Physiological Measures of Risk Perception in Highly Automated Driving

Jaume R. Perello-March, Christopher G. Burns, Stewart A. Birrell, Roger Woodman and Mark T. Elliott

***Abstract*— Highly automated driving will likely result in drivers being out-of-the-loop during specific scenarios and engaging in a wide range of non-driving related tasks. Manifesting in lower levels of risk perception to emerging events, and thus affect drivers' availability to take-over manual control in safety-critical scenarios. In this empirical research, we measured drivers' (N = 20) risk perception with cardiac and skin conductance indicators through a series of high-fidelity, simulated highly automated driving scenarios. By manipulating the presence of surrounding traffic and changing driving conditions as long-term risk modulators, and including a driving hazard event as a short-term risk modulator, we hypothesised that an increase in risk perception would induce greater physiological arousal. Our results demonstrate that heart rate variability features are superior at capturing arousal variations from these long-term, low to moderate risk scenarios. In contrast, skin conductance responses are more sensitive to rapidly evolving situations associated with moderate to high risk. Based on this research, future driver state monitoring systems should adopt multiple physiological measures to capture changes in the long and short term, modulation of risk perception. This will enable enhanced perception of driver readiness and improved availability to safely deal with take-over events when requested by an automated vehicle.**

***Index Terms*— Driver state monitoring, highly automated driving, monitoring request, take-over request, risk perception**

I. Introduction

During Highly Automated Driving (HAD) (i.e. SAE Levels 3 and 4) [1], drivers will not be required to monitor or engage in the driving task during predefined use cases. When reaching the limits of these use cases, take-over requests (TORs) can be issued so drivers can take over manual control. Drivers will likely be engaged in non-driving related tasks (NDRTs) [2], [3] or even sleeping [4] while HAD is activated; and thus, would possibly be out-of-the-loop (OOTL) –i.e. not in physical control of the vehicle, and not monitoring the driving situation [5]. Technologies for risk mitigation and support for impaired drivers will therefore be essential in such scenarios to guarantee successful and safe take-overs.

Current driver state monitoring (DSM) systems, mostly built on eye-tracking parameters, have proven to be an effective and unobtrusive solution to detect hazardous driving behaviours and states, such as fatigue, inattention, distractions, or drowsiness [6]–[9]. Existing DSM research provides a robust and reliable ground from which the next generation of multimodal DSM systems and cooperative intelligent transportation systems for enhancing driving automation safety functions will be developed. As with current DSM systems, future systems must compensate for inappropriate engagement or human errors, enhance driving safety and comfort.

A new range of in-vehicle possibilities and NDRTs will likely become available during HAD. Disengagement from the driving task will also entail other human factors issues not present during manual driving, such as OOTL states or overtrust in automation. Thus, one of the essential functions of future DSM systems will be detecting drivers' availability to take over and drive manually [10]. Drivers' arousal combined with gaze behaviour can provide real-time indicators of several psychophysiological states relevant for driving readiness, such as attentional capability, wakefulness, sleepiness, mental workload or stress [6], [9], [11].

Arousal -or alertness- is a physiological reaction to internal or external stimuli which the brain perceives as potentially hazardous. High levels of arousal increase stress to prepare the body to "fight-or-flight". This process is essentially controlled by the autonomic nervous system and leads to several measurable changes in the body, including an increase in heart rate, breathing rate, pupil dilation, muscle contraction or sweating [12]. Up to a certain level, arousal increase is associated with better performance [13], [14], but beyond that, arousal increases result in performance decrement [15]. This inverted U-curve trend is often referred to as the *Yerkes-Dodson law* [16], although this association has been criticised [17]. Certain NDRTs may increase arousal and mental workload levels, decreasing performance [18]–[21]. On the contrary, other tasks may induce under-arousal levels and hamper performance [14], [22]. Classifying among several arousal states will become crucial to know if the driver is currently capable of manual driving, and for this purpose, algorithms classifying and estimating drivers' states should be developed. Consequently, the next generation of DSM systems will likely be required to perform multimodal, unobtrusive, and real-time arousal assessments to deploy adequate warning and handover strategies.

II. Background

The higher the automation level, the greater the chance drivers engage in NDRTs [2], [3]. Hence, we could expect that during HAD, drivers will most probably be involved in several NDRTs; and their situation awareness for the driving task will be diminished [23], [24]. Especially after a long engagement in alternative activities, it can be challenging to take over manual control. For example, for drivers waking from a nap, the transition from sleep to wakefulness is characterised by “hypovigilance, confusion, disorientation of behaviour, and impaired cognitive and sensory-motor performance” [25], and drivers in such a state would likely be impaired for taking over manual control. Another case could be those drivers that have been engaged in a mentally demanding task (e.g. playing video games, on a phone call or a videoconference). In such a case, mental overload and high-stress levels could impair their take-over performance [26]. This applies to the previously mentioned *Yerkes-Dodson law* relating arousal levels and performance. The concept of driver availability (or readiness) refers to the time required for the transition from the out-of-the-loop state to the in-the-loop state [10], [27]. Arousal levels will be critical to informing the DSM system of the current OOTL state, which will depend on the nature of the ongoing NDRT and each individual. In addition, arousal levels will also inform the optimal in-the-loop -target- state required to perform a safe and successful take over, which will depend on the task complexity, mental resources required, and the available time budget [10]. Before taking over, the system needs to ensure the driver is ready. Here is where multimodal DSM systems, relying on psychophysiological measures, could determine driver readiness.

Several authors have stressed the importance of preparing drivers for taking over, to guarantee an adequate situation awareness before the actual TOR is presented [10], [28], [29]. Indeed, [29] compared drivers’ take-over performance when adding a “monitoring request” seven seconds before a TOR in a simulated scenario involving pedestrians crossing the road. Their results indicate that compared to the TOR-only condition, participants responded to the monitoring request successfully and showed better take-over performance, with shorter response time and longer minimum time to collision. The monitoring request could be directed to bring the driver back on the loop [5] and for the DSM to start monitoring the driver until achieving an appropriate attentional level before the TOR. As proposed in a recent framework [10], monitoring the entire transition process from preparation prior to the take-over to the stage after resuming manual control, could assist in ensuring a safe and successful transition performance. In this current research, we have focused on the first stage, where drivers’ psychophysiological and cognitive states will be determinant towards the successive stages. DSM systems could be used to determine drivers’ availability to take control, with sufficient time before the defined HAD use case is due, and even provide real-time feedback to an in-vehicle interface to prepare the driver for the transition.

DSM systems will require multimodal data and automatic learning techniques to create individual user profiles for such a complex task. The data required will not only involve eye-tracking but also environmental, behavioural, sociodemographic, psychophysiological, and possibly neural parameters in hybrid monitoring systems [7], [11], [30]. The nature of new in-vehicle possibilities (e.g. rotating seats, augmented reality head-up displays, holograms) [31], and NDRTs (e.g. reading, video gaming, conferencing, eating) [32], will strongly determine the integration of these DSM measurements into future vehicles, as well as reliability, privacy, invasiveness, and acceptability issues. Such new in-vehicle setups and their range of possibilities will challenge current eye-tracking based DSM systems. To date, eye tracking is the most suitable technique for driver monitoring, as it is ubiquitous, unobtrusive, and provides a multitude of information about the driver state [33]. However, HAD will also require combining multimodal sensors to ensure effective and comprehensive monitoring. Physiological measures are robust and reliable sources of drivers’ state factors as stress, workload, fatigue, drowsiness, or attention, and are not affected by eye-trackers main limitations, such as weather or lighting conditions [9] [11].

Recent driving simulator experiments, using wearable electrodermal activity (EDA) and electrocardiogram (ECG) sensors, have shown a promising alternative to overcome eye-trackers major drawbacks and towards the development of multimodal DSM systems [34]–[37]. For example, simply detecting if the driver is taking a nap with the seat reclined, awake and listening to music while looking through the window, or fully immersed in a videogame may imply radically different preparation to take-over strategies to get the driver back *in the loop*. Importantly, none of these scenarios would suit an eye-tracker to monitor. A solution could be a multimodal system combining eye-trackers with ECG and EDA devices.

ECG is an indicator of cardiovascular electrical activity [38]. The heart is innervated by the sympathetic (SNS) and parasympathetic (PNS) nervous systems, both branches of the autonomic nervous system. SNS is tied to stress and fight-or-flight responses and thus tends to increase cardiac activity (i.e. heart rate, HR), whereas PNS is tied to vagal and rest-and-digest behaviours and tends to decrease heart rate [19], [39]. However, modes of autonomic control are multidimensional, and the SNS is active at a basic level even during periods of rest to support homeostasis. SNS and PNS can be co-activated, co-inhibited, reciprocally active or independent [19]. Heart rate variability (HRV) is often used to derive metrics, especially from the PNS, which is relevant for psychophysiology in many aspects like self-regulation mechanisms linked to cognitive or affective states [40]. For example, PNS -vagal- withdrawal is required for SNS activation during certain executive functions like attention or emotional processing [19], [41].

EDA is an indicator of changes in skin conductivity resulting from the activation of sweat glands controlled by the SNS which, as stated above, prepares the body for fight-or-flight responses [42], [43]. Either background tonic (Skin Conductance Level: SCL) or rapid phasic components (Skin Conductance Responses: SCRs) can be extracted from the EDA signal. Notably, skin conductivity is not influenced by the PNS, and hence it is considered a direct indicator of psychophysiological arousal, and by extension, cognitive and affective states [42], [43].

Both ECG and EDA have proven their validity in detecting drivers' stress or mental workload in naturalistic driving [44] and driving simulators [37], [45], sleepiness [46], and discomfort [34]–[36]. In particular, increasing stress levels have been found under complex driving conditions [44], [47], possibly due to increased perceived risk. However, in previous work measuring psychophysiological responses to uncomfortable (challenging) driving situations in a driving simulator, the authors found arousal variations during manual driving, but there were no effects in physiological metrics between the autonomous driving controllers during discomfort periods [34].

Relatedly, in a driving simulator study exploring the effect of opposing levels of trust in automation across several highly automated driving scenarios -including a mentally demanding NDRT and a risky manoeuvre [37], we could not find any significant arousal variations derived from traffic complexity or an event-related risk scenario. We expected the NDRT would generate comparable physiological activation to that from the risk event. However, effects in HR, LF/HF ratio, and RMSSD indicated the task elicited a more robust response than the risk event, which only generated a moderate physiological reaction to the hazardous situation. Although a plausible explanation for this phenomenon could be that HRV metrics could not accurately measure rapid and event-related phasic arousal reactions due to the high decay time in the cardiac signal. skin conductance metrics should have captured these event-related reactions.

Notwithstanding, in a series of driving simulator studies [35], [36], the authors captured rapid event-related phasic cardiac reactions under several autonomous driving scenarios eliciting discomfort. They found HR and RMSSD decreasing only under high discomfort periods. Notably, they reported that longer lasting and slowly evolving situations with moderate to low discomfort events did not produce observable physiological changes, as these could only be observed during rapid events associated with moderate to high discomfort.

As drivers tend to be less alert under HAD conditions, compared to when driving manually [20], [23], [45], [48], these findings suggest that risk perception might be lowered under autonomous driving and consequently challenge the sensitivity of physiological indicators. Risk perception has been defined as "the likelihood and consequences of error" [49] and thus, is considered to play a major role in the modulation of trust in automation [49]–[53]. Essentially because trust implies assuming a position of vulnerability in an uncertain situation -i.e., the risk of delegating the control of the situation to an agent. However, perceived risk also relates to the perception, comprehension and projection of the elements in the environment, for which risk perception also has strong ties with situation awareness [54]. The relationship between risk perception and arousal was proposed by [55] when describing two processes for risk assessment that would drive the situational trust decision-making. De Visser identified the recognition-based threat assessment as a rapid affective evaluation of the information available. Conversely, the experience-based threat assessment was a slower, more deliberate evaluation of the risk and benefit based on available and observable information.

DSM systems could therefore make use of these affective cues for "threat assessment" to inform drivers' situational risk perception, and hence if the driver will be available for take-over. To enable an enhanced understanding of whether arousal indicators could be useful to detect risk perception during HAD, we have designed a driving simulator experiment with several autonomous driving situations aiming to produce different levels of perceived risk. The gaps in the literature this present research will investigate are whether:

- Longer-lasting and slowly evolving situations with low to moderate risk produce observable changes in physiology.

- And whether rapid events associated with moderate to high risk produce observable changes in physiology.

III. Research Hypotheses

The purpose of this research was to evaluate whether arousal indicators could be used to measure risk perception under simulated HAD conditions –i.e., applicable to both SAE L3 and L4 use cases. Our objective was two-sided:

First, to investigate whether slowly evolving situations with low to moderate risk would produce observable effects on arousal, we manipulated the presence of surrounding traffic as a between-group variable during a highly autonomous drive involving driving conditions changing slowly.

Second, to determine whether rapid events with moderate to high risk would generate observable arousal variations, we included a hazardous driving situation as a within-subjects variable at the end of the trial.

According to the literature previously discussed, increased risk perception should induce increased arousal to prepare the SNS for "fight-or-flight" reactions. Thus, we expect increased risk perception to reduce vagal HRV metrics and trigger SNS indicators (skin conductance responses). Conversely, lower levels of risk perception should induce vagal "rest-and-digest" low arousal states, observed by increased HRV metrics and reduced SNS indicators (SCRs).

Hence, we propose the following hypotheses:

H1: The presence of Traffic will increase the perception of risk and produce group differences in arousal.

H2: Changes across the driving conditions will progressively increase risk perception, and thus, arousal levels will vary among participants.

H3: The Driving Hazard event should rapidly increase risk perception and produce a greater arousal response than baseline and recovery period.

IV. Method

A. Participants

Twenty volunteers (10 male and 10 female, mean age 24.60 years, SD = 3.91) were recruited to participate in this research. All had held a UK-EU driving license for an average of 5.30 years (SD = 4.18) and possessed an average driving experience of 6780 miles/year (SD = 6140.08). Participants were recruited within the University of Warwick (UK), including undergraduate and postgraduate students and professionals. Recruitment and data collection procedures received approval from the University of Warwick's Biomedical and Scientific Research Ethics Committee. Participants were free to withdraw at any point and did not receive compensation.

Participants were randomly divided into two groups of ten. One group experienced the simulated driving scenario with surrounding traffic, and the other experienced the same scenario without traffic. Both groups were instructed to sit in the driver´s seat, but they were not explicitly asked to monitor the environment. Instead, they were asked to not engage with the driving task. The rationale for doing this was that they were about to test a highly automated vehicle that they did not need to drive manually nor would be requested to take over. Participants were not free to perform any NDRTs either as this could disrupt their situation awareness or affect their arousal.

B. *Apparatus*

This research was carried out using WMG's 3xD driving simulator at the University of Warwick. The 3xD is a fixed-base high-fidelity driving simulator, equipped with a full-body Range Rover Evoque and eight projectors generating a 360º image, projected into a cylindrical screen eight meters in diameter and three meters in height (Fig. 1). The simulated vehicle automation is capable of lateral and longitudinal control, adapting to speed limits, queuing leading vehicles, maintaining safe distances, emergency braking, and overtaking slower/stopped vehicles. The simulation also generated road motion vibration through the seats and environmental sound.

FIGURE 1 HERE

ECG and EDA data were recorded using BIOPAC MP160 with wearable remote Bio-Nomadix amplifiers. The MP160 base station was mounted behind the driver´s seat inside the simulator to achieve the best quality signal. Three ECG electrodes were fitted to each participant, following a standard 3-lead configuration on the participant's torso. The EDA device comprised two electrodes on the medial phalanx region on the index and middle fingers of the participant's non-dominant hand to minimise movement artefacts.

Subjective measures included a bespoke risk perception questionnaire comprising two items which were asked after the entire drive:

(1) Did you feel any sensation of risk or threat from the whole scenario?

(2) Did you feel any sensation of risk or threat from the traffic accident at the end?

These were rated on a Likert scale ranging from 1 (not at all) to 7 (extremely). There are no existing validated tools for risk perception assessment in the driving context to the authors' knowledge. [56] had a similar problem when assessing risk perception associated with trusting in automation in a driving simulator study and used the scale developed by [57]. Even though both studies reported significant effects on risk perception, the reasons for not using this scale were that it has not been validated and that the scale measures perceived situational risk and perceived relational risk. Our research was interested in comparing perceived situational risk from the Driving Conditions with the Driving Hazard scenario. Thus, this would have implied reporting perceived risk at the end of each condition, which we considered was contraindicated, due to our continuous driving scenario design and as the hazardous event occurred immediately after the autonomous driving conditions. Stopping the scenario immediately before the Driving Hazard event could have potentially affected the realism of the scenario and any psychophysiological reactions.

C. *Autonomous driving scenario*

The trial lasted a total of 11 minutes and 30 seconds. This began with four minutes of baseline/resting data, four minutes of autonomous driving scenarios, thirty seconds of the hazardous event, and two minutes of post-event recovery. The four minutes of autonomous driving were split into two scenarios: an initial two-minute suburban driving scenario labelled as Driving Condition 1 (DC1 for cardiac measures and DC1.1, DC1.2, DC1.3 and DC1.4 for SCR analyses which were split into 30-second segments, see Fig. 2).

This scenario started with the ego-vehicle stopped at a red traffic light at a five-lane roundabout which carries traffic to and from the highway to the suburbs and the city centre. The ego vehicle took the third roundabout exit leading to a straight dual carriageway, separated by a central reservation. Speed was limited from 30 to 50mph. Surrounding traffic levels were very low (< 5 road users per minute) at this point for the Traffic group, and weather conditions were cloudy for both groups.

Approximately one minute later, the ego vehicle entered the suburbs. This layout consisted of two lanes passing through residential areas at a maximum speed of 30mph, including several left and right turns and give-way exits. Oncoming traffic (i.e., for the Traffic group) increased to medium levels (< 20 road users per minute), now including pedestrians, cyclists, and parked cars, on the roadside and in driveways.

The simulation then continued with a two-minute city centre scenario, denoted as Driving Condition 2 (DC2 for cardiac measures and DC2.1, DC2.2, DC2.3 and DC2.4 for SCR analyses which were split into 30 seconds segments, see Fig. 2). In this scenario, the ego vehicle entered the city centre, and the layout changed to a "high street" area surrounded by commercial buildings, signs, and billboards. For the Traffic group, this also implied higher levels of moving pedestrians and vehicles, including vans and buses, stopped on the roadside, which the vehicle had to overtake, and T-junctions with traffic approaching from both directions (between 20 and 40 road users per minute). The speed limit was 30mph, and the weather conditions shifted to heavy rain, degrading the visual range.

Finally, the Driving Hazard event occurred when leaving the city centre to enter the suburbs again, on the approach of a T-junction, in a residential area from a straight two-way lane. This event was the sudden appearance of a heavy single-cabin semitrailer truck, which accelerated into the scene at high speed (60mph) from the left side of the T-junction ahead, moving sideways and headed directly towards the ego vehicle. The ego vehicle performed a sudden evasive manoeuvre to avoid the trailer, steered to the right side and eventually collided with a garden fence. This whole action sequence (i.e., from leaving DC2 through to the end of the crash) occurred in 30 seconds. The semitrailer truck was the only vehicle present in this scenario for the Traffic group to ensure the condition was equal for both groups.

D. Procedure

Upon arrival, participants were guided into the simulator control room, where the room temperature was set at 21±2ºC to control for room temperature affecting EDA or ECG recordings (the simulator buck's ventilation system could also be adjusted to participants' requirements). Participants were briefed on lab safety procedures and then filled in the consent form and demographic inventories. Once all physiological sensors were connected, participants were instructed to be careful in applying any pressure to the sensors or stretching the cables to avoid signal spikes and artefacts. Following this, data telemetry from the wearable amplifiers were checked to ensure signal stabilisation and good quality data acquisition.

Participants were then guided inside the driving simulator and were informed that the experiment would start by recording their physiological state baseline for 4 minutes, and after that, the driving scenario would begin. Participants were asked to remain seated in the driver's seat, not to move excessively, breathe normally, and stay relaxed during the baseline recording. Participants were advised that the experimenter would inform them of the start and end of the baseline recording. The driving simulator lights were switched off, the room was silent, and driving scenarios were not projected on the screen. Once the baseline was recorded, the autonomous driving trial began and lasted approximately 5 minutes. After the hazardous event, participants remained in the vehicle with the scenario displayed on-screen for 2 minutes to record a post-event recovery. After that, the experimenter entered the simulator and accompanied them back into the control room to fill in the risk perception scale.

E. Analysis

For each participant, a total of 11 minutes and 30 seconds of continuous data were extracted for analysis. This comprised four minutes of baseline/resting data, four minutes of autonomous driving scenarios, thirty seconds of the hazardous event, and two minutes of post-event recovery. One minute of transition was left between baseline and the first segment of automated driving to allow the signal to stabilise and was not included in the analysis. One participant was excluded from EDA analysis due to substantial artefacts on the raw signal, with N = 19 participants analysed. For ECG data, all participants were analysed with no missing cardiac data.

Data were segmented into epochs of 120 seconds for HRV analyses and 30 seconds for SCR analyses. HRV analyses used four epochs, comprising baseline (BL), Driving Condition 1 (DC1), Driving Condition 2 (DC2), and the final epoch (Hazard-&-Recovery), including the Driving Hazard event and the post-event recovery time. SCR analyses used eleven epochs of 30s each (see Fig. 2). Data were extracted using the automated data analysis routines from Biopac's ACQKnowledge software (CA, USA; version 5.0.2). EDA data were sampled at 62.5 Hz and low-pass filtered to a frequency cut-off fixed at 1 Hz, following standardised guidelines [43], [58]. Phasic EDA features were extracted using a high pass filter at 0.05 Hz, and the skin conductance response (SCR) threshold level was set at 0.03 µS, with a rejection rate set to 10%.

FIGURE 2 HERE

Phasic features extracted were SCR count (i.e. the total number of SCR events within each epoch), SCR amplitude, and SCR magnitude. SCR amplitude represents the delta value from the offset to the peak of the SCRs. According to [43], [59], amplitudes below 0.03 µS were rejected from the analysis. The common practice for normalising these values is applying the square root transformation [42], [58], [60]. SCR magnitude is obtained from the same delta value, but non-response accounts for a zero for the final mean. In this case, the Log+1 transformation is applied to correct for the presence of skewness and kurtosis [42], [58], [60]. Finally, these three SCR features were standardised for parametric statistical analysis to T-scores (M = 50, SD = 10) to allow for inter-individual comparisons. Means and standard deviations used for the T-scoring were obtained from each individual to control inter-individual variability [43], [60].

ECG data were sampled at 2000 Hz and filtered applying Biopac's recommendations, using a bandpass filter with a 35 Hz high-frequency cut-off and a low-frequency cut-off at 0.5 Hz. Cardiac features extracted were heart rate (HR; as in beats per minute) and those heart rate variability (HRV) metrics that better reflect vagal tone. We extracted the high frequency (HF) band (between 0.15 and 0.40 Hz) in the frequency domain. Following the recommendations in [40], we coupled this metric with a time-domain parameter indexing vagal tone: the root mean square of successive differences (RMSSD). This robust metric reflects the vagal tone and is relatively less affected by respiration [40]. Cardiac features were standardised to T-scores following the same method described for SCRs.

This research evaluates whether variations in perceived risk from either slowly evolving or rapidly evolving HAD conditions would produce observable changes in physiology. H1 tested the effect of the grouping variable of Traffic on Driving Conditions (i.e., Suburbs [DC1] and City Centre [DC2]). A 2 x 3 mixed ANOVA (Traffic, No Traffic x BL, DC1, DC2) was performed on HR/HRV measures. Similarly, we ran a 2 x 9 mixed ANOVA on SCR metrics (Traffic, No Traffic x BL, DC1.1, DC1.2, DC1.3, DC1.4, DC2.1, DC2.2, DC2.3, DC2.4). As traffic was not present during the BL condition, and the BL condition would allow further control over pre-existing group differences in arousal prior to experimental manipulations, we included BL in this analysis.

To test for the effect of changing driving conditions within-participants (H2), we run a repeated-measures ANOVA with three levels for HR/HRV metrics comparing BL with DC1 and DC2. A similar test was used for SCR metrics, with 9 levels (BL, DC1.1, DC1.2, DC1.3, DC1.4, DC2.1, DC2.2, DC2.3, DC2.4). As H2 tested for the effect(s) of identical driving conditions for each group (i.e. DC1 and DC2), we merged these groups after finding no main effects or interaction effects during the analysis in H1.

To analyse the effect of the rapidly evolving Driving Hazard event (H3), we ran a repeated-measures ANOVA with two levels for HR/HRV metrics comparing BL with Driving Hazard-&-Recovery. A similar repeated-measures ANOVA with three levels for SCR metrics compared BL with Driving Hazard and Recovery (see Fig. 2). We rationalised this decision as both groups experienced the same Driving Hazard condition.

The Shapiro-Wilk's test ($p \geq 0.05$) was used to assess normality assumption violations, and Mauchly's test was used to assess the assumption of sphericity. Main effects and interactions were followed-up by pair-wise comparisons corrected by the Bonferroni method.

V. Results

A. Hypothesis 1

The first hypothesis tested whether the slowly evolving increase of traffic presence would modulate perceived risk and produce observable differences in arousal between groups during driving conditions.

Among all cardiac measures, mixed ANOVA results revealed that the effect of Traffic was not significant between-groups [HR: ($F(1, 18) = 0.589$, $p = 0.453$, $\eta^2_p = 0.032$); HRV-HF: ($F(1, 18) = 0.024$, $p = 0.878$, $\eta^2_p = 0.001$); HRV-RMSSD: ($F(1, 18) = 0.322$, $p = 0.577$, $\eta^2_p = 0.018$)], and there were no interaction effects for Driving Conditions which supported this hypothesis either [HR: ($F(2, 36) = 1.935$, $p = 0.159$, $\eta^2_p = 0.097$); HRV-HF: ($F(2, 36) = 0.949$, $p = 0.397$, $\eta^2_p = 0.05$); HRV-RMSSD: ($F(2, 36) = 0.451$, $p = 0.641$ $\eta^2_p = 0.024$)].

FIGURE 3 HERE

Skin conductance response analyses did not show any group effects of Traffic [SCR count: ($F(1, 17) = 2.012$, $p = 0.174$, $\eta^2_p = 0.106$); SCR amplitude: ($F(1, 17) = 0.088$, $p = 0.771$, $\eta^2_p = 0.005$; SCR magnitude: ($F(1, 17) = 4.206$, $p = 0.056$, $\eta^2_p = 0.198$)], or interaction effects for Driving Conditions [SCR count: ($F(8, 136) = 0.478$, $p = 0.870$, $\eta^2_p = 0.027$); SCR amplitude: ($F(8, 136) = 0.434$, $p = 0.810$, $\eta^2_p = 0.025$); SCR magnitude: ($F(8, 136) = 1.064$, $p = 0.385$, $\eta^2_p = 0.059$)] that would support our hypothesis either. Trends for all three SCR features analysed are displayed in Fig. 4.

FIGURE 4 HERE

In accordance with these results, a Mann- Whitney *U* test on self-reported risk perception did not show any differences between groups for Traffic ($U = 60.500$, $p = 0.436$. These results seem to indicate that the presence of traffic had no effects on the perceived risk between groups.

B. Hypothesis 2

This hypothesis tested whether slow changes across driving conditions would produce arousal variations within-participants.

There was a main effect of Driving Conditions on HRV-RMSSD ($F(2, 38) = 4.497$, $p = 0.018$, $\eta^2_p = 0.191$), with post-hoc tests indicating a lowered vagal tone from baseline ($M = 55.679$, $SD = 7.562$) to DC1 ($M = 47.147$, $SD = 7.088$, $p = 0.004$) (see Fig. 3). There was also a main effect of Driving Conditions on HRV-HF power ($F(2, 38) = 3.490$, $p = 0.041$, $\eta^2_p = 0.155$); however, this effect diminished in post-hoc tests (see Fig. 3). HR did not report any main effects of Driving Conditions ($F(2, 38) = 0.837$, $p = 0.441$, $\eta^2_p = 0.042$).

SCR measures did not report any main effects within-participants that would support hypothesis 2 [SCR count: ($F(8, 144) = 1.129$, $p = 0.348$, $\eta^2_p = 0.059$); SCR amplitude: ($F(8, 144) = 1.419$, $p = 0.228$, $\eta^2_p = 0.073$); SCR magnitude: ($F(8, 144) = 0.396$, $p = 0.851$, $\eta^2_p = 0.022$)].

These results suggest that HRV features captured slowly evolving arousal variations across Driving Conditions, but SCRs were not. We will discuss these findings further in section VI.

C. Hypothesis 3

The third hypothesis investigated whether the rapidly evolving Driving Hazard event would produce observable effects in arousal indicators within participants.

A repeated measures ANOVA revealed a main effect of Driving Hazard on HRV-RMSSD ($F(1, 19) = 5.815$, $p = 0.026$, $\eta^2_p = 0.234$), with follow-up tests showing a significantly lower vagal tone during Hazard-&-Recovery ($M = 48.499$, $SD = 7.984$), compared to during baseline ($M = 55.679$, $SD = 7.562$) as expected (see Fig. 5). No effects were observed on HR ($F(1, 19) = 4.389$, $p = 0.050$, $\eta^2_p = 0.188$) or HRV-HF power ($F(1, 19) = 1.673$, $p = 0.211$, $\eta^2_p = 0.081$).

FIGURE 5 HERE

Strong evidence in favour of this hypothesis was found across all SCR measures (Fig. 6). SCR count revealed a main effect for Driving Hazard ($F(2, 36) = 10.465$, $p < 0.001$, $\eta^2_p = 0.368$), with a significantly greater SCR count during the hazardous event ($M = 63.156$, $SD = 13.092$), compared to baseline ($M = 48.293$, $SD = 9.343$, $p = 0.004$), and Recovery ($M = 48.984$, $SD = 9.964$, $p = 0.012$).

FIGURE 6 HERE

SCR amplitude showed similar effects ($F(2, 36) = 22.415$, $p < 0.001$, $\eta^2_p = 0.555$), with post-hoc tests indicating a significantly greater amplitude during the Driving Hazard event ($M = 60.870$, $SD = 8.417$), than during baseline ($M = 49.769$, $SD = 4.142$, $p < 0.001$), and Recovery ($M = 48.898$, $SD = 2.655$, $p < 0.001$).

SCR magnitude was also aligned with the previous two SCR measures and revealed a main effect for Driving Hazard event ($F(2, 36) = 177.834$, $p < 0.001$, $\eta^2_p = 0.908$). Follow-up tests showed a significantly greater amplitude during the event ($M = 76.275$, $SD = 3.975$) compared to baseline ($M = 47.633$, $SD = 7.325$, $p < 0.001$), and Recovery ($M = 46.373$, $SD = 3.988$, $p < 0.001$).

Finally, a Wilcoxon signed-rank test reported a main effect for self-reported risk perception ($Z = 194.5$, $p = 0.001$), with perceived risk during the Driving Hazard event ($Mdn = 5.50$, $IQR = 3$) being significantly higher than during Driving Conditions ($Mdn = 3.00$, $IQR = 3$).

Overall, these results suggest the Driving Hazard event had a greater effect on skin conductance measures than on cardiac ones, but instead, HRV was more sensitive to the slowly evolving effect of Driving Conditions than SCRs. Further interpretations of these results will be discussed in the next section.

VI. Discussion

This empirical research aimed to investigate whether slowly evolving autonomous driving situations with low to moderate perceived risk; and rapid driving events with moderate to high risk would produce observable changes in physiological indicators of sympathetic and parasympathetic activity.

A. *Hypothesis 1*

The first hypothesis predicted that the presence of Traffic would slowly increase the perception of risk in one group across driving conditions and result in significant group differences in arousal. The presence of Traffic did not have the expected effect on the perceived risk, however, since no between-group or interaction effects were reported for any of the physiological indicators or self-reported perceived risk.

Whereas the reduced group size has likely contributed to this absence of effects, the lack of arousal differences due to the presence of Traffic seems to indicate this variable would not strongly modulate perceived risk during autonomous driving, as opposed to when manually driving [34], [44], [47]. Regardless, we would instead remain cautious about making such inferences because a genuine criticism of driving simulators is that they lack real risk, and participants are aware of this. Perhaps real-world trials would obtain different results, and therefore, future research should investigate perceived risk from naturalistic autonomous driving conditions. Besides, the influence of traffic on perceived risk might be tightened to individual differences such as personality traits, age or gender, especially when the sample size is small.

B. *Hypothesis 2*

The second hypothesis investigated whether long-term evolving Driving Conditions with low to moderate risk would produce observable changes in arousal within participants –i.e. from BL to DC2. Considering the median of self-reported perceived risk during Driving Conditions was 3 out of a maximum of 7, we would assume moderate levels of risk during these.

Evidence in favour of this hypothesis was observed in HRV, which reported a vagal tone decrease between baseline and suburbs (i.e. Driving Condition 1, DC1). Previous research in the driving context has associated uncoupled vagal withdrawal (i.e. with unaltered sympathetic activity) from baseline resting with increased monitoring during single-driving tasks requiring perceptual-central processing [19], [61]. It would therefore provide some evidence in favour of this hypothesis. In particular, RMSSD would have captured such long-term and slowly evolving arousal fluctuation associated with low to moderate risk from resting to driving through suburbs (DC1). The other HRV parameter, HF-power, coupled these results, but effects diminished with post-hoc comparisons. Given the nature of the HRV signal, perhaps evolving manipulations in our driving scenarios with relatively short periods and without enough recovery time between may have contributed to this.

Skin conductance responses (SCRs) did not report any effects for Driving Conditions in hypothesis 2 either. It might be that these measures were less sensitive to moderate perceived risk levels (note that median self-reported perceived risk during Driving Conditions was 3 out of a maximum of 7). This would also relate with previous work in which we observed SCRs trends increasing along with driving complexity but not reporting any significant effects [37]. However, this lack of significant effects could also be interpreted as SCR features being less sensitive to long-term arousal variations. Our SCR measures here were over a 30-second epoch, and like [35] who used similar short epochs for SCL, we did not observe any changes in SCR either. Longer epochs would likely be more sensitive to long-term changes in arousal. However, in previous research [37], we did not observe any effects on SCR features over two-minute epochs either, so perhaps future work should also consider exploring other electro-dermal activity features for driver state monitoring of slowly evolving risk changes. For example, endosomatic electro-dermal activity measures such as skin potential level or skin potential responses, measure the nervous impulses that activate sweat glands without directly applying current to the body. Unlike SCRs and SCLs, endosomatic measures are less affected by electrode artefacts [43]. Skin potential responses have been successfully used in the driving context for detecting sympathetic reactions to unexpected events [62], suggesting they could be a promising alternative to traditional exosomatic features, worth exploring for its integration in DSMs. Notwithstanding, skin potential measures have certain drawbacks for recording and evaluation as described in [43], which must be carefully considered.

The lack of arousal variations between the two Driving Conditions of suburbs (DC1) and city centre (DC2) may suggest that autonomous driving users might be less aware of slowly changing driving conditions. Indeed, in our previous work in the driving simulator using a different sample, our cardiac features captured long-term arousal variations due to mental workload from an NDRT, but not due to driving conditions [37]. Similarly, [34] detected arousal variations associated with discomfort towards abnormal driving behaviours during manual driving, but not through several autonomous driving controllers. However, as stated earlier in H1, these assumptions must be cautiously considered, and further research is needed.

C. *Hypothesis 3*

The third hypothesis predicted that a rapid driving event associated with moderate to high-risk perception would produce arousal variations compared to baseline resting and post-event recovery resting. Considering that self-reported perceived risk was significantly greater during this event than during Driving Conditions, substantial evidence favouring this hypothesis was found particularly throughout skin conductance response features. All three indices reported robust statistical effects indicating that SCRs were particularly sensitive to arousal fluctuations from a rapid driving event with moderate to high risk associated. These findings are consistent with previous literature reporting greater SCRs with increased mental workload [47], stress [44], or discomfort [34] due to complex driving conditions.

Even though SCRs are well known for being particularly sensitive to phasic arousal changes due to unexpected environmental stimuli [42], [43], the novelty of these results lies in that these would be physiological indicators of a rapid increase of perceived risk. Aside from indicating we successfully generated and measured a greater perception of risk during the event, these findings add evidence to those from [56] in that abnormal traffic behaviours are also perceived as risky during HAD. However, the mere presence of surrounding traffic behaving normally may not strongly influence risk perception as H1 suggests.

Results from cardiac parameters were less robust, however. Only HRV-RMSSD reported a significant vagal tone decrement between baseline and the epoch comprising driving Hazard-&-Recovery, as it was expected that the hazardous event would produce a startle response activating the sympathetic system and deactivating the parasympathetic branch [19]. Nevertheless, heart rate (HR) and HRV-HF did not report any supporting effects, and on the contrary, trends indicated a higher rate during baseline resting than during driving Hazard-&-Recovery, suggesting participants may have been more aroused during baseline resting than during Hazard-&-Recovery. Such a non-reciprocally coupled mode of autonomic control would indicate a co-activation during baseline because sympathetic activation exceeded parasympathetic activation; and a co-inhibition during Hazard-&-Recovery, which may occur because a parasympathetic inhibition exceeded sympathetic inhibition [19], [39], [61].

For this co-activation, likely, our participants were not completely relaxed during the baseline, and thus, it was not necessarily a low arousal state. This was possibly due to the novelty of the experiment. Therefore, we could recommend future research to collect baseline recordings in a more familiar and duller environment, but it could also be argued that such an environment would lack ecological validity. In contrast, co-inhibition observed during Hazard-&-Recovery was possibly due to the hazardous event was followed by a recovery period, which was essentially a state of rest. Therefore combining the hazardous event and post-event recovery into one epoch could have contributed to this effect, although it was necessary to combine these periods to meet the minimum epoch length recommended for HRV measurements [40], [63].

As a whole, our results provide important insights into risk perception and its measurement with physiological indices during HAD. It is evident that HAD implies a driving task reallocation and changes the way users perceive and interact with the vehicle. As stated in section II, perceived risk modulates trust in automation and situation awareness, and consequently, driver readiness towards a take-over request (TOR). Whilst this phenomenon may be mitigated in SAE-L4 with driving automation acting as a fallback, it may become a safety-critical aspect in SAE-L3, where the driver is the fallback user when a TOR is issued. Inadequate situation awareness because of lowered perceived risk is likely to increase trust in automation inappropriately, and that, as has already been explicitly evidenced in several road accidents in the US, may lead to fatal consequences [64]–[67]. Now that *autonowashing* [68] is a hot topic in the automotive industry, it is vital that these vehicles integrate DSM systems for risk mitigation of impaired or unavailable drivers. In this sense, we agree with [28], [29] that monitoring requests to get the driver back on the loop towards the upcoming TOR will be strongly required.

Based on the model of affective risk assessment proposed by De Visser [55], our results suggest a promising pathway for measuring drivers perceived risk with arousal indicators. Hence, situational trust in automation and current situation awareness when issuing a monitoring request towards a planned take-over. Even though arousal indices alone cannot infer such complex psychological states, artificial intelligence and machine learning algorithms could potentially be trained to do so [69]. Furthermore, the effect of individual factors such as personality traits (e.g., sensation seeking [70] or locus of control [71]), age and gender [72] cannot be ignored. Multimodal DSM systems combining eye-tracking data with arousal indices could classify driver states and monitor the transition process until reaching driver availability for take-over.

SCR data derived from the EDA signal would be instrumental in detecting whether the driver has perceived the monitoring request warning issued in non-safety critical TOR. Because SCRs indicate sympathetic activation and thus increased alertness level, they are particularly sensitive to startle stimuli -as we observed in the hazardous event-, which a TOR can be to an OOTL driver. Driver awareness of the monitoring request could be complemented with gaze behaviour indicators (e.g., driver glancing at the interface screen issuing the visual warning) and even with an active behavioural confirmation (e.g., pressing a button or verbally confirming).

Whilst HR/HRV data from the ECG signal would be beneficial for detecting variations in tonic arousal levels. For example, vagal withdrawal and heart rate increases would be good indicators of optimal alertness when monitoring the preparation for take-over and the take-over transition itself, as modes of autonomic control can indicate the source of attentional demands during task performance [19]. Although we only find two effects for cardiac features in the present research, substantial evidence supports the robustness of HR/HRV in detecting arousal from driving conditions of stress [9], [12], or mental workload [19], [39], supports this claim. Moreover, we encourage future work to explore whether shorter or fine-sliding windows are sensitive to shorter-term arousal variations, especially with RMSSD, which allows so [40], and has shown the most robust effects for HRV here. Eye-tracking data could also complement this by ensuring the driver is back on the loop [5] (e.g. actively seeking information on the road ahead, side mirrors or checking the navigation system).

It is essential to bear in mind that psychophysiological variations are relative to a previous state, and this state may vary depending on the NDRT performed before the monitoring request. As mentioned, the arousal state from a driver sleeping will widely differ from, e.g., one playing videogames. Whereas the former may need to increase its alertness level, the latter may need to reduce it to make effective decision-making (the reader may remember the inverted U-shape relating to arousal and performance) [17].

Future research should test the sensitivity of SCRs to monitoring requests. The latency of the signal is between 1-3 seconds after issuing the request, and thus, SCR could inform that the driver has perceived the warning. It would also be crucial to explore how different levels of mental workload induced by NDRTs may affect the perception of monitoring request warnings. Relatedly, it would also be relevant to explore transitions from either resting versus mentally demanding NDRTs after issuing the monitoring request.

VII. Conclusion

Findings from this research demonstrate that future driver state monitoring systems for highly automated driving will need to combine multiple data sources to overcome current eye-tracking-based systems' main drawbacks. Our empirical research provided evidence on how heart rate (HRV) and skin conductance (SCR) features provide valuable additional real-time data to determine drivers' perceived risk, which can be used to indicate their availability to take over control. Overall, our findings indicate that:

- Low to moderate risk perception should be measured with HRV features, which are more sensitive to longer-term changes in arousal levels due to environmental and traffic-related factors than SCRs,
- Moderate to high perceived risk should be measured with SCRs, indicative of the short-term changes to rapidly evolving, safety-critical driving events.

Furthermore, HRV and SCR measures face significant challenges, such as detecting moderate arousal levels required for take-over, false positives due to high tonic arousal, or even false negatives due to highly arousing non-driving related tasks – all of which could mask actual physiological indicators needed for safe take-over of control. Eye-tracker data could potentially be used in tandem in such cases to alleviate these limitations of physiological data alone. Ultimately, future DSM systems will benefit from the knowledge generated by this research through the development of machine learning methods used for determining when drivers would be on the loop, and their availability for optimal take-over performance.

References

[1] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," 2018. doi: https://doi.org/10.4271/J3016_201806.

[2] O. Carsten, F. Lai, Y. Barnard, A. H. Jamson, and N. Merat, "Control Task Substitution in Semiautomated Driving," *Hum. Factors*, vol. 54, no. 5, pp. 747–761, 2012, doi: 10.1177/0018720812460246.

[3] F. Naujoks, C. Purucker, and A. Neukum, "Secondary task engagement and vehicle automation - Comparing the effects of different automation levels in an on-road experiment," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 38, pp. 67–82, 2016, doi: 10.1016/j.trf.2016.01.011.

[4] J. Wörle *et al.*, "Sleep Inertia Countermeasures in Automated Driving: A Concept of Cognitive Stimulation," *Inf.*, vol. 11, no. 7, pp. 1–15, 2020, doi: 10.3390/INFO11070342.

[5] N. Merat *et al.*, "The 'Out-of-the-Loop' concept in automated driving: proposed definition, measures and implications," *Cogn. Technol. Work*, vol. 21, no. 1, pp. 87–98, 2019, doi: 10.1007/s10111-018-0525-8.

[6] S. Begum, "Intelligent driver monitoring systems based on physiological sensor signals: A review," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2013, pp. 282–289, doi: 10.1109/ITSC.2013.6728246.

[7] H. B. Kang, "Various approaches for driver and driving behavior monitoring: A review," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 616–623, 2013, doi: 10.1109/ICCVW.2013.85.

[8] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, 2011, doi: 10.1109/TITS.2010.2092770.

[9] A. Nemcova *et al.*, "Multimodal Features for Detection of Driver Stress and Fatigue: Review," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–20, Mar. 2020, doi: 10.1109/tits.2020.2977762.

[10] C. Marberger, H. Mielenz, F. Naujoks, J. Radlmayr, K. Bengler, and B. Wandtner, "Understanding and applying the concept of 'Driver availability' in automated driving," in *Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing*, 2018, vol. 597, pp. 595–605, doi: 10.1007/978-3-319-60441-1_58.

[11] M. Lohani, B. R. Payne, and D. L. Strayer, "A Review of Psychophysiological Measures to Assess Cognitive

States in Real-World Driving," *Front. Hum. Neurosci.*, vol. 13, no. 57, pp. 1–27, 2019, doi: 10.3389/fnhum.2019.00057.

[12] M. Singh and A. Queyam, "Stress Detection in Automobile Drivers using Physiological Parameters: A Review," *Int. J. Electron. Eng.*, vol. 5, no. 2, pp. 1–5, 2013, [Online]. Available: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Stress+Detection+in+Automobile+Drivers+using+Physiological+Parameters:+A+Review#0.

[13] D. Miller *et al.*, "Distraction becomes engagement in automated driving," in *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*, Sep. 2015, pp. 1676–1680, doi: 10.1177/1541931215591362.

[14] C. Collet, C. Petit, A. Priez, and A. Dittmar, "Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation," *Biol. Psychol.*, vol. 69, no. 2, pp. 195–203, 2005, doi: 10.1016/j.biopsycho.2004.07.003.

[15] G. Durantin, J. F. Gagnon, S. Tremblay, and F. Dehais, "Using near infrared spectroscopy and heart rate variability to detect mental overload," *Behav. Brain Res.*, vol. 259, pp. 16–23, 2014, doi: 10.1016/j.bbr.2013.10.042.

[16] R. M. Yerkes and J. D. Dodson, "The Relation of Strenght of Stimulus to Rapidity of Habit Formation," *J. Comp. Neurol. Psychol.*, vol. 18, pp. 459–482, 1908, doi: 10.5860/choice.47-0570.

[17] K. H. Teigen, "Yerkes-Dodson: A Law for all Seasons," *Theory & Psychology*, vol. 4, no. 4. pp. 525–547, 1994, doi: 10.1177/0959354394044004.

[18] N. R. Bailey and M. W. Scerbo, "Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust," *Theor. Issues Ergon. Sci.*, vol. 8, no. 4, pp. 321–348, 2007, doi: 10.1080/14639220500535301.

[19] J. K. Lenneman and R. W. Backs, "Cardiac autonomic control during simulated driving with a concurrent verbal working memory task," *Hum. Factors*, vol. 51, no. 3, pp. 404–418, Jun. 2009, doi: 10.1177/0018720809337716.

[20] S. Sibi, H. Ayaz, D. P. Kuhns, D. M. Sirkin, and W. Ju, "Monitoring Driver Cognitive Load Using Functional Near Infrared Spectroscopy in Partially Autonomous Cars," *IEEE Intell. Veh. Symp.*, no. Iv, 2016.

[21] V. Melnicuk, S. Thompson, P. Jennings, and S. Birrell, "Effect of cognitive load on drivers' State and task performance during automated driving: Introducing a novel method for determining stabilisation time following take-over of control," *Accid. Anal. Prev.*, vol. 151, pp. 1–14, Mar. 2021, doi: 10.1016/j.aap.2020.105967.

[22] J. Wörle, B. Metz, I. Othersen, and M. Baumann, "Sleep in highly automated driving: Takeover performance after waking up," *Accid. Anal. Prev.*, vol. 144, pp. 1–9, 2020, doi: 10.1016/j.aap.2020.105617.

[23] J. C. F. De Winter, R. Happee, M. H. Martens, and N. A. Stanton, "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.*, vol. 27, no. PB, pp. 196–217, 2014, doi: 10.1016/j.trf.2014.06.016.

[24] M. R. Endsley, "From Here to Autonomy: Lessons Learned from Human-Automation Research," *Hum. Factors*, vol. 59, no. 1, pp. 5–27, 2017, doi: 10.1177/0018720816681350.

[25] M. Ferrara and L. De Gennaro, "The sleep inertia phenomenon during the sleep-wake transition: Theoretical and operational issues," *Aviat. Sp. Environ. Med.*, vol. 71, no. 8, pp. 843–848, 2000, Accessed: Sep. 22, 2020. [Online]. Available: https://www.researchgate.net/publication/12367344.

[26] K. Izzetoglu, S. Bunce, B. Onaral, K. Pourrezaei, and B. Chance, "Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks," *Int. J. Hum. Comput. Interact.*, vol. 17, no. 2, pp. 211–227, 2004, doi: 10.1207/s15327590ijhc1702.

[27] M. H. Martens and A. P. Van Den Beukel, "The road to automated driving: Dual mode and human factors considerations," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2013, pp. 2262–2267, doi: 10.1109/ITSC.2013.6728564.

[28] C. Gold, L. Lorenz, D. Damböck, and K. J. Bengler, "Partially Automated Driving as a Fallback Level of High Automation," *6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren.* pp. 1–5, 2013, Accessed: Oct. 07, 2020. [Online]. Available: https://mediatum.ub.tum.de/doc/1187198/file.pdf.

[29] Z. Lu, B. Zhang, A. Feldhütter, R. Happee, M. Martens, and J. C. F. De Winter, "Beyond mere take-over requests: The effects of monitoring requests on driver attention, take-over performance, and acceptance," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 63, pp. 22–37, May 2019, doi: 10.1016/j.trf.2019.03.018.

[30] V. Melnicuk, S. Birrell, E. Crundall, and P. Jennings, "Towards hybrid driver state monitoring: Review, future perspectives and the role of consumer electronics," in *IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 1392–1397, doi: 10.1109/IVS.2016.7535572.

[31] S. Damiani, E. Deregibus, and L. Andreone, "Driver-vehicle interfaces and interaction: Where are they going?," *Eur. Transp. Res. Rev.*, vol. 1, no. 2, pp. 87–96, 2009, doi: 10.1007/s12544-009-0009-2.

[32] J. Navarro, "A State of science on Highly Automated Driving," *Theor. Issues Ergon. Sci.*, 2018, doi: 10.1080/1463922X.2018.1439544.

[33] T. Hecht, A. Feldh, J. Radlmayr, Y. Nakano, Y. Miki, and C. Henle, "A Review of Driver State Monitoring Systems in the Context of Automated Driving," in *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, 2019, vol. 1, pp. 398–408, doi: 10.1007/978-3-319-96098-2.

[34] V. Radhakrishnan *et al.*, "Measuring drivers'

physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort," *Inf.*, vol. 11, no. 8, p. 390, Aug. 2020, doi: 10.3390/INFO11080390.

[35] M. Beggiato, F. Hartwich, and J. Krems, "Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving," *Front. Hum. Neurosci.*, vol. 12, p. 338, 2018, doi: 10.3389/fnhum.2018.00338.

[36] M. Beggiato, F. Hartwich, and J. Krems, "Physiological correlates of discomfort in automated driving," *Transp. Res. Part F Traffic Psychol. Behav.*, vol. 66, pp. 445–458, 2019, doi: 10.1016/j.trf.2019.09.018.

[37] J. Perello-March, C. Burns, R. Woodman, M. T. Elliott, and S. Birrell, "Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–11, 2021, doi: 10.1109/TITS.2021.3050518.

[38] B. Cowley *et al.*, "The Psychophysiology Primer: A Guide to Methods and a Broad Review with a Focus on Human–Computer Interaction," *Found. Trends® Human–Computer Interact.*, vol. 9, no. 3–4, pp. 151–308, 2016, doi: 10.1561/1100000065.

[39] R. W. Backs, J. K. Lenneman, J. M. Wetzel, and P. Green, "Cardiac Measures of Driver Workload during Simulated Driving with and without Visual Occlusion," *Hum. Factors*, vol. 45, no. 4, pp. 525–538, 2003, doi: 10.1518/hfes.45.4.525.27089.

[40] S. Laborde, E. Mosley, and J. F. Thayer, "Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting," *Front. Psychol.*, vol. 8, no. 213, pp. 1–18, 2017, doi: 10.3389/fpsyg.2017.00213.

[41] H. G. Kim, E. J. Cheon, D. S. Bai, Y. H. Lee, and B. H. Koo, "Stress and heart rate variability: A meta-analysis and review of the literature," *Psychiatry Investig.*, vol. 15, no. 3, pp. 235–245, 2018, doi: 10.30773/pi.2017.08.17.

[42] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," in *Handbook of Psychophysiology, Fourth Edition*, Cambridge University Press, 2016, pp. 217–243.

[43] W. Boucsein, *Electrodermal Activity*, 2nd ed. New York, New York, USA: Springer, 2012.

[44] J. A. Healey and R. W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, 2005, doi: 10.1109/TITS.2005.848368.

[45] V. Melnicuk, S. Birrell, E. Crundall, and P. Jennings, "Employing consumer electronic devices in physiological and emotional evaluation of common driving activities," in *IEEE Intelligent Vehicles Symposium (IV)*, 2017, pp. 1529–1534, doi: 10.1109/IVS.2017.7995926.

[46] J. Wörle, B. Metz, C. Thiele, and G. Weller, "Detecting sleep in drivers during highly automated driving: The potential of physiological parameters," *IET Intell. Transp. Syst.*, vol. 13, no. 8, pp. 1241–1248, Aug. 2019, doi: 10.1049/iet-its.2018.5529.

[47] H. J. Foy and P. Chapman, "Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation," *Appl. Ergon.*, vol. 73, pp. 90–99, 2018, doi: 10.1016/j.apergo.2018.06.006.

[48] F. N. Biondi, M. Lohani, R. Hopman, S. Mills, J. M. Cooper, and D. L. Strayer, "80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal," in *Proceedings of the Human Factors and Ergonomics Society*, Sep. 2018, vol. 3, no. 1, pp. 1878–1882, doi: 10.1177/1541931218621427.

[49] V. Riley, "Operator Reliance on Automation: Theory and Data," in *Automation and Human Performance Theory and Applications*, Mustapha Mouloua and Raja Parasuraman, Eds. Boca Raton: Taylor & Francis Group, 1996, pp. 19–35.

[50] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, 2015, doi: 10.1177/0018720814547570.

[51] J. D. Lee and N. Moray, "Trust, control strategies and allocation of function in human-machine systems," *Ergonomics*, vol. 35, no. 10, pp. 1243–1270, 1992, doi: 10.1080/00140139208967392.

[52] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 46, no. 1, pp. 50–80, 2004, doi: 10.1518/hfes.46.1.50_30392.

[53] B. M. Muir, "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems," *Ergonomics*, vol. 37, no. 11, pp. 1905–1922, 1994, doi: 10.1080/00140139408964957.

[54] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Hum. Factors*, vol. 37, no. 1, pp. 32–64, 1995, doi: 10.1518/001872095779049543.

[55] E. J. de Visser, "The World Is Not Enough: Trust in Cognitive Agents," George Mason University Fairfax, VA, 2012.

[56] M. Li, B. E. Holthausen, R. E. Stuck, and B. N. Walker, "No Risk No Trust: Investigating Perceived Risk in Highly Automated Driving," in *AutomotiveUI '19: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2019, vol. 21–25, no. September, pp. 177–185, doi: 10.1145/3342197.3344525.

[57] B. Rajaonah, N. Tricot, F. Anceaux, and P. Millot, "The role of intervening variables in driver-ACC cooperation," *Int. J. Hum. Comput. Stud.*, vol. 66, no. 3, pp. 185–197, Mar. 2008, doi: 10.1016/j.ijhcs.2007.09.002.

[58] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe, "A Guide for Analysing Electrodermal Activity (EDA) and Skin Conductance Responses (SCRs) for

Psychological Experiments," Birmingham, UK, 2015. [Online]. Available: https://www.biopac.com/wp-content/uploads/EDA-SCR-Analysis.pdf.

[59] W. Boucsein *et al.*, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, 2012, doi: 10.1111/j.1469-8986.2012.01384.x.

[60] J. J. Braithwaite and D. G. Watson, "Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs)," 2015.

[61] R. W. Backs, J. Rohdy, and J. Barnard, "Cardiac control during dual-task performance of visual or auditory monitoring with visual-manual tracking," *Psychologia*, vol. 48, no. 2, pp. 66–83, Jun. 2005, doi: 10.2117/psysoc.2005.66.

[62] P. Zontone *et al.*, "Car Driver's Sympathetic Reaction Detection through Electrodermal Activity and Electrocardiogram Measurements," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 12, pp. 3413–3424, 2020, doi: 10.1109/TBME.2020.2987168.

[63] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Front. Public Heal.*, vol. 5, no. September, pp. 1–17, 2017, doi: 10.3389/fpubh.2017.00258.

[64] NTSB, "Collision Between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016, NTSB/HAR-17/02," Washington, DC, 2017.

[65] NTSB, "Rear-End Collision Between a Car Operating with Advanced Driver Assistance Systems and a Stationary Fire Truck, Culver City, California, January 22, 2018, NTSB/HAB-19/07," Washington, DC, 2019.

[66] NTSB, "Collision Between Car Operating with Partial Driving Automation and Truck-Tractor Semitrailer, Delray Beach, Florida, March 1, 2019, NTSB/HAB-20/01," Washington, DC, 2020.

[67] NTSB, "Collision Between a Sport Utility Vehicle Operating With Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018, NTSB/HAR-20/01," Washington, DC, 2020. [Online]. Available: https://data.ntsb.gov/Docket?ProjectID=96932.

[68] L. Dixon, "Autonowashing: The Greenwashing of Vehicle Automation," *Transportation Research Interdisciplinary Perspectives*, vol. 5. Elsevier Ltd, May 01, 2020, doi: 10.1016/j.trip.2020.100113.

[69] J. Perello-March, C. Burns, M. T. Elliott, and S. Birrell, "Integrating Trust in Automation into Driver State Monitoring Systems," in *Human Interaction and Emerging Technologies. IHIET 2019. Advances in Intelligent Systems and Computing*, 2020, vol. 1018, pp. 344–349, doi: 10.1007/978-3-030-25629-6.

[70] B. A. Jonah, "Sensation seeking and risky driving: A review and synthesis of the literature," *Accid. Anal. Prev.*, vol. 29, no. 5, pp. 651–665, 1997, doi: 10.1016/S0001-4575(97)00017-1.

[71] I. Montag and A. Comrey, "Internality and Externality as Correlates of Involvement in Fatal Driving Accidents," *J. Appl. Psychol.*, vol. 72, no. 3, pp. 339–343, 1987, doi: 10.1037/0021-9010.72.3.339.

[72] H. J. Foy, P. Runham, and P. Chapman, "Prefrontal cortex activation and young driver behaviour: A fNIRS study," *PLoS One*, vol. 11, no. 5, pp. 1–19, 2016, doi: 10.1371/journal.pone.0156512.

Jaume R. Perello-March is a PhD student in Human Factors within the Intelligent Vehicles group at WMG, University of Warwick, since 2018. His research is mainly focused in using psychophysiology and neuroimaging techniques for driver state monitoring, being particularly interested in trust in automated vehicles. He received his BA in Psychology in 2014 and MSc in Human Evolution and Cognition in 2016 from the University of the Balearic Islands, Spain. He has presented his work at international conferences and holds previous automotive research and industrial experience in leading European centres, such as the Swedish National Road Transport Research Institute (VTI) and Galician Automotive Technology Centre (CTAG).

Dr Christopher G. Burns joined WMG's Human Factors section in 2017 as a post-doctoral Research Fellow on the UK Autodrive Project, embedded with a research team at Jaguar Land Rover. Chris's involvement in UK Autodrive, consisted of designing and refining methodologies, then conducting and analysing practical experiments in passenger and pedestrian experiences of a prototype autonomous low-speed electric vehicle (L-SATS) operating in a controlled arena. These studies included user attitudes and intentions-to-use, trust formation, technology acceptance and internal/external human-machine interface evaluations, using both quantitative and qualitative methods. Chris holds a BA (Hons), a Masters by research (MRes) and a PhD in Psychology (University of Edinburgh), with a specialist background in quantitative research methods and statistics in individual differences with psychometric and psychophysiological methods. Previously, Chris has worked on projects involving driving simulation, sustained attention, mental workload, and emotional and attitudinal responses. Chris has also previously worked on ultrasonographic simulation and medical training in teaching contexts.

Dr Stewart A. Birrell is a Professor of Human Factors for Future Transport within the National Transport Design Centre (ntdc) at Coventry University. He received his PhD in Ergonomics from Loughborough University, UK in 2007, and first-class degree in Sport Science in 2002. Stewart has spent the previous 15 years working within the transportation sector within industry and academia, with expertise ranging from driver behaviour and distraction, multimodal warnings, user state monitoring and information requirements – all underpinned by the design of in-vehicle information systems, and their evaluation using driving simulators, virtual reality (VR) and field operational trials. Currently, he applies innovative Human Factors Engineering methodologies to enable real-world and virtual evaluation of user interaction with Connected and Autonomous Vehicle (CAV), Electric Vehicle (EV) and Urban Air Mobility (UAM) technologies

and services. Professor Birrell has over 100 journal and conference papers, book sections and articles published in his field to date, and is an Editor of the internationally renowned, Q1/4* journal IEEE Transactions on Intelligent Transportation Systems.

Dr Roger Woodman is an Assistant Professor at WMG, University of Warwick, where he leads the Human Factors research, within the Intelligent Vehicles group. He received his BSc in Computer Science from the University of Gloucestershire in 2006, MSc in Robotics from the University of the West of England in 2008, and PhD from the Bristol Robotics Laboratory in 2013. He has more than six years of industrial experience, working in leading manufacturing, defence, and software companies. He joined the University of Southampton in 2013 as a Research Fellow, specialising in biomedical imaging. In 2017, he joined WMG at the University of Warwick as a Research Fellow, investigating human factors of low-speed autonomous transport. Among his research interests, are shared mobility, provable AI, transport optimisation, micromobility, and last-mile logistics. He has several scientific papers published in the field of autonomous vehicles and robotics. He lectures in the field of human-technology interaction and is a Fellow of the HEA.

Dr Mark T. Elliott is an Associate Professor at the Institute of Digital Healthcare, WMG, University of Warwick. He completed his PhD at Aston University in 2007, developing intelligent systems to discriminate between different walking patterns. He subsequently spent a number of years as a Research Fellow within the Sensory Motor Neuroscience Laboratory at the University of Birmingham, modelling multisensory integration in the context of human movement coordination. Mark's core research focuses on measuring health, wellbeing and behaviour through data-driven approaches. This primarily involves analysing and modelling data from wearable and mobile devices that capture movement and physiological responses. Much of Dr Elliott's research is highly applied and involves collaborating with commercial and public-sector partners.

Fig. 1 3xD Driving Simulator

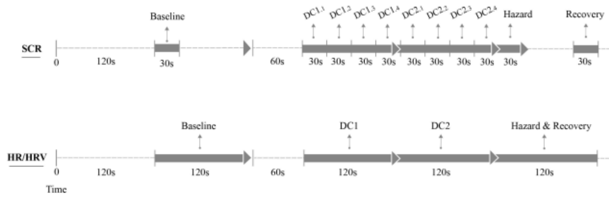

Fig. 2 Experimental timeline. The top timeline represents epochs extracted for SCR analyses (30 s), and the lower timeline epochs for HR/HRV analyses (120s).

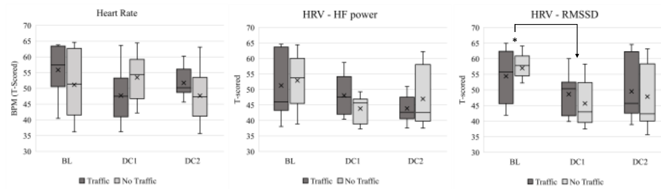

Fig. 3 T-scored cardiac features for the traffic factor on driving conditions. Heart Rate (left), HRV-HF (centre) and HRV-RMSSD (right).

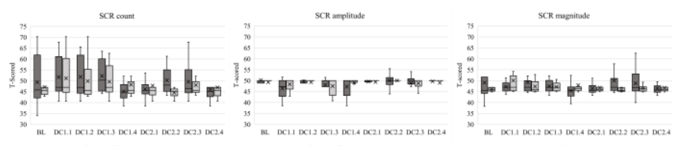

Fig. 4 T-scored Skin Conductance Response (SCR) features for the traffic factor on driving conditions. SCR count (left), amplitude (centre), and magnitude (right).

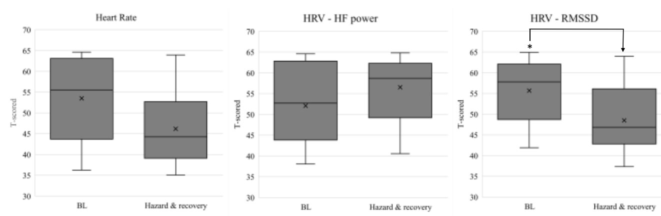

Fig. 5 T-scored cardiac features for the driving hazard event. Heart Rate (left), HRV-HF (centre) and HRV-RMSSD (right).

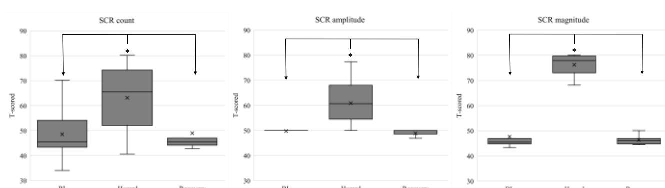

Fig. 6 T-scored Skin Conductance Response (SCR) features for the driving hazard event. SCR count (left), amplitude (centre), and magnitude (right).