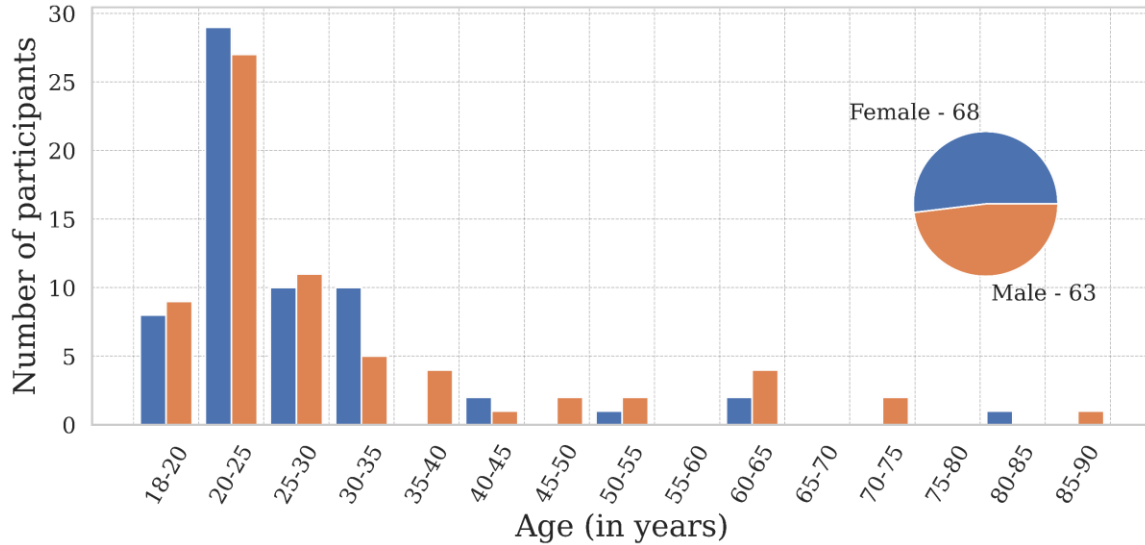


Figure 2 Age and gender distribution of the participants

Table 3 Valid data counts for the TOPI and physiological indicators

TOPI	Measures			Valid counts	
	pre-NHR & pre-HRV	Δ NHR & Δ HRV	OGR & RAR	Runs	Participants
✓				321	131
	✓			250	103
		✓		227	100
✓	✓			216	102
✓	✓	✓		193	99
			✓	219	79
✓			✓	194	79
✓	✓		✓	128	62
✓	✓	✓	✓	114	60

Note: ✓ denotes valid data for the corresponding measures

3. Statistical analyses

As stated earlier, the study investigates the differences in driver's SA and mental stress over time (i.e., successive experiment runs) and their impacts on takeover performance in conditionally automated vehicles. To do so, we use the following physiological measures as indicators of driver state: OGR and RAR for SA during the automated driving, pre-NHR and pre-HRV for pre-warning mental stress, and Δ NHR and Δ HRV for change in mental stress after the takeover warning. We use the TOPI to analyze the effects of driver state on the overall takeover performance. In addition, we analyze the impacts of other factors, including novelty in takeover experience (i.e., an indicator variable for the first run) and type of takeover warning (uncertainty alert and TOR), on takeover performance using the TOPI. We also analyze the correlations between these physiological indicators to investigate the relationship between SA and mental stress in conditionally automated vehicles.

We estimate linear mixed models (LMMs) to analyze the differences in physiological indicators over successive experiment runs. LMM is used for two key reasons: (i) it accounts for the correlated errors that arise due to panel data (multiple runs from the same participant) by modeling them as random effects, and (ii) it provides unbiased results without the need to exclude the entire data (all three runs) for a participant if a single data point (single run) is missing or invalid. Therefore, it enables us to use data for participants who did not take over manual control of the vehicle (i.e., TOPI is not calculated) or did not have valid physiological data in all three runs (see Table 3).

In this study, we estimate 12 LMMs, six to analyze the differences in each physiological indicator over successive runs and another six to analyze the effect of each physiological indicator on the TOPI. The reason for analyzing the impacts of physiological indicators on the TOPI separately is the loss of data for combinations of physiological indicators (see Table 3). The basic form of the estimated LMMs to analyze the differences in physiological indicators is given by Eq. (3) and to analyze their impacts on the TOPI is given by Eq. (4).

$$PI = \beta_0 + \beta_{run} + \gamma + \varepsilon \quad (3)$$

$$(TOPI)^2 \sim \beta_0 + \beta_{PI} + \beta_{nov} + \beta_{TOR} + \gamma + \varepsilon \quad (4)$$

In Eq. (3), PI denotes the physiological indicator and is the dependent variable for the corresponding analysis. We analyze the differences in the physiological indicators over time by using a discrete variable for experiment runs represented by β_{run} . We treat β_{run} as a discrete variable instead of a continuous variable to avoid assuming that the differences in physiological indicators over time are linear. In Eq. (4), $(TOPI)^2$ denotes the square of the TOPI value and is the dependent variable. We transform (squared) TOPI to meet the model assumption of normality of residuals. β_{PI} , β_{nov} , and β_{TOR} denote the effect of the physiological indicator PI, novelty in takeover experience (indicator variable for the first run), and the presence of TOR, respectively. In both equations, β_0 , γ , and ε denote the model intercept, random effects coefficient for the panel data, and normally distributed error term, respectively.

In our previous study (Agrawal and Peeta, 2020), we found a strong positive stepwise trend in driver's takeover performance due to learning effects after their first experience with a takeover situation (i.e., novelty effects). We also reported a strong negative effect of the presence of TOR on takeover performance in the same study. Thus, we model these effects as covariates in the current study to accurately analyze the effects of physiological indicators on takeover performance.

A likelihood-ratio test (LRT) is used to assess the goodness-of-fit of two or more nested models (i.e., one model has all the parameters of the other) (Bollen, 1989). In this study, we assess each LMM with its corresponding null model (i.e., linear model without the random effects) using the LRT. If the LRT does not indicate a better goodness-of-fit of the LMM compared to its null model, the linear model is used to analyze the corresponding effects. The models were estimated using the "lme4 v1.1.17" package in R 4.0.0 (Bates et al., 2015; R Core Team, 2020). The model assumption of normality of residuals is tested using visual inspection of a quantile-quantile (QQ) plot. Further, the statistical significance of differences in the physiological indicators over experiment runs is tested using ANOVA (analysis of variance) for the estimated LMMs. Satterthwaite's formula is used to approximate the effective degrees-of-freedom to account for different sample variances in ANOVA. Post-hoc multiple comparison Tukey's test with Holm-Bonferroni correction for family-wise error rate is also performed to pairwise compare these differences between runs. The correlations between the physiological indicators are estimated using repeated measures correlation analysis to account for the panel data (Bakdash and Marusich, 2017).

4. Results

4.1. Descriptive results

In this experiment, though participants were informed that an uncertainty alert does not entail manual intervention (see Section 2.1.2), most of them (119 out of 131 participants with a valid TOPI) took over manual control of the ADS after an uncertainty alert but before a TOR in at least one of the three runs. We observed crashes in 3 out of 321 runs, one in each of the three runs and by three different participants. The mean and standard deviation of takeover time (i.e., the time elapsed after the issuance of an uncertainty alert before the participant intervened) in this experiment were 3.92 seconds and 1.34 seconds, respectively.

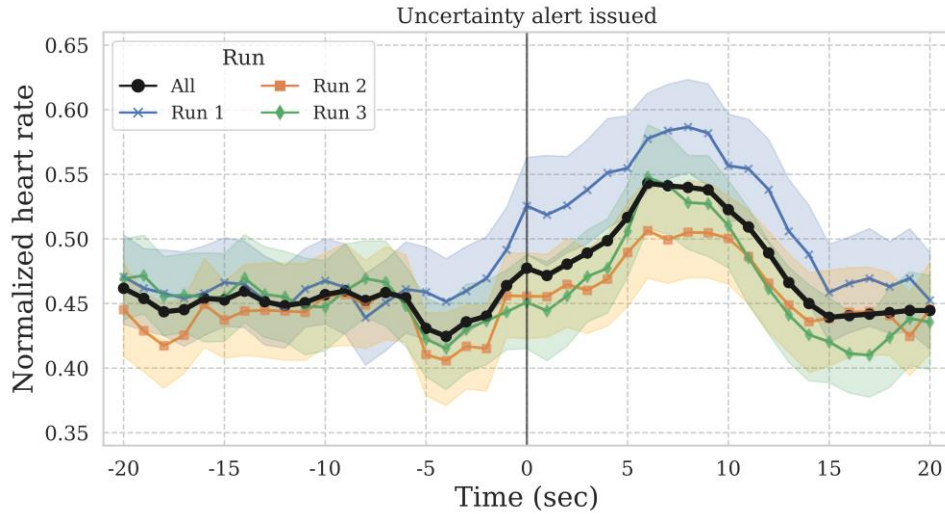


Figure 3 Normalized heart rates before and after the uncertainty alert with 95% confidence interval bands for each run

Figure 3 shows the average NHRs before and after the uncertainty alert for all participants with valid HR data for all three runs and their average. It illustrates that the NHR after the warning increases for about 6 to 8 seconds, which is similar to the time budget for the uncertainty alert used in this experiment (see Table 1), and then returns to normal in about 15 seconds from the uncertainty alert. This is most likely due to the increase in mental stress as participants approach the blocking vehicle, which starts reducing (for most participants) after they have successfully maneuvered around the blocking vehicle. Further, the figure shows that the NHR after the warning is higher in Run 1 compared to the other two runs. This suggests that the participants experienced higher mental stress in the first run that decreased in the later runs, most likely because they became more comfortable with the driving simulator and the overall experiment setup as the experiment progressed.

4.2. Physiological indicators

This section discusses the results of the estimated LMMs and post-hoc analyses to evaluate the differences in driver's SA and mental stress over successive experiment runs using physiological indicators. A visual inspection of the QQ plots suggests that the residuals of all the estimated models are normally distributed. It also analyzes the correlations between these physiological indicators. Figure 4 illustrates the mean and standard deviation of the six physiological indicators across experiment runs.

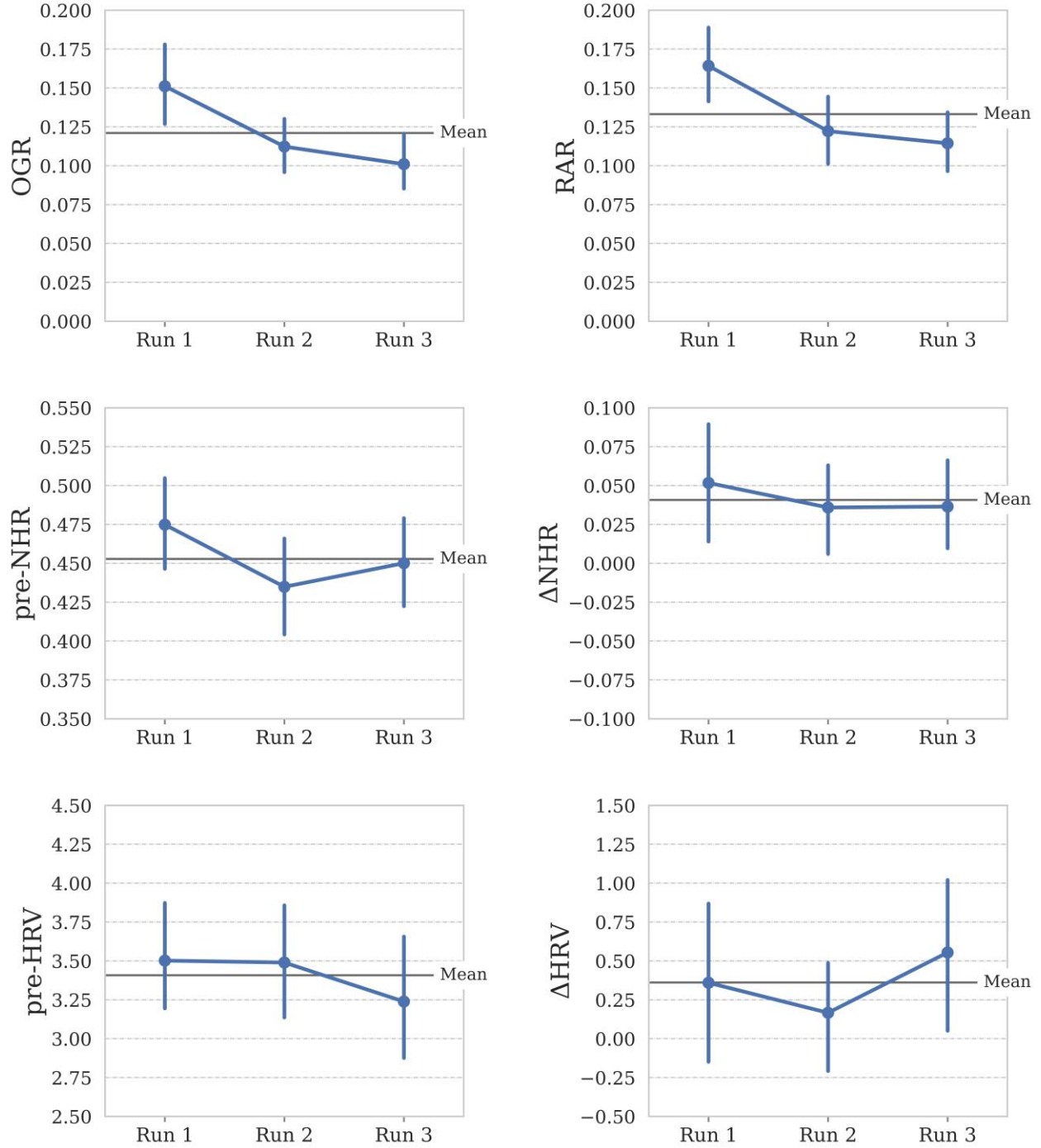


Figure 4 Mean and standard deviation of the six physiological indicators, by run

4.2.1. On-road glance rate (OGR)

The model estimation results for differences in OGR over runs are presented in Table 4. A likelihood-ratio test ($\chi^2(1) = 84.41, p < 0.001$) illustrates a significantly better goodness-of-fit of the LMM ($N = 219, \text{groups} = 79$) compared to its corresponding null model. Model results show significant differences in OGR for different runs ($F(2, 142.7) = 16.241, p < 0.001$). A multiple comparison Tukey's test indicates that OGR is significantly higher in Run 1 compared to Run 2 and Run 3. This suggests that the drivers monitored the road more frequently in the first run, which could be attributed to their first real

experience with the ADS in the driving simulator. In this context, while they practiced driving with the ADS in the practice run, it was in a simplified driving environment with no traffic (see details in Section 2.3). However, no significant differences in OGR were found between Run 2 and Run 3. These results suggest that drivers were quick to adjust the frequency of on-road glances that they believe is sufficient to maintain their desired levels of SA after their first run.

Table 4 Model results for on-road glance rate (OGR)

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	<i>N</i> (runs)	219			
	Groups (participants)	79			
Fixed Effects	Intercept	0.101	0.010	130.6	< 0.001
	Run 1	0.050	0.009	142.2	< 0.001
	Run 2	0.013	0.009	142.3	0.167
Random Effects	Group variance	0.005			
	LRT (χ^2)	84.41		1	< 0.001
ANOVA	Run (<i>F</i>)	16.241		(2, 142.7)	< 0.001
Multiple Comparison	Run 2 – Run 1	-0.038	0.009		< 0.001
	Run 3 – Run 1	-0.050	0.009		< 0.001
Tukey's test	Run 3 – Run 2	-0.013	0.009		0.165

4.2.2. Road attention ratio (RAR)

The model estimation results for differences in RAR over runs are presented in Table 5. A likelihood-ratio test ($\chi^2(1) = 123.18$, $p < 0.001$) illustrates a significantly better goodness-of-fit of the LMM ($N = 219$, groups = 79) compared to its corresponding null model. LMM results show significant differences in RAR for different runs ($F(2, 141.5) = 24.111$, $p < 0.001$). A multiple comparison Tukey's test indicates that RAR is significantly higher in Run 1 compared to Run 2 and Run 3. No significant differences in RAR were found between Run 2 and Run 3. Similar to OGR, this suggests that drivers monitored the road for longer duration, on average, in the first run, but were quick to adapt their road monitoring behavior to maintain their desired levels of SA.

Table 5 Model results for road attention ratio (RAR)

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	<i>N</i> (runs)	219			
	Groups (participants)	79			
Fixed Effects	Intercept	0.113	0.011	113.8	< 0.001
	Run 1	0.054	0.008	141.1	< 0.001
	Run 2	0.012	0.008	141.2	0.140
Random Effects	Group variance	0.007			
	LRT (χ^2)	123.18		1	< 0.001
ANOVA	Run (<i>F</i>)	24.111		(2, 141.5)	< 0.001
Multiple Comparison	Run 2 – Run 1	-0.042	0.008		< 0.001
	Run 3 – Run 1	-0.054	0.008		< 0.001
Tukey's test	Run 3 – Run 2	-0.012	0.008		0.138

4.2.3. Pre-warning normalized heart rate (pre-NHR)

The model estimation results for differences in pre-NHR over runs are presented in Table 6. A likelihood-ratio test ($\chi^2(1) = 76.09$, $p < 0.001$) illustrates a significantly better goodness-of-fit of the LMM ($N = 250$, groups = 103) compared to its corresponding null model. Model results show marginally significant differences in pre-NHR for different runs ($F(2, 155.1) = 2.928$, $p = 0.056$). A multiple comparison Tukey's test indicates that pre-NHR is marginally higher in Run 1 compared to Run 2. These results suggest that drivers experienced marginally higher mental stress before the takeover warning in the first run compared to the second run. This is consistent with other driving simulator studies that have reported a decrease in HR due to driving stress as the experiment progresses (time-on-task effects) (de Winter et al., 2014). It happens because the participants become more comfortable with the experiment setup over time. However, no significant differences in the pre-NHR were observed between Run 3 and other runs. This could be due to the presence of two counteracting factors: (i) expected reduction in stress with an increase in time-on-task, and (ii) an increase in the level of stress after experiencing a mandatory takeover situation in the second run (fixed order of runs).

Table 6 Model results for pre-warning normalized heart rate (pre-NHR)

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	N (runs)	250			
	Groups (participants)	103			
Fixed Effects	Intercept	0.453	0.015	175.5	< 0.001
	Run 1	0.023	0.014	156.2	0.099
	Run 2	-0.010	0.014	153.7	0.472
Random Effects	Group variance	0.012			
	LRT (χ^2)	76.09		1	< 0.001
ANOVA	Run (F)	2.928		(2, 155.1)	0.056
Multiple	Run 2 – Run 1	-0.033	0.014		0.053
Comparison	Run 3 – Run 1	-0.023	0.014		0.194
Tukey's test	Run 3 – Run 2	0.010	0.014		0.471

4.2.4. Pre-warning heart rate variability (pre-HRV)

The model estimation results for differences in pre-HRV over runs are presented in Table 7. A likelihood-ratio test ($\chi^2(1) = 37.78$, $p < 0.001$) illustrates a significantly better goodness-of-fit of the LMM ($N = 250$, groups = 103) compared to its corresponding null model. ANOVA did not show any statistically significant differences in pre-HRV across runs ($F(2, 145.8) = 1.677$, $p = 0.191$). Hence, a post-hoc multiple comparison Tukey's test is not performed. This indicates that the pre-HRV did not significantly differ across three runs.

Table 7 Model results for pre-warning heart rate variability (pre-HRV)

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	N (runs)	250			
	Groups (participants)	103			
Fixed Effects	Intercept	3.289	0.184	186.3	< 0.001
	Run 1	0.357	0.196	147.3	0.070
	Run 2	0.193	0.191	143.97	0.315

Random Effects	Group variance	1.580		
	LRT (χ^2)	37.78	1	< 0.001
ANOVA	Run (F)	1.677	(2, 145.8)	0.191

4.2.5. Change in normalized heart rate (ΔNHR)

The model estimation results for differences in ΔNHR over runs are presented in Table 8. A likelihood-ratio test ($\chi^2(1) = 42.57, p < 0.001$) illustrates a significantly better goodness-of-fit of the LMM ($N = 227$, groups = 100) compared to its corresponding null model. ANOVA did not show any statistically significant differences in ΔNHR across runs ($F(2, 147.0) = 1.147, p = 0.320$); thus, a post-hoc multiple comparison Tukey's test is not performed. The results indicate that ΔNHR during the takeover event did not significantly differ for drivers between the three runs. This suggests that drivers' prior experiences with takeover events may not have a significant impact on the increase in stress levels, which could have negative implications for road safety and drivers' trust in the ADS. However, these results may be valid only in the short-term as we only used three experiment runs, and therefore, further research is needed to examine the long-term effects.

Table 8 Model results for change in normalized heart rate (ΔNHR)

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	N (runs)	227			
	Groups (participants)	100			
Fixed Effects	Intercept	0.035	0.015	189.6	0.019
	Run 1	0.024	0.017	146.9	0.157
	Run 2	0.003	0.016	146.2	0.837
Random Effects	Group variance	0.009			
	LRT (χ^2)	42.57		1	< 0.001
ANOVA	Run (F)	1.147		(2, 147.0)	0.320

4.2.6. Change in heart rate variability (ΔHRV)

The model estimation results for differences in ΔHRV over runs are presented in Table 9. A likelihood-ratio test ($\chi^2(1) = 5.94, p = 0.015$) suggests that the LMM ($N = 227$, groups = 100) has a significantly better goodness-of-fit compared to its corresponding null model. We did not observe any statistically significant differences in ΔHRV across runs ($F(2, 146.6) = 0.940, p = 0.393$). Hence, a post-hoc multiple comparison Tukey's test is not performed. Similar to the ΔNHR results, these results indicate statistically non-significant differences in the increase in mental stress during the takeover event across all three runs.

Table 9 Model results for change in heart rate variability (ΔHRV)

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	N (runs)	227			
	Groups (participants)	100			
Fixed Effects	Intercept	0.541	0.217	218.5	0.013
	Run 1	-0.213	0.294	147.1	0.469
	Run 2	-0.384	0.281	144.1	0.174
Random Effects	Group variance	0.806			
	LRT (χ^2)	5.94		1	0.015

ANOVA	Run (<i>F</i>)	0.940	(2, 146.6)	0.393
--------------	-----------------------	--------------	-------------------	--------------

4.2.7. Correlation analysis

The estimated repeated measures correlations between the physiological indicators are presented in Table 10. We found a strong positive correlation between OGR and RAR, which is expected as both of them are indicators of driver's SA from monitoring of the road. We also found a strong negative correlation between pre-NHR and Δ NHR and between pre-HRV and Δ HRV, which can be attributed to their functional relationship. We found that RAR has a significant positive correlation with pre-HRV, indicating that drivers with higher SA during automated driving may have lower pre-warning mental stress. However, no statistically significant correlations were found between eye-tracking measures (OGR and RAR) and NHR measures (pre-NHR and Δ NHR). Overall, these results suggest that driver's SA during automated driving may or may not affect their mental stress significantly in conditionally automated vehicles. Moreover, although both NHR and HRV measures are associated with mental stress, they did not have a strong correlation.

Table 10 Repeated measures correlation matrix for physiological indicators

Indicators	OGR	RAR	pre-NHR	pre-HRV	Δ NHR	Δ HRV
OGR	-					
RAR	0.498 (139, < 0.001)	-				
pre-NHR	0.023 (84, 0.831)	0.079 (84, 0.468)	-			
pre-HRV	0.143 (84, 0.188)	0.249 (84, 0.021)	0.141 (146, 0.088)	-		
ΔNHR	-0.125 (72, 0.288)	-0.010 (72, 0.936)	-0.487 (126, < 0.001)	-0.107 (126, 0.227)	-	
ΔHRV	-0.214 (72, 0.067)	-0.118 (72, 0.317)	-0.063 (126, 0.478)	-0.725 (126, <0.001)	0.151 (126, 0.089)	-

Note: Degrees of freedom and p-value are shown in parentheses

4.3. Takeover performance

The effect of Δ NHR ($F(1, 160.8) = 5.964$, $p = 0.016$) on the TOPI was found to be statistically significant. However, a likelihood-ratio test ($\chi^2(1) = 1.659$, $p = 0.198$) shows that the LMM ($N = 193$, groups = 99) did not have a better goodness-of-fit compared to its corresponding null model. Thus, takeover performance is evaluated using a linear model with no random effects. A visual inspection of the QQ plot suggests that the residuals are normally distributed. The model results of the estimated linear model are presented in Table 11. An adjusted R-squared value of 0.281 indicates a good model fit. Model results show significant negative effects of the novelty in takeover experience ($F(1, 189) = 70.320$, $p < 0.001$) and the presence of TOR ($F(1, 189) = 21.950$, $p < 0.001$) on the TOPI. These results are consistent with the findings of our previous work that discussed their effects in detail (Agrawal and Peeta, 2020). The results also indicate a significant negative effect of Δ NHR on the TOPI ($F(1, 189) = 3.946$, $p = 0.048$). This suggests that the change in driver's heart rate after the takeover warning significantly affects their takeover performance. Previous studies have linked an increase in heart rate with the surprise emotion in unexpected events (Jang et al., 2015) and with an increase in task-induced mental stress (Alrefaie et al., 2019). In our context, the unexpected event is the takeover situation, and the task is taking over vehicle control. Past studies have also shown that the effects of mental stress on decision-making and performance are task-specific (Starcke and Brand, 2012). This suggests that the increase in stress (i.e., higher Δ NHR)

during a takeover event may have led to poor instantaneous decision-making by drivers resulting in poor takeover performance. However, it should be noted that the increase in stress after the takeover warning may have a bi-directional relationship with takeover performance, as inadequate or improper driving responses during a rapidly evolving takeover situation could cause more stress to the driver. Note that the dependent variable is the square of the TOPI; hence, the estimated results represent a quadratic relationship between the independent variables and the TOPI instead of a linear relationship. This may affect the size of the (marginal) effect size, but since the TOPI value lies between 0 and 1, the sign of the effect remains unaffected.

The effects of OGR ($F(1, 122.9) = 0.052, p = 0.820$), RAR ($F(1, 122.1) = 0.354, p = 0.553$), pre-NHR ($F(1, 179.1) = 0.172, p = 0.679$), and pre-HRV ($F(1, 200.6) = 0.003, p = 0.954$) on the TOPI were not found to be statistically significant. These results suggest that the driver's SA during automated driving and their level of stress before the takeover warning do not have a significant impact on takeover performance. This is interesting as previous studies have suggested that increasing driver's SA could improve their takeover performance (Olaverri-Monreal et al., 2018; Petersen et al., 2019). There are two possible explanations for this. First, those studies do not use a comprehensive takeover performance metric and, thereby, observed improvements in only some aspects of the takeover performance (e.g., reaction times) but not the overall takeover performance. Second, our study participants maintained a sufficient level of SA, and hence it did not significantly affect their takeover performance. Unlike Δ NHR, the effects of Δ HRV ($F(1, 191.2) = 0.936, p = 0.335$) on the TOPI were found to be statistically non-significant.

Table 11 Model results for the effects of Δ NHR on the TOPI

	Parameter	Estimate/Value	Std. Error	DF	Sig.
Data	<i>N</i> (runs)	193			
Fixed Effects	Intercept	0.393	0.016		< 0.001
	Novelty	-0.214	0.025		< 0.001
	TOR	-0.119	0.025		< 0.001
	Δ NHR	-0.150	0.075		0.048
Model	Model fit (<i>F</i>)	26.030		(3, 189)	< 0.001
	Adjusted R-squared	0.281			

Novelty: Novelty in takeover experience
TOR: Presence of TOR

5. Discussion

The study results show a significantly higher SA in the first run compared to the later runs, measured by eye-tracking indicators (i.e., OGR and RAR), as illustrated in Table 4 and Table 5. However, it was not found to be significantly different among the later runs. This suggests that drivers initially expended more visual resources to scan the road environment, most likely due to the novelty in the automated driving experience, but they were quick to adjust their road monitoring behavior to maintain a level of SA during automated driving that they deemed to be sufficient (i.e., lower level of SA in runs 2 and 3 compared to run 1). Note that the two practice runs were designed to acclimatize drivers with the driving simulator (first practice run) and make them feel comfortable interacting with the ADS and NDRT (second practice run). However, the practice runs had oversimplified traffic conditions, and hence did not provide drivers with a realistic automated driving experience. This enables us to assess driver state during automated driving in realistic traffic environments over time, including their first experience (i.e., Run 1). A key advantage of

1 driver monitoring systems is that they can track driver state in real-time, which the vehicle can use to alert
2 the driver or take certain maneuvering actions.

3 Previous studies have reported that the driver's HR reduces and HRV increases over time in partially
4 automated vehicles (Heikoo et al., 2018, 2017). These trends in HR measures are attributed to a decrease
5 in mental stress, and an increase in passive fatigue and mental underload during automated driving.
6 However, our results show only a marginal decrease in pre-NHR across different runs, as illustrated in
7 Table 6. This suggests that the HR and, thereby, mental stress may not reduce significantly over time under
8 conditional automation. However, since we did not compare between driving in partially automated
9 vehicles and conditionally automated vehicles, it is possible that the heart rate is initially lower (i.e., lower
10 mental stress) under conditional automation. In addition, we did not find significant differences in Δ NHR
11 and Δ HRV between runs (see Table 8 and Table 9). This suggests that the time spent in automated driving,
12 as well as the number of takeover experiences, may not affect the change in the driver's mental stress during
13 an unexpected takeover situation. Further, the lack of significant correlations between NHR and HRV
14 measures (see Table 10) illustrates the challenges of estimating driver state accurately using such
15 physiological indicators. This suggests the need to consider multiple sources of physiological data in in-
16 vehicle driver monitoring systems to better estimate the driver state under conditional automation.

17 The lack of correlations between eye-tracking measures and post-warning change in HR measures
18 (Δ NHR and Δ HRV) indicates that the level of SA during automated driving has a non-significant impact
19 on the change in mental stress after the takeover warning. However, the significant and non-significant
20 correlations of RAR with pre-NHR and pre-HRV, respectively, do not provide conclusive evidence of
21 correlation between SA during automated driving and pre-warning mental stress. The characteristics of
22 NDRT in this study (requires visual attention but low mental effort) may also have affected the relationship
23 between them. These insights suggest that driver monitoring system designers should consider using
24 different sources of physiological data to assess the driver state as a single source may not be sufficient to
25 estimate different aspects of the driver state.

26 The results also illustrate a significant negative effect of Δ NHR on the TOPI. This suggests that
27 unexpected takeover events that require immediate safety-critical responses from drivers increase their
28 mental stress, which could further lead to poor instantaneous decision-making while taking control of the
29 vehicle from the ADS. In addition, a bi-directional relationship is possible whereby improper driving
30 responses lead to a more dangerous situation that increases driver's mental stress. Future studies can
31 investigate this potential bi-directional relationship between the increase in stress and takeover
32 performance. Further, as discussed earlier, we did not find any significant differences in Δ NHR over time
33 (successive experiment runs). This indicates that prior experiences with takeover situations may not affect
34 the change in driver's mental stress during an unexpected takeover situation over time, suggesting that
35 learning from prior takeover experiences may be overshadowed by the surprise factor with the unexpected
36 event. Previous studies have reported that absence of crashes during automated driving help increase user's
37 trust in the system (Gold et al., 2015), and that higher trust promotes the adoption of automated vehicles
38 (Choi and Ji, 2015). In this context, the study insights can aid policymakers to develop operator training
39 programs that focus on mental stress management during takeover situations so as to enhance driver's trust
40 in conditionally automated vehicles and promote their adoption. They also suggest that it may be useful for
41 auto manufacturers to monitor the changes in driver's mental stress during takeover situations through
42 driver monitoring systems, and adjust the ADS and takeover warning system parameters to reduce mental
43 stress increase to enhance user experience and road safety.

44 Most drivers intervened directly after an uncertainty alert (i.e., before a TOR) in this study. This
45 takeover behavior could result from a combination of the tendency to intervene when provided with a

sufficient time budget (Epple et al., 2018) or with a lower level of trust in the ADS (Körber et al., 2018). In line with our previous work (Agrawal and Peeta, 2020), the results illustrate significant negative impacts of novelty in takeover experience and the presence of TOR on takeover performance. Note that since the experiment runs were not randomized, the impacts of novelty in takeover experience on the TOPI could also be due to scenario-specific factors (i.e., time budget or the difficulty of the takeover situation). Nevertheless, these findings have key implications for policymakers to develop operator training and driver licensing strategies as well as for auto manufacturers to design fallback procedures, as discussed in our previous work. We did not find significant impacts of driver's SA during the automated driving (OGR and RAR) and pre-warning mental stress (pre-NHR and pre-HRV) on their takeover performance. This suggests that the driver's SA did not significantly impact their takeover performance under conditional automation in this study. However, we did not manipulate SA in this study, and hence the lack of significant effects of SA may be because the drivers maintained a sufficient level of SA that is needed in conditionally automated vehicles. The lack of significant effects of pre-warning mental stress on the TOPI is consistent with the findings in our previous work that analyzed the effects of driver's pre-warning cognitive state on the TOPI using EEG indicators (Agrawal and Peeta, 2020). However, in our previous study, we did find the effects of certain pre-warning cognitive states like fatigue, drowsiness, passive attention, and low alertness on the TOPI. Although eye-tracking and heart rate measures do not provide a very detailed profile of the driver state that EEG-based measures would, unlike EEG they can be measured using in-vehicle driver monitoring systems and have real-world applications.

6. Concluding Comments

This study analyzes the differences in driver state (in terms of SA and mental stress) over time and their impacts on takeover performance using driving simulator experiments. Driver's SA and mental stress are estimated using physiological indicators that can be measured by in-vehicle driver monitoring systems in the real world. Estimating driver state using eye-tracking and heart rate measures have certain limitations. For example, certain eye-tracking indicators are correlated with SA, but these measures only capture the perception element of SA and not the comprehension and projection of future state elements, as described by Endsley (1988). However, although these indicators do not provide a detailed profile of driver state, they can be used by auto manufacturers to adjust the ADS and takeover warning system parameters based on the driver state to enhance road safety and user experience.

The study results show that drivers quickly adjust their road monitoring behavior and consequently their SA after the first automated driving experience. They also show that drivers experience a similar increase in mental stress during a takeover event over time, which becomes a critical issue from the road safety perspective as a higher increase in mental stress results in reduced takeover performance. Thus, the results illustrate the importance of considering driver state characteristics such as mental stress in conditionally automated vehicles to evaluate driver's takeover performance and take appropriate actions in real-time to enhance road safety by leveraging in-vehicle driver monitoring systems. The study findings provide valuable insights to auto manufacturers and policymakers to design integrated in-vehicle driver monitoring and warning systems, operator training programs, and driver licensing strategies for conditionally automated vehicles, as discussed in the previous section.

The study has the following limitations that can be addressed in future research. First, the study sample is skewed towards younger adults (less than 35 years of age). Further experimentation with a more representative sample of the general population can improve the generalizability of the results. Second, there was a significant data loss due to invalid physiological data, which could also affect the generalizability of the data and results. Future efforts can focus on designing data collection protocols more

carefully. Third, the order effects of experiment runs were not considered in the analysis. Future studies can randomize the experiment runs to control these studies, which may provide better insights regarding the impacts of time-on-task and type of takeover situation on the driver state and takeover performance. Fourth, the characteristics of the takeover situation (e.g., time budget) affects takeover performance in automated vehicles (Gold et al., 2013). Thus, the study findings are, to some extent, contingent on the tested takeover scenarios. The generalizability of the study findings can be increased by collecting more data under different takeover situations.

Some potential future directions are as follows. First, other data that can be collected using in-vehicle driver monitoring systems such as skin conductance and facial expressions can be used to augment the eye-tracking and heart rate measures for evaluating takeover performance. Second, future research can identify optimal size of time windows for computing HR variables to gauge their impacts on takeover performance. Third, the effects of NDRT on physiological indicators, especially eye-tracking measures, and their impacts on takeover performance can be explored. Fourth, future studies can focus on developing integrated in-vehicle driver monitoring and takeover warning systems for conditionally automated vehicles that use the real-time estimation of driver state to predict driver's response in a takeover situation and provide timely and appropriate warnings to prepare them for the approaching road safety hazards.

Acknowledgments

The authors would like to thank Dustin Souders for his efforts in designing the driving simulator experiments and Irina Benedyk for her help with running the experiments. The authors would also like to thank the students from the NEXTRANS Center, Purdue University for their assistance in conducting the experiments. This research was funded through the Center for Connected and Automated Transportation University Transportation Center (USDOT grant no. 69A3551747105).

References

- Advanced Brain Monitoring, 2017. B-Alert X24 - Advanced Brain Monitoring [WWW Document]. URL <http://www.advancedbrainmonitoring.com/xseries/x24/> (accessed 2.21.17).
- Agrawal, S., Peeta, S., 2020. Evaluating the impacts of driver's pre-warning cognitive state on takeover performance under conditional automation. under Rev.
- Alrefaie, M.T., Summerskill, S., Jackson, T.W., 2019. In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles. *Accid. Anal. Prev.* 131, 180–190. <https://doi.org/10.1016/j.aap.2019.06.011>
- AVSimulation, 2020. AVSimulation [WWW Document]. URL <https://www.avsimulation.com/> (accessed 1.10.20).
- B-Alert, 2009. B-Alert User Manual.
- Bakdash, J.Z., Marusich, L.R., 2017. Repeated measures correlation. *Front. Psychol.* 8, 1–13. <https://doi.org/10.3389/fpsyg.2017.00456>
- Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bollen, K.A., 1989. *Structural Equations with Latent Variables*. John Wiley & Sons, Inc., Hoboken, NJ, USA. <https://doi.org/10.1002/9781118619179>

- Choi, J.K., Ji, Y.G., 2015. Investigating the Importance of Trust on Adopting an Autonomous Vehicle. *Int. J. Hum. Comput. Interact.* 31, 692–702. <https://doi.org/10.1080/10447318.2015.1070549>
- Choi, M., Koo, G., Seo, M., Kim, S.W., 2018. Wearable device-based system to monitor a driver's stress, fatigue, and drowsiness. *IEEE Trans. Instrum. Meas.* 67, 634–645. <https://doi.org/10.1109/TIM.2017.2779329>
- de Winter, J.C.F., Happee, R., Martens, M.H., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transp. Res. Part F Traffic Psychol. Behav.* 27, 196–217. <https://doi.org/10.1016/j.trf.2014.06.016>
- Dogan, E., Honnêt, V., Masfrand, S., Guillaume, A., 2019. Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* 62, 494–504. <https://doi.org/10.1016/j.trf.2019.02.010>
- Endsley, M.R., 1988. Design and evaluation for situation awareness enhancement. *Proc. Hum. Factors Soc. Annu. Meet.* 32, 97–101. <https://doi.org/10.1177/154193128803200221>
- Epple, S., Roche, F., Brandenburg, S., 2018. The sooner the better: Drivers' reactions to two-step take-over requests in highly automated driving, in: *Proceedings of the Human Factors and Ergonomics Society*. pp. 1883–1887. <https://doi.org/10.1177/1541931218621428>
- Ergoneers, 2018. Dikablis Glasses 3 [WWW Document]. URL <https://www.ergoneers.com/en/hardware/dikablis-glasses/> (accessed 5.3.20).
- Franz, B., Haccius, J.P., Stelzig-krombholz, D., Pfromm, M., Kauer, M., 2015. Evaluation of the SAGAT method for highly automated driving, *Proceedings 19th Triennial Congress of the IEA*.
- Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. Take over! How long does it take to get the driver back into the loop?, in: *Proceedings of the Human Factors and Ergonomics Society*. pp. 1938–1942. <https://doi.org/10.1177/1541931213571433>
- Gold, C., Körber, M., Hohenberger, C., Lechner, D., Bengler, K., 2015. Trust in Automation – Before and After the Experience of Take-over Scenarios in a Highly Automated Vehicle. *Procedia Manuf.* 3, 3025–3032. <https://doi.org/10.1016/j.promfg.2015.07.847>
- Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated vehicles in complex traffic situations. *Hum. Factors* 58, 642–652. <https://doi.org/10.1177/0018720816634226>
- Heikoop, D.D., de Winter, J.C.F., van Arem, B., Stanton, N.A., 2018. Effects of mental demands on situation awareness during platooning: A driving simulator study. *Transp. Res. Part F Traffic Psychol. Behav.* 58, 193–209. <https://doi.org/10.1016/j.trf.2018.04.015>
- Heikoop, D.D., de Winter, J.C.F., van Arem, B., Stanton, N.A., 2017. Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study. *Appl. Ergon.* 60, 116–127. <https://doi.org/10.1016/j.apergo.2016.10.016>
- Jang, E.-H., Park, B.-J., Park, M.-S., Kim, S.-H., Sohn, J.-H., 2015. Analysis of physiological signals for recognition of boredom, pain, and surprise emotions. *J. Physiol. Anthropol.* 34, 1–12. <https://doi.org/10.1186/s40101-015-0063-5>

- 1 Kim, J., Kim, H.S., Kim, W., Yoon, D., 2018. Take-over performance analysis depending on the drivers'
2 non-driving secondary tasks in automated vehicles. 9th Int. Conf. Inf. Commun. Technol. Conver.
3 ICT Conver. Powered by Smart Intell. ICTC 2018 1364–1366.
4 <https://doi.org/10.1109/ICTC.2018.8539431>
- 5 Körber, M., Baseler, E., Bengler, K., 2018. Introduction matters: Manipulating trust in automation and
6 reliance in automated driving. Appl. Ergon. 66, 18–31. <https://doi.org/10.1016/j.apergo.2017.07.006>
- 7 Körber, M., Cingel, A., Zimmermann, M., Bengler, K., 2015. Vigilance decrement and passive fatigue
8 caused by monotony in automated driving, in: 6th International Conference on Applied Human Factors
9 and Ergonomics. pp. 2403–2409. <https://doi.org/10.1016/j.promfg.2015.07.499>
- 10 Kreuzmair, C., Gold, C., Meyer, M.L., 2017. The influence of driver fatigue on take-over performance in
11 highly automated vehicles, in: 25th International Technical Conference on the Enhanced Safety of
12 Vehicles (ESV) National Highway Traffic Safety Administration. pp. 1–7.
- 13 Large, D.R., Burnett, G., Morris, A., Muthumani, A., Matthias, R., 2018. A longitudinal simulator study to
14 explore drivers' behaviour during highly-automated driving, in: Advances in Intelligent Systems and
15 Computing. pp. 583–594. https://doi.org/10.1007/978-3-319-60441-1_57
- 16 Louw, T., Merat, N., 2017. Are you in the loop? Using gaze dispersion to understand driver visual attention
17 during vehicle automation. Transp. Res. Part C Emerg. Technol. 76, 35–50.
18 <https://doi.org/10.1016/j.trc.2017.01.001>
- 19 Lu, Z., Zhang, B., Feldhütter, A., Happee, R., Martens, M., De Winter, J.C.F., 2019. Beyond mere take-
20 over requests: The effects of monitoring requests on driver attention, take-over performance, and
21 acceptance. Transp. Res. Part F 63, 22–37. <https://doi.org/10.1016/j.trf.2019.03.018>
- 22 Merat, N., Jamson, A.H., Lai, F.C.H., Carsten, O., 2012. Highly automated driving, secondary task
23 performance, and driver state. Hum. Factors 54, 762–771. <https://doi.org/10.1177/0018720812442087>
- 24 Moore, K., Gugerty, L., 2010. Development of a Novel Measure of Situation Awareness: The Case for Eye
25 Movement Analysis, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
26 Sage CA: Los Angeles, CA: SAGE Publications, pp. 1650–1654.
27 <https://doi.org/10.1177/154193121005401961>
- 28 Nguyen, T., Lim, C.P., Nguyen, N.D., Gordon-Brown, L., Nahavandi, S., 2019. A review of situation
29 awareness assessment approaches in aviation environments. IEEE Syst. J. 13, 3590–3603.
30 <https://doi.org/10.1109/JSYST.2019.2918283>
- 31 Olaverri-Monreal, C., Kumar, S., & Díaz-Álvarez, A., 2018. Automated driving: Interactive automation
32 control system to enhance situational awareness in conditional automation. *IEEE Intelligent Vehicles*
33 *Symposium, Proceedings, 2018-June*, 1698–1703. <https://doi.org/10.1109/IVS.2018.8500367>
- 34 Petersen, L., Robert, L., Yang, X.J., Tilbury, D., 2019. Situational awareness, driver's trust in automated
35 driving systems and secondary task performance. SAE Int. J. Connect. Autom. Veh. 2, 1–26.
36 <https://doi.org/10.4271/12-02-02-0009>
- 37 R Core Team, 2020. R: A language and environment for statistical computing [WWW Document]. URL
38 <https://www.r-project.org/> (accessed 6.13.20).

- 1 Radlmayr, J., Briich, K., Schmidt, K., Solbeck, C., Wehner, T., 2018a. Peripheral monitoring of traffic in
2 conditionally automated driving. *Proc. Hum. Factors Ergon. Soc.* 3, 1828–1832.
3 <https://doi.org/10.1177/1541931218621416>
- 4 Radlmayr, J., Gold, C., Lorenz, L., Farid, M., Bengler, K., 2014. How traffic situations and non-driving
5 related tasks affect the take-over quality in highly automated driving. *Proc. Hum. Factors Ergon. Soc.*
6 2014-Janua, 2063–2067. <https://doi.org/10.1177/1541931214581434>
- 7 Radlmayr, J., Ratter, M., Feldhütter, A., Körber, M., Prasch, L., Schmidler, J., Yang, Y., Bengler, K.,
8 2018b. Take-overs in level 3 automated driving - Proposal of the take-over performance score (TOPS),
9 in: *Congress of the International Ergonomics Association*. pp. 436–446. [https://doi.org/10.1007/978-3-](https://doi.org/10.1007/978-3-319-96074-6_46)
10 319-96074-6_46
- 11 SAE J3016, 2018. Surface vehicle recommended practice. Taxonomy and definitions for terms related to
12 driving automation systems for on-road motor vehicles, Society of Automotive Engineers.
- 13 Solís-Marcos, I., Kircher, K., 2018. Event-related potentials as indices of mental workload while using an
14 in-vehicle information system. *Cogn. Technol. Work* 1–13. [https://doi.org/10.1007/s10111-018-0485-](https://doi.org/10.1007/s10111-018-0485-z)
15 z
- 16 Souders, D., Agrawal, S., Benedyk, I., Guo, Y., Li, Y., Peeta, S., 2020. Highlighting the driver's
17 responsibilities when using conditional driving automation: Effects on attitudes, monitoring, and take-
18 over Performance. under Rev.
- 19 Spector, P.E., 1994. Using self-report questionnaires in OB research: A comment on the use of a
20 controversial method. *J. Organ. Behav.* 15, 385–392. <https://doi.org/10.1002/job.4030150503>
- 21 Starcke, K., Brand, M., 2012. Decision making under stress: A selective review. *Neurosci. Biobehav. Rev.*
22 36, 1228–1248. <https://doi.org/10.1016/j.neubiorev.2012.02.003>
- 23 Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2008). Influence of mental stress on heart rate
24 and heart rate variability. *IFMBE Proceedings*, 22, 1366–1369. [https://doi.org/10.1007/978-3-540-](https://doi.org/10.1007/978-3-540-89208-3_324)
25 89208-3_324
- 26 Wechsler, D., 1997. Wechsler memory scale - Third edition administration and scoring manual. San
27 Antonio, TX Psychol. Corp.
- 28 Wiedemann, K., Naujoks, F., Wörle, J., Kenntner-Mabiala, R., Kaussner, Y., Neukum, A., 2018. Effect of
29 different alcohol levels on take-over performance in conditionally automated driving. *Accid. Anal.*
30 *Prev.* 115, 89–97. <https://doi.org/10.1016/j.aap.2018.03.001>
- 31 Yoon, S.H., Ji, Y.G., 2019. Non-driving-related tasks, workload, and takeover performance in highly
32 automated driving contexts. *Transp. Res. Part F Traffic Psychol. Behav.* 60, 620–631.
33 <https://doi.org/10.1016/j.trf.2018.11.015>
- 34 Young, M., Stanton, N., 2002. Malleable Attentional Resources Theory: A new explanation for the effects
35 of mental underload on performance. *Hum. Factors* 44, 365–375.
- 36 Zeeb, K., Buchner, A., Schrauf, M., 2015. What determines the take-over time? An integrated model
37 approach of driver take-over after automated driving. *Accid. Anal. Prev.* 78, 212–221.
38 <https://doi.org/10.1016/j.aap.2015.02.023>

- 1 Zhang, B., de Winter, J., Varotto, S., Happee, R., Martens, M., 2019. Determinants of take-over time from
- 2 automated driving: A meta-analysis of 129 studies. *Transp. Res. Part F Traffic Psychol. Behav.* 64,
- 3 285–307. <https://doi.org/10.1016/j.trf.2019.04.020>