Towards Multimodal Driver State Monitoring Systems for Highly Automated Driving

By

Jaume R. Perello-March

A thesis submitted in partial fulfilment of the requirements for the degree of:

Doctor of Philosophy in Engineering

University of Warwick, WMG

January 2022

Table of Contents

Tables and Illustrated Material

List of Tables

List of Figures

Acknowledgements

Four years ago, I was rushing on a PhD application late at night; the deadline was the day after. For some reason, the application never made it to the responsible academic team, so after a few days of not hearing back from them, I decided to email the lead contact straightaway.

Stewart kindly replied to my email and asked me to forward him my details to consider my application. A few weeks later, I was offered this PhD position. I am deeply grateful to you, Stewart, for giving me this opportunity and for staying involved even after assuming a new position in a different university. I would also like to express my most sincere gratitude to Mark; I could not expect more involvement and hard work from my second supervisor initially and now as a first supervisor. Thank you for setting the bar high, for your comments and guidance during the tough times, especially during the COVID-19 pandemic breakthrough in my last year of PhD. Chris, more than a colleague, you acted as a supervisor too – even though you had no formal responsibilities upon this thesis. Thanks for your perseverance, for teaching me to be rigorous, sceptical, and methodical. Thank you for those long discussions about the interpretations of the results, and thank you for teaching me to be level-headed when dealing with inevitable challenging setbacks. Essentially thank you for teaching me to be a scientist; it was all about it. Finally, thanks, Roger, for your supervision during the second half of this PhD, for helping in the set-up for data collection in both studies, enhancing the papers with your feedback, and for your advice in wrapping up this thesis.

I would also like to express my most sincere appreciation to my teammates, especially to Arun and my mentor in the distance, Nacho. Thank you both for taking the time to read the papers and bits of this thesis and for your advice and insightful comments, which have enhanced this work in many aspects. You are such an inspiration to me. My deepest gratitude is also to all those who have been next to me in the good times during this journey and supported me in the rough times; you have been my second family. You all know who you are.

Last but not least, none of this would have been ever possible without the unconditional support of my family, my partner, and my closest ones, so this one will be in our language:

Als meus pares, aquesta tesi no és més que un reflex de l'èxit en la vostra tasca com a pares. L'esforç i sacrificis que hem fet tots tres ara tenen sa seva recompensa. Gràcies per cuidar-me, per ser-hi present sempre, per deixar-me perseguir els meus somnis i per donar una mà desinteressadament sempre que ho he necessitat.

A ti Dani, que has sido mi apoyo incondicional estos últimos tres años, gracias por tu paciencia, por hacerme crecer y madurar. Gracias por todas esas agotadoras horas escuchándome hablar de fNIRS, del simulador o del ritmo cardiaco. Gracias por obligarme a salir de la cueva a respirar, por empujarme a seguir adelante y a levantarme cuando la motivación escaseaba, tú me has hecho este camino mucho más fácil.

Declaration and Inclusion of Material from a Prior Thesis

I hereby declare that this thesis is the result of my own work and has not been submitted for a degree at another university. Materials contained in this thesis which had been published are indicated in the next section below.

Inclusion of Published Work

This PhD project has contributed to knowledge with two Q1 journal publications and one international conference paper:

- Perello-March, J. R., Burns, C. G., Woodman, R., Elliott, M. T., & Birrell, S. A. (2021). Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios. *IEEE Transactions on Intelligent Transportation Systems*.
- Perello-March, J. R., Burns, C. G., Birrell, S. A., & Woodman, R., Elliott, M. T. (In press). Physiological Measures of Risk Perception in Highly Automated Driving. *IEEE Transactions on Intelligent Transportation Systems, Special Issue on: Technologies for Risk Mitigation and Support of Incapacitated Drivers*.
- Perello-March, J., Burns, C., Elliott, M., & Birrell, S. (2019, August). Integrating trust in automation into driver state monitoring systems. In *International Conference on Human Interaction and Emerging Technologies* (pp. 344-349). Springer, Cham.

Abstract

Real-time monitoring of drivers' functional states will soon become a required safety feature for commercially available vehicles with automated driving capability. Automated driving technology aims to mitigate human error from road transport with the progressive automatisation of specific driving tasks. However, while control of the driving task remains shared between humans and automated systems, the inclusion of this new technology is not exempt from other human factors-related challenges. Drivers' functional states are essentially a combination of psychological, emotional, and cognitive states, and they generate a constant activity footprint available for measurement through neural and peripheral physiology, among other measures. These factors can determine drivers' functional states and, thus, drivers' availability to safely perform control transitions between human and vehicle.

This doctoral project aims at investigating the potential of electrocardiogram (ECG), electrodermal activity (EDA) and functional near-infrared spectroscopy (fNIRS) as measures for a multimodal driver state monitoring (DSM) system for highly automated driving (i.e., SAE levels 3 and 4). While current DSM systems relying on gaze behaviour measures have proven valid and effective, several limitations and challenges could only be overcome using eye-tracking in tandem with physiological parameters. This thesis investigates whether ECG, EDA and fNIRS would be good candidates for such a purpose.

Two driving simulator studies were performed to measure mental workload, trust in automation, stress and perceived risk, all identified as modulators of drivers' functional states and that could eventually determine drivers' availability to take-over manual control. The main findings demonstrate that DSM systems should adopt multiple physiological measures to capture changes in functional states relevant for driver readiness. Future DSM systems will benefit from the knowledge generated by this research by applying machine learning methods to these measures for determining drivers' availability for optimal take-over performance.

Abbreviations

ADAS: Advanced Driving Assistance Systems

ANS: Autonomic Nervous System

DSM: Driver State Monitoring

ECG: Electrocardiogram

EDA: Electrodermal Activity

fNIRS: functional Near-Infrared Spectroscopy

HAD: Highly Automated Driving

HbO: Oxygenated haemoglobin

HbR: Deoxygenated haemoglobin

HbT: Total haemoglobin

HR: Heart Rate

HRV: Heart Rate Variability

NDRT: Non-Driving Related Task

PNS: Parasympathetic Nervous System

SA: Situation Awareness

SAE: Society of Automotive Engineers

SCR: Skin Conductance Response

SNS: Sympathetic Nervous System

TiA: Trust in Automation

TOR: Take Over Request

1 Introduction and motivation

1.1 Introduction

Worldwide social, political, and economic changes during the 20th and 21st centuries have brought about an increasing level of traffic and consequently an increased number of traffic accidents (ETSC, 2020; EuropeanComission, 2018). The increasing number of fatalities has raised concerns about enhancing vehicles' safety. Indeed, human error is still the main factor causing around 90% of traffic-related accidents (Singh, 2015). An attempt to mitigate this issue has motivated the industry to develop advanced driving assistance systems (ADAS) and the need to satisfy market demands, as sales of vehicles including ADAS evidence (Vitale et al., 2017). Such rapid development of ADAS has brought driving automation technology and related new terminology, ambiguities and great misinterpretations, which seem to exaggerate public expectations of current driving automation capabilities (Payre et al., 2020). For example, a global survey commissioned by Euro NCAP revealed that more than 70% of car drivers believed they could already purchase a car that drives itself (EuroNCAP, 2018). Such phenomenon has been defined as "autonowashing" and refers to the gap between public expectations of driving automation capabilities and the actual system capabilities and limitations (Dixon, 2020). Overestimated automated driving expectations are becoming a significant problem in recent years, especially now that partial driving automation (SAE Level 2) (SAE International, 2018) is being pervasively integrated into the automotive market and has already resulted in several fatal crashes in the US (NTSB, 2017, 2019b, 2020a, 2020b). Remarkably, the accident reports suggest these crashes had two critical aspects in common: all were the result of drivers' inattention due to overreliance, and all involved the same carmaker. Other collisions have also involved other automated driving technology providers (NTSB, 2019a; Waymo, 2018), but these were related to developmental or testing prototypes, not production vehicles, and accident reports indicate that were not due to operator's over-reliance, but automation complacency or under-reliance in automation. Manufacturers are responsible for setting an appropriate level of expectation of ability since there are still issues with this technology being at its early stage. It is irresponsible then to advertise a system requiring constant supervision and in which the driver holds total responsibility of the driving task (namely an SAE Level 2 ADAS) as, e.g., "Autopilot". Alternatively, even worse, an SAE Level 2 system that still requires the driver to be ready for taking over manual control when issued, advertised as, e.g., "Full Self Driving". The German government, for example, has already taken measures banning the use of such branding to prevent customers' misunderstanding and misuse (Reuters, 2016).

Marketing motivations apart, these crashes are just the tip of the iceberg, and "autonowashing" is more than a carmaker-related phenomenon like the EuroNCAP survey mentioned earlier indicates. Until reaching full driving automation capability (SAE Level 5), intermediate levels of driving automation will entail -to a more or lesser extent- human intervention or supervision upon the system. Thus, human factors will still modulate these interactions. For example, as with the collisions mentioned above, trust in automation is a critical factor for correct usage, engagement and acceptance of driving automation technology (Hoff & Bashir, 2015; Lee & See, 2004), but other factors such as driver availability, motion sickness, mode confusion, mental workload or situation awareness will also play a determinant role in the human-driving automation interaction (Martens & Van Den Beukel, 2013; Navarro, 2018; Smyth et al., 2019).

To tackle this, US road transport administrations will soon require that all new passenger vehicles with SAE Level 2 are equipped with a driver state monitoring (DSM) system that will minimise driver disengagement, prevent automation complacency, and account for foreseeable misuse of the automation (NHTSA, 2022). Whilst Europe does not have a specific policy yet, the European Council announced that by mid-2022 all new motor vehicles put on the EU market would have to be equipped with advanced safety features, including DSM systems for inattention detection (EuropeanCouncil, 2019). Moreover, from 2020, Euro NCAP has begun requiring driver monitoring systems for any new on-road vehicle to achieve the highest safety rating (EuroNCAP, 2017). Aligned with this strategy, since 2019, the European Commission has been working with automotive partners in the development of legislation for future vehicle safety measures (TRL, 2020), including driver monitoring for:

- Driver drowsiness and attention monitoring and warning,
- Advanced distraction recognition, and
- Driver readiness monitoring for automated driving

Such driver monitoring systems are currently mainly camera-based technology for gaze behaviour. As will be later introduced in section 2.3.1, such systems are adequate for detecting, e.g., fatigue, drowsiness or distractions with contemporary driving technology (i.e., up to SAE L2) (Dong et al., 2011). However, eye-tracking systems have certain inherent drawbacks (e.g., static cameras, require drivers facing forward, certain lighting conditions) that will be magnified with the inclusion of future HAD technology (see section 2.3.2). The detection of human factors relevant for take-over control transitions during HAD, such as mental workload, stress, trust and risk perception, may also challenge current eye-tracking system capabilities. Driver readiness monitoring technology for HAD will require a multimodal approach that combines several sources of driver data to determine driver readiness for taking over manual control of the driving task. Following this novel approach, physiology has been proposed as a potential candidate to complement eye-tracking technology (Lohani et al., 2019). Physiology parameters derived from cardiac, electrodermal or brain activity have extensively been used in driver behaviour research to assess several functional states such as attention levels, drowsiness, fatigue or mental workload. They can be measured with research-grade ECG, EDA and fNIRS devices, respectively. Particularly in the HAD context, previous work has used physiology-based wearable metrics to measure drivers' discomfort throughout several complex and uncertain traffic situations in a series of driving simulator studies (Beggiato et al., 2018, 2019). These studies show the potential utility of integrating physiology in DSM systems, but overall, the usage of consumer-grade devices has limited their results. Furthermore, the effects on drivers' physiology of several human factor-related constructs such as mental workload, stress, trust and risk perception during HAD still remains unclear.

Physiology has the benefit of being a ubiquitous source of data directly linked to emotional and cognitive functional processes, and thus, could be an indicator of driver functional states. Regardless of the advances in non-invasive physiology measuring technology, the applications of this technology for driver readiness monitoring remain unclear due to several limitations related to its effectiveness for detecting certain factors (e.g., trust, mental workload, or risk perception) or the possibilities for an inconspicuous and unobtrusive in-vehicle integration.

With this regard, the European Commission has funded the "AutoConduct Project", focused on developing multimodal DSM systems for levels 3-4 driving automation (EuropeanCommission, 2019). Initial outcomes from this project highlight the importance of incorporating these systems to mitigate incorrect usage of driving automation technology or assist drivers during take-over requests (Coeugnet et al., 2021; Hidalgo-Muñoz et al., 2019). It seems evident then that further research in this field is required during the upcoming years if driving automation is to be deployed throughout European roads.

This thesis will focus on the use of physiology for DSM to address the need to shed light on several research gaps and contribute to the development of multimodal DSM systems for driver readiness detection for HAD.

1.2 Aims and Objectives of the thesis

This thesis investigates whether neuro and psychophysiological metrics could be used as real-time indicators of human factors-related parameters, including mental workload, stress, trust and risk perception, which can affect driver availability to take over during the usage of a highly automated driving (HAD). As stated in the previous section, DSM systems will be required for production vehicles shortly and research has suggested that physiology-based technology has the potential of measuring driver functional states in real-time. However, it remains unclear whether during HAD, driver functional states can be affected by different levels of mental workload, stress, trust and risk perception, and hence, whether physiological indices can detect it. The novelty of the present work is to investigate with research-grade devices the usage of three physiological measures selected from the literature review as possible candidates for a multimodal DSM system. In addition, it is the first time these three measures are used in tandem to assess mental workload, trust, stress and risk perception derived from HAD in real-time. Likewise, this work aims to prove future work with a solid methodological grounding to develop multimodal DSM systems using real-time objective measurements. The aim will be achieved by:

- Emulating realistic driving conditions using an immersive driving simulator whilst maintaining experimental control.
- Generating simulated driving situations that elicit the target driver states.
- Assessing drivers' physiological indicators continuously during the driving scenarios.
- Comparing physiological indicators against self-reports to validate the findings.
- Providing a methodological ground for future work in developing multimodal DSM systems.

Research questions:

- Can electrocardiogram (ECG) measures be used as indicators of factors modulating driver availability state?

- Can electrodermal activity (EDA) measures be used as indicators of factors modulating driver availability state?
- Can functional near-infrared spectroscopy (fNIRS) measures be used as indicators of factors modulating driver availability state?
- Considerations for integrating physiology measures into a multimodal driver state monitoring (DSM) system

1.3 Structure of the thesis

The present work will measure the physiological indicators of those critical factors modulating drivers' availability state during Highly Automated Driving (SAE Levels 3 and 4). An overview of the structure of this thesis is as follows:

- The first chapter will introduce the motivation underpinning this PhD project, its research goals, and its core structure.
- The second chapter will review those most relevant human factors influencing driver availability state during High Driving Automation and their implications on physiology and take-over behaviours. Moreover, a parallel review will address the state of the art of current driver state monitoring systems based on driving performance and eye-tracking measures and potential alternatives for future multimodal monitoring systems for automated driving according to existing literature.
- The third chapter will describe the methods, materials, devices and facilities used in the studies conducted for this PhD project. General data extraction procedures, data pre-processing and analysis methods will also be described.
- The fourth chapter will present and discuss the findings from the first driving simulator study. Overall, this study has focused on measuring neuro and psychophysiological indicators of drivers' mental workload, stress and trust in automation derived from a series of HAD scenarios.
- The fifth chapter will present and discuss the second driving simulator study findings. This study has explored drivers' perceived risk through neuro and psychophysiological indicators during HAD conditions varying in risk levels. Measuring drivers' risk perception is relevant towards developing effective planned

take-over strategies since it affects situation awareness and trust in automation, and thus, drivers’ availability.

- The sixth and final chapter will be a general discussion considering all findings and current evidence from the literature. This chapter will also provide directions for future human factors studies in this research scope, and guidelines for developing driver state monitoring systems.

2 Literature Review

This literature review will present an overview of the state of art concerning driving automation technology and human intervention to resume control. The scope will then be narrowed down to those human factors that play a pivotal role in the driver availability for resuming control. Finally, a parallel review will discuss current DSM systems and their current usage and possible methods for developing the next generation of DSM systems to detect driver availability state.

2.1 Background

The inclusion of automated driving technology implies a progressive control task substitution from partial (SAE L2) to total (SAE L5, see Figure 1 for details), changing drivers' task allocation, and thus, switching attentional allocation to non-driving related tasks (NDRTs) (Carsten et al., 2012; Naujoks et al., 2016). Drivers can progressively engage longer in other tasks not related to driving and stay less aware of the ongoing driving situation with increasing driving automation levels (Endsley, 2017). Whilst for partial automation (SAE L2), drivers must constantly be supervising the automated driving system and ready to take over control instantly, for conditional automation (SAE L3), drivers may momentarily disengage from driving but need to remain alert to take-over requests (TOR) from the vehicle and ready to regain manual control in a short period when the system requires so. Finally, high driving automation (SAE L4) will be the last level where drivers may be required at some point. At this level, the automated driving system will perform the entire driving task within predefined use cases (e.g., city driving, highway, parking) and even assume the driving task fallback. That is, when reaching the use case boundaries, detecting a system failure or not being able to deal with a particular situation, the vehicle will perform a minimal risk manoeuvre and stop the vehicle safely if the driver is not able or ready to take over manual control (SAE International, 2018). Although this would not be the desirable scenario, and the automated driving system would probably issue a monitoring request (Gold et al., 2013; Lu et al., 2019) to get the driver back *on the loop* (Merat et al., 2019), followed by a TOR to engage manual control safely.

Regarding the driving task control loop, Merat and colleagues have proposed the following definitions for the different driver functional states (Merat et al., 2019):

- **In the loop**: In physical control of the vehicle and monitoring the driving situation.
- **On the loop**: Not in physical control of the vehicle but monitoring the driving situation.
- **Out of the loop**: Not in physical control of the vehicle, and not monitoring the driving situation, OR in physical control of the vehicle but not monitoring the driving situation.

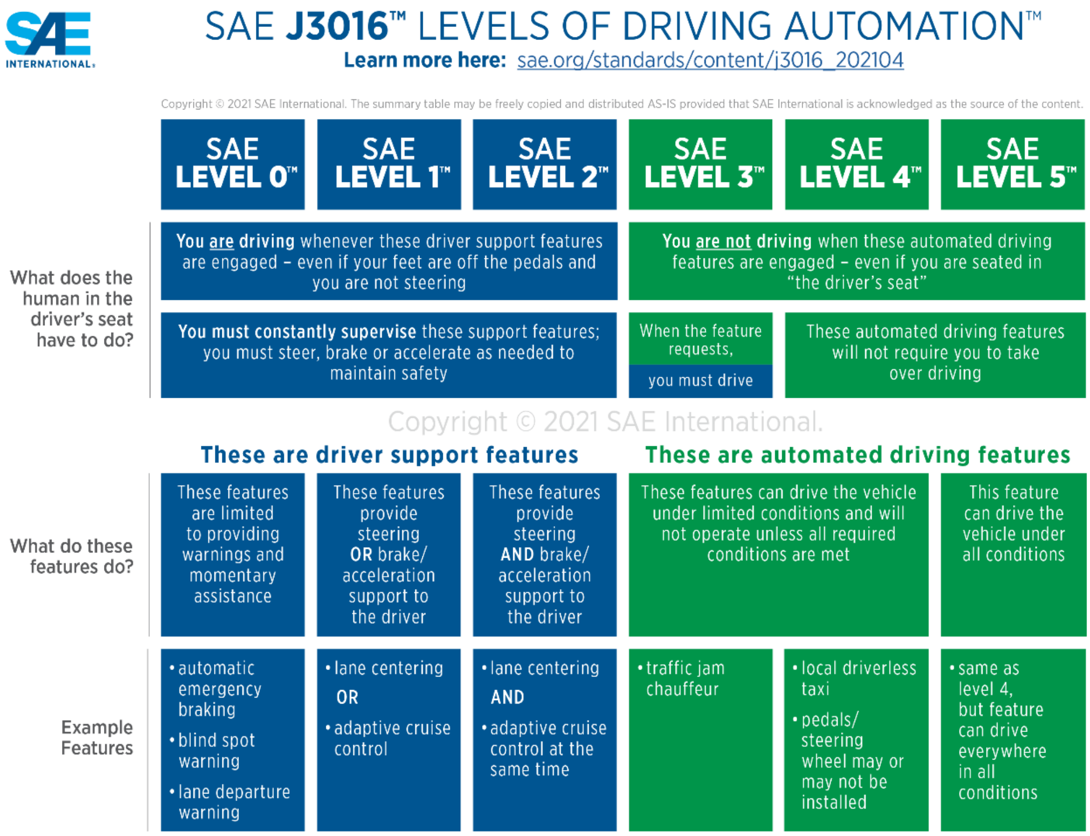


Figure 1 SAE International taxonomy for driving automation levels (SAE International, 2018).

Therefore, the SAE L4 system must ensure the driver is available for manual driving before taking over control. Marberger and colleagues proposed that DSM systems should supervise the whole transition process from (1) the monitoring request, ensuring drivers achieve adequate attentional levels for (2) an optimal take-over performance, and then (3) monitoring whether after the take-over such attentional level is sustained (Marberger et al., 2018). Attention, activation and alertness levels are strongly related to arousal (Kahneman, 1973). Arousal results from several cognitive and emotional processes controlled by the autonomic nervous system (ANS). The ANS is responsible for either (1) rest-and-digest states through the activation of the parasympathetic nervous system (PNS) or (2) fight-or-flight reactions to stressors by activating the sympathetic nervous system (SNS).

An inverted U-curve describing a non-linear relationship between task performance and arousal is often referred to as the Yerkes-Dodson law (Yerkes & Dodson, 1908). Although the original work by Yerkes-Dodson described the relationship between stimuli strength and habit-formation for different tasks varying in difficultness instead of a relationship between arousal and task performance; thus, the association of the arousal-performance relationship with their work has been criticised (Teigen, 1994). This relationship between arousal levels and performance, however, describes an essential phenomenon for understanding the concept of driver availability state: task performance improves with moderate arousal levels up to a certain point (Collet et al., 2005; Miller et al., 2015), but after that, arousal increases to the detriment of performance (Durantin et al., 2014). Adequate attentional levels for optimal performance within this spectrum could therefore indicate drivers' availability for take-over.

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 (Marberger et al., 2018; Martens & Van Den Beukel, 2013); and will raise importance in SAE L3 and L4, where drivers are allowed to be out-of-the-loop, are likely to engage in NDRTs, and thus, will be impaired for resuming manual control straight away most of the times.

Drivers' availability will be partly conditioned by the amount of time immersed in the NDRT performed before the take-over and the task itself. For those use cases involving long distances and thus, prolonged periods with HAD engaged, drivers' situation awareness (SA) will be enormously diminished (De Winter et al., 2014; Endsley, 2017). Sleeping, videoconferencing, playing videogames, reading, or even eating may be a bit more commonplace for some people while HAD is engaged; and thus, entailing a wide range of stress, mental workload, alertness or attention allocation levels, with all of these modulating arousal levels derived from ANS activity. For drivers to re-engage driving manually, they need to transit from being *out-of-the-loop* to *on-the-loop* (Merat et al., 2019). Hence, detecting when the driver is capable and ready to take over manual control among various driving states will be challenging but still needed because drivers will be required to engage manual control at any point.

Other than that, availability to drive may not necessarily be exclusively tied to NDRTs. Section 1.1 has provided examples of how user attitudes and credibility expectations towards the automated driving system can be modulated, mitigating risk perception and leading to over-reliance on automation and complacency. Both behaviours are interrelated and motivated by overtrust in automation (Parasuraman & Manzey, 2010). Complacency is a psychological state that has been associated with deficient SA (Endsley, 1995b, 2017), and thus impaired vigilance (Bailey & Scerbo, 2007) and increased reaction times to system failures (Payre et al., 2016). Overreliance is usually the resultant behaviour of overtrust in automation (Lee & See, 2004) and is characterised by abuse or misuse of the system (Parasuraman & Riley, 1997) that can eventually derive in a complacency state (Parasuraman & Manzey, 2010).

Driver availability state then can be described as a construct dependant on the interaction between mental workload, trust in automation and situation awareness. Such an assumption was depicted in the Human–Autonomy System Oversight (HASO) model (Endsley, 2017), according to which take-over performance will be dependent on SA and workload levels and moderated by trust in automation:

> "The operator must have sufficient situation awareness to realize that the present situation is outside of the bounds of automation capabilities, or that the automation is performing incorrectly for the present situation, in order to decide that an intervention is needed. Further, the operator must have sufficient time and resources to be able to make the intervention. […], as moderated through operator trust."

Further experimental evidence supporting this model was found in a longitudinal driving simulator study using conditional driving automation (SAE L3). As trust increased, the likelihood to disengage did so, resulting in a decrease in situational awareness (Large et al., 2019).

These constructs are inherently mediated by either cognitive (e.g. attentional, motivational or executive) and affective (i.e., emotional) components (Endsley, 1995b; Lee & See, 2004; Wickens, 2008), hence modulating the ANS modes of activation and consequently producing a visible psychophysiological activity footprint. With this in mind, in the next section, we will look into these concepts and their relation to driver availability state in the context of HAD.

2.2 Human factors modulating driver availability state in HAD

2.2.1 Trust in automation

A determinant factor for automation acceptance, reliance and usage is the level of trust in a system (Lee & Moray, 1994; Sheridan, 1989). Trust in automation (TiA) has been defined as:

> "a history-dependent attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Khastgir et al., 2018).

Within the automated driving domain, TiA becomes extremely important because all efforts put in developing automated driving systems would be wasted if users do not trust them, and hence, do not use them as intended or do not use them at all. The main problem with TiA is that it is a multidimensional psychological construct that may not always result in observable behaviours (i.e., reliance). It is multifaceted (e.g. dispositional, situational or learned) and changes over time with experience (Hoff & Bashir, 2015; Lee & See, 2004). As a psychological construct (i.e., an attitude), TiA is considered the result of a complex decision-making process influenced by both cognitive and emotional processes (Drnec et al., 2016) (see Figure 2 for an overview).

The extent of the contribution of each, along with time availability and expertise during the decision-making process, will determine the trustworthiness of the automated system in a particular situation (Lee & See, 2004). For example, experts would determine with a calculated judgement their degree of TiA upon the systems' capabilities and limitations to deal with a particular situation (i.e., decision-making led by an analytic process). A decision-making process led by heuristics or expectations based on previous experiences with similar systems would be classified as an analogic process. Finally, quick judgements based essentially on feelings or emotions rather than any calculated reasoning would be an affective decision-making process (Lee & See, 2004). The three dimensions of TiA can also be extracted from this taxonomy. The former (analytic process) would be an example of calibrated or appropriate TiA (Khastgir et al., 2018). Analogic processes based on heuristics or expectations could lead to overtrust in automation, as reflected with the accidents discussed in the introduction (Lee & See, 2004; Parasuraman & Riley, 1997). In contrast, the latter (i.e., the affective process) could be a case leading to distrust and possibly disuse (Lee & See, 2004; Parasuraman & Riley, 1997).

The TiA decision-making process is also affected by the current cognitive resources available (i.e. mental workload). For example, a situation of mental overload will hamper the performance of the cognitively demanding judgements required for the analytic TiA process (Hoff & Bashir, 2015). Stress derived from the driving situation may also affect attentional capacity, induce mental overload and thus, impair the decision making (Kahneman, 1973). Hence, risk perception has been depicted as a critical factor across the TiA literature. As will be later discussed in section 5.2.2, perceived risk acts as a nexus between SA and TiA (Endsley, 2017), and will affect resultant monitoring behaviours such as overreliance and complacency (Parasuraman & Riley, 1997).

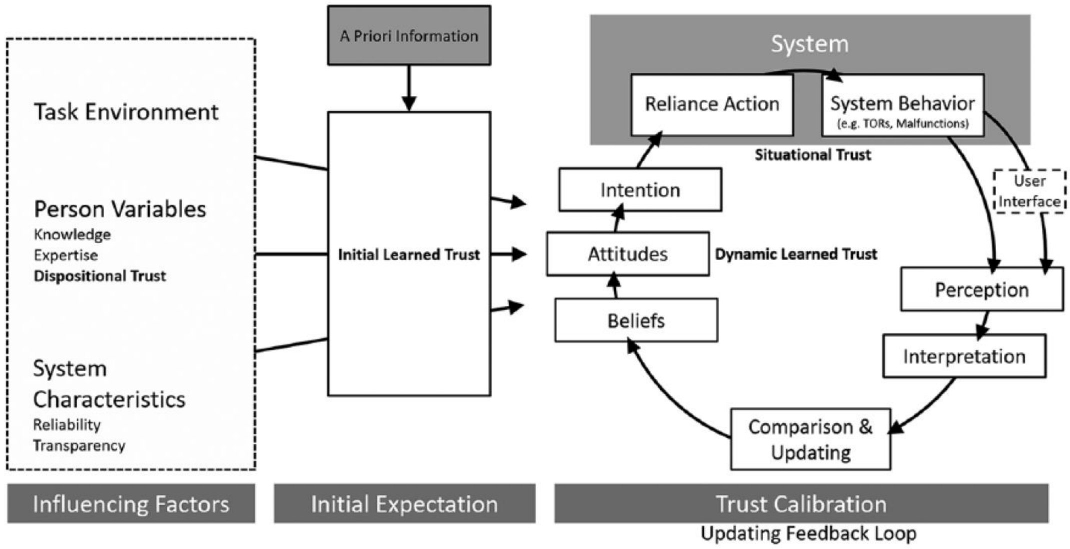


Figure 2 The process of TiA formation extracted from Kraus, Scholz, Stiegemeier, & Baumann (2020).

Evidence from neuroscience in experimental economics and neuroeconomics support these associations. These studies have explored reciprocal social exchanges (Krueger et al., 2007), seller profile's trustworthiness (Dimoka, 2010), and trustworthiness evaluations of online offers (Hubert et al., 2018). Overall, the main findings can be summarised as:

- Trust and distrust are distinct yet related constructs regarding the nature and type of neural responses involved (Dimoka, 2010).
- Trust builds slowly over time through careful deliberation, whilst distrust is quick and episodic, based on emotional cues (Dimoka, 2010).
- The neural mechanisms of trust and distrust involve emotional and cognitive structures to a different extent. Trust is more dependent on intentional, calculated decision-making, whereas distrust depends on autonomic emotional processes (Dimoka, 2010; Hubert et al., 2018; Krueger et al., 2007).

To our knowledge, little work has been undertaken to prove the transferability of these findings to the automated driving domain. Interpersonal trust and trust in automation share commonalities, but they are, in fact, different constructs (Madhavan & Wiegmann, 2007). For example, higher interpersonal trust involves social engagement, whereas higher TiA implies disengagement from the driving tasks, and conversely, low trust increases engagement with the driving task (Seet et al., 2020). Even though the resultant behaviour may be different, the neural mechanisms managing trust-based decisions would probably share commonalities, because in evolutionary terms, elapsed time since the first interactions between humans and automation have not been sufficient to modify the responsible brain structures.

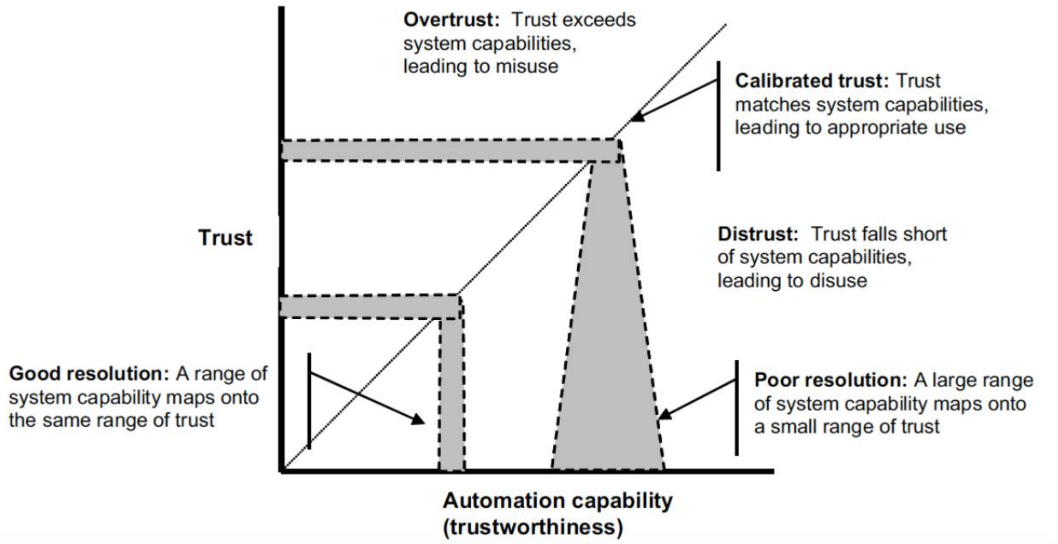


Figure 3 Relationship between TiA levels and automation capability. Extracted from Holthausen (2020).

Therefore, in light of the findings from experimental economics and neuroeconomics, we could expect the process of analytic TiA calibration to recruit cognitive resources from the brain areas responsible for intentional, calculated decision-making and attentional allocation towards the automated driving system performance as in de Visser et al. (2018). Calibration here refers to the level of trust which matches the actual system capabilities and limitations, fluctuating over time along with the accumulative knowledge and experience with the system (see Figure 3) (Khastgir et al., 2018; Kraus et al., 2020). Overtrust is often related to inattention and complacency, primarily observable through gaze allocation and reduced monitoring behaviours (Bailey & Scerbo, 2007; Walker et al., 2019). Although since it involves a lack of risk perception and SA (Bailey & Scerbo, 2007; Parasuraman & Manzey, 2010), it could be related to a *rest-and-digest* psychophysiological state, thus potentially measurable with cardiac parasympathetic activation. On the contrary, distrust resultant from affective processes should prepare the autonomic nervous system towards a *fight-or-flight* reaction, observable primarily with quick and episodic skin conductance changes in sympathetic activity as found in (Morris et al., 2017; Walker et al., 2019).

However, although the evidence suggests that TiA is a dynamic psychological construct (i.e., an attitude) rooted in behaviour and physiology, it is mainly measured exclusively through subjective tools (Holthausen, 2020; Jian et al., 2000; Körber, 2018). Only a few studies have attempted to measure TiA from behaviour (e.g., Bailey & Scerbo, 2007; Walker et al., 2019) and physiology (e.g., de Visser et al., 2018; Morris et al., 2017; Walker et al., 2019). Particularly from what concerns here, studies using physiological indicators have mainly relied on neurophysiological techniques such as EEG (Ajenaghughrure et al., 2020; de Visser et al., 2018; Seet et al., 2020). Aside from brain metabolic data, gaze behaviour, EDA and ECG are the physiological measures most frequently used (Ajenaghughrure et al., 2020; Morris et al., 2017; Walker et al., 2019).

2.2.2 Mental workload

Mental workload is a multidimensional construct that refers to the cognitive effort, and attention allocation required to meet concurrent task(s) demands (Wickens, 2008); and which can be modulated by several factors in the driving context (De Waard, 1996):

- Driver state factors such as monotony, fatigue, sedative drugs, or alcohol consumption.
- Driver trait factors such as experience, age, or strategy.
- Environmental factors including road type, traffic, vehicle ergonomics, automation level or feedback.

The primary feature for understanding mental workload is that cognitive resources are limited (Kahneman, 1973). Henceforth, the theory of multiple resources (Wickens, 2008) states the existence of four resource pools (or dimensions), consisting of three components related to the demand, resource overlap, and allocation policy. By definition, mental workload essentially relates to the first of these, characterising the demand imposed by tasks on the human's limited mental resources. The four dimensions of resources are classified by dichotomies of information processing (Figure 4):

- Stages of processing: perceptual and cognitive activity (e.g., working memory) versus motor and action-oriented activities.
- Codes of processing: spatial versus verbal information.
- Modalities: auditory versus visual processes.
- Visual channels: focal versus ambient vision.

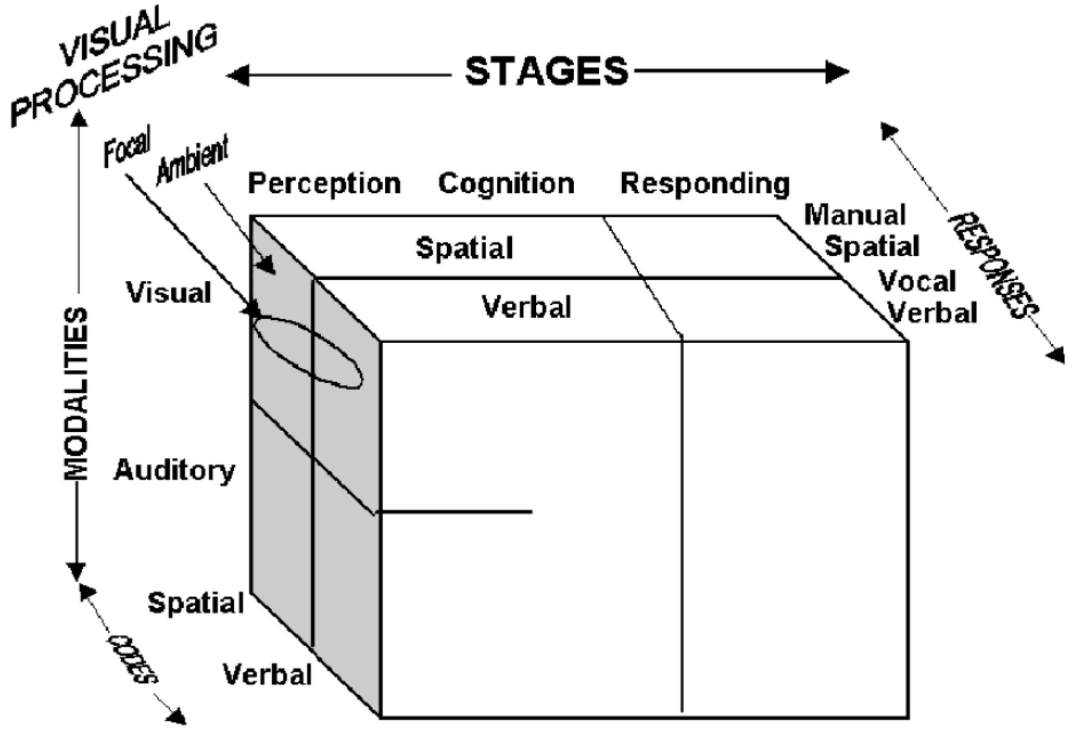


Figure 4 The four dimensions of the multiple resources model. Extracted from Wickens (2008).

When task demand, resource overlap or allocation policy exceed the resources available in any particular dimension, task performance will break down (Wickens, 2008). Phone calls occurring during manual driving are probably among the most studied examples of the detrimental effect on task performance due to cognitive overload. A phone conversation requires perceptual-cognitive processing of auditory and verbal information, which compete with the visual-spatial information processing required for the motor and action-oriented resources required for the driving task. The result is poor performance with increased driving task demands while immersed in the conversation or otherwise interrupting the conversation to meet driving demands (De Waard, 1996; Kircher, 2007; Strayer et al., 2011). Compensatory behaviours such as reducing the speed attempt to mitigate such overload effects (De Waard, 1996).

HAD entails a drastic decrease in drivers' mental workload compared to manual driving (De Winter et al., 2014). Although as stated earlier, freeing drivers from the driving task implies they will potentially engage in other activities (Carsten et al., 2012; Naujoks et al., 2016), which paradoxically will encompass an increase in mental workload to match these other tasks demands (e.g. reading, videoconferencing, video-gaming). It has been argued that engagement in such NDRTs could be beneficial to keep drivers' alertness levels at an optimal level for take-over performance (Miller et al., 2015) and even to inform them regarding the current time budget for take-over within task transitions (Wintersberger et al., 2018). Although SA would be strongly impaired and therefore, could lead to other situations prone to cognitive overload under HAD. For example, a time-constrained decision-making process in an unexpected situation, would lead to automation complacency (Bailey & Scerbo, 2007). This decision-making would imply an active comparison of possible foreseen scenarios resulting in a mental workload increase, especially if risk perception and stress are high. In addition, given that higher automation levels are often complex systems and a complete understanding of system capabilities and limitations might be unrealistic, system transparency through the user interface design could modulate users' mental workload in such situations (Hoff & Bashir, 2015; Kraus et al., 2020; Oliveira et al., 2020).

Mental workload fluctuations and the theory of multiple resources have a neurophysiological basis (Wickens, 2008). As its definition suggests, the mental workload is essentially an increase of neural activity to meet the requirements of concurrent tasks and thus, also produces a psychophysiological footprint. Like muscular activity, increased neural activity due to mental workload requires more blood oxygenation and the withdrawal of deoxygenated blood, which means increasing cardiac activity. The ANS is responsible for that process through the activation/deactivation of sympathetic and parasympathetic branches (De Waard, 1996). Within the scope of human factors in the driving context, functional Near-Infrared Spectroscopy (fNIRS) is a widely used neurophysiology technique for measuring mental workload through changes in blood oxygenation in naturalistic experimental setups (Durantin et al., 2014; Foy & Chapman, 2018; Lohani et al., 2019; Sibi et al., 2016, 2017; Unni et al., 2017). As the reader could foresee, measures of cardiac activity extracted from electrocardiogram (ECG) are also often used for measuring mental demands in the driving context, either in isolation on combined with other psychophysiological indicators (Backs et al., 2003; De Waard, 1996; Lenneman & Backs, 2009; Lohani et al., 2019; Paxion et al., 2014; Verwey & Veltman, 1996). Besides fNIRS, other neurophysiological techniques like an electroencephalogram (EEG) are also standard and established indicators of mental workload in the driving context (De Waard, 1996; Lohani et al., 2019; Paxion et al., 2014; Solís-Marcos & Kircher, 2018). Finally, the measurement of pupil size and reactivity, and gaze-behaviour metrics such as blink rate, duration and latency via eye-tracking are other measures indicative of cognitive demands (De Waard, 1996; Kahneman, 1973; Lohani et al., 2019). However, like other arousal-related measures (e.g. electro-dermal activity), the pupil diameter and gaze behaviour cannot be considered a diagnostic measure of mental workload as they can be strongly affected by other factors, e.g., changes in ambient light or air quality (De Waard, 1996).

2.2.3 Situation awareness

Situation awareness (SA) is (1) the perception of the elements in the environment within a volume of time and space, (2) the comprehension of their meaning, and (3) the projection of their status in the near future (Endsley, 1995b). These three dimensions are modulated by several properties inherent to the human brain, such as attention, perception, working memory, long-term memory, heuristics, motivation, task complexity, workload or stress (Endsley, 1995b). Therefore, it is the individual's state of knowledge about a dynamic environment resulting from a combination of top-down and bottom-up processes, affecting alertness levels, decision-making, and task performance (Figure 5).

As earlier discussed in sections 2.1, 2.2.1 and 2.2.2, SA interplays with mental workload and TiA bidirectionally. For example releasing the driver from the driving task can improve SA, thus allowing the allocation of mental resources used for driving towards the elements in the environment (De Winter et al., 2014). However, this assumption must be cautiously considered since humans are poor supervisors with limited sustained attention capabilities (Kahneman, 1973). Reasonably, it is no surprise then, that SA is greater during manual driving than while supervising partial/conditional driving automation levels (SAE Levels 2-3) (Jamson et al., 2013; Merat et al., 2012). It could either be due to driving automation levels that require sustained supervision of the driving task without physical input control (SAE Levels 2-3), can elicit even greater attentional effort and mental workload than manual driving (Sibi et al., 2017; Young et al., 2007), impairing appropriate SA. Alternatively, in realistic conditions, drivers tend to engage more in NDRTs as the driving automation level increases (Carsten et al., 2012; Naujoks et al., 2016), which consequently leads to a loss of SA (Endsley, 2017). This loss of SA while automated driving is engaged can notably jeopardise drivers' take-over performance to unexpected hazardous situations or even for planned transitions

(Endsley, 2017). Particularly when poor SA leads to low perceived risk and consequently complacency or overreliance in automation (Endsley, 2017; Parasuraman & Riley, 1997). Good SA is essential for undertaking safe take-over transitions, thus, detecting levels of SA will be crucial for determining the driver availability state in HAD.

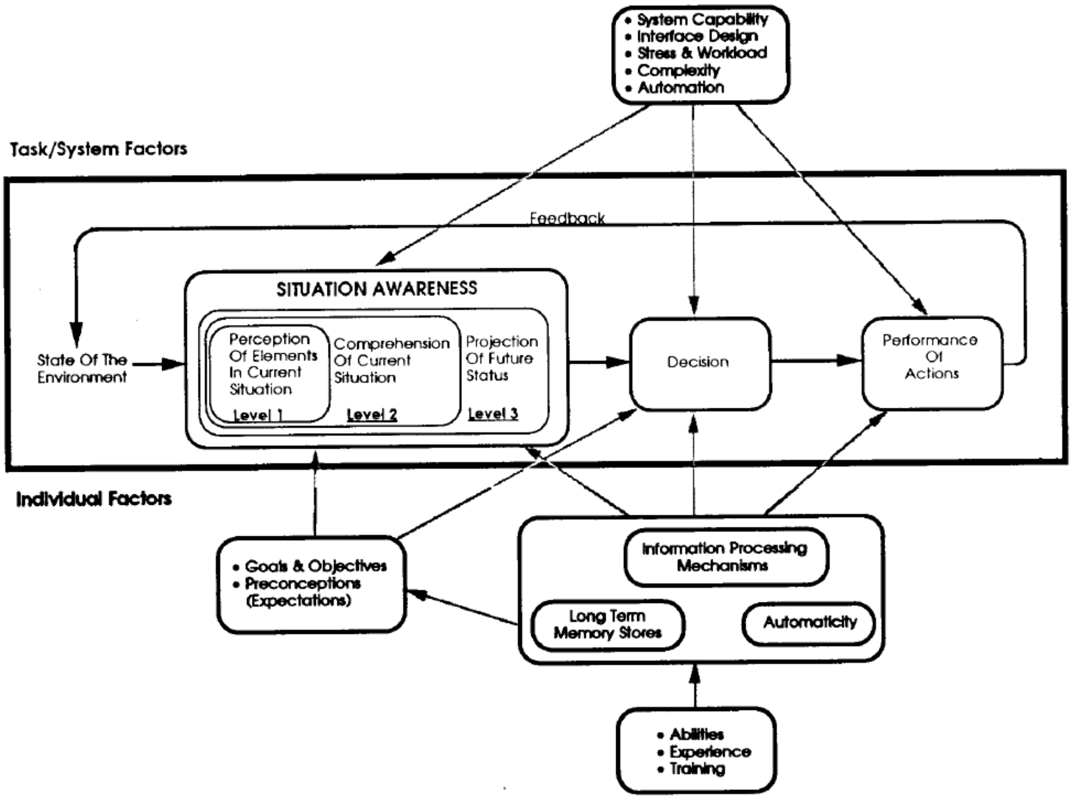


Figure 5 The model of situation awareness in dynamic decision making, extracted from Endsley (1995b).

Because SA has strong ties with perception, attention allocation, decision-making and memory, several psychophysiological indicators can be used to infer SA fluctuations to achieve driver availability to take-over control:

- **Gaze behaviour** can provide multiple information regarding drivers' situation awareness and availability. It is often used to detect whether drivers are attentive or drowsy (Dong et al., 2011), but also to detect when drivers are actively seeking information in their surroundings or the road ahead (e.g. checking side mirrors or navigation system) (De Winter et al., 2014; Dong et al., 2011). Pupilometry can also provide real-time information regarding attention allocation on salient stimuli, as well as current mental workload levels (Lohani et al., 2019). Hence, eye-tracking metrics would be a potential tool for measuring driver availability during the state transition from out-of-the-loop to in-the-loop. Although described later in section 2.3.2, eye-tracking technology during HAD faces some limitations that may require its combination with psychophysiological or neurophysiological tools to overcome.

- **Arousal levels** will also be crucial to inform the current out-of-the-loop state, which will depend on the nature of the ongoing NDRT and each individual. As well as to 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 (Marberger et al., 2018). Using attentional resources for SA increases cognitive workload and produces changes in arousal in stress levels (Endsley, 1995b). Hence, cardiac metrics (i.e., HR and HRV) are also an indirect but reliable indicator of driver states related to SA (De Winter et al., 2014; Lohani et al., 2019). Skin conductance metrics could be an excellent complement to cardiac measures because of their sensitiveness to wakefulness/sleepiness, workload and stressful events (De Waard, 1996; Healey & Picard, 2005; Lohani et al., 2019; Paxion et al., 2014). For example, state transition strategies may vary significantly from out-of-the-loop drivers with a vagal predominance (e.g. sleeping), to those with sympathetic dominance (e.g., video gaming) where the former may take longer to reach driver availability than the latter (Miller et al., 2015).
- **Brain activity** directly indicates cognitive states related to SA, such as attention or mental workload in the driving context (Bracken et al., 2021). It is usually measured through a neuro-electrical activity using electroencephalogram (EEG) (Bracken et al., 2021; Kästle et al., 2021), or blood oxygenation changes with functional near-infrared spectroscopy (fNIRS) (Bracken et al., 2021; Lohani et al., 2019). Both measures have been used for discriminating among variations in neurophysiological arousal related to mental workload (Borghini et al., 2014; Sibi et al., 2016; Solís-Marcos & Kircher, 2018; Unni et al., 2017), fatigue (Borghini et al., 2014; Li et al., 2009; Lin et al., 2020), alertness versus drowsiness (Chen et al., 2015), and sleep (Wörle et al., 2019). Both EEG and fNIRS are good indicators of under-arousal and over-arousal states, for which are a potential tool for measuring driver availability to take control, at least in experimental contexts.

Overall, this section has discussed how trust in automation, mental workload and situation awareness can modulate drivers' physiological state in the context of HAD, hence pinpointing those indicators that may be used for determining driver availability to take over in real-time with objective metrics. The following section will describe how some of these indicators are already being used in commercially available vehicles, their current limitations for their integration into future vehicles equipped with HAD technology and propose how these limitations could be surpassed with the combination with other indicators.

2.3 Driver state monitoring systems

2.3.1 Current driver state monitoring systems

Impaired mental or physical states make drivers prone to errors and hazardous behaviours such as distractions or drowsiness, a common cause of human error-related road accidents. Given that drivers' state is constantly changing (i.e., people get sleepy, bored, anxious or annoyed), real-time assessment of their state has been found helpful for preventing and mitigating such hazardous driving behaviours (Begum, 2013; Dong et al., 2011; Kang, 2013). The concept of a driver state monitoring (DSM) system was first introduced in the 90s as part of the DETER project to detect driver impairment mainly through behavioural and driving performance cues (i.e., lateral position handling, steering wheel handling, speed management, headway control and pedal use) (Brookhuis, 1993; Fairclough et al., 1993). DSM systems are already available in some current production vehicles and will be a required standard feature for new passenger vehicles within the upcoming years in the US and the EU (EuroNCAP, 2017; EuropeanCouncil, 2019; NTSB, 2020a). These systems are mainly based on driving performance and gaze behaviour metrics.

Driving performance includes those metrics derived from human inputs on vehicle control such as acceleration, speed, steering wheel angle, vehicle position, throttle and brake pedal positions, often obtained from the vehicle's Controller Area Network (Dong et al., 2011; Melnicuk et al., 2016). Impaired steering ability and car following, unintended speed changes or lane departures would indicate cognitive and visual distractions. In contrast, the reduced frequency of steering correction manoeuvres, jerky motion, low-speed steering or impaired lane tracking ability are indicators of drowsiness and fatigue (for a review, see Dong et al., 2011). For example, Mercedes-Benz in 2009 launched its *Attention Assist* system, which only uses driving parameters (e.g. speed, longitudinal and lateral acceleration, angle of the steering wheel, indicators and pedals usage or specific driver control actions) to create a unique driver profile and prevent from drowsiness with visual and audible alerts (Daimler AG, 2009).

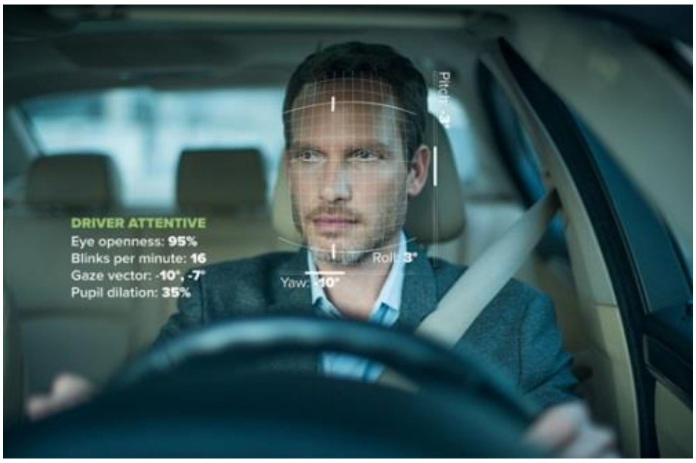


Figure 6 Example of a current DSM system from Cipia Technologies for the detection of driver attentiveness using eye-tracking and head positioning parameters.

Eye-tracking or gaze behaviour systems use video cameras and image processing technologies for measuring several gaze behaviour parameters indicative of drivers' attentional states such as fatigue, distractions, drowsiness, inattention or mental workload Figure 6 and Figure 7). Blinking frequency, fixations, saccades or pupil dilation are some of the parameters often used to classify driver states (Dong et al., 2011; Hecht et al., 2019; Lohani et al., 2019). With the cognitive distraction, drivers have been found to narrow the inspection of the driving environment, the instruments and mirrors with slower saccades, reduced glance frequency at traffic signals or increased blink frequency (Dong et al., 2011). Whilst fatigue and drowsiness have also been associated with increased blink frequency and increased PERCLOS – i.e., the percentage of time the eye is more than 80% closed (Dong et al., 2011). An example of a commercially available system featuring this technology is the *Driver Monitoring System* implemented by Toyota (and Lexus). Consisting of a near-Infrared camera on top of the steering column can determine the position of the driver's head and extract gaze activity features. This system can detect drowsiness and distractions and, in conjunction with the *Advanced Pre-crash Safety system*, mitigate them by activating pre-crash warnings or even applying the brakes to alert the driver (Toyota Motor Corporation, 2008). However, previous work has argued that near-infrared light-based eye-trackers may cause permanent damage to the eye after prolonged exposure, especially under night conditions, because the eye will not be able to regulate the artificial infrared light intake by closing the iris and protect itself, as it would do under daylight conditions (Leonhardt et al., 2018). Indeed, variations of daylight conditions represent one of the main limitations of optical camera-based monitoring systems (Leonhardt et al., 2018).

Figure 7 Example of a current DSM system from Cipia Technologies for the detection of driver drowsiness using eye-tracking and head positioning parameters.

Current DSM systems have proven to be an effective and unobtrusive solution for mitigating hazardous driving behaviours in real-world conditions for manual driving and lower levels of driving assistance (SAE Levels 0-1) by mitigating distractions, sleepiness or fatigue (Begum, 2013; Dong et al., 2011), providing a robust ground for developing the next generation of multimodal DSM systems for driving automation. Notwithstanding, the emergence of next-generation production vehicles with higher levels of driving automation (SAE Levels 2-4) will entail a wide range of different human-vehicle interactions (Carsten et al., 2012; Naujoks et al., 2016) and new human factors (Endsley, 2017), thus challenging the capabilities of current DSM systems, as will be discussed in the following section.

2.3.2 Future hybrid driver state monitoring systems

As with currently available DSM systems described in the previous section, future systems must compensate for inappropriate engagement or human errors and enhance automated driving safety functions and comfort. Since the automated driving system will progressively enhance its driving performance, the usefulness of this measure will diminish as automated driving functions increase. Although human driving performance may still be used during transitions to manual driving, eye tracking seems the most suitable technique for driver monitoring when automated driving is engaged because it is ubiquitous, unobtrusive, and provides a multitude of information about the driver state (Hecht et al., 2019).

However, this may be true only up to conditional driving automation (SAE L3), where the driver may momentarily disengage from the driving task to perform any non-driving related activities but still assumes the dynamic driving task fallback and must be ready for taking over at any time. A wide range of in-vehicle possibilities and non-driving related tasks will likely emerge during conditional driving automation, especially during high driving automation (SAE L4), where the automated driving system assumes the fallback responsibility. The higher the automation level, the greater the chance drivers engage in NDRTs (Carsten et al., 2012; Naujoks et al., 2016). It could be expected that while conditional or high driving automation is engaged, drivers will most probably be involved in several NDRTs, their situation awareness for the driving task will decrease (De Winter et al., 2014). Sleeping, reading, watching a video, playing with the console, videoconferencing will likely become everyday in-vehicle activities (Navarro, 2018). Likewise, car manufacturers will explore new interior design layouts adapting to these new in-vehicle possibilities to enhance users' comfort and experience – e.g., augmented reality head-up displays, holograms, rotating seats (Figure 8). (Damiani et al., 2009). Furthermore, these will unavoidably challenge eye-tracking systems capabilities.

Figure 8 In-vehicle possibilities during HAD will challenge current static camera-based DSM systems.

With automated driving engaged, one of the essential functions of future DSM systems will be detecting drivers' availability to take over and drive manually (Marberger et al., 2018). Either towards a planned take-over request (TOR) or an unplanned voluntary system deactivation. In the first case, planned TORs will likely be preceded by a monitoring request to adequate drivers' attentional level towards the transition (Gold et al., 2013; Lu et al., 2019; Marberger et al., 2018). It will be challenging to detect with eye-tracker whether 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. Each case scenario may imply radically different preparation to take-over strategies to get the driver back in the loop. None of these scenarios would suit an eye-tracker to monitor and assess drivers' availability to take over. Even though during high driving automation, the system will perform a minimal risk stop manoeuvre if the driver does not respond to a TOR; the public acceptance of a vehicle leaving their occupants stranded in the middle of the journey at some parking spot instead of ensuring an effective transition will be at the very least, questionable. Hence, achieving optimal attentional levels for a safe take-over performance will be required to implement high driving automation technology successfully.

Other than that, drivers may wish to voluntarily deactivate the automated driving function for whatever reasons and take manual control, even if a TOR is not issued and manual driving is not required at that moment. There is a chance they could overestimate their availability for manual driving because higher levels of driving automation often entail lesser SA (Endsley, 2017) and possibly, lower risk perception. Especially after a long engagement in alternative activities, it can be challenging to take over manual control straight after. Following the same example above, 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" (Ferrara & De Gennaro, 2000). Hence, drivers in such a state would possibly 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 this case, mental overload and high-stress levels would impair their take-over performance (Izzetoglu et al., 2004). In such a case, the driver will probably already be in a manual driving position (i.e., staring at the road out front), but still, eye-trackers would also benefit from other sources of driver state data to guarantee the driver is back on the loop (Merat et al., 2019) and available for taking over control. As mentioned in the earlier chapter, being awake does not necessarily mean being situationally aware – i.e., mental overload or daydreaming may induce false positives with eye-tracking systems.

Therefore, drivers' arousal could be considered a potential complement to eye-tracking for measuring attentional levels. Indeed, follow-up work from the DETER project in the early 90s already considered the potential of including psychophysiology indicators for measuring workload and stress (Fairclough, 1993; Fairclough & Hirst, 1993). Arousal is a real-time indicator of several emotional and cognitive states relevant for driving readiness, such as attentional capability, wakefulness, sleepiness, mental workload, or stress (Begum, 2013; Lohani et al., 2019; Nemcova et al., 2020). Arousal -or alertness- is a physiological reaction to internal or external stimuli which the brain perceives as potentially hazardous. This process is essentially controlled by the autonomic nervous system (ANS). The activation of the parasympathetic nervous system (PNS) branch is responsible for both rest-and-digest states and engaging in tend-and-befriend behaviours and essential for energy conservation. In contrast, the sympathetic nervous system (SNS) activation occurs during fight-or-flight reactions towards stressing stimuli (Laborde et al., 2017; Shaffer & Ginsberg, 2017). Activation of one branch often leads to a reciprocally coupled deactivation of the other. However, the autonomic space is multidimensional, and this activation can be nonreciprocally coupled (i.e., coactivation or coinhibition) or uncoupled (i.e., activation changes of one branch are not coupled with changes in the other branch) (Backs et al., 2003; Lenneman & Backs, 2009). Different modes of autonomic control generate several measurable changes in the body, including heart rate, breathing rate, pupil diameter or changes in skin conductance (Singh & Queyam, 2013).

Importantly, related to the earlier mentioned *Yerkes-Dodson law*, engagement in demanding NDRTs might increase arousal and mental workload levels to the detriment of performance (see Figure 9) (Bailey & Scerbo, 2007; Lenneman & Backs, 2009; Melnicuk et al., 2021; Sibi et al., 2016). On the contrary, arousal levels can be mitigated by relaxing NDRTs, sleepiness, or complacency due to overtrust in automation (Bailey & Scerbo, 2007; Parasuraman & Manzey, 2010). These findings indicate that under-aroused drivers will likely perform a take-over poorly, but over-aroused drivers will do so as well (Lohani et al., 2019). Hence, detecting and achieving optimal arousal levels will be necessary for the safe take-over performance and the primary role of next-generation DSM systems.

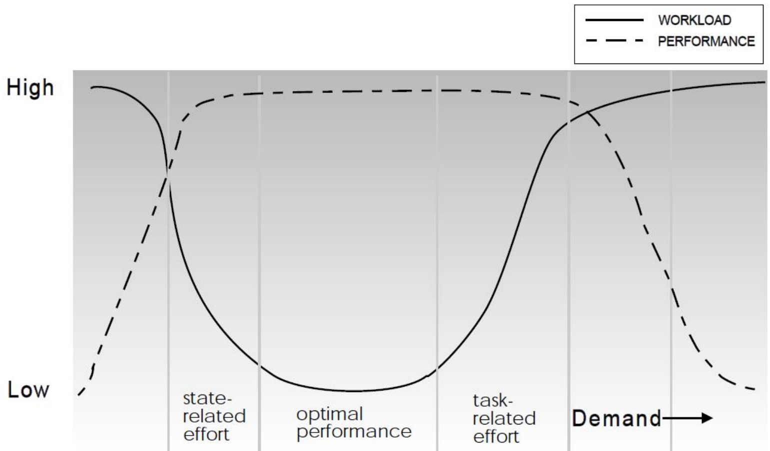


*Figure 9 The relationship between mental workload and performance. Extracted f*rom De Waard (1996).

Researchers have been exploring the usage of several psychophysiological arousal indicators for driving behaviour psychology and human factors engineering over the last twenty years. Driving simulators provide ideal conditions for realistic experimentation without compromising research participants' safety, as the literature published indicates relating to, e.g. stress (Nemcova et al., 2020; Singh & Queyam, 2013), mental workload (Backs et al., 2003; Lenneman & Backs, 2009), inattention (including distractions, fatigue and sleepiness) (Dong et al., 2011; Nemcova et al., 2020). Real-world experimentation has been less frequent, but still, a few studies have been conducted to validate the usage of hybrid psychophysiological arousal indicators in naturalistic settings. In an open road driving study, Healey & Picard (2005) tested a method for measuring and classifying stress levels in real-time based on multimodal physiological indicators, including ECG, EMG, EDA, and respiration. Cable-bound sensors were attached to participants' bodies and connected to a computer located on the rear seat. The classification algorithm achieved an overall accuracy of 97.4% using 5-minute intervals of data, with heart rate and skin conductance providing the highest correlations with driver stress levels. A similar study was conducted by Baek et al. (2009), although embedding non-intrusive ECG, PPG, EDA, and respiration sensors on the steering wheel, driver's seat, and seat belt. Drivers' stress levels were assessed off-line during night driving, cold environment, hot environment, noisy environment, narrow alleys and simultaneous mental loads (arithmetic operations).

Notwithstanding, before its implementation in mass production vehicles, multimodal DSM systems including psychophysiological indicators should overcome at least two main pitfalls:

- First, these indicators should be implemented with sensors as ubiquitous and unobtrusive as current eye-tracking monitoring systems to reach sufficient public acceptance.
- Second, these indicators should be practical and valid to measure new human factors related to automated driving.

Regarding the former point, most of the existing literature has used medical-grade devices to measure these arousal indicators in the driving context. Whereas these devices provide high-quality data, strengthening the validity of their results, they are limited for experimental or medical diagnostic use, and thus, could not realistically be integrated into a passenger vehicle. Moreover, these typically require electrodes or sensors attached to the individual, hampering its integration for daily public use. In an attempt to develop such a DSM system for public use in passenger vehicles, Melnicuk and colleagues have proposed and tested a potential alternative for a DSM system based on non-invasive and affordable consumer-grade wearable devices currently used for sport-related activities (Melnicuk et al., 2016, 2017). With a wireless watch-like wristband (Empatica E4) for measuring skin conductance levels and a chest heart monitor (Polar H7) for heart rate/heart rate variability, they successfully measured arousal related to mental workload fluctuations from several driving simulator scenarios. The E4 wristband (Figure 10) was also used in other driving simulator studies to measure variations in skin conductance level (Walker et al., 2019) or heart rate variability (Wintersberger et al., 2017), both associated with trust in automation levels. Another similar wristband, namely the Microsoft Band 2, was used in a series of driving simulator studies for measuring drivers' discomfort through heart rate/ heart rate variability and tonic (i.e., skin conductance level) and phasic (i.e., skin conductance responses) electrodermal activity (Beggiato et al., 2018, 2019). These studies show the potential for developing ubiquitous and unobtrusive DSM systems, but overall, weak – or even the lack of – statistical effects were common among them, suggesting this technology needs to be further refined. Hence, Walker et al. (2019) claimed that future research should investigate whether more precise psychophysiological measurements lead to more substantial effects. To date, studies using psychophysiological indicators (either medical or consumer-grade) to measure human factors related to driving automation are limited, and in some cases with questionable, unclear methods for data collection, processing or analysis, which hinders the interpretation, replication and generalisation of results (Lohani et al., 2019).

Figure 10 The Empatica E4 wristband allows the real-time measurement of EDA, blood volume pulse or skin temperature among other features.

On the second point, psychophysiological indicators should be effective in measuring (and classifying with the aid of artificial intelligence or machine learning techniques) arousal levels associated with those human factors arising from the interaction with automated driving technology. Previous work in the air traffic control domain has successfully applied machine learning techniques based on physiological indicators from brain activity, ECG, EDA and gaze behaviour to determine automation operators' functional state in real-time (Verdière et al., 2018; Wilson & Russell, 2003, 2004, 2007). In the automated driving domain, only a few works have addressed the use of physiology with such a purpose. For example, Pakdamanian et al. (2021) used multimodal data from cardiac, skin conductance and gaze behaviour to feed a deep neural network that predicts take-over intention, time and quality with an accuracy of 96%, 93%, and 83%, respectively. Other authors have used data from single inputs such as gaze behaviour to predict SA (Zhou et al., 2021) or fNIRS to predict mental workload (Unni et al., 2018). Multimodal inputs have also been proposed for real-time prediction of TiA (Perello-March et al., 2020) using machine learning techniques.

Broadly, the current state-of-the-art suggests that – aside from EEG and eye-tracking – fNIRS, ECG, and EDA are also strong candidates for assing human factors derived from HAD but often overlooked as they might be less informative when used in isolation. However, used in tandem can provide an accurate real-time picture of drivers' functional state. We aim to investigate their sensitivity to different human factors parameters known to modulate drivers' functional state such as mental workload, trust in automation and situation awareness across several driving conditions, and any considerations or recommendations for its integration in multimodal DSM systems. Therefore, the aims of the thesis (see section 1.2) will be achieved by conducting a series of simulator studies to evaluate different aspects of driver availability state during HAD and measure its physiological indicators. The next chapter will discuss the methodologies and equipment adopted to conduct these experiments, and the data extraction and processing techniques used.

Table 1 Expected relationship between the physiology measures selected and the main concepts presented. Legend: (-) Slight decrease, (--) Moderate decrease, (---) Strong decrease, (+) Slight increase, (++) Moderate increase, (+++) Strong increase.

		ECG		**EDA**	**fNIRS**	
		HR	HRV	SCR	HbO/HbT	HbR
Trust in	Trust	-	+	-	-	+
Automation	Distrust	+++	---	+++	+++	---
Mental Workload	Low	--	++	--	--	++
	Moderate	+	-	+	+	-
	High	+++	---	+++	+++	---
Stress	Low	--	++	--	--	++
	Moderate	+	-	+	+	-
	High	+++	---	+++	+++	---
Risk perception	Low	--	++	--	--	++
	Moderate	+	-	+	+	-
	High	+++	---	+++	+++	---

3 Methods and experimental equipment

3.1 3xD Driving simulator

Studies were conducted using WMG's 3xD driving simulator at the University of Warwick (Figure 11). The 3xD is a fixed-base high-fidelity driving simulator equipped with a full-body Range Rover Evoque and eight projectors generating a 360◦ image, projected into a cylindrical screen eight meters in diameter and three meters in height. 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 11 WMG 3xD Driving Simulator.

Simulated driving scenarios were created *ad-hoc* to meet the specific requirements of our studies. The process of drafting driving scenarios started with creating a map-based file, also known as "world", in which the basic road layouts are added-up together as single jigsaw pieces. The software only allowed several pre-defined basic road-layouts (e.g., rural, highway, urban, dual-carriageway), with several variants for each (e.g., rural single road straight, rural double road with a left bend, rural road with detached houses). To ease the trial launch, the primary "world" had to allow a single-run drive with all the scenarios we intended to test. Each scenario also had to last above the minimum time required for physiological epoch recording, requiring several weeks of testing the primary world layout to ensure all driving scenarios had a similar duration.

Once the map-based world file was completed, a navigation route was created throughout the map. This route creation started with setting the “ego-vehicle” (i.e., the vehicle users would drive on during the experiment) at a point in the map and manually drafting a navigation route until the final driving scenario. Designing this route implied manually drafting a dotted line the ego vehicle would follow from the start to the end of the map, in which every single dot allowed its driving speed. This feature enabled accelerations (which could be smoothened with more expansive inter-dot spaces and progressive speed increases, e.g., from 20 mph to 60 mph in 30 seconds) and decelerations (which allowed heavy breaking with minimum inter-dot space and significant speed jumps, e.g., from 60 mph to 1mph in 5 seconds). Particular care had to be taken in bends, roundabouts, intersections and traffic lights. In each case, speed had to be carefully manipulated to generate a smooth, comfortable and realistic driving sensation, respecting speed limits and traffic code rules.

Figure 12 Example of a city centre driving simulator scenario from the participants' point of view.

Aside from the ego vehicle navigation, the world had to be filled with traffic, pedestrians and other road users. To control the effect of surrounding road users, we removed any random interaction, so the movement of every single vehicle, cyclist and pedestrian on the map were manually generated. This process involved adding a selected number of vehicles, pedestrians and cyclists during each scenario and manually inputting their behaviours (e.g., driving at a certain speed, braking at give-ways, turning left or right) and dotting their navigation through the map. We used several map triggers for efficiency, so they would only start moving around when the ego vehicle would pass through the trigger. These triggers also required extensive testing to ensure all interactions were realistic (e.g., pedestrians crossing the street after a bus stopped, vehicles parking in driveways, vehicles giving way to other vehicles). See Figure 12 for a detail of an urban section.

3.2 Biopac MP160

Psychophysiological data from electrocardiogram (ECG) and electro-dermal activity (EDA) measures were recorded using a BIOPAC MP160 with wearable remote Bio-Nomadix amplifiers (Figure 14). The MP160 base station was mounted behind the driver's seat inside the simulator to achieve the best quality signal. Data were extracted and analysed using the automated data analysis routines from Biopac's ACQKnowledge software (CA, USA; version: 5.0.2).

Three ECG electrodes were placed following a 3-lead configuration on the participant's torso (see Figure 13). ECG data were sampled at 2000Hz and filtered, applying Biopac's recommendations using a bandpass filter with a 35Hz high-frequency cut-off and a low-frequency cut-off at 0.5Hz. ECG is an indicator of cardiovascular electrical activity (Cowley et al., 2016). Cardiac features extracted were heart rate (HR; beats per minute) and heart rate variability (HRV) metrics from both frequency and time domain.

1) In the frequency-domain, the low frequencies/high frequencies (LF/HF) ratio is a combination of sympathetic and vagal (parasympathetic) physiological activity, and thus, a low LF/HF ratio reflects parasympathetic dominance, whilst a high LF/HF ratio indicates sympathetic dominance (Laborde et al., 2017; Shaffer & Ginsberg, 2017). The absolute power of the high frequency (HF) band (between 0.15 and 0.40 Hz) was also extracted as a frequency-domain indicator of parasympathetic modulation.

2) In the time-domain, the root mean square of successive differences between normal heartbeats (RMSSD) is obtained by first calculating each successive time difference between heartbeats in milliseconds. Then, each of the values is squared, and the result is averaged before the square root of the total is obtained. RMSSD was obtained based on Laborde et al. (2017) recommendations of coupling frequency-domain parameters with a time-domain parameter indexing vagal modulation (Laborde et al., 2017; Shaffer & Ginsberg, 2017).

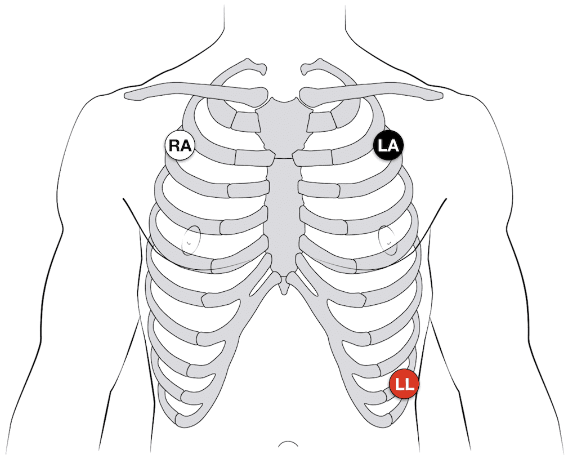


Figure 13 3-lead configuration used for ECG data recording. Source: https://litfl.com/ecg-lead-positioning/

The EDA device comprised two electrodes on the medial phalanx region on the first and second fingers of the participant's non-dominant hand to minimise movement artefacts. EDA data were sampled at 62.5 Hz and low-pass filtered to a frequency cut-off fixed at 1 Hz, following standardised guidelines (Boucsein, 2012; Braithwaite et al., 2015). EDA is an indicator of changes in skin conductivity resulting from the activation of sweat glands controlled by the SNS, which prepares the body for fight-or-flight responses (Boucsein, 2012; Dawson et al., 2016). Both background tonic (skin conductance level: SCL) and rapid phasic components (skin conductance responses: SCRs) can be extracted from the EDA signal. Notably, skin conductivity is not influenced by the PNS, and thus it is considered a direct indicator of psychophysiological arousal, and by extension, cognitive and affective states (Boucsein, 2012; Dawson et al., 2016). 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%. Phasic features extracted were:

1) SCR count. The total number of SCR events within each epoch.
2) SCR amplitude. It represents the change in (i.e., delta) value from the offset to the peak of the SCRs. According to Boucsein (2012) and Boucsein et al. (2012), amplitudes below 0.03 µS were rejected from the analysis. The common practice for normalising these values is applying the square root transformation (Braithwaite et al., 2015; Braithwaite & Watson, 2015; Dawson et al., 2016).
3) SCR magnitude. It is obtained from the same delta value, but a 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 (Braithwaite et al., 2015; Braithwaite & Watson, 2015; Dawson et al., 2016).

Figure 14 Biopac MP160 base station and wearable remote Bio-Nomadix amplifiers attached to the wrist with EDA sensors. Source: www.biopac.com

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 (Boucsein, 2012; Braithwaite & Watson, 2015).

3.3 NIRSport

Neurophysiological data was obtained from the pre-frontal cortex region with a NIRSport CW-NIRS device (NIRx Medical Technologies LLC, USA) (Figure 15). Data were extracted using NIRStar acquisition software (CA, USA; version 15.0) running on a Microsoft Surface. NIRSport is a non-invasive wearable device consisting of eight sources and seven detectors sampling at a frequency of 7.8125 Hz. The sources simultaneously emit infrared signals of two distinct wavelengths, 760nm and 850nm, allowing quantification of oxygenated haemoglobin (HbO), deoxygenated haemoglobin (HbR), and total haemoglobin (HbT = HbO + HbR). Both chromophores can be differentiated when light attenuation is measured at two or more wavelengths due to their differential absorption spectra in the near-infrared spectrum (600–950 nm). This differentiation would be analogue to the fact that fully oxygenated (arterial) and partially deoxygenated (venous) blood differ in colour (Obrig, 2010). A particular issue we found is that our fNIRS montage was incompatible with infrared light based corneal reflection eye trackers due to infrared light interference. The infrared illumination directed towards the participant's face was saturating the fNIRS optodes. Perhaps a different fNIRS montage (e.g., on temporo-parietal or occipital regions) without direct infrared light interference would not have such issues. Another alternative could be using a wearable eye-tracking system (e.g., Tobii Pro Glasses), which would not interfere with fNIRS.

Plastic spacers located at a distance of 3 cm between each source and detector pair constitute a recording channel. Our montage consisted of a total of 8 sources and 7 detectors, thus resulting in 22 recording channels (see Figure 15 and Table 3**Error! Reference source not found.**). Channels were mounted within the Montreal Neurological Institute (MNI) coordinate space for consistency across head size variation (Zimeo Morais et al., 2018). These coordinates allow subsets of fNIRS channels down to those directly measuring particular regions of interest (ROIs).

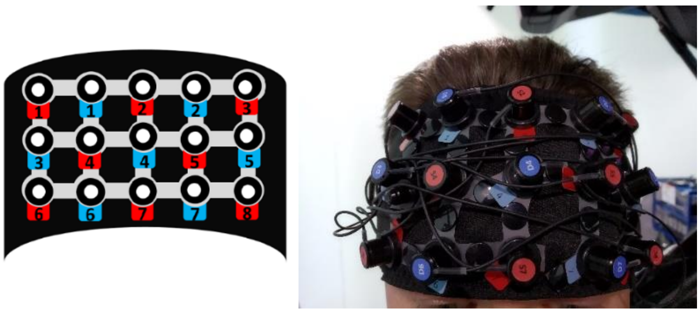


Figure 15 Headset montage used for fNIRS data collection in the pre-frontal cortex consisting of 8 sources (in red) and 7 detectors (in blue).

Raw fNIRS data were pre-processed using HomER 3 (Huppert et al., 2009) scripts running on MATLAB R2019a (Mathworks Inc.), and following the current recommendations for pre-processing fNIRS data (Pinti et al., 2019; Yücel et al., 2021) (see Table 2 for more details). Corrected optical density data were then converted into HbO, HbR and HbT concentrations using the modified Beer-Lambert law (Yücel et al., 2021). Once optical density concentrations were calculated, data was block-averaged and exported as Hemodynamic Response Function (HRF) means.

Block averaged HbO and HbR values from HomER 3 were exported as Excel files containing HRF means for each channel, condition, and participant. The underlying ROIs were determined using the NIRS Brain AnalyzIR toolbox (Santosa et al., 2018) to calculate the corresponding anatomical labels for each position. The toolbox creates a variable that lists the channels and Brodmann Areas (BAs) covered by the probe and relative 'weights' for each channel and BA. The weights for each BA add up to 1. The channel with the most sensitivity to a BA has the highest weight for that area. The relative weight is a helpful metric, but it does not give the whole picture, so we also extracted a 'depth' value for each channel and BA.

Table 2 Procedure for pre-processing, filtering and extracting block averaged HbO, HbR, and HbT concentrations from HomER 3.

Step	Description	Function	Input values
1	Removes channels in which the signal was too weak, too strong, or their standard deviation was too great	hmrRPruneChannels	dRange = 1e+04 1e+07 SNRthresh = 5 SDrange = 0.0 to 45.0
2	Transforms fNIRS raw data into optical density	hmrRIntensity2OD	None
3	Identifies and corrects motion artefacts	hmrRMotionCorrectPCArecurse	tMotion = 0.5 tMask = 1 STDthresh = 20 AMPthresh = 5 nSV = 0.97 maxIter = 5
4	Eliminates noise from physiological activity, low frequency signal drifts or machine noise.	hmrRBandPassFilt	hpf = 0.01 lpf = 0.5
5	Converts optical density to concentrations	hmrROD2concNew	ppf = 1 (760nm) 1 (850nm)
6	Calculates the block average of the given conditions	hmrR_BlockAvg	None

Depth values represent the distance on average between the channel and the BA. The further the distance, the lower the likelihood that the channel captures that BA. Up to three channels accounting for at least a combined relative weight of 0.80 (i.e., covering at least 80% of a particular BA) and the lowest combined depth value (i.e., the smallest combined distance on average) were used. The rationale for not averaging all channels together with a relative weight greater than 0 for a given BA is that some of these values are far too low, and if too many channels are averaged together, responses will be negated. Following Wiggins, Anderson, Kitterick, & Hartley (2016), we established a criterion of averaging together only up to 3 channels. The most highly sensitive channels were grouped into ROIs, whereas the least sensitive channels were discarded. For example, BA08 in the right hemisphere was the result of grouping channels 1, 3 and 8. This process led to 10 ROIs: Bilateral BAs 08, 09, 10 and 46, and left BA44 and 45 (see Table 3).

Table 3 Brodmann areas (i.e., ROIs) captured by the fNIRS montage.

Source	**Receptor**	**Channel**	**Brodmann Area**
1	1	1	8 right
1	3	2	-
2	1	3	8 right
2	2	4	8 left
2	4	5	8 left
3	2	6	8 left
3	5	7	44 left
4	1	8	8 right
4	3	9	9 right / 46 right
4	4	10	9 left / 9 right
4	6	11	9 right
5	2	12	-
5	4	13	9 left
5	5	14	44 left / 45 left
5	7	15	10 left
6	3	16	46 right
6	6	17	10 right/ 46 right
7	4	18	9 left
7	6	19	10 right
7	7	20	10 left
8	5	21	45 left
8	7	22	45 left / 46 left

Having grouped the relevant channels into ROIs, values were averaged within each ROI for each experimental condition, resulting in a single mean concentration value per participant. For example, the mean HbO concentration for participant ID01 in BA08 left during baseline was the average of channels 4, 5 and 6 during baseline, and so on. These concentration values were then standardised to enable inter-individual and intra-individual comparisons using Z-scores (M = 0; SD =1). Each single mean concentration value was then transformed into Z-scores against the mean group baseline value and its standard deviation (i.e., Z = (X – baseline mean) / baseline SD) (see Table 4 for details). Data standardisation is a common procedure among fNIRS studies to allow for inter-individual comparisons in parametrical statistical analysis using block averaged values (Durantin et al., 2014; Leon-Dominguez et al., 2014; Lin et al., 2020; Minematsu et al., 2018; Roche-Labarbe et al., 2008; Tanida et al., 2004; Verdière et al., 2018).

Table 4 Method used for determining the Regions of Interest and standardising data for parametric analysis.

Step	Description	Function	Criteria
1	Determining underlying ROIs for each channel	nirs.util.converlabels2roi	
2	Determining sensitivity for each channel using relative weights and depth values	nirs.util.depthmap	
3	Determining most sensitive channels combined for each ROI		Up to three channels with the highest relative weight and depth values
4	Averaging most sensitive channels for each ROI and for each experimental condition	Arithmetic mean	
5	Standardising individual mean concentrations to allow for inter-individual comparisons	Z-score = (mean concentration value – mean baseline) / SD of the baseline	

The General Linear Model is the standard approach for analysing and interpreting hemodynamic responses (Monti, 2011; Pinti et al., 2019). Among the range of possibilities this approach offers, the analysis of variance (ANOVA) is a common technique to determine localised brain activation based on changes in simultaneous HbO and HbR concentrations in repeated measures designs (Tak & Ye, 2014). Although it is common in the related literature to report only HbO, HbR or HbT – i.e., the combination of both –, the hemodynamic is a bi-dimensional response and both chromophores, HbO and HbR, usually correlate negatively during brain stimulation. The rationale underlying this correlation is that increased blood flow produces an increase in oxygenated haemoglobin and a decrease in deoxygenated haemoglobin (Fallgatter & Strik, 1998; Mehagnoul-Schipper et al., 2002; Schroeter et al., 2002; Taga et al., 2003). Nonetheless, since these features may not necessarily be always reciprocal, several authors have argued that interpretations based exclusively on one chromophore would be incomplete and advocate in favour of reporting both features in tandem (Liu et al., 2016; Obrig, 2010; Quaresima et al., 2012). Therefore, following these recomendations, we will perform ANOVAs to determine changes in haemoglobin concentrations on each chromophore, and report both features in tandem (and HbT) to interpret hemodynamic responses.

3.4 Device synchronisation and setup

Synchronisation of both BIOPAC MP160 and NIRSport physiological devices was enabled via an *ad hoc* script running in a Microsoft Surface which involved remotely pressing the "F1" key and setting a marker every 5 seconds on the BIOPAC data acquisition software and every 15 seconds on the fNIRS software. The Biopac laptop and the fNIRS Surface were connected via an ethernet cable to the other Surface running the script (see Figure 16. Both scripts used can be found in Appendix 3.

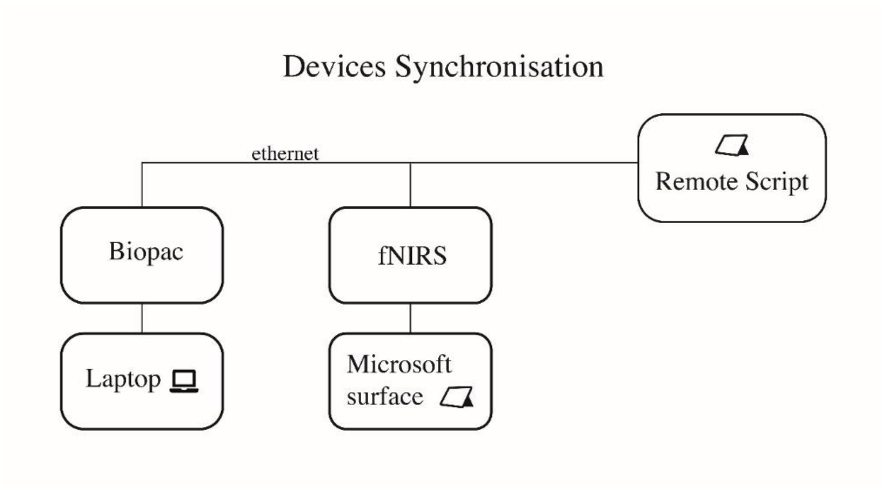


Figure 16 Diagram describing the synchronisation set up for both Biopac and fNIRS devices.

4 Study 1 – Measuring factors modulating driver availability state with physiological indices

The work in this chapter has been published in: Perello-March, J. R., Burns, C. G., Woodman, R., Elliott, M. T., & Birrell, S. A. (2021). Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios. *IEEE Transactions on Intelligent Transportation Systems*.

4.1 Introduction

As discussed in the literature review (see section 2), the inclusion of HAD entails a dramatic change in task allocation because under specific scenarios, manual control will not be required, and users will engage in NDRTs (SAE International, 2018). Unlike current DSM systems, the next generation will need to adapt to drivers' temporary disengagement from the driving task but still be able to monitor drivers' state transition from out-of-the-loop to being available for taking over manual control (Marberger et al., 2018).

Whereas certain NDRTs may facilitate the take-over preventing drivers' drowsiness by maintaining a suitable arousal level before take-over, other factors such as the traffic complexity or automation credibility expectations may also affect the driver state during HAD. The present study evaluates the modulating effects of three human-centric parameters on drivers' state measured through neuro and psychophysiological indicators and self-reports. Hence, (1) mental workload, (2) stress and (3) trust in automation will be elicited through task engagement, traffic complexity and automation credibility expectations, respectively.

4.2 Background

4.2.1 Task engagement and mental workload

Releasing drivers from the driving task during HAD will paradoxically entail a mental workload increase to match the demands of other non-driving related activities (e.g., reading, videoconferencing, video gaming). Aside from a likely activity, engagement in NDRTs could be beneficial to keep drivers' alertness at an optimal level for take-over performance (Miller et al., 2015) and even inform them regarding the current time budget for take-over within task transitions (Wintersberger et al., 2018). Indeed, Alrefaie et al. (2019) highlighted the potential of using drivers' physiological changes to determine the quality of a take-over request in an SAE L3 driving simulator study. Results showed that increased pupil diameter and heart rate while performing an NDRT were predictors of response time and, consequently, the take-over quality. However, if task demands exceed mental resources, the mental overload will work to the detriment of take-over performance (Melnicuk et al., 2021; Radlmayr et al., 2014).

Mental workload derived from task engagement in automated driving is often elicited through controlled and standardised methods. The Surrogate Reference Task (SuRT) is a visual task used to simulate eyes-off road NDRTs. In a driving simulator study, Walker et al. (2019) used the SuRT to explore the relationship between task engagement and trust in automation. Results indicated that higher self-reported trust was associated with more task engagement. In another experiment exploring the effects of induced automation credibility expectations on trust in automation, Körber et al. (2018) used the SuRT with the same purpose as Walker and colleagues. They found similar results, too, with higher task engagement and attention allocation towards the NDRT. Radlmayr et al. (2014) aimed to compare the effects on the take-over process of a mainly visual task (the SuRT) with a cognitive task (the N-Back task). They found both tasks showing similar effects on the take-over process, although the SuRT resulted in a higher total number of collisions in high traffic density scenarios, probably due to the visual distraction worsening SA.

The N-Back is another standardised method for eliciting mental workload in the driving context to simulate mind-off road. This method primarily uses verbal working memory resources, although it has experienced some variations among the literature. In essence, as *N* increases, the amount of working memory processing resources required to perform the task increase as well, and thus eliciting a greater mental workload. For example, Lenneman & Backs (2009) asked their participants to decide whether the currently presented letter matched the previously presented letter (i.e., 1-Back) in a series of sequentially presented letters. The difficulty in this experiment increased up to a 3-Back level. Unni et al. (2018) used a number-based task instead, and participants compared the currently presented number with the number presented *N* steps back (up to a 4-Back level) and provided a response when both numbers were the same. Melnicuk et al. (2021) used another variant of the N-Back where participants were asked to verbally recall the number they saw on a smartphone's screen with up to a 2-Back level. Gable et al. (2015) used a similar method asking their participants to recall the numbers, but these were presented in audio blocks of 10 digits.

Whereas the SuRT task is similar to a naturalistic interaction with any in-vehicle interface, the N-Back may lack such ecological validity. However, since it involves verbal working memory, it relates to other naturalistic tasks such as a phone conversation or talking to other passengers, but the N-Back allows experimental control on the verbal mental workload elicited. Previous work using different N-Back levels has found it predictive of mental workload demands using fNIRS (Unni et al., 2018) or cardiac measures (Lenneman & Backs, 2009; Melnicuk et al., 2021) in naturalistic simulated driving conditions.

Aside from mental workload, traffic density and complex scenarios can also modulate driver availability and even accentuate such adverse effects on take-over performance. For example, Radlmayr et al. (2014) found that high traffic density magnified the adverse effects on take-over performance due to visual and mental high workload, resulting in greater time to take-over and increasing the number of collisions during a simulator study. Similar results were also reported by Gold et al. (2016) when examining the role of traffic density and a verbal NDRT on take-over performance. They found the presence of traffic to play a significant role in impairing take-over performance than the verbal task did, leading to longer take-over times, shorter time to collision and more collisions. Techer et al. (2019) found that high traffic urban scenarios had adverse effects on subjectively reported drivers' emotional state and attitudes, motivating take-over behaviours. However, none of these studies measured whether high mental workload and high stress from traffic produce a similar response on arousal indicators, and thus, could be disentangled from a driver availability perspective. The following section will discuss how the role of traffic complexity has been found to impact drivers' arousal.

4.2.2 Traffic complexity and stress

Traffic complexity or driving environment demands affect drivers' stress levels. Traffic complexity refers to the combination of several environmental features such as traffic volume, flow and lane change presence among other road users (Teh et al., 2014). For example, a traffic jam on a busy six-lane motorway with vehicles barely moving could be considered less complex than a two-lane motorway with vehicles merging-in and overtaking at high speed. The effect of increasing traffic density and complex driving layouts during manual driving has been studied. For example, in a naturalistic driving study monitoring drivers stress using ECG, electromyogram, EDA, and respiration, the authors observed an increase of stress measures during high traffic density and urban scenarios (Healey & Picard, 2005). Similar findings were observed in a driving simulator study which recorded drivers' prefrontal hemodynamic responses, gaze behaviour, heart rate and skin responses across several road layouts of changing complexity (Foy & Chapman, 2018). They also found that complex road layouts (i.e., city centre and suburbs) increased physiological activity compared to a dual-carriageway and inter-urban road. Thus, urban scenarios increased oxygenated haemoglobin concentration (HbO), skin conductance level (SCL), skin conductance responses (SCRs), horizontal spread of search and a decrease of fixation duration. Further evidence was found by Melnicuk et al. (2017) in a driving simulator study combining different road layouts and including an automated driving scenario. Complex driving scenarios (i.e., urban and rural) reported the highest activation levels as shown by a reduced HRV (LF/HF ratio and RMSSD) and increased heart rate and skin conductance levels compared to the less demanding scenarios.

In the context of HAD (including SAE L3 & L4), only a series of driving simulator studies have directly measured the effect of traffic complexity on stress metrics. Beggiato et al. (2018, 2019) evaluated heart rate, HRV, blink ratio, pupil diameter, body motion and SCL variations of drivers reporting discomfort when facing several complex and uncertain situations. They found heart rate to decrease during discomfort, returning to the prior level approximately five seconds after the reported discomfort (Beggiato et al., 2018, 2019). HRV (RMSSD) showed a u-shaped tendency also decreasing during the discomfort intervals (Beggiato et al., 2018). In a related driving simulator study, Radhakrishnan et al. (2020) evaluated drivers' discomfort under different vehicle controllers (i.e., manual driving and four automated driving controllers) across several scenarios varying in traffic complexity and layout. They found that rural scenarios were more arousing than urban environments with a significant increase of SCRs/min, which the authors attributed to the higher speed limits, narrower roads and tighter curves in these environments.

Overall, these studies suggest that high stress from traffic conditions elicits a stress response activating the SNS and deactivating the PNS. That is, challenging traffic layouts will increase drivers' alertness, both in manual and automated driving conditions. However, it remains unclear whether different levels of trust in automation during automated driving conditions can accentuate this alertness, and hence if physiological indices can detect it.

4.2.3 Automation credibility expectations and trust in automation

In the introduction (see chapter 1), it has been discussed how users' misconceived beliefs regarding the driving automation actual capabilities and limitations may lead to overreliance or complacency resultant from overtrust in automation. *Autonowashing* suggests that automation performance or credibility expectations can be modulated in naïve or inexperienced users, affecting user engagement with the automated system. Several studies have found empirical evidence supporting the fact that TiA can be induced in naïve users through beliefs or expectations.

For example, in an EEG study conducted by de Visser et al. (2018), the authors manipulated participants' credibility expectations on a computer algorithm by inducing different levels of credibility on their participants with a written story. Event-related potential components were used to infer miscalibrated trust among participants with induced computer algorithm credibility (expected performance) while monitoring algorithm credibility (actual performance). Their results indicated that greater attentional orienting responses to unexpected errors from a reliable algorithm were positively correlated with self-reported trust. Thus, participants quickly calibrated their trust toward the actual algorithm performance, ignoring the credibility expectation provided. Körber et al. (2018) manipulated participants' credibility on system performance using an introductory video on automated driving performance, followed by a written text and a short practice. They found trust-promoted participants being less situationally aware (i.e., fewer glances at the road and more engagement with NDRTs) and more reliant on the automated driving system than the trust-lowered group. In Walker et al. (2019), participants were only presented with videos of dichotomised automated driving performance depending on the group assigned to promote or denote trust. Their results were similar to those from Körber and colleagues, with greater task engagement and lesser road monitoring behaviours on participants reporting higher trust. In addition, they found that higher self-reported trust was associated with lower skin conductance levels than participants with lowered trust. Others have used the experimenter's role to induce such beliefs verbally by simply describing the automated driving system performance to each group before starting the trials (Li et al., 2019). These authors used self-reported metrics to evaluate whether induced beliefs affected TiA, risk perception and perceived automation credibility. Results indicated that trust-promoted participants reported the higher trust, perceived automation credibility, and the lowest level of perceived risk when driving in a low-risk situation.

Evidence from these studies suggests that trust-promoted drivers would, in general, display mitigated alertness and arousal levels (↓SNS, ↑PNS). In contrast, distrust-promoted drivers should instead remain more vigilant and aroused (↑SNS, ↓PNS), especially during complex driving conditions with the greater risk involved. Understanding the relationship between TiA and arousal is essential because it could be used to infer whether drivers do not perceive the actual situational risk due to misconceived expectations, and hence are not available for take-over because of improper situation awareness. In other words, if trust-promoted drivers do not show a startle response and stay alert during a hazardous driving event, this may be an indicator of miscalibrated trust.

4.3 Objectives and hypotheses

The present study evaluates the effect of task engagement, traffic complexity and automation credibility expectations on drivers' states by measuring several neural and psychophysiological parameters in a realistic simulated driving setup. These three variables are expected to elicit mental workload, stress and trust in automation, respectively, which are multidimensional and likely to coexist in realistic driving conditions. Investigating how these variables affect each other and drivers' state indicators is essential for developing future DSM systems as they will contribute towards the operationalisation of the concept driver availability state for take-over. The aim will be achieved by:

- Performing a mentally demanding non-driving related task (2-Back task) during a low traffic complexity scenario to generate mental workload (i.e., as a control measure so that traffic density does not add additional mental workload to task demands).
- Presenting several highly automated driving simulated scenarios with varying traffic complexity levels, road layouts and a risky event. It will primarily investigate the effect of traffic complexity on stress and its effects on promoted or lowered trust.
- Inducing opposing driving automation credibility expectations to each group of participants to manipulate trust in automation. It should make the trust-lowered group more alert and physiologically arouse than the trust-promoted group, especially during the high traffic complexity scenarios.

The following hypotheses are proposed:

1. Performing the 2-Back task will induce a higher mental workload, increase brain activity and arousal.

2. High traffic complexity scenarios will elicit greater stress – i.e., increased sympathetic activity and parasympathetic withdrawal- than pre-drive baseline and lower traffic complexity conditions.

3. Induced driving automation credibility levels will result in greater self-reported trust in automation on the trust promoted group and lower trust in the trust lowered group. It consequently will generate group differences in both neuro and psycho-physiology metrics, with the trust-lowered group reporting higher brain activity during high traffic complexity scenarios (e.g., city centre and risk).

4.4 Method

4.4.1 Sample recruitment

Thirty-four participants were recruited within the University of Warwick (UK), including undergraduate students, postgraduate students, university staff and other professionals. All of them held a UK-EU driving license. Seven participants withdrew due to motion sickness and their data were excluded from analysis. A total of twenty-seven participants completed the trials and were included for data analysis. This sample size was considered acceptable based on the samples from the related literature. Recruitment and data collection methods received approval from the Biomedical and Scientific Research Ethics Committee from the University of Warwick. Participants voluntarily agreed to participate in this experiment and were free to withdraw at any point. All of them received a £10 voucher after the experiment.

Participants were randomly assigned to each group of automation credibility expectations. To the low credibility (LC) group (N = 12), the HAD performance was described as an early prototype system. Although not entirely reliable yet, a system capable of self-driving and adapting to most road conditions since it was still under development. Conversely, to the high credibility (HC) group (N = 15), the HAD system was described as an entirely reliable system, capable of driving through any scenario and adjusting to all road conditions effectively. Importantly, vehicle-driving performance was equal for both groups across all driving conditions. Both groups were instructed not to attempt to take control of the vehicle under any circumstances to generate the vulnerability required for TiA (Hoff & Bashir, 2015; Lee & See, 2004).

4.4.2 Demographic data

Sociodemographic data were gathered before the experiment to have a detailed description of the recruited sample. With this purpose, variables concerning participants' age and gender were collected, and other variables such as occupation, driving experience (i.e., years holding a valid EU-UK driving license) and average mileage per year. Other variables relating to participants' driving behaviour, locus of control and personality traits were also collected.

From the twenty-seven participants recruited, twenty were male, and seven were female, from which eleven males and one female were randomly assigned to the LC group (11-1), whilst nine males and six females were assigned to the HC group (9-6). To the base of my knowledge, consistent gender differences have not been described as a modulator of trust, mental workload or stress during driving conditions yet, and thereby there is no reason to suspect these would influence our results. Participants were mostly between 18 and 35 years old (85.19%). More details on age distribution and frequencies by each group can be found in Figure 17.

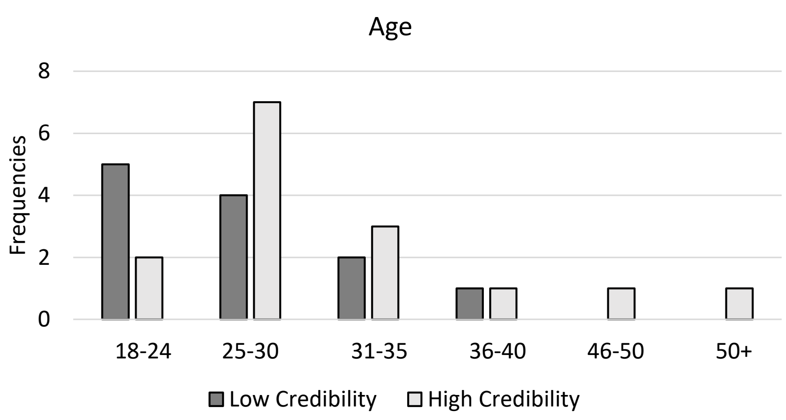


Figure 17 Frequencies with participants' age from each group.

Twenty were students, and seven were professionals or in managerial roles. The distribution per group was: LC = 10 students + 2 professional/managerial; HC = 10 students + 5 professional/managerial. Despite their young age, participants were relatively experienced, with seventeen of them (63%) holding a driving license for more than six years (Figure 18), and thirteen of them (48%) driving an average of more than 10k miles a year (Figure 19).

Figure 18 Frequencies with participants driving experience in years for each group.

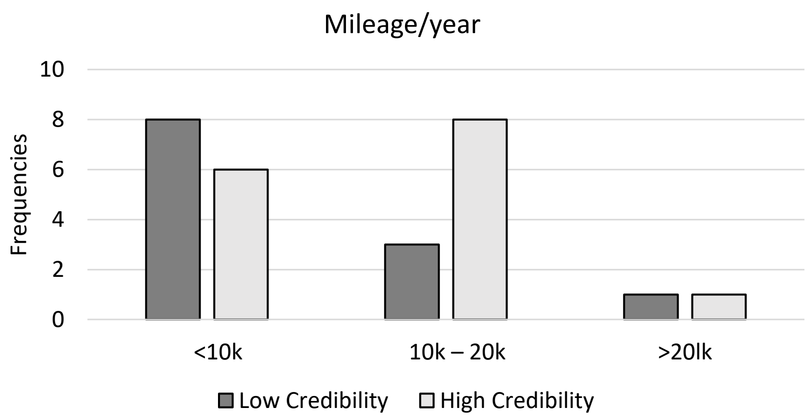


Figure 19 Frequencies with participants' mileage per year in each group.

The Driver Behaviour Questionnaire (DBQ) (Reason et al., 1990) was selected to inform the sample profile regarding their propensity to get involved in several driving errors and violations. Errors represent the failure of a planned action and reflect driver performance limitations in terms of perception, attention and information processing abilities. In contrast, violations are a deliberate infringement of the driving code of behaviour and represent the driving style and habits established with driving experience. The questionnaire used is a shorter and refined version from Parker, Reason, Manstead, & Stradling (1995). This included 24 items describing a variety of driving errors, lapses and violations, in which participants indicate how often they get involved in such situations during the last year in a scale ranging from 0 (never) to 5 (nearly all the time). It is an established tool for accident prediction due to hazardous driving behaviours. It has been validated across several studies (for a review, see De Winter & Dodou, 2010), for which we considered could provide valuable information on drivers perceptions of riskiness and hazards of the HAD driving behaviour.

Drivers locus of control was measured through the Driving Internality and Driver Externality Scale (DIDES) (Montag & Comrey, 1987). This scale measures drivers' attitudes towards the attribution of causality by rating between 0 (Disagree very much) to 5 (Agree very much), their level of agreement with 30 statements as to causes of accidents. Each factor consists of 15 items, with items 6 to 10, 16 to 20, and 26 to 30 forming the DI scale, and items 1 to 5, 11 to 15, and 21 to 25 on the DE scale. Both scales are related to road accident risk and have been used to evaluate drivers attitudes towards driving automation systems (Payre et al., 2014; Stanton et al., 2007).

Personality traits were evaluated using the Extra Short 5 (XS5) personality inventory (Konstabel et al., 2017). This inventory is a reduced version of the "Short Five" (Konstabel et al., 2012), which is based on the subscales of the NEO PI-R (Costa Jr. & McCrae, 2008), a well-known conceptualisation of the Five-Factor Model -a.k.a. "Big Five"- of personality traits in psychology (Costa Jr. & McCrae, 2008; Konstabel et al., 2012). Therefore, this inventory is constituted of five factors, namely Neuroticism (items 1, 15, 20, 24, 26, 28), Extraversion (items 2, 5, 8, 11, 17, 21), Agreeableness (items 4, 16, 23, 25, 27, 29), Openness to experience (items 3, 6, 9, 12, 18, 22), and Conscientiousness (items 7, 10, 13, 14, 19, 30), represented in 30 statements. Participants are asked to read the statements and indicate to what extent they apply to them, ranging from -3 (Completely wrong) to 3 (Completely right). Each factor is obtained by first reverse coding the negatively keyed items and computing the mean.

One-way ANOVAs were performed to test for between-participants differences in driver behaviour (DBQ), locus of control (DIDES), and personality (XS5), revealing no statistically significant differences between the high and low credibility groups in terms of driving behaviours from the DBQ either for lapses, errors and violations. There were no differences in locus of control (DIDES) either for driving externality or driving internality, and no differences were observed for the big-five personality traits (XS5).

4.4.3 Materials and equipment

The trials were conducted using WMG's 3xD driving simulator at the University of Warwick (see section 3.1). Psychophysiological data, including ECG and EDA measures, were recorded using a BIOPAC MP160 with a wearable remote Bio-Nomadix amplifier (see section 3.2). Neurophysiological data was obtained from the prefrontal cortex with a NIRSport CW-NIRS device (NIRx Medical Technologies LLC, USA) (see section 3.3).

The Driving Activity Load Index (DALI) (Pauzié, 2008) assessed drivers' mental workload. This scale is an adaptation to the driving context of the well-known NASA-TLX (Hart, 2006). The DALI uses the same rating procedures as NASA-TLX with six pre-defined factors: the effort of attention, visual demand, auditory demand, temporal demand, interference, situational stress. Each factor is rated on a 9-point Likert scale with endpoints low to high. Ratings from all six factors are averaged for each participant to compute the overall workload rating. The scale has been validated (Pauzié, 2008) and extensively used in the driving context to assess subjective mental workload (Birrell et al., 2012; Melnicuk et al., 2021; Paxion et al., 2014; Spiessl & Hussmann, 2011). Therefore, we considered it would be a better choice than the TLX for measuring the mental workload associated to the driving task. Although the HAD scenarios did not involve manual driving, the aim was measuring mental workload derived from the driving context and the DALI has its factors framed around driving. See Appendix 1 for more details concerning this scale.

The Self-Assessment Manikin (SAM) (Bradley & Lang, 1994) is a standard tool used to determine valence (i.e., the positive or negative feeling associated with an object or event) and arousal (i.e., the low or high level of excitement associated with an object or event). The SAM is a picture-based tool assessing each emotional dimension in a 9-point Likert scale, ranging from a smiling happy figure to a frowning unhappy figure for the valence dimension; and from an excited, wide-eyed figure to a relaxed, sleepy figure for the arousal dimension. This tool has been widely used for measuring emotions in the human-computer interaction domain (Bandara et al., 2018; Egger et al., 2019; Hirshfield et al., 2014, 2019). Given that this tool aims to measure arousal, we considered it a good complement of physiological indicators for detecting stress from traffic complexity. See Appendix 1.

Self-reported trust in automation was collected using the Trust in Automated Systems Scale (TASS) (Jian et al., 2000). This scale is comprised of 12 items with a 7-point Likert scale. Items 1 to 5 assess the construct of distrust, and items 6 to 12 assess trust. A total score can also be obtained by reverse scoring those items corresponding to distrust. It is an established scale widely used in research to measure operators' trust in automated systems (Banks & Stanton, 2016; Satterfield et al., 2017; Zhang et al., 2018). See Appendix 1.

4.4.4 Experimental conditions and automated driving scenarios

This study included a total of seven experimental conditions, including a pre-drive resting baseline, the 2-back task, highway, interurban, suburbs, city centre, and risk, in order of occurrence.

The **highway scenario** was a relatively straight, triple-lane road, with high-speed limits of 60 to 80 mph and opposite traffic separated by a central reservation. Traffic density was bidirectional, low and regular (< 5 road users per minute), so no braking or overtaking was needed. This scenario included relatively few signs -including overhead gantries-, and no pedestrians, pedal cyclists or buildings along the roadside. Weather conditions were set to sunny and clear. Details concerning all scenarios can be found in Figure 20.

The **inter-urban scenario** carried traffic to and from the highway to the suburbs and city centre in a straight line with two roundabouts and two lanes per way separated by a central reservation. Speed was limited to 30 to 50mph, and medium levels of oncoming traffic (< 20 road users per minute). Weather conditions changed to cloudy.

The **suburbs scenario** began within a layout defined as two lanes passing through residential areas at a 30mph limit, including several left and right turns, give-ways and with a medium volume of oncoming traffic, pedestrians, cyclists and parked cars on the roadside.

The **city centre scenario** passed through an area with commercial buildings, signs, billboards, and with the highest levels of moving and parked vehicles (i.e., vans, motorcycles, buses, trucks and emergency vehicles) and pedestrians compared to the other scenarios (between 20 and 40 road users per minute). Some vehicles were parked in driveways; others were parked on the street, and buses were waiting at a bus stop with pedestrians running to them whilst inappropriately crossing the street. The speed limit was 30mph, and the participant's automated vehicle had to overtake these stopped vehicles with traffic approaching ahead and deal with T-junctions with traffic approaching from both directions. Additionally, the simulated weather conditions shifted to heavy rain, degrading the visual range.

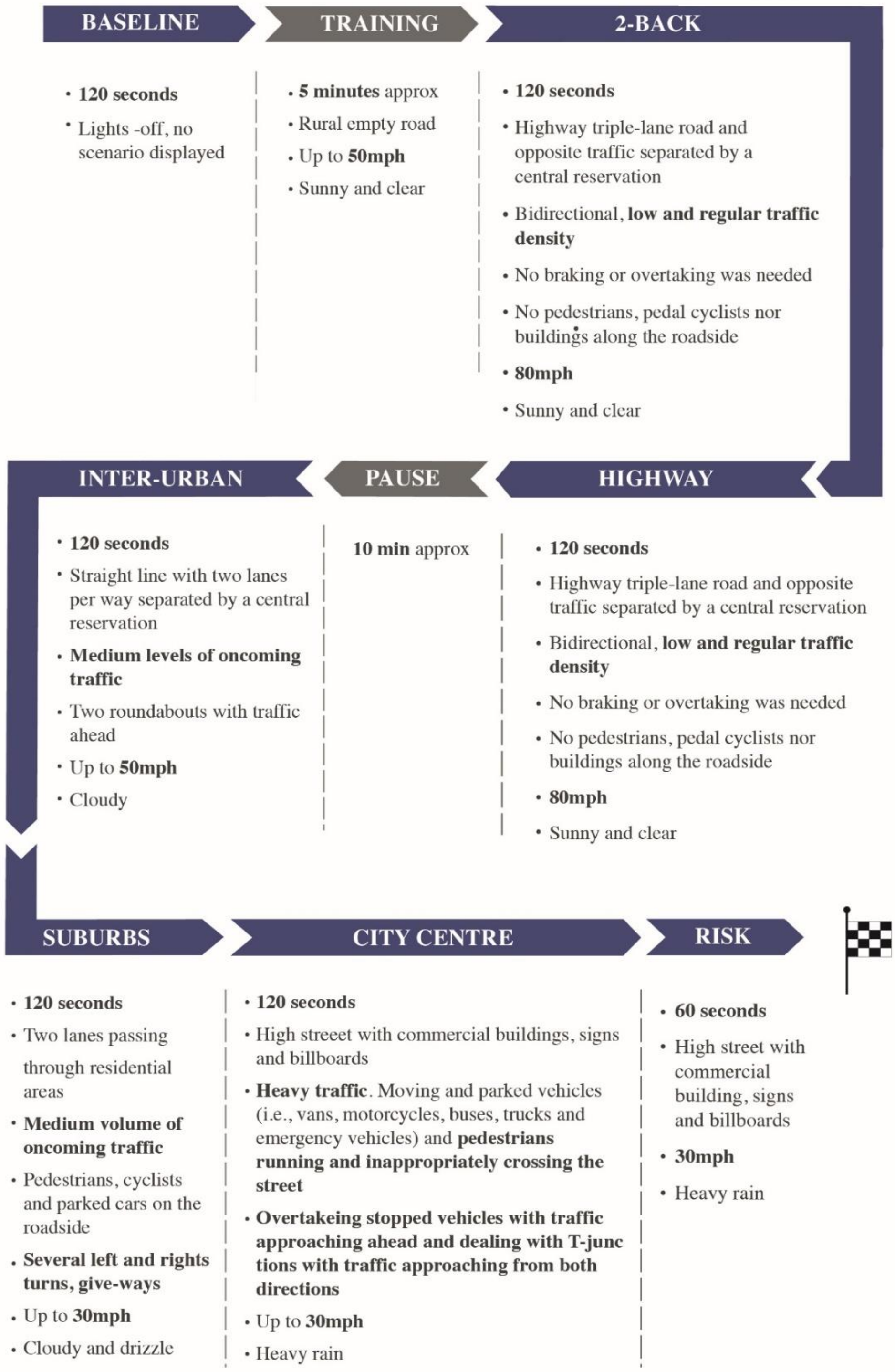


Figure 20 Flowchart with scenarios layout and description.

The **risk scenario** occurred within the city centre environment and involved the ego-vehicle following a van which, immediately after a left bend, both encountered a cyclist and proceeded to overtake while approaching a junction with the right-of-way. Immediately after the van passed the junction, and while the ego-vehicle overtook the cyclist, an ambulance with emergency lights and a siren moved into view at high speed from the left side of the junction. The ego vehicle had to perform an emergency braking and evasive manoeuvre to avoid crashing into the ambulance, and immediately after, a police vehicle followed the ambulance, so the ego vehicle had to brake again. Overall, this event lasted for approximately 60 seconds.

4.4.5 Procedure

Upon their arrival, participants were guided into the simulator control room, briefed on lab safety procedures and advised to follow the experimenter's instructions at all times. Consent forms and demographics questionnaires were filled in the week before the trial. Once all sensors were connected, approximately five minutes were allowed for electrodes and conductance gels to stabilise before data recording began. During this time, participants were instructed to be particularly careful in not applying any pressure to the sensors or stretching the cables to avoid signal spikes and artefacts. The room temperature was set at $21 \pm 2^{\circ}$C to control for EDA or ECG signal artefacts. Following this, participants were briefed on the 2-back task and performed a short practice session. After the 2-back training, the wearable amplifiers' data telemetry was checked to ensure good quality data acquisition.

Participants were guided inside the driving simulator and asked to remain seated in the driver's seat while the NIRsport headband was attached to their forehead without causing pain or discomfort. Participants' baseline for brain and psychophysiological activity was recorded for 2 minutes with the lights switched off and without projecting the driving scenario. Baseline was followed by a pilot trial with manual driving across empty rural roads, which acted as a familiarisation run to minimise the impact of motion sickness (Smyth, Jennings, et al., 2021). Participants were instructed to drive cautiously to gain familiarity, up to 20 mph, respecting UK Highway Code rules. The vehicle had an automatic gearbox, so they only used the accelerator, brakes and steering wheel. The manual driving trial eventually led to a roundabout connecting to a highway. Here participants were previously instructed to engage the automated driving function by pressing a button on the centre console after hearing the appropriate audio cue.

Experimental scenarios began once HAD was engaged. Two minutes after engaging HAD, participants heard an audio cue announcing they were about to start a 2-back task and providing the instructions concerning the task again. This was the first experimental condition and lasted for four sets of 30 seconds each. After performing the 2-back, the highway HAD scenario continued for five more minutes until reaching a highway exit. A two-minute epoch was extracted from this period forming the second experimental condition, namely highway scenario. The vehicle stopped at a red traffic light in the highway exit roundabout.

At this point, the simulation was paused for a short break, as long exposures to driving simulators tend to increase the risk of simulator sickness (Smyth, Jennings, et al., 2021). Participants left the vehicle and went into the control room to fill in the TASS, SAM and DALI scales. Physiological signals were calibrated once again before resuming the scenarios.

Upon resumption, the scenario began with HAD engaged from the same stopping point, leading to an interurban drive with low traffic complexity for 2 minutes. After this, the vehicle entered the suburbs, where traffic complexity slightly increased throughout the scenario. Two minutes later, traffic density increased, leading to a 2-minute city centre scenario. The experiment ended with the ego vehicle performing an evasive manoeuvre and stopping at the next T-junction, referred to as the risk scenario, which lasted approximately 60 seconds. After this, participants left the driving simulator and filled in the TASS, SAM and DALI scales.

4.4.6 Data Analysis

Seven epochs, one per experimental condition, were extracted for data analysis, being these: pre-drive (baseline), 2-back, highway, inter-urban, suburbs, city centre, risk. All epochs were of a length of 120 seconds, except the last condition (risk), for which the epoch length was 60 seconds. Due to its event-related singularity, the risk scenario had to be designed as a rapid, sudden and unexpected event to generate an artificial perception of risk. Otherwise, it would have been tough to keep a sustained perception of risk for 120 seconds in a simulated environment. A full description of each condition is provided in section 4.4.4.

Psychophysiological parameters from ECG and EDA signals were extracted using the automated data analysis routines from Biopac’s ACQKnowledge software (CA, USA; version: 5.0.2). Cardiac features extracted comprised heart rate (beats per minute) and two typically used frequency and time-domain heart rate variability (HRV) parameters, namely the low frequencies/high frequencies (LF/HF) ratio and Root Mean Square of Successive Differences (RMSSD). Skin conductance features extracted were the number of SCRs per epoch (SCR count), SCR amplitude and SCR magnitude (see section 3.2 for details on these metrics). A mixed repeated-measures analysis of variance (ANOVA) was conducted on each feature with 7 within-subjects factors (baseline, 2-back, highway, interurban, suburbs, city centre, and risk) by 2 between-subjects factors (low credibility/high credibility).

Neurophysiological parameters from fNIRS were extracted as block averaged chromophores (i.e., HbO and HbR values) from HomER 3 and exported in excel files containing the hemodynamic response function means for each channel, each condition and each participant. Having grouped the relevant channels into ROIs, values were averaged within each ROI for each experimental condition. That results in seven means (one per experimental condition) per participant, per each ROI and each chromophore. According to this rationale, the analyses explore the differences between groups and within conditions for each chromophore individually. With these in mind, a mixed repeated-measures analysis of variance was conducted for 2 groups (low credibility/high credibility) by 7 experimental conditions (baseline, 2-back, highway, interurban, suburbs, city centre, and risk). Thus, 10 ANOVAs were conducted for each ROI individually (i.e., BAs 8,9,10 and 46 bilateral, plus BAs 44 and 45 on the left hemisphere) for 3 chromophores (Hbo, HbR and HbT).

Finally, regarding self-reported data, the TASS and SAM were reported three times across the study: before grouping expectations on vehicle credibility were provided (PRE), during the mid-study pause at the end of the highway scenario (MID), and once the study was completed (AFTER). The DALI was reported twice: during the mid-study pause and after the trial because there was no mental workload to rate before the trial.

Data analyses were run on IBM SPSS Statistics 26 software (NY, USA). The Shapiro-Wilk's test ($p > 0.05$) 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 Bonferroni.

4.5 Results

This study aimed at investigating the effect of three independent variables on HAD users' neuro and psychophysiological states: task engagement, traffic complexity and automation credibility expectations. The following analyses all took the form of a mixed ANOVA, unless stated otherwise, with two factors: trust as the between-groups factor and driving condition as the within subjects factor. With trust having two levels –i.e. trust promoted vs trust lowered, and driving conditions having seven levels –i.e. pre-drive baseline, 2-Back task, highway, inter-urban, suburbs, city centre, and risk event. Participants experienced the same conditions in the same order but with their a priori expectations differing according to each group. The standard level of significance was used (i.e., $p < 0.05$) so anything reported as being non-significant had a p-value > 0.05.

Eight participants were excluded from the EDA analysis either because of missing data or substantial artefacts on the raw signal, with N = 19 participants analysed (10 in LC + 9 in HC). Two participants were excluded from analysis for ECG data due to missing data and excessive artefacts (N = 25; 11 in LC + 14 in HC). On fNIRS data, three participants were discarded for analysis due to artefacts (i.e., excessive noise or movement spikes), thus resulting in data from 24 participants included for analysis (12 participants per group).

4.5.1 Hypothesis 1 – Task engagement

It was hypothesised that performing the 2-Back task would produce a higher mental workload, resulting in increased brain activity, sympathetic response, and a parasympathetic withdrawal, comparable to that from the risk event.

A one-way ANOVA was used to test for the effect of self-reported mental workload (i.e., DALI) during mid-trial compared to post-trial. Although the DALI was not reported straight after performing the 2-back task, participants reported experiencing greater mental workload during the mid-study pause than after the trial ($F (1, 25) = 10.157$, $p = 0.004$, $\eta^2_p = 0.289$), thus suggesting that the 2-back had the expected effect eliciting higher mental workload.

ECG data reported main effects for heart rate ($F (6,138) = 34.470$, $p < 0.001$, $\eta^2_p = 0.600$; Figure 21), with follow-up pairwise comparisons indicating that the 2-back condition generated significantly higher rate (M = 78.125, SD = 10.881) than highway (M = 69.664, SD = 10.357, $p < 0.001$), inter-urban (M = 66.501, SD = 10.705, $p < 0.001$), suburbs (M = 67.154, SD = 10.725, p < 0.001), city centre (M = 68.556, SD = 9.907, p < 0.001), and risk (M = 67.800, SD = 9.153, p < 0.001).

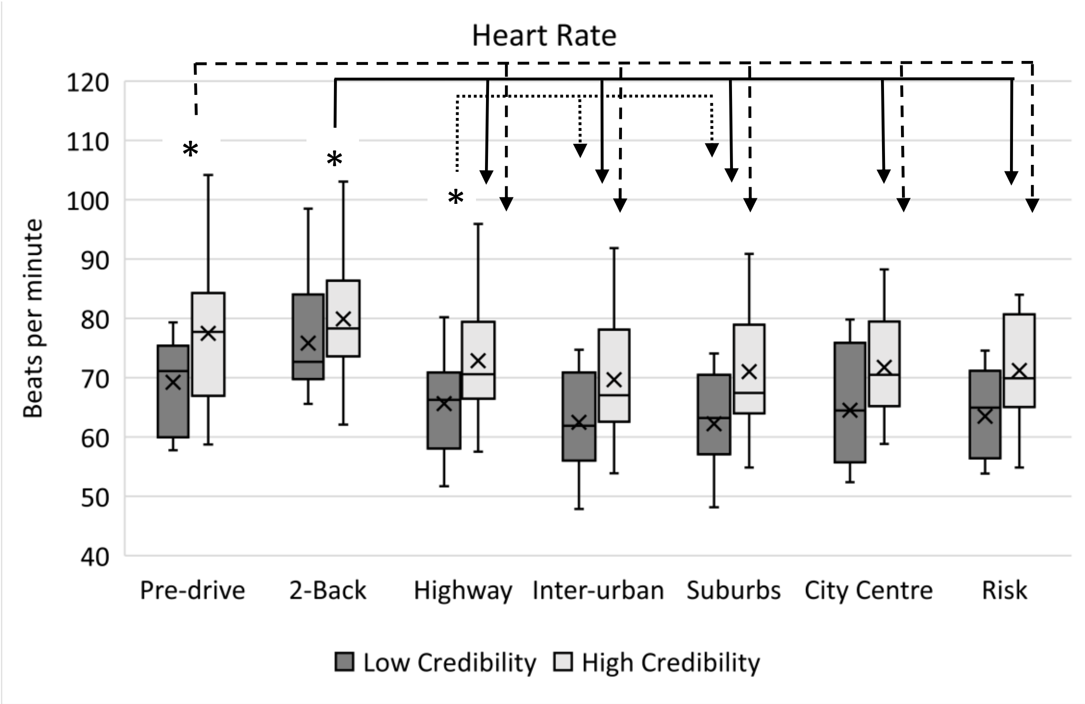


Figure 21 Heart Rate in beats per minute across driving conditions for each group. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

HRV-LF/HF ratio was aligned with these findings also reporting main effects ($F(6,138) = 4.935$, $p = 0.003$, $\eta^2_p = 0.177$; Figure 22), revealing a higher ratio during the 2-back condition (m = 3.297, SD = 3.093) than pre-drive (m = 1.779, SD = 1.912, p < 0.001), highway (m = 1.942, SD = 1.857, p = 0.001), suburbs (m = 1.992, SD = 2.418, p < 0.001), and risk (m = 1.251, SD = 1.032, p = 0.002).

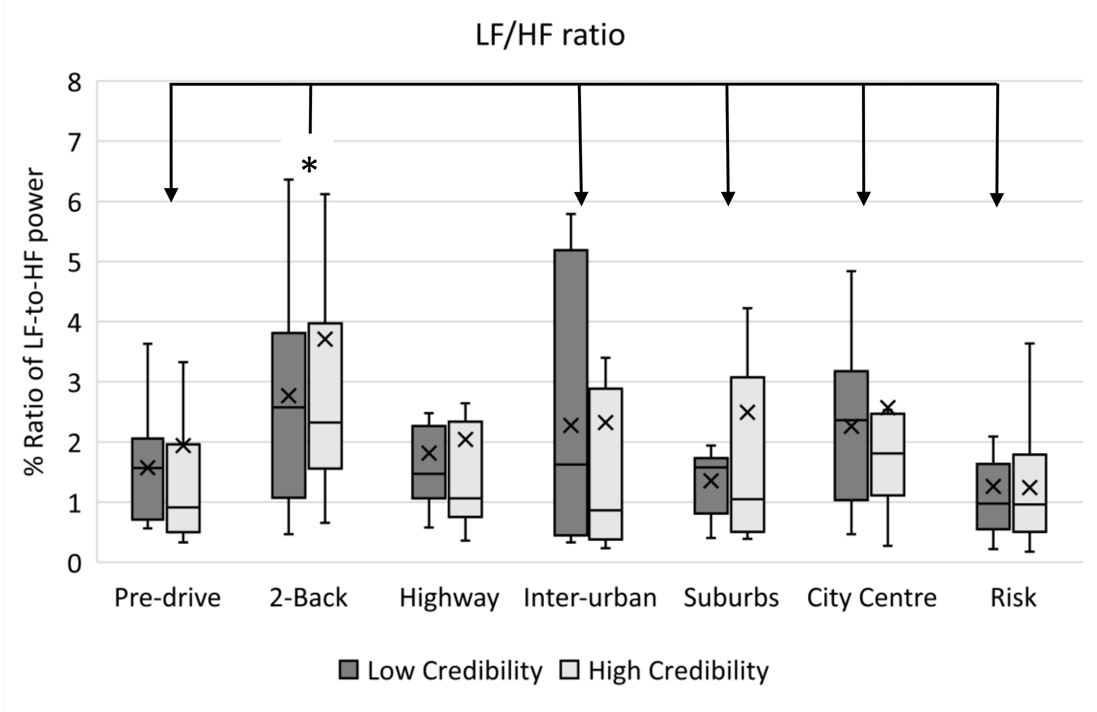


Figure 22 Heart Rate Variability as in LF/HF ratio across driving conditions for each group. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

As expected, HRV-RMSSD coupled these results with main effects ($F(6,138) = 7.048$, $p = 0.001$, $\eta^2_p = 0.235$; Figure 23), with the 2-back scenario showing significantly lower time variability ($M = 46.320$, $SD = 24.284$) relative to inter-urban ($M = 64.3649$, $SD = 43.222$, $p = 0.001$), suburbs ($M = 60.032$, $SD = 37.647$, $p = 0.002$), city centre ($M = 58.243$, $SD = 35.269$, $p = 0.001$) and risk ($M = 59.243$, $SD = 33.413$, $p < 0.001$). No interaction or between-subjects effects were reported for ECG measures.

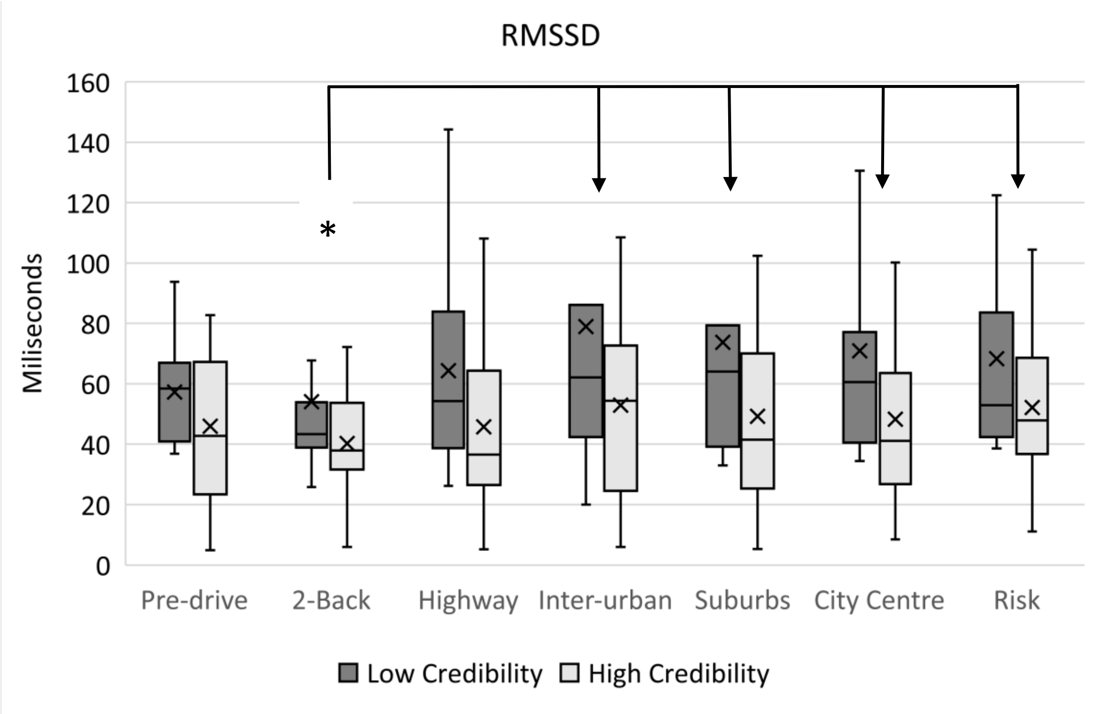


Figure 23 Heart Rate Variability as in RMSSD across driving conditions for each group. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

Other than cardiac data, EDA data analysis also supported this hypothesis. SCR count varied across scenarios ($F(6, 102) = 7.034$, $p < 0.001$, $\eta^2_p = 0.293$; Figure 24), showing more skin conductance responses (SCR count) during the 2-back condition ($M = 59.978$, $SD = 8.608$) than highway ($M = 44.464$, $SD = 5.743$, $p < 0.001$) and suburbs ($M = 47.381$, $SD = 7.183$, $p = 0.007$) scenarios. No interaction or between-subjects effects were reported for SCR count, and no effects were observed for SCR amplitudes and magnitudes either.

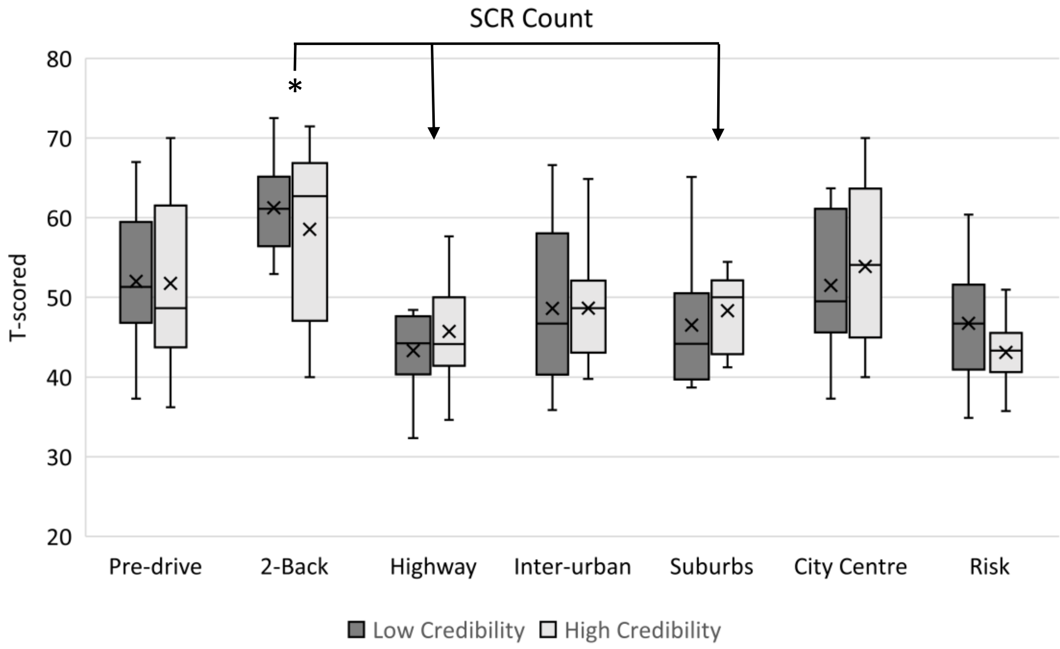


Figure 24 Skin Conductance Response count across driving conditions for each group. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

Overall, these effects observed on cardiac data indicate a vagal withdrawal and sympathetic activation, which can be described as a non-reciprocally coupled mode of autonomic control expected for high mental workload during the 2-back task.

Neurophysiological data from the fNIRS signal did not report any within-subjects effects that would support this hypothesis. However, between-subjects effects were observed, and since these were due to the grouping variable (i.e., automation credibility expectations), these will be described in section 4.5.3.

4.5.2 Hypothesis 2 – Traffic complexity

High traffic complexity scenarios (i.e., suburbs, city centre and risk event) were expected to elicit a greater stress response than pre-drive baseline and lower traffic complexity conditions (i.e., highway and inter-urban). An increment in sympathetic activity and a withdrawal of parasympathetic activity was expected as the driving conditions become more complex.

A one-way ANOVA on the Self-Assessment Manikin (SAM) ratings did not report any main effects within-subjects for arousal or valence that would support this hypothesis. Aligned with this, no statistical effects were observed supporting this hypothesis either for ECG or EDA features. On the contrary, main effects for heart rate ($F(6,138) = 34.47$, $p < 0.001$, $\eta^2_p = 0.600$; Figure 21) revealed a higher rate during pre-drive baseline than all other experimental conditions, but for the 2-back; and for the highway scenario compared to inter-urban and suburbs.

Notwithstanding, the risk condition reported one of the highest mean SCR amplitudes and magnitudes among experimental conditions, only comparable with those observed during the 2-Back scenario (see Figure 25 and Figure 26). These findings will be interpreted in the discussion section.

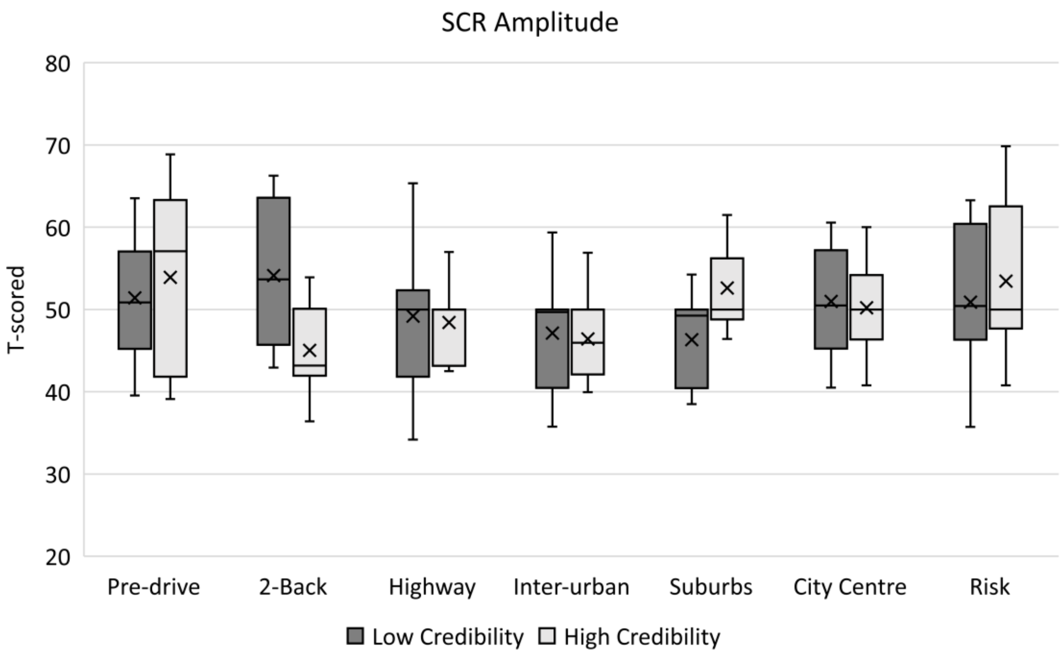


Figure 25 Skin Conductance Response amplitude across driving conditions. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

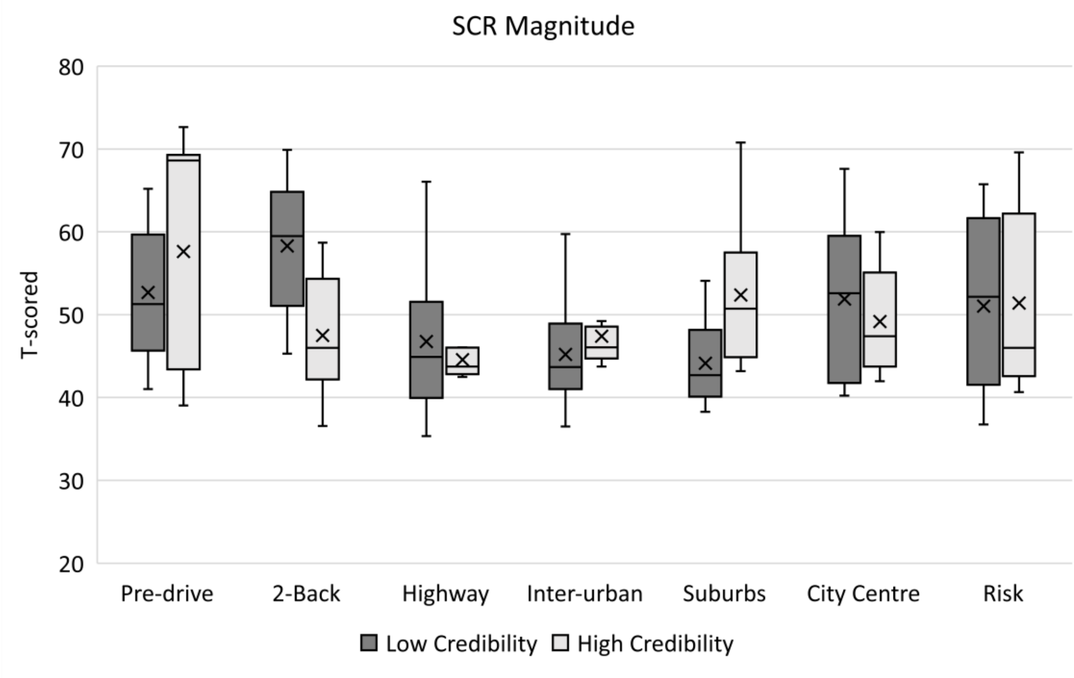


Figure 26 Skin Conductance Response magnitude across driving conditions for each group. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

4.5.3 Hypothesis 3 – Credibility expectations

It was hypothesised that induced driving automation credibility expectations would result in greater self-reported trust in automation on the trust promoted group and lower trust in the trust lowered group. By manipulating trust in automation levels through induced automation credibility expectations, group differences in neuro and psycho-physiology metrics were expected due to the additional alertness and mental resources required for automation performance monitoring in the trust-lowered group.

Self-reported TiA

A mixed 2 x 3 ANOVA was performed on the Trust subscale ratings and reported interaction effects ($F (2, 50) = 4.823$, $p = 0.012$, $\eta^2_p = 0.162$, Figure 27), but these diminish after follow-up tests. Interaction effects were also observed for the distrust subscale (ratings ($F (2, 50) = 4.961$, $p = 0.011$, $\eta^2_p = 0.166$) indicating that the low credibility (LC) group ($M = 3.333$, $SD = 1.086$) developed more distrust than the high credibility (HC) group ($M = 2.253$, $SD = 0.787$, $p = 0.006$) during the mid-study pause of the experiment. This trend remained after the experiment, evidencing the effect of induced LC expectations (M = 3.617, SD = 1.146) compared to the HC group (M = 2.627, SD = 1.289, p = 0.048) for distrust. These findings were corroborated by the Total score (F (2, 50) = 6.136, p = 0.004, η^2_p = 0.197), highlighting the detrimental effect of LC expectations (M = 4.646, SD = 0.944) compared to HC expectations (M = 5.411, SD = 0.682, p = 0.022) during the mid-study pause.

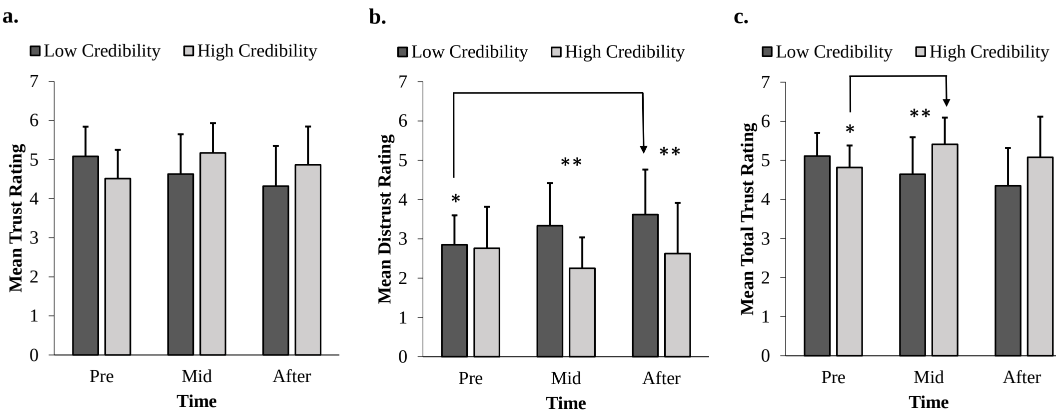


Figure 27 Mean Trust (a.), Distrust (b.), and Total (c.) ratings took pre-study, mid-study, and after completing the study for each group. Double asterisks (**) indicate interaction effects, single asterisks (*) indicate main effects.

Aside from the expected differences between groups, main effects indicating a decrement on trust and increase of distrust were observed for the LC group (F (2, 50) = 4.961, p = 0.011, η^2_p = 0.166). Particularly, distrust varied significantly within PRE (M = 2.850, SD = 0.749) and AFTER (M = 3.617, SD = 1.146, p = 0.041). Main effects were also reported for total trust ratings within the LC group (F (2, 50) = 6.136, p = 0.004, η^2_p = 0.197), but these diminished for post-hoc pairwise comparisons.

On the contrary, the opposite effects were observed within the HC group. Total trust ratings varied significantly across the experiment for the HC group (F (2, 50) = 6.136, p = 0.004, η^2_p = 0.197), having increased at the MID-study pause (M = 5.411, SD = 0.682) compared to PRE-experiment (M = 4.817, SD = 0.564, p = 0.031).

These findings indicate that TiA levels were aligned with credibility as expected, and henceforth, it can be assumed that TiA levels were manipulated successfully. Group differences in physiological data were found in brain activity from fNIRS but not in cardiac or skin conductance data. Therefore, the following results present these findings from fNIRS data between both groups.

Oxygenated haemoglobin concentrations (HbO)

Between-subjects effects for oxygenated haemoglobin concentrations (HbO) were observed in the right orbitofrontal cortex (BA10 right) ($F(1, 22) = 9.096$, $p = 0.006$, $\eta^2_p = 0.293$, Figure 28). Follow-up pairwise comparisons revealed greater HbO concentrations for the LC group during city centre ($M = 2.049$, $SD = 1.306$, $p = 0.001$) and risk scenario ($M = 2.444$, $SD = 2.865$, $p = 0.011$); compared to the HC group ($M = 0.458$, $SD = 0.597$; and $M = -0.003$, $SD = 1.094$, respectively). A further exploration also revealed significant HbO variations within the LC group between baseline ($M = 0.001$, $SD = 1$ $p < 0.001$) and city centre.

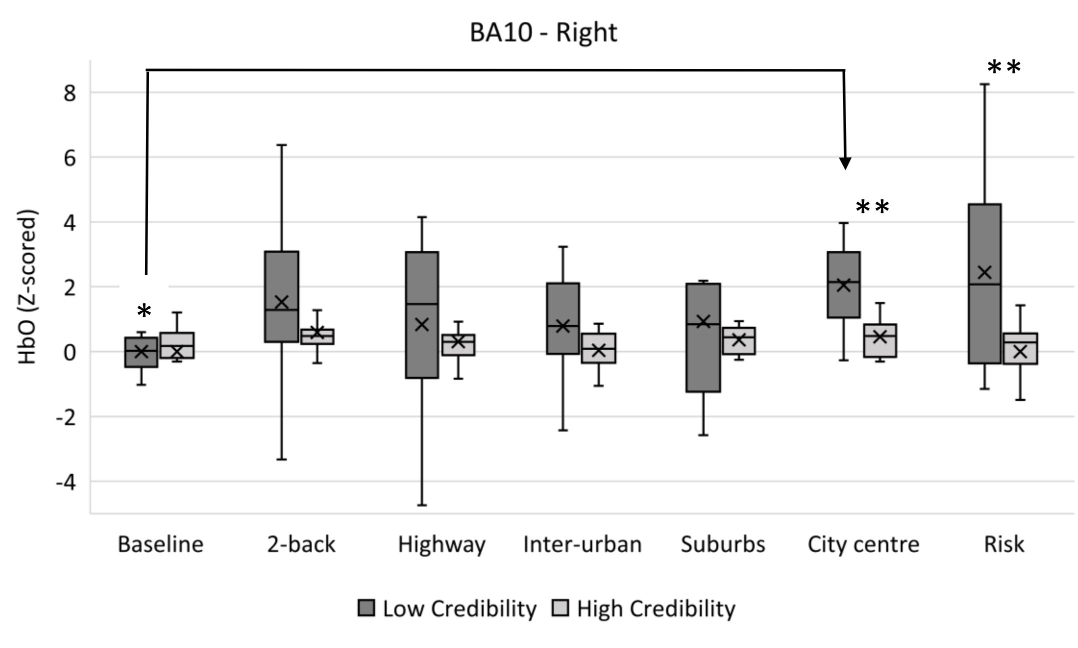


Figure 28 HbO levels in BA10 - Right across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

A condition by group interaction effect was observed for HbO in the left ventrolateral prefrontal cortex (BA44) ($F(6, 132) = 2.767$, $p = 0.034$, $\eta^2_p = 0.112$, Figure 29), followed by between-subjects effects ($F(1, 22) = 11.243$, $p = 0.003$, $\eta^2_p = 0.338$. These effects revealed higher HbO concentrations for the LC group during 2-back ($M = 1.401$, $SD = 2.174$, $p = 0.020$), inter-urban ($M = 0.442$, $SD = 1.027$, $p = 0.038$) and risk ($M = 1.302$, $SD = 2.636$, $p = 0.019$)

conditions compared to the HC group (in order: M = -1.088, SD = 2.655; M = -0.469, SD = 0.997; M = -1.082, SD = 1.898).

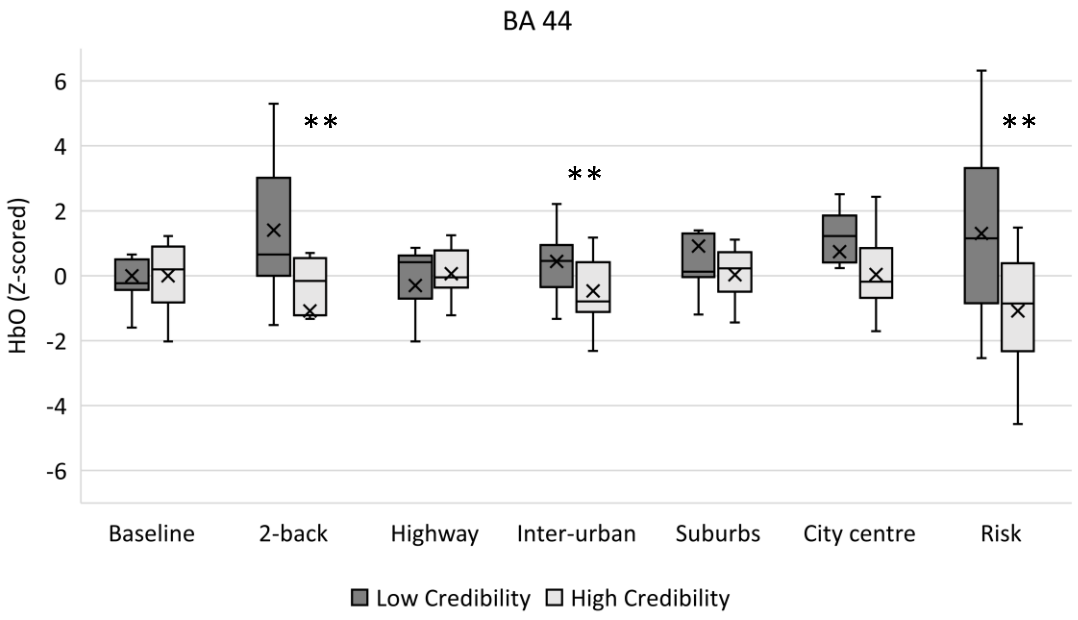


Figure 29 HbO levels in BA44 across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

Interaction effects for HbO levels were also observed in the right dorsolateral prefrontal cortex (BA46-R) (F (6,132) = 2.461. p = 0.027, η^2_p = 0.101, Figure 30), followed by between-subjects effects (F (1, 22) = 4.819, p = 0.039, η^2_p = 0.180), and revealing higher HbO concentrations for LC group (M = 1.488, SD = 1.572, p = 0.003) during the risk event compared to the HC group (M = -0.765, SD = 1.732).

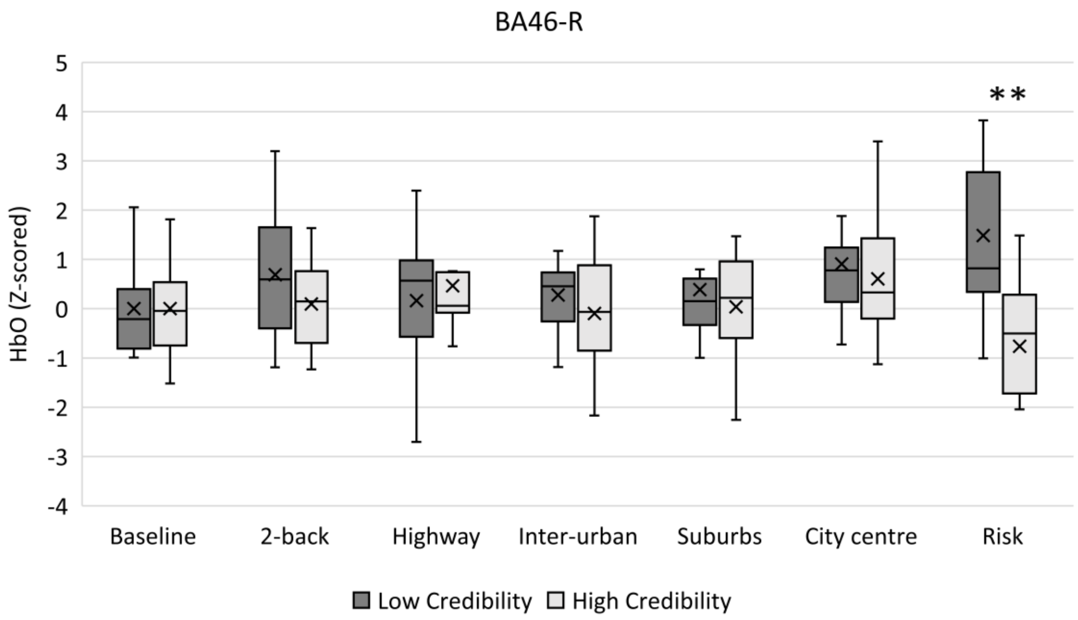


Figure 30 HbO levels in BA46 - Right across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

Deoxygenated haemoglobin concentrations (HbR)

Between-subjects effects were observed for deoxygenated haemoglobin (HbR) in the left dorsolateral cortex (BA09 - L) (F (1, 22) = 4.577, p = 0.044, η^2_p = 0.172), but these diminished in follow-up pairwise comparisons. Similar group effects were also observed in the right orbitofrontal region (BA10 – R) (F, (1, 22) = 16.322, p < 0.001, η^2_p = 0.470, Figure 31), reporting higher HbR concentrations for the LC group during inter-urban (M = 0.394, SD = 0.870, p = 0.026) and city centre (M = 1.050, SD = 1.495, p = 0.010) scenarios, compared to the HC group (inter-urban M = -0.247, SD = 0.336; M = -0.192, SD = 0.279). Follow-up tests also revealed HbR variations within the LC group from suburbs (M = -0.064, SD = 1.498, p = 0.031) to city centre.

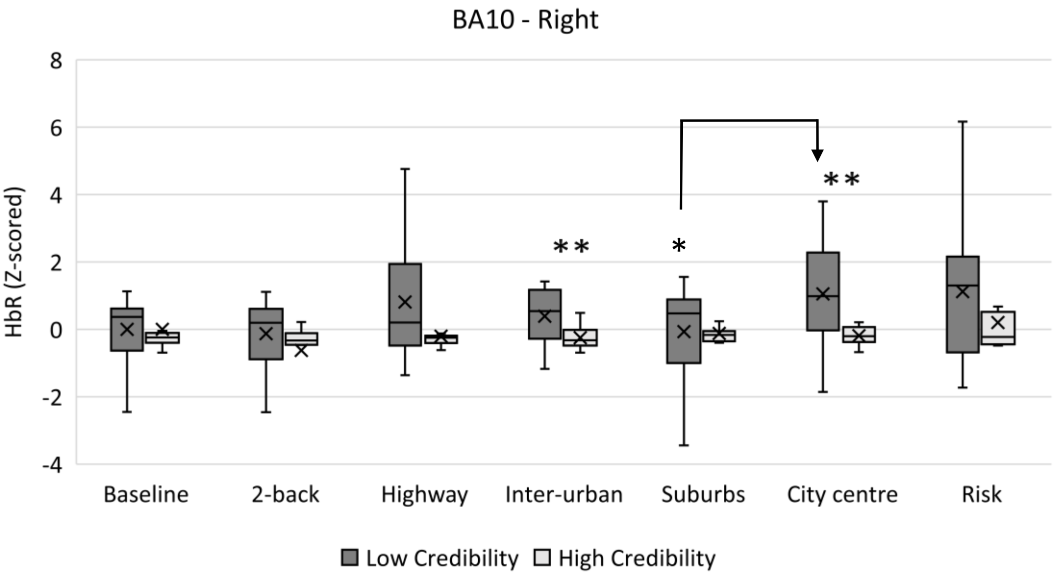


Figure 31 HbR levels in BA10 - Right across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

Total haemoglobin concentrations (HbT)

Between-subjects effects were reported for total haemoglobin (i.e., HbO + HbR) concentrations in the right dorsolateral prefrontal cortex (BA09 – R) (F (1, 22) = 4.777, p = 0.040, η^2_p = 0.178, Figure 32) during the risk event, with the LC group showing higher total haemoglobin concentrations (M = 1.412, SD = 2.001) than the HC group (M = -0.364, SD = 2.133, p = 0.047). Concomitant effects were also observed in the neighbour region (BA46 – R) (F (1, 22) = 5.489, p = 0.029, η^2_p = 0.200, Figure 33) during the risk event, with higher HbT in the LC group (M = 1.432, SD = 1.469) compared to the HC group (M = -0.371, SD = 1.766, p = 0.013).

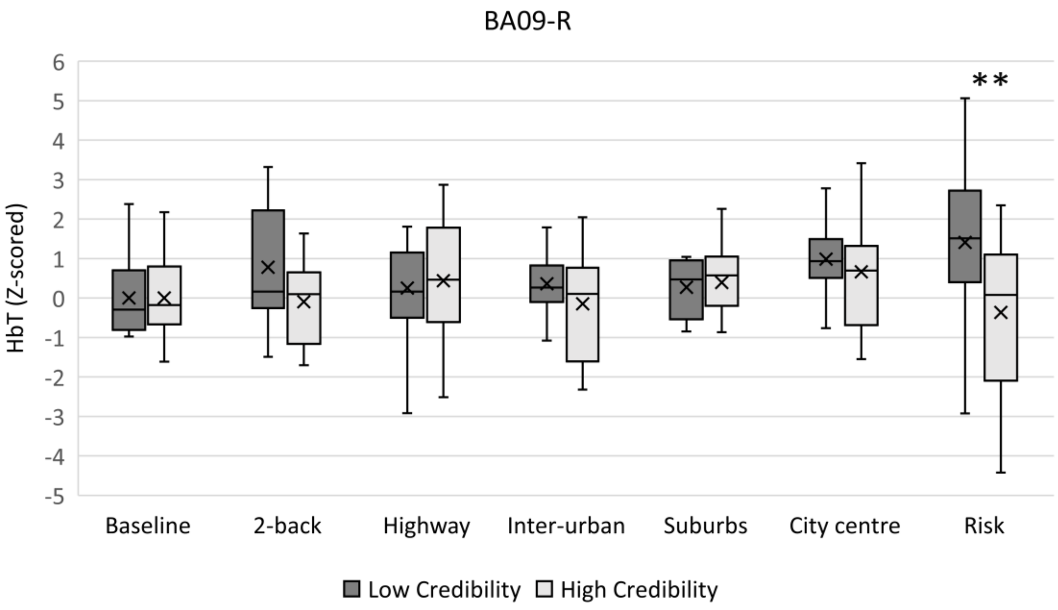


Figure 32 HbT levels in BA09 - Right across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

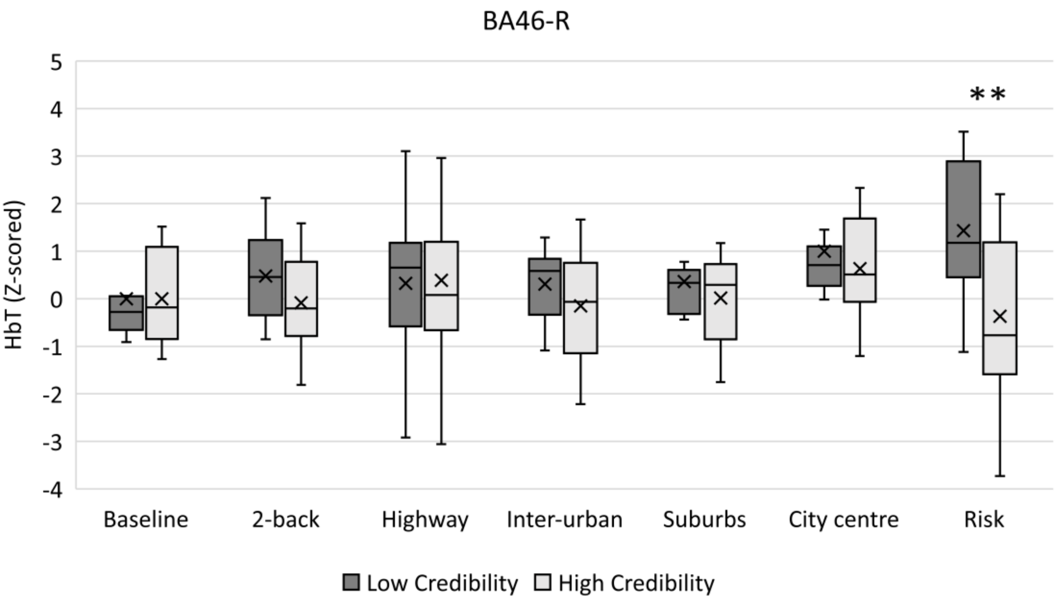


Figure 33 HbT levels in BA46 - Right across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

In addition, and consistent with findings from HbO, between-subjects effects for HbT were observed in the right orbitofrontal region (BA10 – R) ($F(1, 22) = 10.252$, $p = 0.004$, $\eta^2_p = 0.318$, Figure 34). *Post-hoc* pairwise tests showed the LC group ($M = 2.023$, $SD = 1.147$) showing higher HbT concentrations than the HC group (M = 0.641, SD = 1.301, p = 0.011) in the city centre scenario. Moreover, HbT levels varied within the LC group from baseline (M = 0.001), SD = 1.001) to city centre (M = 2.023, SD = 1.146, p = 0.002).

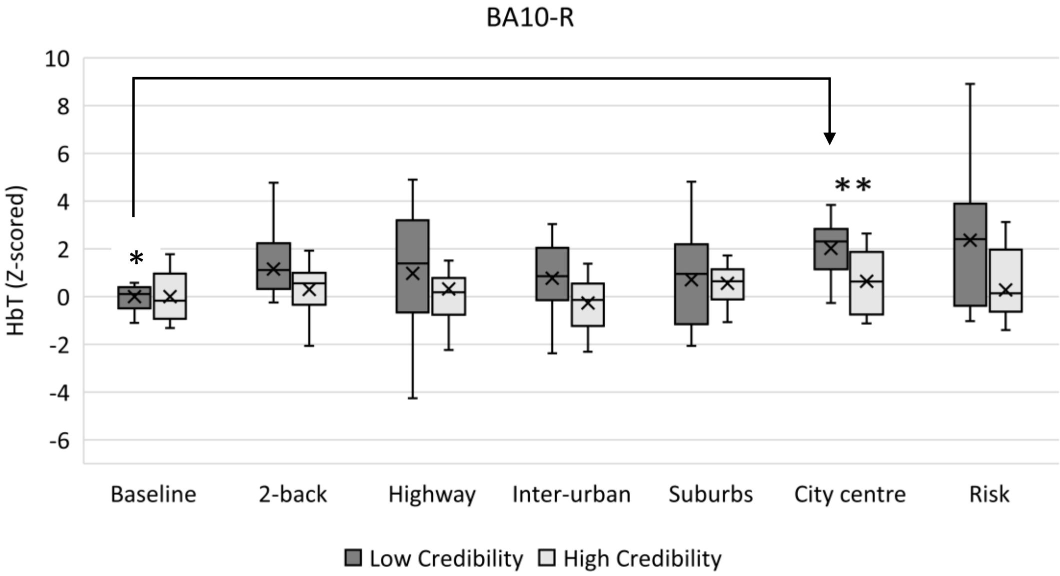


Figure 34 HbT levels in BA10 - Right across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

Finally, between-subjects effects were also found in HbT concentrations in the left ventrolateral cortex (BA44) ($F(1, 22) = 12.937$, $p = 0.002$, $\eta^2_p = 0.370$, Figure 35); with follow-up tests revealing group differences in HbT levels during inter-urban (LC M = 0.438, SD = 0.957; HC M = -0.443, SD = 1.047, p 0.043) and risk scenarios (LC M = 1.519, SD = 2.605; HC M = -1.019, SD = 2.064, p = 0.015).

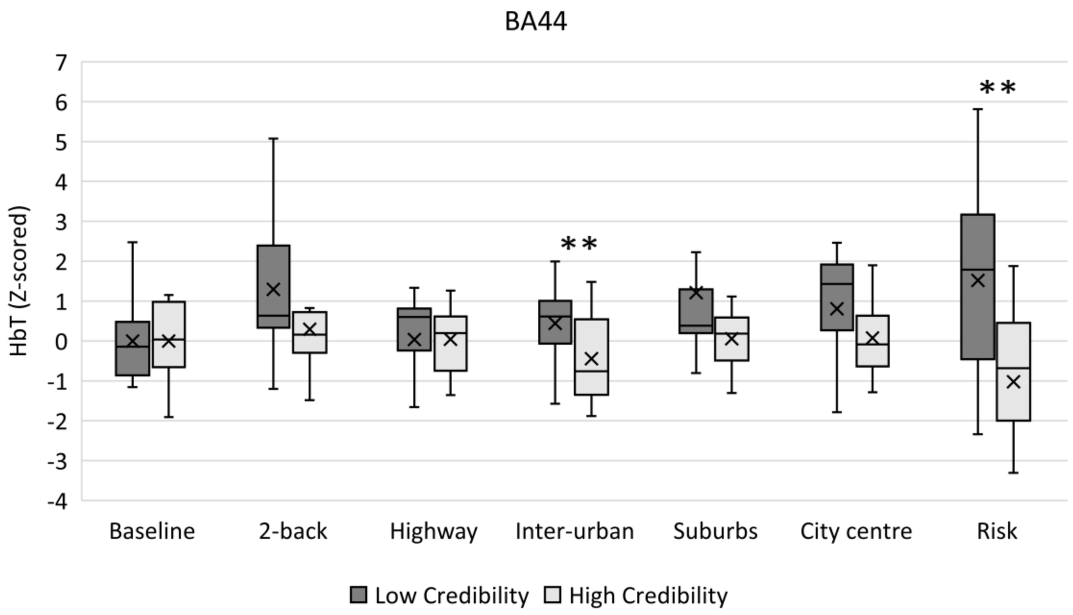


Figure 35 HbT levels in BA44 across driving conditions for each group. Single asterisks (*) indicate significant effects within participants and arrows their direction. Double asterisks (**) indicate significant effects between groups. Mean is indicated by the (X).

4.5.4 Summary

The main findings from this study are summarised as follows:

- The mental workload elicited from the 2-back task was captured with all cardiac features and produced a greater arousal response than the risk event.
- Traffic complexity did not elicit any effects on peripheral physiology metrics within participants.
- The trust lowered group reported more distrust across the study and showed increased brain activity during inter-urban, city centre and risk scenarios compared to the trust promoted group, which reported higher trust and showed lower brain activation during these scenarios.

4.6 Discussion

The present study examined the effects of task engagement, traffic complexity, and automation credibility expectations on drivers' mental workload, stress and trust in automation, respectively, during a simulated HAD cruise; by using a combination of self-reported scales, neuro and psychophysiological parameters.

4.6.1 Hypothesis 1 – Task engagement

Hypothesis 1 predicted that performing the 2-back task would produce a higher mental workload, resulting in increased brain activity, sympathetic response and parasympathetic withdrawal.

Substantial evidence favouring H1 was found in self-reported data and supported by all cardiac measures as expected. Sympathetic dominance during the 2-back scenario was indicated by greater LF/HF ratio and Heart Rate (HR) and reciprocally coupled by the deactivation of the parasympathetic branch as RMSSD indicated (see Figure 21, Figure 22, and Figure 23). This mode of autonomic control measured through cardiac indicators while performing mental tasks has also been found in previous research in a driving simulator (Lenneman & Backs, 2009; Melnicuk et al., 2021) and agree with the broader literature that the HR and LF/HF ratio increases with greater stress, whilst RMSSD decreases (Berntson & Cacioppo, 2004; Cowley et al., 2016; Kageyama et al., 2007; Kim et al., 2018).

However, it appears to contradict other driver behaviour studies that have used LF/HF ratio as well but found the opposite result under high stress driving scenarios (Brookhuis et al., 2009; Melnicuk et al., 2017). The LF/HF ratio is recognised to be a controversial measure, with variations in studies showing both increases and decreases in value for high-stress scenarios (Berntson & Cacioppo, 2004; Laborde et al., 2017; Shaffer & Ginsberg, 2017). LF/HF ratio is not a pure index of sympathetic activity but a combination of sympathetic and vagal activity, which is known to be a non-linear and non-reciprocal relationship (Laborde et al., 2017; Shaffer & Ginsberg, 2017).

A greater number of skin conductance responses was also observed during the 2-back task supporting the effects from cardiac indices. EDA is not commonly used as an index of mental workload, although it is known to be an indicator of sympathetic activity (Boucsein, 2012; Dawson et al., 2016), which in this case could be interpreted as the SNS activation to meet the mental demands from the 2-back task. However, considering the 2-back was an auditory task requiring verbal responses, the increment in the number of SCRs could also be associated with an event-related response towards an external stimulus (i.e., the audio cue used for the task) or either to the participant's verbal responses (Boucsein, 2012; Braithwaite et al., 2015).

Finally, contrary to what was expected, fNIRS data did not report any effects within-subjects for the mental workload. However, between-subjects effects were observed in the left ventrolateral prefrontal cortex (BA44) during 2-back, indicating group differences in HbO levels (see Figure 29). HbO levels increased during the 2-back to meet task demands for the LC group, as found in previous research (Unni et al., 2018). However, for the HC group, HbO levels decreased unexpectedly. Some authors from fMRI research have associated this mirrored trend with neural suppression and blood flow redistribution during task execution from the reallocation of cognitive processing resources, also known as the "steal effect" (Obrig, 2010; Quaresima et al., 2012; Wenzel et al., 2000). Remarkably, this mirrored trend occurred in Broca's area, mainly known for language processing (Quaresima et al., 2012), during a verbal working memory task. The researcher ensured that participants from both groups kept engaged in the task. Hence perhaps the observed localised deactivation in BA44 during the 2-back task for the HC group was presumably due to a reallocation of cognitive resources towards other neural regions rather than a signal of poor task performance.

Overall, these results agree with the wider literature that cardiac features (HR/V) are particularly sensitive for measuring mental workload fluctuations derived from NDRTs during HAD and thus appear to be a promising complement for eye-tracking in multimodal DSMs. Moreover, our results suggest that users involved in verbal working memory tasks –i.e., a demanding phone conversation or talking with other passengers may develop high arousal levels, and this may be useful for preventing drivers' drowsiness states towards the preparation for a take-over request (Miller et al., 2015). However, this interpretation must be carefully taken since such tasks may be counter-indicated and overload drivers in certain demanding conditions.

4.6.2 Hypothesis 2 – Traffic complexity

Compared to pre-drive baseline and lower traffic complexity conditions, it was expected that scenarios with greater traffic complexity would increase drivers' stress – i.e., increased sympathetic activity and parasympathetic withdrawal-

No effects were observed supporting this hypothesis for ECG or EDA features. Nonetheless, overall means suggest a progressive increase of arousal during city centre and risk scenarios, where the greatest SCR amplitudes and magnitudes were recorded. These would be aligned with previous findings (Beggiato et al., 2018, 2019; Foy & Chapman, 2018; Healey & Picard, 2005; Melnicuk et al., 2017; Radhakrishnan et al., 2020) suggesting that driving environment demands may be a factor modulating stress levels in HAD, although not as strongly as in manual driving. Additional evidence favouring this assumption was observed in right orbitofrontal (i.e., BA10-R) activity variations reported within driving conditions only for the LC group. This finding was consistent among all three chromophores suggesting this brain region might be involved in trust calibration due to environmental changes (Palmer et al., 2020).

However, the robustness of these findings is limited since significant effects in ECG or EDA were not found. Further research should elucidate whether the scenarios' length and order, the number of instructions provided, the manual driving induction, or the 2-back task affected physiological activity. Another possibility is considering whether the driving simulator scenarios were not sufficiently realistic to induce any risk perception and stress.

The lack of effects might present another problem for implementing DSM systems, where changes in arousal levels from low to moderate or low to high can be detected with these metrics, but not from moderate to high arousal scenarios. Indeed, (Beggiato et al., 2019) noticed a similar issue but in the opposite direction as:

> "Physiological reactions could be observed for situations with specific events that provoke moderate to high discomfort". […] Longer lasting and slowly evolving situations with moderate to low reported discomfort did not show associated changes in physiology and can therefore hardly be detected by these parameters." pp. 454-455.

Perhaps in their case, this was due to the nature of their event-related design in contrast to the long-lasting and slowly evolving situations here, apart from the risk event, which, noticeably, generated a similar outcome. Machine learning methods could provide a potential solution to overcome these issues. Otherwise, the limitations of ECG and EDA methods could also be supplemented with eye-tracking techniques. Further research is needed to determine the effectiveness of ECG and EDA indicators in detecting stress fluctuations due to traffic conditions under driving automation conditions, and particularly those intermediate variations from low to moderate or high to moderate because moderate levels of arousal would be those required for performing safe take-overs.

4.6.3 Hypothesis 3 – Credibility expectations

Induced driving automation credibility was expected to promote trust in one group and lower trust in the other group. In the trust-lowered group, increasing electrodermal, cardiac and brain activity metrics due to the additional alertness and mental resources required for automation performance monitoring.

Results from the self-reported TiA scale indicate that induced driving automation credibility had the expected effect on TiA. This is in agreement with previous studies, which have also provided information regarding the driving automation as an independent variable to manipulate TiA (de Visser et al., 2018; Khastgir et al., 2019; Körber et al., 2018; Li et al., 2019; Walker et al., 2019).

On a related note, it was expected that trust levels would modulate users' engagement and monitoring behaviours and thus, induce group differences in electrodermal, cardiac and brain activity levels.

Contrary to previous studies (Morris et al., 2017; Walker et al., 2019), no differences in electrodermal activity or cardiac activity were observed that could be associated with trust levels. Several factors may explain this lack of findings. Perhaps prior expectations or knowledge become irrelevant in highly arousing events like the risk scenario. Another plausible explanation may be that the method of inducing driving automation credibility expectations was not optimal for eliciting observable differences between credibility groups in arousal levels. For example, Walker et al. (2019), who used a descriptive video to induce driving automation capabilities on their participants, found effects in EDA that correlated with self-reported TiA.

Notwithstanding, previous work has also reported similar findings to those from the present study. In a driving simulator study with a vehicle running HAD, Du et al. (2019) programmed the vehicle to provide explanations either after or before its actions. They found TiA to increase with provided explanations, but no effects were observed in (self-reported) anxiety and workload. Indeed, it has been found that self-reported TiA may not necessarily be reflected on participants' physiology or behaviour (Kircher et al., 2014; Miller et al., 2016), and thus, further research should explore the correlation between self-reported TiA and psychophysiological responses to understand this phenomenon better.

It is worth mentioning here a series of studies measuring drivers' discomfort through psychophysiological indices across several driving automation simulated scenarios; because they share numerous similarities with the present study, and their findings could also be potentially transferable to the TiA literature, in particular to distrust:

- First, they state that discomfort could lead to safety-critical situations in automated driving, mainly due to unnecessary take-overs (Beggiato et al., 2018; Radhakrishnan et al., 2020). Within the TiA literature, unnecessary disengaging from an automated system in critical situations is known as *disuse* and is usually produced by distrust (Muir, 1987; Parasuraman & Riley, 1997; Riley, 1996).
- Second, several concepts associated with discomfort in Beggiato et al. (2018, 2019) and Radhakrishnan et al. (2020) are extraordinarily similar to those historically attributed to TiA (Hoff & Bashir, 2015; Lee & See, 2004). E.g., “the definition of comfort is rather broad, and shows similarities and overlap with related concepts of stress, mental workload, alertness, anxiety, fear, motion sickness or anger. […] As the human role in automated driving shifts from active driver to a user, additional psychological determinants of driving comfort are discussed, such as apparent safety, trust in the system, feelings of control, the familiarity of driving manoeuvres, and information about system states and actions.” (Beggiato et al., 2019) p. 446.
- Third, Beggiato et al. (2018, 2019) also instructed their participants not to take over manual control to induce driver vulnerability. Vulnerability has been identified as a determinant factor leading to trust/distrust (Khastgir et al., 2017; Lee & See, 2004).
- Fourth, they also assume that drivers’ arousal will increase along with the complexity and unpredictability of the situation and the uncertainty about the vehicle capability to deal with the task (Beggiato et al., 2018, 2019; Radhakrishnan et al., 2020). As done in the present study with credibility expectations, uncertainty about the system capability is another key factor leading to trust/distrust (Khastgir et al., 2017; Lee & See, 2004). Moreover, it is also hypothesised here that distrust would be analogous and increase under stress situations, and we have observed an increase of distrust/decrease of trust after the experiment, along with the higher traffic complexity conditions.
- Finally, yet importantly, the commonalities between our driving scenarios are equally remarkable. We all have followed an almost identical approach creating traffic complexity through infrastructure-related factors (i.e., complex intersections, roundabouts, highway exits), ambiguous behaviours from other road users -some of them vulnerable such as children inappropriately crossing and bicycles-, or unpredictable behaviours from the ego vehicle like avoiding obstacles, overtaking

buses at the bus-stop with traffic ahead. Likewise, we manipulated external factors like adverse weather conditions in some scenarios.

Overall, this makes the results from all four studies transferable and suggests that discomfort and distrust may manifest similarly on psychophysiology, or perhaps that the burdens between both constructs are not so clear.

Finally, regarding the results observed from neurophysiological indices obtained with fNIRS, group differences across several brain areas suggest different cortical activation patterns, which could be attributed to the opposing TiA levels induced.

Increased oxygenated haemoglobin (HbO) concentrations in the orbitofrontal cortex (i.e., BA10 right) have been associated with the uncertainty of judging the credibility of an unmanned vehicle (Palmer et al., 2020). The orbitofrontal and anterior cingulate cortex have been found to play a crucial role in intentional engagement (Dimoka, 2010). Hence suggesting the LC group in our study was possibly judging the credibility of the driving automation, calibrating their TiA, and maybe even had the intention of taking-over manual control during city centre and risk scenarios. This statement would agree not only with self-reported distrust (see Figure 27) but also with the variations observed exclusively within this group from baseline to city centre, indicating an increase in brain activity (↑HbO and ↑HbT) in BA10-right, possibly due to the uncertainty generated by increased traffic complexity (see Figure 28 and Figure 34).

Similarly, results from Palmer et al. (2020) also found that increased HbO in the ventrolateral prefrontal cortex (including BA44) could be implicated in the development of distrust as a result of poor decision making, as earlier noted by Hubert et al. (2018). Likewise, this area has also been associated with suspicion during computer malfunctions (Hirshfield et al., 2014), deliberate deception – lying – (Bunce et al., 2005), and frustration during automated driving (Damm et al., 2019). The ventrolateral prefrontal cortex is very proximal to the insular cortex, an area triggered by intense emotions such as fear and anticipation of negative consequences, and which has been associated with distrust (Dimoka, 2010; Hubert et al., 2018). In the present study, BA44 in the ventrolateral prefrontal cortex was found more active (↑HbO and ↑HbT) during inter-urban and risk scenarios for the Low Credibility group (see Figure 29 and Figure 35), which especially in the case of the risk scenario, would agree with the broader literature linking this area with distrust and intense negative emotions.

Finally, a lateralised activation (↑HbO and ↑HbT) in the right dorsolateral prefrontal cortex (BA09-R and BA46-R) was also observed in the LC group only during the risk scenario (see Figure 30, Figure 32, and Figure 33). Once more, this seems to agree with those findings from Palmer et al. (2020). They also found increased HbO in the right DLPC, particularly in BA46, when participants judged the credibility of the unmanned vehicle's abilities under assisted manual and assisted automated control. Relatedly, Hubert et al. (2018) associated the DLPC with reflective processes and deliberate decision-making during the evaluation of trustworthiness. The right DLPC has also been critical for visuospatial working memory, visuomotor mapping and vigilance while driving (Anderson et al., 2007; Bruno et al., 2018). In particular, Bruno et al. (2018) found bilateral DLPC increases in HbO during incongruent vehicle dynamics, thus supporting the role of the right DLPFC in judging vehicle performance and possibly the calibration of situational TiA.

Overall, these findings agree with the broader literature that distrust is quick and episodic (i.e., event-related) and linked to emotional brain mechanisms. Lateralised orbitofrontal and dorsolateral prefrontal structures play an active role in trust calibration during uncertain situations, and thus, that situational calibration of trust increases attention and monitoring. These results are promising for future research investigating TiA processes since they build up evidence favouring the two separate processes for trust and distrust. Where distrust being event-related and strongly tied to affective mechanisms might be measurable with EDA metrics as previous literature suggests, and thus potentially integrated into multimodal DSM systems. Even though our results were not robust enough to support such a statement, it might be due to other variables like the sensitiveness of the metrics used (as discussed in the previous section) or the realism of the driving simulation. Trust calibration processes seem to increase monitoring and working memory, supporting the view that TiA and situation awareness are strongly related during driving automation usage (Endsley, 2017; Large et al., 2019), but to date, evidence is not sufficient to suggest it could be detected with eye-trackers or ECG in a DSM system. Finally, in this study, we have not measured overtrust/over-reliance since take-over was not required.

4.7 Limitations and future work

Findings from this study have answered some questions, but more importantly, have raised more. In the pathway towards the next generation of DSM systems for HAD, these findings highlight the potential of cortical activity, skin conductance and cardiac-based monitoring devices for detecting various driving states determinants for driver availability.

This first study has also allowed us to identify several drawbacks which should be considered for further research to enhance their findings. Physiological readings are prone to motion artefacts, and the experimenter must take careful considerations to avoid or mitigate them during the experiment. Adequate management of participants before starting the experiment is also fundamental to avoid artificial baseline readings and is perhaps often overlooked. In this sense, an interesting issue arose with the generally high levels of physiological activity when participants were seated in the simulator vehicle prior to the scenarios starting – i.e., pre-drive baseline. It is reasonable to think that a combination of the Hawthorne effect (i.e., the overall novelty of the experience) and the number of initial instructions delivered to participants (i.e., instructions for manual driving and transition to automated driving, 2-back task, recommendations for mitigating motion artefacts) has been the principal cause of such high arousal levels. The complexity of this experiment, however, required detailed and extensive instructions. Thus, we would encourage further research requiring similar procedures to consider more extended baseline periods when measuring physiology in driving simulators before starting the trials to avoid similar problems. Recommendations for future work would be to extend the period of this 'active baseline' to at least 10 minutes, allowing approximately 5 minutes for participants to relax and then start recording baseline for the resting 5 minutes.

Comparing epochs with different lengths might be problematic, especially for HRV parameters (Laborde et al., 2017). In our case, our epochs were two-minute-long because this is the minimum required for adequate cardiac readings (Laborde et al., 2017; Shaffer & Ginsberg, 2017). However, the risk scenario had to be shorter because maintaining risk perception for more than a minute in a driving simulator is troublesome. Perhaps the ideal would have been recording for an additional minute after the risk event, as a post-event recovery, to achieve a two-minute epoch. Hence, we would strongly recommend other authors collecting cardiac data to use equal epochs of at least two minutes of duration for repeated-measures designs.

In addition, and related to cardiac features, we would encourage researchers to standardise (e.g., transformation to z-scores) raw cardiac parameters, especially in between-subjects designs. Although unlike in the EDA literature (Boucsein, 2012; Braithwaite & Watson, 2015; Dawson et al., 2016), there is a lack of specific recommendations in this sense among the consulted guidelines of cardiac activity currently, which we consider should be explicitly stated even though some may consider this an obvious step for statistical comparison between-subjects.

4.8 Conclusions

General conclusions drafted from these findings can be summarised as:

- Both HR/V features are particularly sensitive to mental workload derived from NDRTs demands during HAD and would be a promising complement for eye-trackers in DSMs.
- HAD users may be less affected by traffic complexity, although this needs to be further investigated.
- ECG and EDA features may be suitable for detecting low to high arousal fluctuations but less sensitive to moderate variations.
- The episodic and strong emotional component of distrust observed here with fNIRS, together with evidence from previous work, suggest that EDA indices might be suitable for measuring distrust.
- Trust calibration increases monitoring and working memory as can be inferred from fNIRS results, but further research is needed to explore ways to assess it with multimodal DSM systems in tandem with eye trackers.

5 Study 2 – Measuring perceived risk from HAD conditions with physiological indicators

This work has been accepted for publication in: Perello-March, J. R., Burns, C. G., Birrell, S. A., & Woodman, R., Elliott, M. T. (2022). Physiological Measures of Risk Perception in Highly Automated Driving. *IEEE Transactions on Intelligent Transportation Systems, Special Issue on: Technologies for Risk Mitigation and Support of Incapacitated Drivers*.

5.1 Introduction

The study discussed in the previous chapter highlighted our lack of understanding of the influence of environmental factors on drivers during HAD and the sensitivity of ECG and EDA indicators to moderate arousal fluctuations. Factors concerning situation awareness (SA), such as perceived risk from the simulated driving situation and the stress associated with it, may have been modulating psychophysiological responses. Because lowered risk perception may lead to overtrust, diminish SA and thus, result in a detriment of driver availability to take-over; understanding whether these variables affected on our measures is vital for drawing interpretations on the validity of these measures to detect arousal variations during HAD, but also to understand better the role of situation awareness on trust in automation.

Hence a second study was planned to evaluate risk perception from a simulated HAD setup. The driving layout aims to generate a driving condition with low to moderate risk associated and another driving condition with moderate to high perceived risk. In addition, the presence or absence of traffic was manipulated as a grouping variable to investigate whether traffic modulates risk perception in HAD. Risk perception was measured through neuro and psychophysiological indicators and a self-reported scale.

5.2 Background

5.2.1 Effects of driving conditions in manual driving vs HAD

Environmental factors from the driving conditions affect drivers' state during manual driving. For example, stress levels have been found to increase during complex traffic conditions (Foy & Chapman, 2018; Healey & Picard, 2005), possibly due to the associated increase in risk perception.

However, recent findings in driving simulator studies suggest that the effect of environmental factors on drivers' state in HAD contexts might be less straightforward (Hidalgo-Muñoz et al., 2019). In a driving simulator study comparing psychophysiological responses to uncomfortable (challenging) driving situations, the authors found arousal variations during manual driving. However, there were no significant differences in physiological metrics between several automated driving controllers during discomfort periods (Radhakrishnan et al., 2020).

In the study presented in section 4, arousal variations could only be associated with increased mental workload derived from an NDRT. No arousal variations were observed due to changes in traffic complexity or driving conditions. Only changes in brain oxygenation in the orbitofrontal cortex (i.e., BA10-R across all three measures extracted) might have been related to the effect of driving conditions, but these were only associated with the trust-lowered group. Moreover, we expected the mental workload elicited by the 2-back task would generate comparable physiological activation to that from the risk event. However, effects in HR, LF/HF ratio, and RMSSD showed that the 2-back task elicited a stronger arousal response than the risk event, which generated a moderate arousal reaction. Although HRV metrics may not be particularly suitable for measuring rapid arousal fluctuations due to the high decay time in the cardiac signal, skin conductance metrics should have captured these rapid sympathetic reactions to the risk event.

Previous work captured rapid variations in cardiac activity under several driving scenarios of high discomfort (Beggiato et al., 2018, 2019), adding more uncertainty to these findings. Noticeably, they reported that only rapid events associated with moderate to high discomfort produced observable changes in physiology. Whereas longer-lasting and slowly evolving situations with moderate to low discomfort events did not generate any observable changes in arousal.

A potential explanation for such inconsistent findings is that drivers tend to be less attentive under HAD conditions compared to when driving manually (Biondi et al., 2018; De Winter et al., 2014; Hidalgo-Muñoz et al., 2019; Melnicuk et al., 2017; Sibi et al., 2016), which has been attributed to a phenomenon known as behavioural adaptation (Jamson et al., 2013; Metz et al., 2021; Rudin-Brown & Parker, 2004). Indeed, a decrement in arousal during HAD compared to manual driving has been reported in a real-world study (Biondi et al., 2018) and in several driving simulator studies (Carsten et al., 2012; De Winter et al., 2014; Melnicuk et al., 2017). From a neurophysiological perspective, this phenomenon was evidenced by variations in HbO concentrations indicating an attentional drift during HAD compared to manual driving (Hidalgo-Muñoz et al., 2019). Relatedly, Hidalgo-Muñoz et al. also observed HAD drivers missing relevant environmental hazardous cues (i.e., brake lights from the leading car) when engaged in a secondary task, suggesting a hampered perception of potential road hazards ahead. These findings favour those reporting lower mental workload with reduced HbO concentrations when HAD users were asked to monitor driving performance compared to when the same users were driving manually (Sibi et al., 2016). Hence, it is likely that drivers may be less aware of changes in driving conditions because of delegated responsibility with the driving task. We suspect this could be related to a lowered perception of risk under automated driving conditions, which would mitigate neuro- and psychophysiological reactions to changes in driving conditions.

5.2.2 Risk perception

Perceived risk has been defined as “the likelihood and consequences of error” (Riley, 1996) and has been historically considered a key element in the literature of trust in automation (Hoff & Bashir, 2015; Lee & Moray, 1992, 1994; Lee & See, 2004; Mayer et al., 1995; Muir, 1987, 1994; Riley, 1996). Essentially, because trust implies assuming a position of vulnerability in an uncertain situation and the willingness of taking risks (e.g., the risk of delegating the control of the situation to an agent), and thus, perceived risk is considered an external modulator of situational trust, a context-dependent layer of TiA (Hoff & Bashir, 2015).

Research exploring risk perception suggests that reliance on automated systems is moderated by the risk associated with its use. Users are reluctant to engage in an automated system when the probability of negative consequences is high, and it also takes longer to reengage in higher-risk than in lower-risk conditions (Riley, 1996). In agreement, Perkins, Miller, Hashemi, & Burns (2010) found that participants reported trusting less the suggestions given by a route planner resembling a GPS in riskier situations. In an experiment using an uncrewed aerial vehicle, Satterfield et al. (2017) found participants to intervene more (i.e., automation disuse) during high-risk conditions, unnecessarily increasing their workload, but found no effects on self-reported trust.

Individual factors and personality traits may also play an important role in drivers' perception of risk. Personality traits such as sensation seeking (Jonah, 1997) or locus of control (Montag & Comrey, 1987) are factors known to modulate driving behaviour and risk perception. However, in HAD, where the responsibility of the driving task is delegated or shared, the influence of these factors is yet unclear. Individual factors such as driving experience have been related to SA (Gugerty, 2011), with experienced drivers perceiving better hazards from the road ahead (Paxion et al., 2015). Age and gender-related factors seem to be other individual factors related to risk-taking behaviours. For example, it has been found that young males seem to have a lower perception of risk, which is often associated with risky driving behaviours (Foy et al., 2016).

In the driving automation domain, only a few studies have investigated the effect of risk perception on trust, although findings seem to agree with those previously mentioned. Morris et al. (2017) found participants to report lower trust while the automated vehicle performed in a risky manner. Li et al. (2019) investigated risk assessments in different driving environments and the effect of introductory information regarding driving automation credibility on TiA and perceived risk. Their participants identified levels of risk based on traffic characteristics and abnormal behaviours from other road users. Other than that, participants reported the highest trust and the lowest perceived risk when using the system described as highly reliable and when driving in a low risky situation.

To sum up, with increased perceived risk, it is likely that TiA would be low and vice versa. For example, for drivers to engage in a secondary task while HAD is activated, their TiA must surpass their perception of potential risks in that situation, so they would be willing to be vulnerable regardless of the risks.

Therefore, understanding how driving automation users perceive risk from the driving context, and how this assessment of potential risk affects TiA is important towards developing DSM systems and deploying adequate warnings for monitoring and take-over requests. With these regards, De Visser (2012) described two processes for risk assessment that would drive the trust decision-making:

- The **recognition-based** threat assessment is a **rapid affective** evaluation of the information available.
- The **experience-based** threat assessment is a **slower, more deliberate** evaluation of the risk and benefit based on available and observable information.

However, it is essential to note here that these risk assessment processes involve the (1) the perception of the information available, (2) the comprehension of its meaning, and (3) the projection of its status in the near future, which happen to be the three dimensions of SA (see section 2.2.3). Hence, risk perception could be well described as a link between SA and TiA. The strong ties between TiA and SA have been extensively described (Endsley, 2017) and are known to modulate TiA-related monitoring behaviours such as overreliance and complacency (Parasuraman & Riley, 1997).

These findings highlight the importance of understanding risk perception under normal and abnormal HAD conditions. Accurate real-time measurement of perceived risk would be crucial for determining drivers' SA before resuming control and would also be essential to support TiA calibration in automated vehicles. Ideally, a DSM system should first determine objective risks and then verify whether drivers perceive that as a hazard and react accordingly (e.g., resuming manual control). In certain scenarios where SA is good, eye-tracking systems may be sufficient to determine that perceived risk is matching the objective risk (e.g., the driver is gazing at the road hazard). However, there might be situations where drivers may be seeing an objective risk on the road but they may not perceive it as a risk if they overtrust the system. Another critical situation could occur with complacent drivers due to poor SA or because of high mental workload. In such situations, physiology measures could serve as real-time indicators of attention allocation, stress increase, trust calibration, decision-making and strong negative emotions. Essentially, indicating the preparation for a fight-or-flight reaction associated with a perceived potential hazard.

Now the question is: how can we measure risk perception in real-time with objective indicators? The threat assessment mechanisms described earlier by De Visser offer a potential pathway for measuring situational perceived risk using both affective and cognitive indicators. Affective and cognitive indicators are accessible through ECG, EDA and fNIRS measurements; and these could be indirect indicators of situational both TiA – as discussed in the driver availability experiment presented in the previous chapter – and of SA (Bracken et al., 2021; De Winter et al., 2014). In the present experiment, we will investigate whether these indicators could be useful to detect perceived risk during several HAD situations with different types of risk associated.

5.3 Objectives and hypotheses

This study aimed (1) to investigate whether slowly evolving situations with low to moderate risk would produce observable effects on arousal and brain activity. A simulated highly automated drive involving slowly-changing driving conditions of increasing complexity was designed for this study. Moreover, the presence of surrounding traffic across these driving conditions was manipulated as a between-group variable. The second aim of this study was (2) to determine whether rapid events with moderate to high risk would produce observable arousal and brain activity variations, so a hazardous driving situation was included at the end of the trial.

According to the literature discussed in chapter 2, the ANS triggers the SNS and deactivates the PNS towards perceiving a threat or hazardous stimuli in the environment. Therefore, increased risk perception should increase sympathetic activity (HR & SCRs ↑) and reduce vagal modulation (HRV ↓). The opposite should be then expected with lowered risk perception.

The following hypotheses are proposed:

1. The presence of Traffic will increase the perception of risk and produce group differences in arousal and brain activity.
2. Changes across the Driving Conditions will progressively increase risk perception, and thus, arousal levels and brain activity will vary within participants.
3. The Driving Hazard event should rapidly increase the perception of risk and produce greater arousal and neural response compared to baseline and recovery period.

5.4 Method

5.4.1 Participants

Twenty-three volunteers were recruited to participate in this study. Three participants were excluded for analysis as they dropped out from the experiment due to motion sickness, with the data of twenty participants analysed (10 male and 10 female, mean age 24.60 years, SD = 3.91). 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.

5.4.2 Materials and equipment

This study used WMG's 3xD driving simulator at the University of Warwick (see 3.1). ECG and EDA data were recorded using BIOPAC MP160 with wearable remote Bio-Nomadix amplifiers (see 3.2), and neural blood oxygenation concentrations were recorded using NIRS.

Subjective measures included a bespoke risk perception questionnaire comprising two items which were asked after the entire drive (see Appendix 2:

(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). To my knowledge, there are no existing validated tools for risk perception assessment in the driving context. Relatedly, Li et al. (2019) used the scale developed by Rajaonah et al. (2008) to assess risk perception associated with trusting in automation. 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. In our study, we were interested in comparing perceived situational risk with the hazardous event itself during the automated driving 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 study design and as the hazardous event occurred immediately after the automated driving conditions. Stopping the scenario immediately before the Driving Hazard event could have potentially affected the realism of the scenario and any psychophysiological reactions.

In addition, the Trust in Automated Systems Scale (TASS, see 4.4.3) (Jian et al., 2000) was included to evaluate the impact of perceived risk in the build-up of trust in automation. The scale was rated before and immediately after the trial was completed.

5.4.3 Experimental conditions and automated driving scenarios

The trial lasted a total of 11 minutes and 30 seconds. This began with four minutes of baseline/resting data, four minutes of automated driving scenarios, thirty seconds of the hazardous event, and two minutes of post-event recovery. The four minutes of automated driving were split into two scenarios. An initial two-minute suburban driving scenario, labelled as Driving Condition 1 –i.e., DC1 for cardiac measures, DC1.1, DC1.2, DC1.3 and DC1.4 for SCR analyses as it was split into 30-second segments, and DC1.1, DC1.2 and DC1.3 for fNIRS split into 30 seconds segments allowing 15 seconds of recovery between each by default (see Figure 36).

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 of 30mph, including several left and right turns and give-way exits. Oncoming traffic increased to medium levels (< 20 road users per minute), now including pedestrians, cyclists, and parked cars, on the roadside and in driveways.

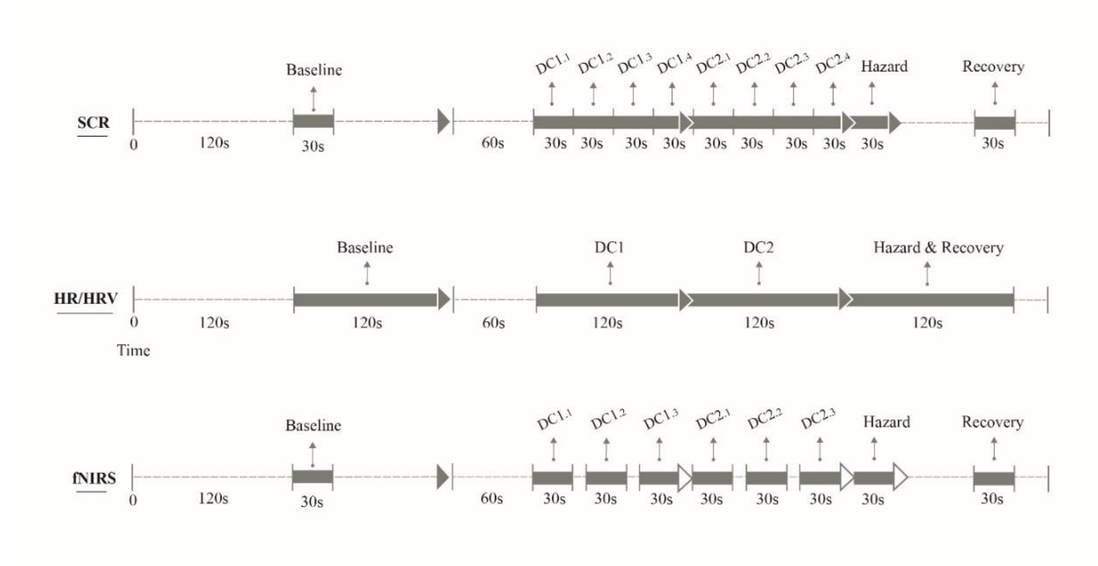


Figure 36 Experimental conditions and epochs extracted for each physiological measure.

The simulation then continued with a two-minute city centre scenario, denoted as Driving Condition 2, i.e., DC2 for cardiac measures, DC2.1, DC2.2, DC2.3 and DC2.4 for SCR analyses which were split into 30-second segments, and DC2.1, DC2.2 and DC2.3 for fNIRS which were split into 30 seconds segments allowing 15 seconds of recovery between each by default (see Figure 36). 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.

Figure 37 Representation of the Driving Hazard event in the Scenario Editor software.

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 (Figure 37). 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 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. To ensure both groups experienced the same event in terms of traffic presence, the semitrailer truck was the only vehicle present in this scenario for the Traffic group.

5.4.4 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 automated 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 both the risk perception and TASS scales.

5.4.5 Data 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 automated 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. Two participants (one from each group) were excluded for fNIRS due to significant noise in raw data (N = 18). 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 fNIRS and SCR analyses. HRV analyses used five 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 30 seconds each (see Figure 36). Data were extracted using the automated data analysis routines from Biopac's ACQKnowledge software (CA, USA; version 5.0.2). Data extraction and signal pre-processing methods can be found in section 3.2. FNIRS used nine epochs of 30 seconds each (see Figure 36).

Cardiac features extracted were heart rate (HR; beats per minute) and those heart rate variability (HRV) metrics that better reflect vagal modulation. Given that LF/HF ratio in the previous study was problematic to interpret, in the frequency-domain, this time we extracted the high frequency (HF) band instead (between 0.15 and 0.40 Hz), which is known to be a robust indicator of vagal modulation (Laborde et al., 2017; Shaffer & Ginsberg, 2017). HF was coupled with a time-domain indicator of vagal modulation according to the recommendations in Laborde et al. (2017), the root mean square of successive differences (RMSSD). In addition, cardiac data were standardised for parametric statistical analyses and to enhance the robustness of inter-individual comparisons, transforming raw data into T-scores ($M = 50$, $SD = 10$) following the same procedure described for standardising SCRs in section 3.2.

Features from Skin Conductance Responses (SCRs) were identical to those extracted in the driver availability experiment (chapter 4) (i.e., SCR count, magnitude and amplitude) and using the same data processing techniques described in section 3.2.

The same features described in section 3.3 were extracted from the fNIRS signal, identical to those used in the driver availability experiment (chapter 4) (i.e., HbO, HbR and HbT), and applying the same data processing techniques as well.

The Shapiro-Wilks 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 investigated post-hoc by pairwise comparisons corrected by the Bonferroni method. Power is indicated for main effects using Partial Eta-Squared.

5.5 Results

This study evaluates whether variations in perceived risk from either slowly evolving or rapidly evolving HAD conditions would produce observable changes in physiology.

5.5.1 Hypothesis 1 – Group differences for traffic

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. For SCR metrics, we ran a 2 x 9 mixed ANOVA (Traffic, No Traffic x BL, DC1.1, DC1.2, DC1.3, DC1.4, DC2.1, DC2.2, DC2.3, DC2.4). Although 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. The first hypothesis tested whether the slowly evolving traffic increase 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, η^2_p = 0.032); HRV-HF: (F (1, 18) = 0.024, p = 0.878, η^2_p = 0.001); HRV-RMSSD: (F (1, 18) = 0.322, p = 0.577, η^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, η^2_p = 0.097); HRV-HF: (F (2, 36) = 0.949, p = 0.397, η^2_p = 0.05); HRV-RMSSD: (F (2, 36) = 0.451, p = 0.641 η^2_p = 0.024)].

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

Brain oxygenation data from the fNIRS signal did not report any group effects for Traffic across any of the three parameters obtained.

Similarly, a Mann- Whitney *U* test on self-reported risk perception did not differ between groups for Traffic (*U* = 60.500, p = 0.436). These results indicate that the presence of traffic had no effects on the perceived risk between groups.

5.5.2 Hypothesis 2 – Within participants variations for Driving Conditions

This hypothesis tested whether slow changes across Driving Conditions would produce arousal variations within participants. To test for the effect of changing driving conditions within participants (H2), we ran 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.

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

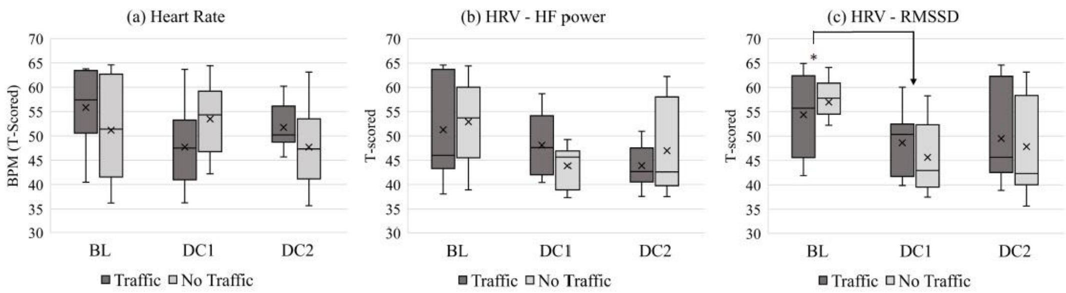


Figure 38 Cardiac features for both groups across baseline (BL) and driving conditions (DC1 and DC2). T-scored heart rate (a), HRV-HF power (b), and HRV-RMSSD (c). Asterisks (*) indicate main effects and their direction. Mean is indicated by the (X).

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, η^2_p = 0.059); SCR amplitude: (F (8, 144) = 1.419, p = 0.228, η^2_p = 0.073); SCR magnitude: (F (8, 144) = 0.369, p = 0.851, η^2_p = 0.022)].

On fNIRS data, a main effect for Driving Conditions was observed in the left orbitofrontal cortex (BA10-L) for HbO (F (6, 96) = 3.309, p = 0.018, η^2_p = 0.171, Figure 39). Post-hoc tests indicated an increase in oxygenation from BL (M = 0.000, SD = 1.000) to DC2.1 (M = 1.067, SD = 0.880). This effect was supported by a main effect for HbT (F (6, 96) = 2.736, p = 0.049, η^2_p = 0.146), although it fades with post-hoc comparisons.

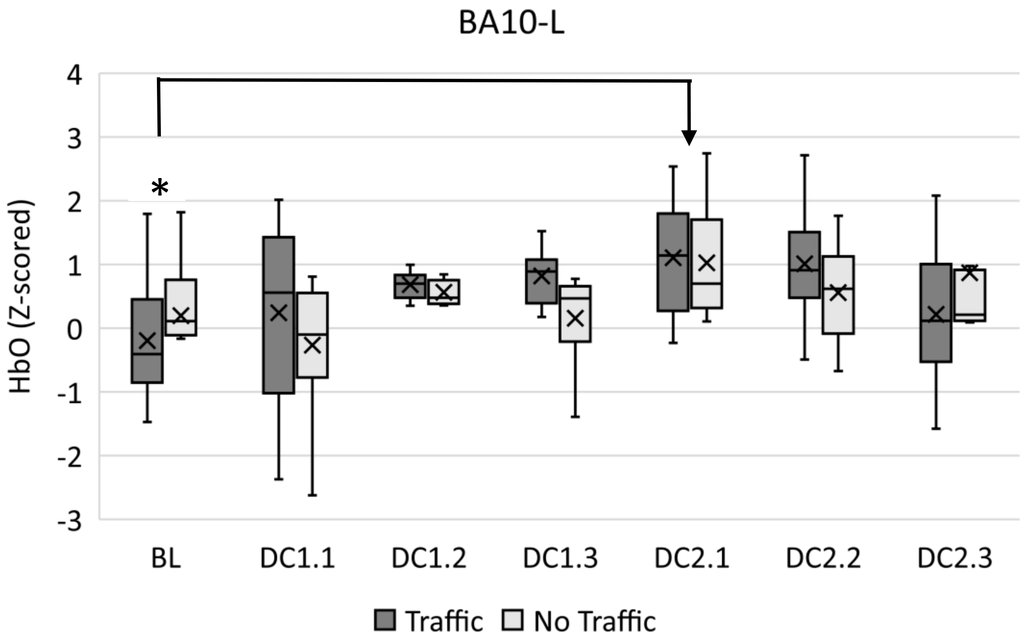


Figure 39 HbO levels in BA10 - Left across driving conditions for each group. Mean is indicated by the (X).

BA45, located in the left ventrolateral cortex, reported a main effect for Driving Conditions on HbT (F (6, 96) = 3.094, p = 0.028, η^2_p = 0.162, Figure 40), with post-hoc comparisons revealing an increase from BL (M = 0.000, SD = 1.000) to DC1.2 (M = 0.829, SD = 0.671) and DC2.1 (M = 1.061, SD = 0.554).

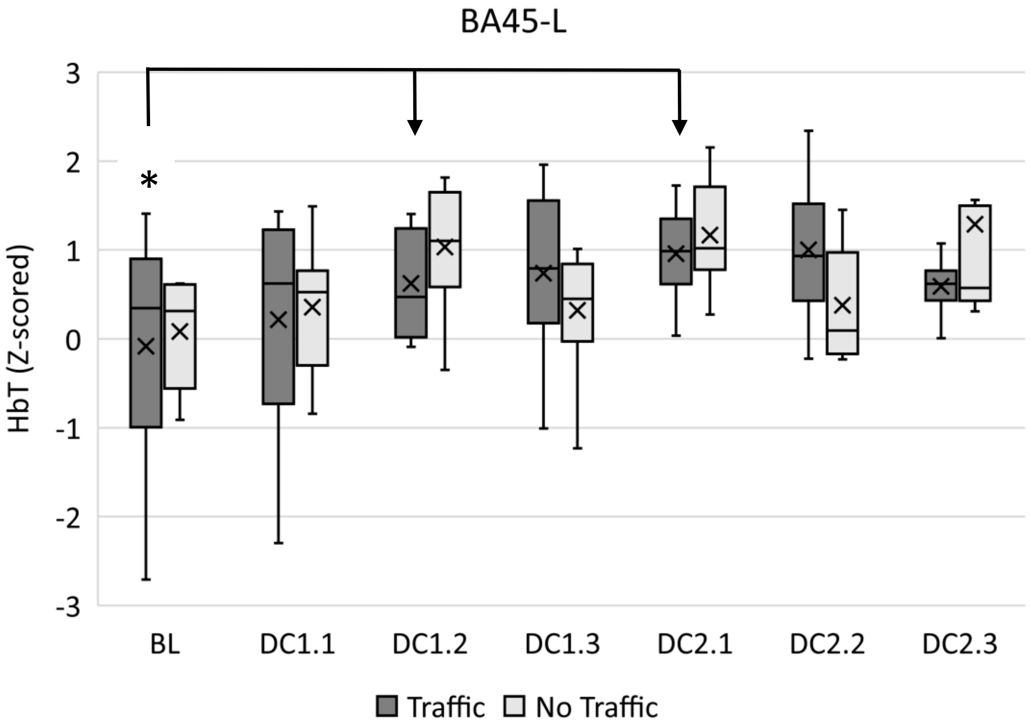


Figure 40 HbT levels in BA45 across driving conditions for each group. Mean is indicated by the (X).

BA46-L, located in the left dorsolateral cortex, reported a main effect for Driving Conditions on HbT ($F(6, 96) = 3.301$, $p = 0.034$, $\eta^2_p = 0.171$, Figure 41), with post-hoc comparisons revealing an increase from DC1.3 ($M = 0.372$, $SD = 0.592$) to DC2.1 ($M = 1.006$, $SD = 0.462$).

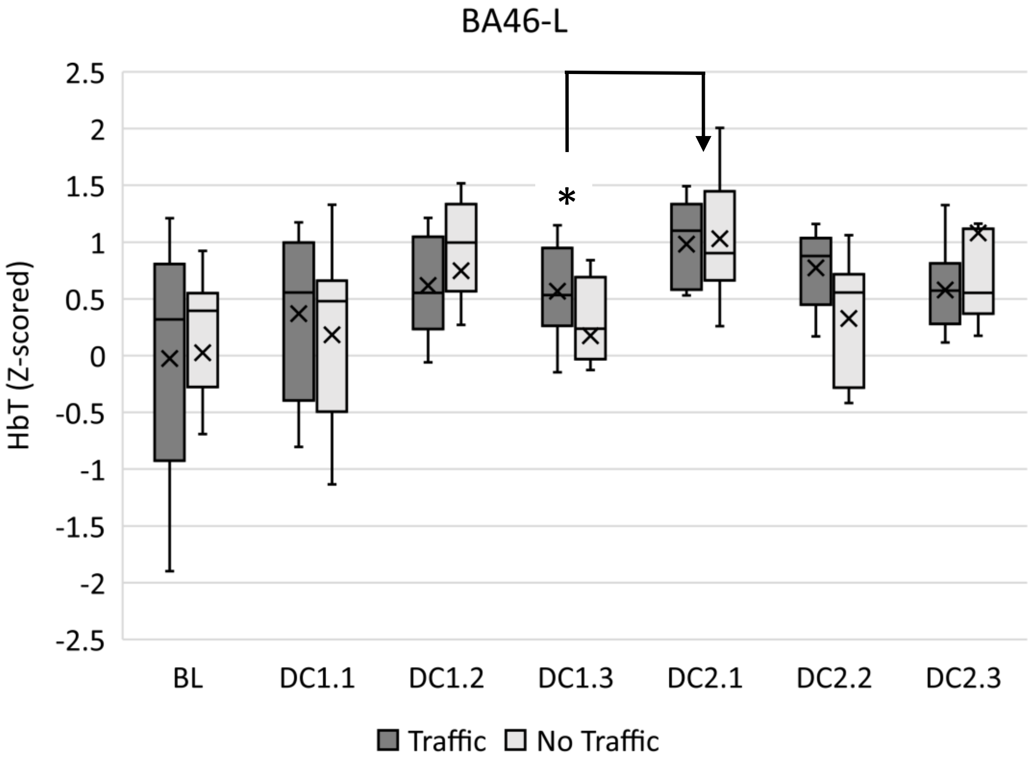


Figure 41 HbT levels in BA46 - Left across driving conditions for each group. Mean is indicated by the (X).

These results suggest that HRV, HbO and HbT features captured slowly evolving arousal variations across Driving Conditions. We will discuss these findings further in section VI.

5.5.3 Hypothesis 3 – Perceived risk from Driving Hazard

The third hypothesis investigated whether the rapidly evolving Driving Hazard event would produce observable effects in arousal indicators within participants. 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 Figure 36). We rationalised this decision as both groups experienced the same driving hazard condition. A similar repeated-measures ANOVA was performed on fNIRS metrics with three levels comparing BL with Driving Hazard and Recovery (see Figure 36).

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 dominance during Hazard-&-Recovery ($M = 48.499$, $SD = 7.984$), compared to during baseline ($M = 55.679$, $SD = 7.562$) as expected (see Figure 42). No effects were observed on HR ($F(1, 19) = 4.389$, $p = 0.050$, $\eta^2_p = 0.188$) or HRV-HF power ($F(1, 19) = 1.673$, $p = 0.211$, $\eta^2_p = 0.081$).

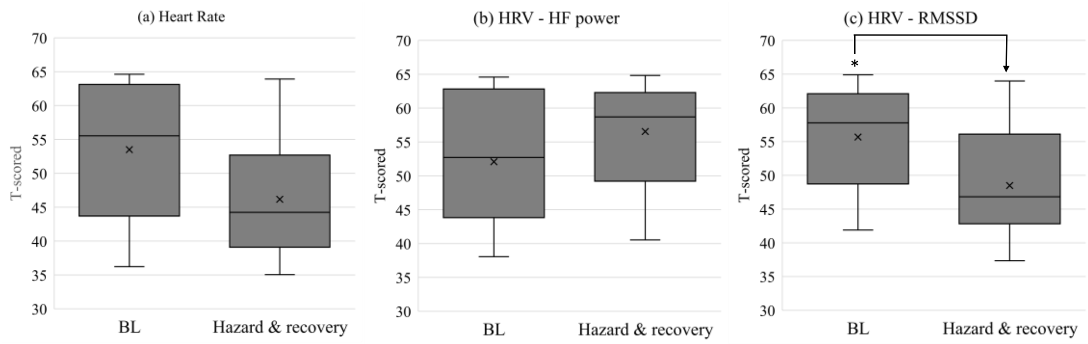


Figure 42 Cardiac features for both groups across baseline (BL) and Hazard & Recovery condition. T-scored heart rate (a), HRV-HF power (b), and HRV-RMSSD (c). Asterisks (*) indicate main effects and their direction. Mean is indicated by the (X).

Strong evidence in favour of this hypothesis was found across all SCR measures. SCR count revealed a main effect for Driving Hazard ($F(2, 36) = 10.465$, $p < 0.001$, $\eta^2_p = 0.368$, Figure 43), 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 ($M = 48.984$, $SD = 9.964$, $p = 0.012$).

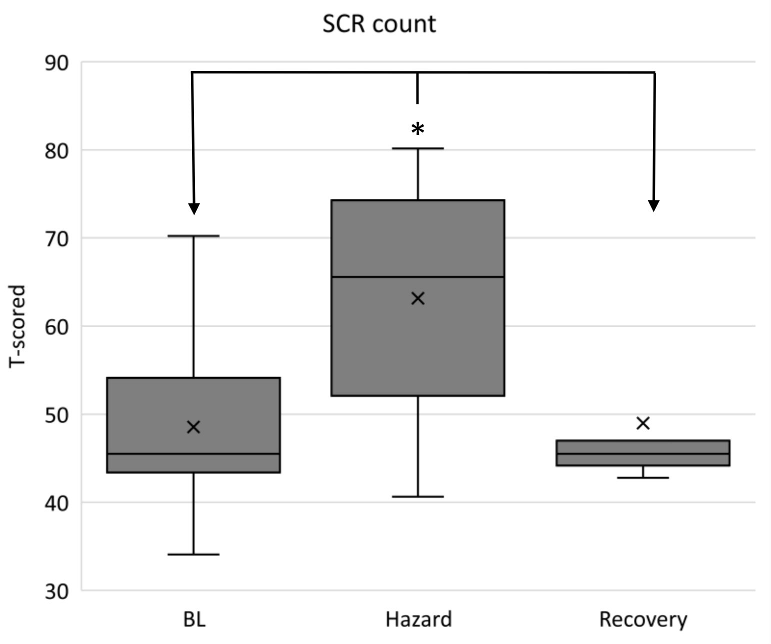


Figure 43 SCR count for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

SCR amplitude showed similar effects ($F\ (2, 36) = 22.415$, $p < 0.001$, $\eta^2_p = 0.555$, Figure 44), 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 ($M = 48.898$, $SD = 2.655$, $p < 0.001$).

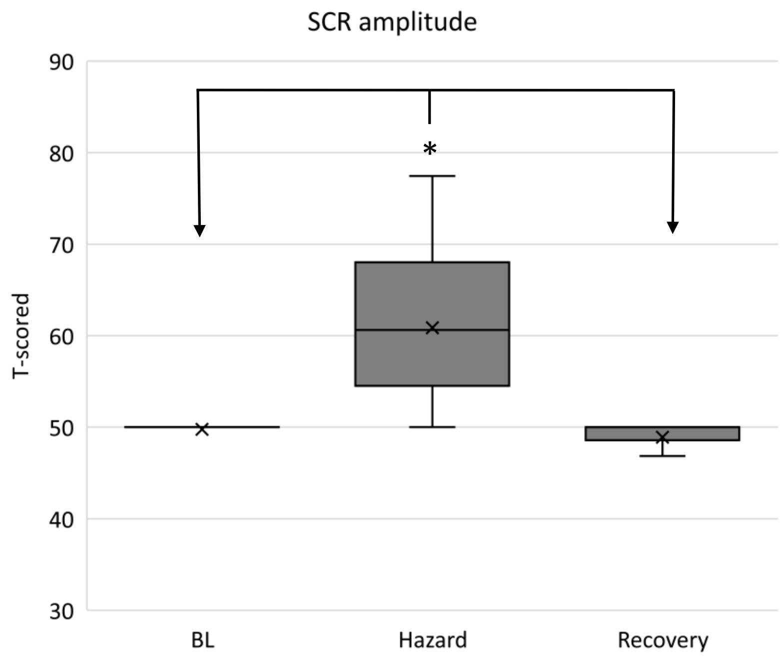


Figure 44 SCR amplitude for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

SCR magnitude was also aligned with the previous two SCR measures and revealed a main effect for Driving Hazard event ($F (2, 36) = 177.834$, $p < 0.001$, $\eta^2_p = 0.908$, Figure 45). 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 ($M = 46.373$, $SD = 3.988$, $p < 0.001$).

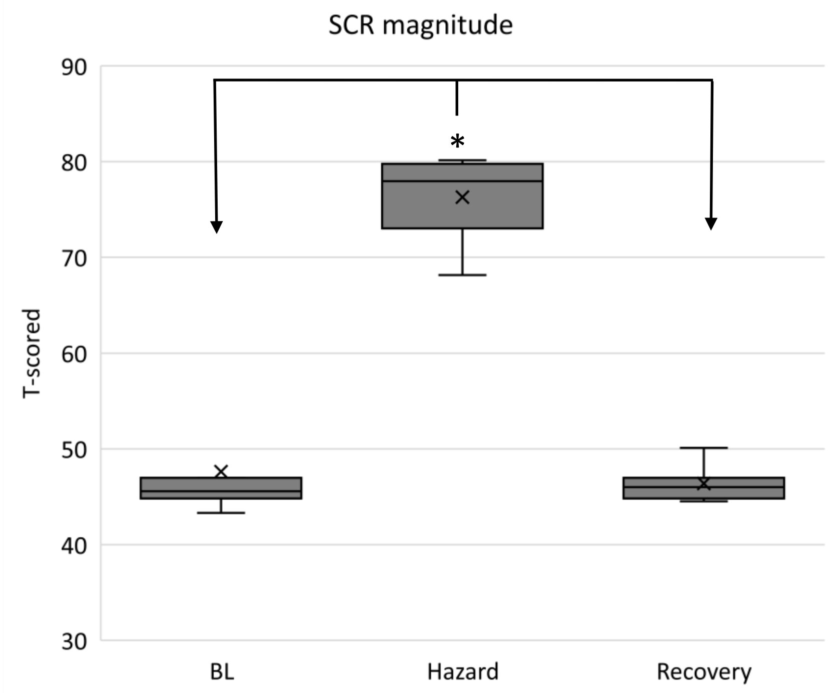


Figure 45 SCR magnitude for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

Strong evidence in favour of H3 was also found across fNIRS metrics of HbO and HbT. A main effect for Driving Hazard on HbO (F (2, 34) = 4.418, p = 0.020, η^2_p = 0.206) and HbT (F (2, 34) = 3.470, p = 0.043, η^2_p = 0.170) was found on BA09-R, but these effects diminished in post-hoc tests.

Bilateral orbitofrontal activation was observed, with BA10-L reporting a main effect of Driving Hazard on HbO (F (2, 34) = 4.663, p = 0.016, η^2_p = 0.215, Figure 46) and post-hoc tests revealing an increase from BL (M = 0.000, SD = 1.000) to Hazard (M = 1.333, SD = 1.768, p = 0.047). BA10-R reported a main effect of Driving Hazard on HbO (F (2, 34) = 5.846, p = 0.007, η^2_p = 0.256, Figure 46), with post-hoc tests indicating a similar increase from BL (M = 0.000, SD = 1.000) to Hazard (M = 1.4507, SD = 1.661, p = 0.021). This effect was also seconded by HbT (F (2, 34) = 4.118, p = 0.025, η^2_p = 0.195), although fading away with post-hoc tests.

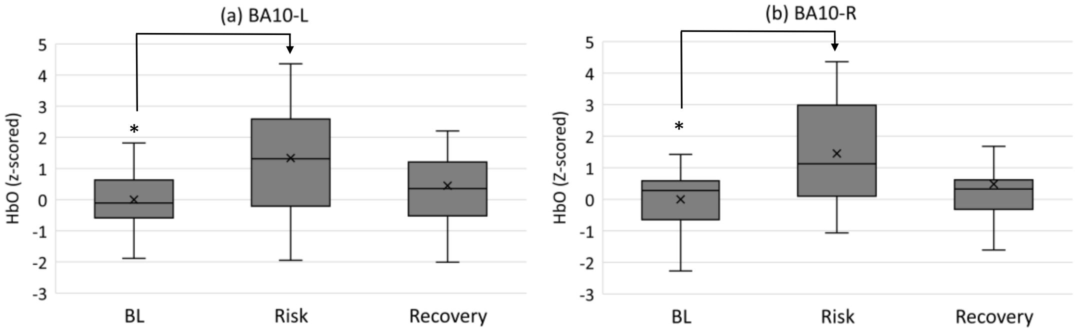


Figure 46 HbO concentrations in BA10-L (a) and BA10-R (b) for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X). Mean is indicated by the (X).

BA44 reported a main effect of Driving Hazard on HbO (F (2, 34) = 6.968, p = 0.007, η^2_p = 0.291, Figure 47) with post-hoc tests revealing an increase from BL (M = 0.000, SD = 1.000) to Hazard (M = 1.145, SD = 1.530, p = 0.019). This effect was seconded by HbT (F (2, 34) = 6.691, p = 0.011, η^2_p = 0.282, Figure 47), with post-hoc tests revealing a similar increase from BL (M = 0.000, SD = 1.000) to Hazard (M = 1.111, SD = 1.579, p = 0.030).

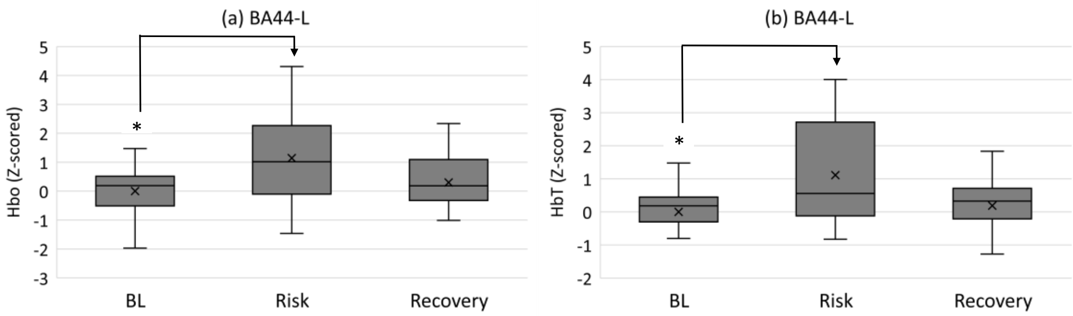


Figure 47 HbO concentrations in BA44-L (a) and HbT (b) for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X). Mean is indicated by the (X).

BA45 reported a main effect of Driving Hazard on HbO (F (2, 34) = 10.950, p < 0.001, η^2_p = 0.392, Figure 48) with post-hoc tests revealing an increase from BL (M = 0.000, SD = 1.000) to Hazard (M = 1.387, SD = 1.455, p = 0.002), and followed by a decrease from Hazard to Recovery (M = 0.436, SD = 1.060, p = 0.038). This effect was seconded by HbT (F (2, 34) = 7.559, p = 0.002, η^2_p = 0.308, Figure 48), with post-hoc tests revealing a similar increase from BL (M = 0.000, SD = 1.000) to Hazard (M = 1.162, SD = 1.503, p = 0.013).

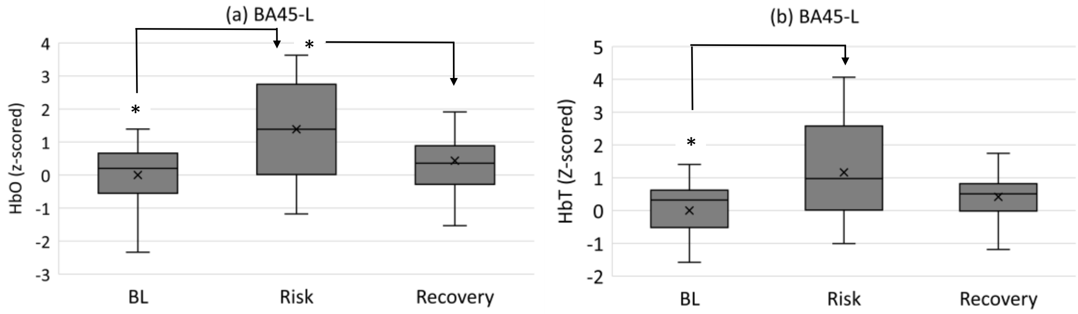


Figure 48 HbO concentrations in BA45-L (a) and HbT (b) for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

The dorsolateral prefrontal cortex also showed bilateral activity. BA46-L reported a main effect of Driving Hazard on HbO ($F(2, 34) = 11.743$, $p < 0.001$, $\eta^2_p = 0.409$, Figure 49) with post-hoc tests revealing an increase from BL ($M = 0.000$, $SD = 1.000$) to Hazard ($M = 1.434$, $SD = 1.227$, $p = 0.002$), and followed by a decrease from Hazard to Recovery ($M = 0.427$, $SD = 0.916$, $p = 0.017$). This effect was seconded by HbT ($F(2, 34) = 9.008$, $p = 0.001$, $\eta^2_p = 0.346$, Figure 49), with post-hoc tests revealing a similar increase from BL ($M = 0.000$, $SD = 1.000$) to Hazard ($M = 1.356$, $SD = 1.386$, $p = 0.009$), and a posterior decrease from Hazard to Recovery ($M = 0.467$, $SD = 0.898$, $p = 0.046$). Whereas BA46-R reported a main effect of Driving Hazard on HbO ($F(2, 34) = 5.760$, $p = 0.017$, $\eta^2_p = 0.253$) although fading away with post-hoc comparisons.

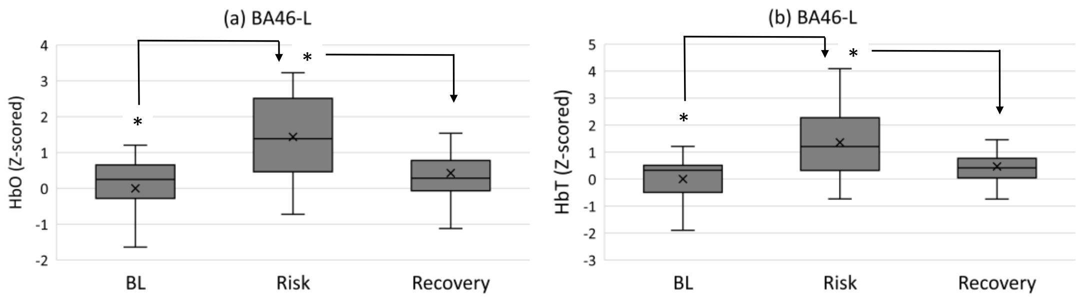


Figure 49 HbO concentrations in BA46-L (a) and HbT (b) for the effect of Driving Hazard on baseline and recovery. Asterisks (*) indicate significant effects and arrows their direction. Mean is indicated by the (X).

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 ($Mdn = 3.00$, $IQR = 3$).

Results from the TASS scale revealed a main effect for trust (F (1, 18) = 5.975, p = 0.025, η^2_p = 0.249, Figure 50), indicating a significant increase in trust ratings after the trial (M = 4.721, SD = 1.207, p = 0.025) compared to before the trial (M = 4.279, SD = 0.893).

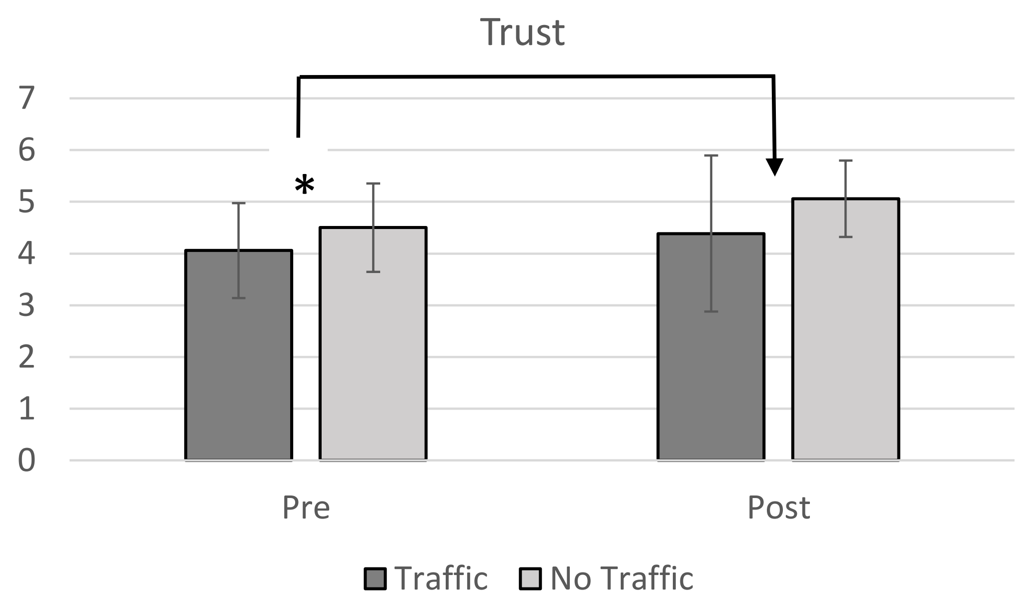


Figure 50 Trust means pre and post-trial for both groups. Asterisks (*) indicate a main effect and their direction

Overall, these results suggest that the Driving Hazard event had a more significant effect on skin conductance and brain oxygenation measures than cardiac findings. Instead, HRV and brain oxygenation were more sensitive to the slowly evolving effect of Driving Conditions than SCRs. Further interpretations of these results will be discussed in the next section.

5.6 Discussion

This second study set out to investigate whether slowly evolving automated driving situations with low to moderate perceived risk and rapid driving events with moderate to high risk associated would produce observable changes in physiological indicators of sympathetic and parasympathetic activity.

5.6.1 Hypothesis 1 – Group differences for traffic

The first question in this study sought to determine whether the presence of Traffic would slowly increase the perception of risk in one group and result in significant group differences in arousal across driving conditions. However, the presence of Traffic did not have the expected effect on perceived risk apparently, because no between-group or interaction effects were reported for any of the physiological indicators or self-reported perceived risk.

There is a chance that the reduced sample size has contributed to the absence of effects between groups. However, the lack of differences in arousal or brain activity due to the presence of Traffic could also suggest this variable would not strongly affect perceived risk during automated driving, as opposed to when manually driving (Foy & Chapman, 2018; Healey & Picard, 2005; Radhakrishnan et al., 2020). Regardless, such inferences should be cautiously considered 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 the effects of traffic on perceived risk from naturalistic automated driving conditions. Besides, the influence of traffic on perceived risk might be linked to individual differences such as personality traits, age or gender, especially when the sample size is small.

5.6.2 Hypothesis 2 – Within-participants variations for Driving Conditions

The second hypothesis investigated whether slowly changing Driving Conditions with low to moderate risk would produce variations in arousal and brain activity within participants –i.e., from BL to DC2. Moderate levels of risk during these were assumed as the median of self-reported perceived risk during Driving Conditions was 3 out of a maximum of 7.

In favour of this hypothesis, HRV reported decreased vagal modulation between baseline and suburbs scenarios (i.e., Driving Condition 1, DC1). Such uncoupled vagal decrement from a baseline resting (i.e., with unaltered sympathetic activity) has been associated with increased monitoring during single-driving tasks requiring perceptual-central processing (Backs et al., 2005; Lenneman & Backs, 2009), and would therefore provide some evidence in favour of this hypothesis. In particular, RMSSD would have been more sensitive to such long-term evolving arousal fluctuation from resting to driving through suburbs (DC1), associated with low to moderate risk. HF-power, the other HRV parameter, coupled these results, but effects were less robust and diminished with post-hoc comparisons. Considering the nature of the HRV signal, perhaps evolving manipulations in our driving scenarios with longer periods and allowing enough recovery time between may have contributed to obtaining more robust effects.

This HRV effect was supported with a very similar HbT increment from BL to suburbs (DC1.2) in the left ventrolateral cortex (BA45) and with additional increments on HbT in the same area (BA45) and HbO in the left orbitofrontal area (BA10-L), both from BL to city centre (DC2.1). Increased oxygenation and total haemoglobin concentrations in the left ventrolateral and orbitofrontal areas would be consistent with increased vigilance and attention towards the automated vehicle's performance during these driving conditions.

The orbitofrontal region has been found to activate when operators are judging the trustworthiness of automated systems (Palmer et al., 2020) and as an indicator of the willingness to intentional engagement (Dimoka, 2010), thus perhaps indicating an intention of taking-over manual control during city centre (DC2.1). A consistent effect on the orbitofrontal area was also observed in our previous study during the city centre scenario, only for the trust-lowered group (see 4.6.3). These results suggest that moderate levels of perceived risk due to changes in driving conditions would be captured with variations in orbitofrontal HbO/HbT, which would indicate a lowering in TiA. This effect would agree with the notion discussed in 5.2.2 that increased perceived risk is aligned with a lowering of trust.

In favour of this argument, increased oxygenation in the ventrolateral cortex has been associated with distrust (Hubert et al., 2018; Palmer et al., 2020), state-level suspicion (Hirshfield et al., 2014), and frustration during automated driving (Damm et al., 2019). In addition, this area was triggered during the risk scenario only for the trust-lowered group in our previous study (see 4.6.3) and is very proximal to the insular cortex (not accessible with fNIRS), which is related to intense negative emotions of distrust, fear and negative consequences (Dimoka, 2010; Hubert et al., 2018).

In contrast, no effects for Driving Conditions were reported from skin conductance responses (SCRs) that would support hypothesis 2. A potential explanation for this lack of effects from SCRs could be that these 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). Indeed, in the previous study (see chapter 4), SCRs trends increased as driving complexity did, but similarly, did not report any significant effects for driving conditions (Perello-March et al., 2021). Another explanation for this phenomenon could be that SCR features might be less sensitive to long-term arousal variations. Indeed, skin conductance levels (SCL) would be a better choice for such a purpose as these measure the tonic component of the signal. For example, (Beggiato et al., 2018) measured discomfort variations from HAD scenarios with short epochs – although using SCLs instead – and could not observe any statistical effects either. Perhaps longer epochs would be more sensitive for long-term changes in arousal, but in the previous study (see chapter 4), no effects were observed over two-minute epochs either. Therefore, future work recommendations would also be to consider other electro-dermal activity features for driver state monitoring of long-term arousal fluctuations. For example, endosomatic electro-dermal activity features such as skin potential level or skin potential responses measure the nervous impulses that activate sweat glands without applying direct current to the skin and are less affected by electrode artefacts than exosomatic features (e.g. SCRs and SCLs) (Boucsein, 2012). In the driving context, skin potential responses have been successfully used for detecting sympathetic reactions to unexpected events (Zontone et al., 2020). It would be worth considering its integration in DSMs as a promising alternative to traditional exosomatic features. Notwithstanding, care must be taken with skin potential measures, because these also have certain drawbacks for recording and evaluation, such as the problems arising from parameterisation and interpretation of the skin potential response amplitude (Boucsein, 2012).

Another aspect worth discussing is the lack of effects between suburbs and city centre scenarios (DC1 vs DC2) across almost all physiological measures. Only HbT reported a variation on BA46-L from suburbs (DC1.3) to the city centre (DC2.1), an area implicated in decision-making from perceptual inputs and which has been associated with trust calibration (Drnec et al., 2016; Hubert et al., 2018; Palmer et al., 2020). This might indicate participants perceived the change in driving conditions (i.e., from suburbs to city centre) and were evaluating the automated vehicle performance and trustworthiness when entering such new driving context, but perceived risk across driving conditions remained low to moderate, as self-reported and the lack of arousal effects showed.

Overall, findings from H2 would indicate an increase in monitoring/alertness from BL to suburbs, and even possibly with a certain level of suspicion towards vehicle trustworthiness when entering the city centre. However, overall perceived risk levels during driving conditions remained moderate.

5.6.3 Hypothesis 3 – Perceived risk from Driving Hazard

The third hypothesis predicted that the rapid event of Driving Hazard associated with moderate to high risk would produce arousal increment compared to baseline resting and post-event recovery resting. Substantial evidence favouring this hypothesis was found as participants reported perceiving significantly greater risk during this event than during Driving Conditions, supported with robust effects throughout all three SCR features extracted. These findings suggest that SCRs were particularly sensitive to arousal fluctuations from a rapid driving event with moderate to high risk associated, which would be unsurprising since SCRs are well known for being sensitive to phasic arousal changes due to unexpected environmental stimuli (Boucsein, 2012; Dawson et al., 2016). Even though the novelty of these results lies in that these would be physiological indicators of a rapid increase of perceived risk. This would be consistent with other research reporting greater skin conductance activity with increased mental workload (Foy & Chapman, 2018), stress (Healey & Picard, 2005), or discomfort (Radhakrishnan et al., 2020) during hazardous driving conditions. Furthermore, these findings add evidence to those from Li et al. (2019), suggesting that HAD users also perceive abnormal traffic behaviours as risky; but from the results of our H1 (see 5.6.1), one could infer that the simple presence of surrounding traffic behaving normally may not have a strong influence on risk perception. Therefore, it could be interesting to explore whether perceived risk from HAD users varies with surrounding traffic behaving normally (e.g., following the traffic code) versus abnormally (e.g., hazardous manoeuvring, racing, skipping traffic lights).

SCR variations were strongly supported by fNIRS metrics, both HbO and HbT. As earlier discussed in the previous chapter (see 4.6.3) and in the present chapter (see 5.6.2), bilateral orbitofrontal (i.e., BA10) activation during Driving Hazard could be attributed to an active judgement of the automated vehicle trustworthiness derived from an increased risk perception (Dimoka, 2010; Palmer et al., 2020). Consistent activation of this cortical area due to factors related to automated driving performance (e.g., changes in traffic complexity and driving conditions) across both studies in this thesis reinforce the concept that the orbitofrontal cortex would play an important role in trust calibration during uncertain situations. It is likely then that the orbitofrontal acts as a “comparator” from perceptual information and vehicle performance reliability (or expectations of credibility), from which situational TiA is eventually derived.

More evidence supporting this statement was found in the left ventrolateral prefrontal cortex (BA44/BA45). Increases in both HbO/HbT from the BL to Hazard scenarios are indicative of attention allocation during visual search (Anderson et al., 2007), increment in distrust (Hirshfield et al., 2014; Hubert et al., 2018; Palmer et al., 2020) and experiencing strong unpleasant emotions (Hirshfield et al., 2014; Hoshi et al., 2011; Hubert et al., 2018).

Likely, our participants were actively seeking visual cues to analyse, understand and anticipate the potential consequences of the sudden Driving Hazard event, which rapidly triggered a strong startle response (as indicated by SCRs) possibly linked with experiencing strong unpleasant emotions such as momentary fear and distrust. Notably, the insular cortex in the inferior frontal gyrus (close to BA45) seems to be the centre for risk perception, often linked to decisions with strong negative emotional components (Dimoka, 2010; Hubert et al., 2018), possibly an evolutionary neural mechanism to prevent from negative interactions and their consequences (Kahneman & Tversky, 1983, 2019; Rangel et al., 2008). However, the insula also seems to be strongly related to a more cognition-based mechanism for risk assessment of contextual information and their appraisal (Hubert et al., 2018; Singer et al., 2009). Because of the proximity of BA44/45 with the insular cortex, these findings would support the role of the insular cortex and provide direct evidence of the two processes for risk assessment that drive trust judgements described by De Visser (2012) (see 5.2.2).

Bilateral activation was also observed in the dorsolateral prefrontal cortex (i.e., BA46-L, BA46-R and BA09-R), reporting HbO/HbT increases from BL to Hazard, followed by decreases from Hazard to Recovery phases in both measures. Dorsolateral prefrontal cortex activity is often attributed to deliberate decision-making and reflective processes related to trust (Dimoka, 2010; Drnec et al., 2016; Hubert et al., 2018), supporting the orbitofrontal cortex in comparing uncertain perceptual (e.g., visuospatial) information (Bruno et al., 2018). Therefore, as in the orbitofrontal cortex, higher activation of the dorsolateral prefrontal cortex would be related to situational trust calibration processes when relevant contextual changes occur, and possibly with increasing distrust too, as observed in 4.6.3 and by Hubert et al. (2018). This would be explained by the role of the dorsolateral prefrontal cortex in emotional regulation and self-control (Hirshfield et al., 2014; Hubert et al., 2018) after experiencing sudden strong negative emotions derived from the Hazard event (indicated by ventrolateral activity and SCRs). In favour of this claim is the significant decrement in HbO and HbT observed after the event during the recovery phase. This oxygenation withdrawal to basal levels could indicate participants successfully regulated their emotions and self-controlled, e.g., not steering or pressing the brakes.

Results from cardiac parameters during the Driving Hazard event were less robust than those from SCRs and fNIRS discussed above. Only HRV-RMSSD reported statistical effects with a vagal component decrement between baseline and the epoch comprising driving Hazard-&-Recovery. This effect was partially in favour of H3, as it was expected that the hazardous event would produce a startle response activating the sympathetic system and deactivating the parasympathetic branch (Lenneman & Backs, 2009). Whilst the effect reported by HRV-RMSSD indicated a deactivation of the parasympathetic branch, heart rate (HR) and HRV-HF did not report any supporting effects (i.e., HR was expected to increase and HF to decrease). On the contrary, trends indicated an HR decrease from baseline resting to Hazard-&-Recovery, suggesting participants may have been more aroused during baseline resting than during Hazard-&-Recovery, but these trends were not statistically significant. This mode of autonomic control is known as non-reciprocally coupled and would be indicative of a co-activation during baseline – i.e., because sympathetic activation exceeded parasympathetic activation; and a co-inhibition during Hazard-&-Recovery – i.e., because a parasympathetic inhibition exceeded sympathetic inhibition (Backs et al., 2003, 2005; Lenneman & Backs, 2009).

A potential explanation for the initial co-activation is that our participants might have been not wholly relaxed during the baseline, and thus, it was not necessarily a low arousal state. Perhaps the novelty of the experiment has contributed to this phenomenon. Recommendations for future research would therefore be to collect baseline recordings in a more familiar and duller environment. However, it could also be argued that a heavily controlled and experimental environment would lack ecological validity (i.e., it is unrealistic to expect future HAD users would lay in the supine position for 5 minutes and without any external stimuli before starting a journey). On the contrary, the co-inhibition observed during Hazard-&-Recovery was most likely because a recovery period of resting followed the hazardous event. Hence, combining the startle response from the hazardous event with the resting state from the post-event recovery into a single epoch would have contributed to this effect. However, combining these periods into a single epoch was necessary to meet the minimum epoch length recommended for HRV measurements (Laborde et al., 2017; Shaffer & Ginsberg, 2017).

Another finding worth mentioning was the observed increase in self-reported trust after the trial. Whereas this finding may seem contradictory to the rationale discussed in section 5.2.2, it can be argued that the TASS scale is measuring propensity to trust (i.e., dispositional trust) rather than situational trust (Adams et al., 2003; Holthausen, 2020). Hence, such increment in dispositional-learned trust could be attributed to a process of familiarisation due to mere exposure. Similar increments in the trust after a short exposure from naïve participants have also been reported in several studies (Dixon et al., 2019; Gold et al., 2015; Kraus et al., 2020; Kundinger et al., 2019; Large et al., 2019; Lee et al., 2021; Muir & Moray, 1996). Dispositional and learned trust are different layers than situational trust (see 2.2.1) and might be modulated by other factors. Whilst a higher risk perception is likely to correlate negatively with lower situational trust, this may not necessarily apply to dispositional and learned trust (Hoff & Bashir, 2015; Holthausen, 2020).

Overall, results from the present study have evidenced the effects of contextual cues on risk perception and its derived physiological indicators. Whereas slow changes in the driving context associated with moderate perceived risk levels were successfully captured with cardiac and brain oxygenation measures, rapid changes associated with high-risk perception were captured with skin conductance and brain oxygenation features. Therefore, cardiac measures –i.e., HR/HRV- seem to be more sensitive to measure slowly evolving tonic arousal changes in the automated driving context. For example, heart rate increments and vagal component decrement are robust indicators of attentional levels, alertness or vigilance states (Lenneman & Backs, 2009; Nemcova et al., 2020; Singh & Queyam, 2013), which would be useful in checking a monitoring request towards a planned takeover has been successfully deployed. By extension, cardiac indices would also be reliable indicators of the driver availability state to take over when combined with, e.g., eye-tracking data, to ensure drivers are back *in the loop* (i.e., actively seeking information on the road ahead, side mirrors or checking the navigation system).

Conversely, skin conductance responses indicate sympathetic activation and thus increased alertness level. They are particularly sensitive to startle stimuli – as observed in the hazardous event –, which a TOR can be to an OOTL driver and could be complemented with gaze behaviour indicators (e.g., driver glancing at the interface screen issuing the visual warning). Last but not least, brain activity measured with fNIRS is not likely to be integrated into DSM systems -at least with current wearables- but has proven to be an essential complement to interpret arousal indices providing important insights into participants' perception, comprehension and projection of changes in the driving scenarios.

5.7 Limitations and future work

This study has provided important insights into perceived risk during automated driving, its measurement with physiological indicators and its implications on human-centric parameters such as TiA and SA.

Regardless of successfully inducing levels of perceived risk across our driving scenarios and the event, the validity of driving simulators is still limited, especially in terms of generating genuine risk because there is no real hazard involved. Although driving simulator realism has improved in several aspects, this issue is still a drawback for current research in human factors in automated driving, and thus, the generalization of the results. Ideally, real-world experiments would mitigate the issue, but even though we would have had the equipment to undertake such experiments, it is unlikely the ethics committee would have approved them. Carrying out controlled, real-world experiments on closed roads or private circuits entails huge experimental costs and potential ethical issues. Therefore, we would encourage future research in driving simulators studying aspects of trust to assess the levels of self-reported risk, and relatedly, we stress the need for developing a validated risk perception scale for the driving context. Self-reports are an essential tool for interpreting physiological results and a valuable complement for enhancing human factors research not to be ignored. With these regards, future work must integrate subjective, physiological, and behavioural data as far as possible.

Other than that, it is important to bear in mind that physiological variations are relative to a previous arousal state, which in real-world contexts may notably vary in length and basal arousal levels depending on the NDRT performed before the monitoring request. Moderate arousal levels will likely be the target state for optimal take-over performance, as argued in section 2.1, and multimodal DSM systems can detect arousal variations from both extremes of the spectrum, either from low to moderate arousal or from high to moderate. Using the same earlier example, the arousal level 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 according to the *Yerkes-Dodson Law*.

Therefore, future research in multimodal DSM systems for HAD should test the sensitivity of both ECG and EDA features in detecting drivers' awareness of monitoring requests and availability to take over in several contexts. It is imperative to mitigate the effects of false positives (i.e., the system interprets high arousal/wakefulness as driver in-the-loop) and false negatives (i.e., the system interprets low arousal as driver out-of-the-loop). Eye-trackers would become a required complement to alleviate such limitations. Hence, it would also be relevant to explore the sensitivity of such DSM systems from either resting states or different NDRTs, and how competition for mental resources in each case may have a detrimental effect towards perceiving and reacting to a monitoring request and subsequent take-over performance.

5.8 Conclusions

Poor situation awareness because of lowered risk perception may result in excessive trust in automation, and that, as has already been explicitly evidenced in several road accidents in the US, may lead to fatal consequences (NTSB, 2017, 2019b, 2020b, 2020a). Especially with *autonowashing* (Dixon, 2020) inducing false automation credibility expectations under marketing strategies. Whilst this issue 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. Hence, we stress the importance of issuing monitoring requests to get the driver back on the loop towards the upcoming TOR as suggested by other authors (Gold et al., 2013; Lu et al., 2019), and for which DSM system will be essential to guarantee optimal driver readiness levels (Marberger et al., 2018).

Although we acknowledge that arousal indices cannot directly infer complex psychological states *per se*, our results have evidenced that risk perception indicators can be objectively measured in real-time with physiological indices. These indicators could be used to train machine learning algorithms to develop multimodal DSM systems (Perello-March et al., 2020), and in combination with eye-tracking, be used to infer driver availability to take over. In short, our findings suggest that:

- Low to moderate risk perception could be measured with HRV features, which would be more sensitive to longer-term changes in arousal levels due to environmental factors,
- Moderate to high perceived risk could be measured with SCRs, which would indicate the short-term changes from rapidly evolving driving events.
- Localised orbitofrontal, ventrolateral, and dorsolateral increases in HbO/HbT would be direct indicators of risk perception both in slowly changing and rapidly evolving conditions.

Future research working towards the development of multimodal DSM systems would benefit from this knowledge for determining when drivers would be back on the loop, as well as their availability for optimal take-over performance.

6 General discussion and conclusions

6.1 Summary of main findings

As a whole, the outcomes from this thesis provide important insights into some of the key human factors-related aspects to consider for the development of multimodal DSM systems for HAD. It is evident that HAD implies a driving task reallocation and changes the way users perceive and interact with the vehicle. Trust, situation awareness and mental workload will be determinants for adequate interactions with this technology and its safe usage, as discussed in this thesis. Finding a way to measure these constructs in real-time with objective indicators is crucial for safely implementing vehicles equipped with HAD technology. This thesis sought to determine whether specific physiological indicators could be a potential solution to use in combination with other driver state monitoring systems, e.g., eye-tracking based, as measures for such key human factors. Some lessons have been learnt from both studies, which make an essential contribution to knowledge in this research domain, raising new questions and confirming previous findings.

From the first study described in Chapter 4, the utility of cardiac measures was evidenced in detecting drivers' engagement in NDRTs and real-time mental workload levels. This result confirms the utility of ECG as a tool for assessing drivers' availability in terms of their current mental workload levels, which will determine the pool of attentional resources available to perform a safe take-over manoeuvre.

Another significant observation from the risk perception experiment (chapter 4) is that induced trust in automation levels from credibility expectations can be experimentally measured with physiological indicators. Whilst this is not the first study inducing different levels of TiA to their participants or even measuring arousal or gaze behaviour associated with it, to our knowledge, it is the first evidence that opposing levels of induced trust in automated vehicles are captured with fNIRS.

Other aspects of TiA that were also confirmed within this first study were the episodic and strong emotional components of distrust, suggesting that skin conductance responses might serve as indirect indicators of situational distrust and high perceived risk, as observed in Chapter 5. On the contrary, we observed that building up and calibrating trust implies a more deliberate, slow and executive component. These findings confirm experimentally former frameworks of TiA formation and calibration at initial stages (Lee & See, 2004).

Regarding other time-related aspects, findings from the study described in Chapter 5 suggest cardiac features would be a more accurate indicator of long-term slowly evolving driver states, whereas skin conductance responses would be particularly sensitive to rapidly evolving driver states. These results were partially aligned with those from the driver availability experiment (see chapter 4) but require further investigation to be confirmed.

Another finding from the risk perception experiment (chapter 5) is that perceived risk can be experimentally generated in controlled driving simulator layouts, and that those physiological indices of arousal and brain activity are sensitive to changes in perceived risk.

Finally, results from both studies indicate that pre-frontal cortex activity during HAD provides valuable real-time information on important aspects of users' engagement in automated driving and their availability to re-gain manual control, particularly in terms of situation awareness, mental workload, trust in automation and risk perception.

6.2 Discussion

With regards to the research questions framed in the first chapter of this thesis, the work undertaken has contributed to clarifying crucial aspects relating to user engagement in HAD and shed light on the gaps in the literature identified. Each research question will be directly addressed in the following subsections.

6.2.1 Can electrocardiogram (ECG) measures be used as indicators of factors modulating driver availability state?

Cardiac activity is directly tied to both branches of the autonomic nervous system: the sympathetic branch responsible for preparing the body for action and the parasympathetic branch responsible for inducing relaxation and rest. Because cardiac activity has proven to be a valuable indicator of several driver states, both in experimental and real-world conditions (Lohani et al., 2019; Nemcova et al., 2020), and is particularly sensitive to mental workload levels (Backs et al., 2003; Lenneman & Backs, 2009), its measurement in real-time offers a window into the drivers' complex emotional and cognitive states.

Our results from the driver availability experiment presented in chapter 4 align with previous literature in evidence that mental workload from a non-driving related task was captured with all three cardiac features we extracted (i.e., heart rate, LF/HF ratio and RMSSD). Our findings contribute to build-up evidence favouring the important role cardiac indicators can play in multimodal driver state monitoring systems. Cardiac measures would capture real-time mental workload variations derived from engagement in non-driving related activities during highly automated driving use cases. As discussed earlier in sections 2.2.2 and 2.3.2, the current mental workload is a crucial indicator of drivers' availability to take over. It will determine the mental resources available and attentional capability for drivers to become situationally aware and take back control of manual driving.

Concerning situation awareness, results from the risk perception experiment presented in chapter 5 suggest that heart rate variability features indicating vagal modulation might be sensitive to moderate arousal variations associated with changes in the driving situation. Although from the driver availability experiment, we showed that these might be masked by mental workload when drivers perform a demanding NDRT. If situation awareness can only be detected when drivers are not engaged in NDRTs (because otherwise, cardiac activity will indicate mental workload from the task), disentangling concurrent levels of mental workload from situation awareness within cardiac data will likely be one of the main challenges to overcome for this data modality.

This challenge may eventually be of lesser relevance because cardiac data could only be required during control transitions, and during these, drivers would disengage from any non-driving related activity. Especially during planned take-overs, control transitions will possibly involve plenty of time for drivers to prepare, for example, attentive user interfaces informing drivers of the current time budget for the subsequent take-over (Wintersberger et al., 2018). Such situations will enable time windows of sufficient duration for cardiac variations to be successfully detected (i.e., around two minutes), as these could include from when drivers are first informed of a planned take-over ahead, to issuing the monitoring request, then the take-over control request, to finally engaging manual control. Evidence from the risk perception study (chapter 5) suggests that HRV is sensitive to slowly evolving changes in arousal and that cardiac activity has the potential to be essential in the assessment of effective transitions from out-of-the-loop states to on-the-loop and ultimately in-the-loop (Merat et al., 2019).

Real-time cardiac monitoring could also be used to measure driver take-over performance along with gaze behaviour and other physiological features (Backs et al., 2003; Hogervorst et al., 2014). With trained machine learning algorithms, individual user profiles could be created to achieve the individual's optimal arousal state for successful take-over performance (Wilson & Russell, 2003). Similar work has already been applied in the military domain with multi-agent systems involving the co-operation, or joint performance of human operators and automated agents, where achieving an "optimal operator functional state" is critical to enhancing the overall joint system performance (Wilson & Russell, 2007). With current driving automation assistance systems, the functional state of the human operator is not monitored during safety-critical situations such as during a take-over, only the technological and engineering aspects of the automated driving system components. Not monitoring the driver's functional state can lead to situations where an automated driving system's performance is good, but human errors are still present (Waymo, 2018). Estimating operator functional state has been successfully used in military and air traffic control domains to determine if and when system intervention is required to assist the operator and enhance joint system performance by using real-time machine learning classifiers (Wilson & Russell, 2004, 2007). In their 2004 paper, Wilson & Russell used a machine learning algorithm based on EEG, ECG and EOG signals to classify pilots' mental workload as acceptable load or overload, achieving a 98% accuracy. Results from Wilson & Russell (2007) built upon their previous study and indicated that physiologically determined assistance significantly improved task performance compared to no-assistance conditions or when assistance was randomly provided. Furthermore, such improvement was even more remarkable when assistance was determined over individual profiles than group-based profiles.

Real-time monitoring of driver availability state (or operator functional state as described in Wilson & Russell's papers) would enable the ongoing assessment of drivers' mental workload and cognitive capability to deal with, e.g., a take-over request. When evaluated in the context of current and predicted system demands, driver availability state assessment may indicate that the driver is overloaded and not capable of successfully dealing with the cognitive demands of the task. In that case, the system may assist the driver by automating some aspects of the task to mitigate the cognitive demands of the driver or even perform a minimum-risk manoeuvre to stop the vehicle safely. An accurate estimation of drivers' functional state is crucial in these situations.

One of the current challenges for implementing this technology into production vehicles is finding a way to embed cardiac (and other) sensors unobtrusively to ensure technology acceptance (Melnicuk et al., 2019; Smyth, Chen, et al., 2021). Alternative contact-based solutions have been proposed to discard the attachment of cable-bound ECG electrodes on individuals' skin because of its obvious impracticality. Locations mainly include electrodes located on the seat base, lumbar or thoracic areas from the backrest, or the steering wheel, but also other locations such as seatbelts, armrests or gearbox shift lever have been explored (Leonhardt et al., 2018). Several combinations of electrode locations have been tested either on the drivers' seat only (e.g., Bhardwaj & Balasubramanian, 2019); or with sensors embedded into the driver's seatbelt and the steering wheel for measuring stress based on multimodal data from heart rate, skin conductance, skin temperature, and respiration, combined with driving performance metrics (Ford, 2012). However, these contact-based sensors may be handicapped by poor contact due to, e.g., drivers' clothes thickness or maladjusted clothing combinations (e.g., polyester–cotton, silk–cotton); or lack of contact due to, e.g., steering movements with changes of the location of gripping, or steering with only one hand, or even hands-free situations when highly automated driving is engaged (Leonhardt et al., 2018).

Other innovative techniques are exploring contactless detection through optical or radar-based methods (Leonhardt et al., 2018). Optical methods rely on static cameras and infrared light emitters facing the driver's face, similar to those used for current eye-tracking monitoring systems (see 2.3.1) or fNIRS technology (see 3.3). For example, Gambi and colleagues used a consumer-grade Microsoft Kinect device to measure cardiac parameters from blood pressure based on Eulerian Video Magnification, photoplethysmography and videoplethysmography (Gambi et al., 2017). Microsoft Kinect uses infrared light to detect cardiac parameters from blood flow. They found this device was nearly as reliable as a commercially available Garmin smartwatch, with heart rate estimation varying between 2%-3.4% to that obtained from the smartwatch. Optical methods are promising since they offer the possibility of image fusion by combining eye-tracking, emotion recognition from facial expression, and physiological measures (e.g., heart rate, skin temperature, blood pressure, respiration rate). However, optical methods are susceptible to certain drawbacks, such as requiring a free line of sight and having the target individual under the field of view or variable light conditions, which may alter light detection by the sensing camera (Leonhardt et al., 2018).

Radar techniques are also a promising growing research field for contactless physiological monitoring (Leonhardt et al., 2018). Pulsed ultra wide-band and continuous wave are the most common methods. These can measure movement information caused by chest displacement due to respiratory and cardiac activity (Cho & Park, 2018; Lazaro et al., 2010; Leonhardt et al., 2018). Similar to optical methods, radar sensors can be integrated across several locations into the cockpit, such as the seat, backrest, seatbelt or steering wheel. The penetration depth of electromagnetic radiation into biological tissue through clothes or other surfaces represents the main advantage compared to optical methods (Cho & Park, 2018). Nevertheless, since this technology derives physiological measures through distance measurement, it may be susceptible to motion artefacts (Leonhardt et al., 2018).

The integration of cardiac signals into DSM systems will enhance their capabilities, providing real-time feedback both to the automated driving system and the driver to identify driver overload, determine driver availability for safe take-over performance and ultimately reduce human error from road transport.

6.2.2 Can electrodermal activity (EDA) measures be used as indicators of factors modulating driver availability state?

Electrodermal activity holds specific attributes which make it a worthy candidate for multimodal DSM systems. Starting by the fact that EDA is a pure indicator of sympathetic activity (SNS) and, contrary to cardiac activity, is not affected by vagal modulation (PNS), it can provide valuable information regarding drivers' functional states. This assumption has received support across our results, particularly those from the risk perception experiment described in chapter 5.

EDA is a valuable driver state indicator because skin conductance responses trigger due to salient environmental stimuli, meaning that SCRs can work as instant automatic "verifiers" of drivers' perception of certain contextual cues or in-vehicle warnings. It has been evidenced with an increase of overall means during complex driving conditions in the driver availability experiment (see section 4.5.2 for more details), and especially during the driving hazard scenario in the risk perception experiment (see section 5.5.3). In combination with eye-trackers, SCRs derived from salient stimuli are direct indicators of information perception, attentional levels, and by extension, could indicate situation awareness levels 1 and 2 (see 2.2.3). Such information provides a real-time assessment of drivers' monitoring behaviour and readiness for manual control as, for example, a verifier of drivers' perception of in-vehicle warnings such as monitoring requests or take-over requests. SCR data indicating the non-perception of these warnings could also be used for triggering minimum risk manoeuvres in Level 4 automated vehicles. Furthermore, although in this research we have not measured skin conductance level (SCL) because we focus on the phasic components of the EDA signal for measuring short-term variations in arousal, SCL is also a valid indicator of long-term SNS fluctuations in the driving context (Healey & Picard, 2005; Hogervorst et al., 2014; Walker et al., 2019).

Another aspect where EDA can be used as an indicator of factors modulating drivers' availability state is trust in automation (TiA). Section 2.2.1 has discussed how trust can affect drivers' readiness to take over in, e.g., automation-induced complacency or overreliance behaviours (Bailey & Scerbo, 2007). In addition, EDA indicators have been used with other physiological measures for real-time trust assessment in joint human-automation collaborative systems in military and air traffic domains (Drnec & Metcalfe, 2016; Metcalfe et al., 2017; Nothwang et al., 2016).

However, it is the episodic and strong emotional component of distrust (Dimoka, 2010; Lee & See, 2004), which should be measurable through skin conductance activity (López-Gil et al., 2016; Rani et al., 2006). Previous work exploring the physiological components of trust support this assumption since they found a higher skin conductance level when participants reported distrusting the driving automation (Morris et al., 2017; Walker et al., 2019). Moreover, distrust arises with increased risk perception (Perkins et al., 2010; Riley, 1996; Satterfield et al., 2017), and our results from the risk perception study (chapter 5) have shown that increased risk perception can be captured with SCRs.

However, results from the low credibility group (LC) (i.e., distrusters) in the driver availability experiment (chapter 4) were not aligned with this assumption. The lack of higher skin conductance from the LC group due to distrust stresses the complexity of TiA and the extensive work still required to disentangle its intricate multiple dimensions and relationships with other human factors. Indeed, from the driver availability experiment, we have confirmed that distrust, aside from a higher perception of risk, can also be induced through knowledge in the form of automation credibility expectations. An explanation for this lack of effects on SCRs for the LC group is that distrust derived from higher risk perception uses the affective process described in Lee & See (2004) linked to the recognition-based threat assessment process described by de Visser (2012) since it is based on on-going information from the driving environment. In contrast, *learned* distrust is derived from induced credibility expectations and may work on analogic or analytic mechanisms (Lee & See, 2004) since it is based on knowledge about the driving automation and activates brain regions associated with distrust and executive processes. Perhaps competition for mental resources due to the NDRT in the driver availability experiment (chapter 4) affected the calibration of trust and subsequent physiological reactivity towards external stimuli, even though participants *knew* the driving automation was untrustworthy, or perhaps learned distrust does not manifest in the form of sympathetic activity or arousal.

Distrust and resultant automation disuse often receive less attention across the literature than their opposites – i.e., overtrust and overreliance –, possibly because only a few examples have been echoed by the media, but distrust can also result in severe consequences. This is especially so in driving automation where drivers are fallback users (e.g., SAE L3) and can take over control as they consider; situational distrust can result in taking over manual control when not appropriate or safe. An example occurred with a prototype vehicle being tested on an open road and supervised by a professional operator (Waymo, 2018). While approaching a traffic jam ahead, the operator took over manual control, thinking the automated vehicle would not brake in time to avoid the collision. The operator steered to the right lane but did not see a motorbike approaching from behind in the right lane and ended up colliding with the motorbike. The accident analysis report revealed the automated driving system had successfully detected the motorbike and, importantly, had the traffic jam ahead also under control. Fortunately, this event resulted in no fatal consequences for the automated vehicle operator nor the motorbike driver but exposed the potential consequences of distrust in automation even for experienced users. In this example, a DSM system combining EDA with gaze behaviour could have likely detected the operator's distrust. Indicators such as a sudden increase in perceived risk increasing the number and the amplitude/magnitude of SCRs, and with an increased number of saccades between the traffic jam ahead and the vehicle HMI, fixations in trying to find out whether the driving automation was going to brake in time, and even possibly a reduction of fixations at the side mirrors. Therefore, the combination of EDA with eye-tracking measures would be necessary because whereas eye-tracking provides relevant attentional information online, EDA provides the emotional arousal deriving from the interpretation of such attentional information in real-time.

Accordingly, previous work has proposed EDA as a modality for the real-time assessment of TiA (Perello-March et al., 2020), relying on several machine learning approaches. Whilst this is just a proposal, and trim work has used EDA for validating its usage in TiA assessment, commercially available smart wristbands such as the Microsoft Band 2 and Empatica E4, which measure EDA, already offer the possibility of measuring specific indicators of trust using wearable devices which are less likely to hamper consumer acceptance (Melnicuk et al., 2019; Smyth, Chen, et al., 2021). Such wristbands and smartwatches are currently capable of measuring cardiac activity, blood pressure, body temperature or skin conductance with acceptable reliability and have already been used with this purpose across several studies in the driving context (Beggiato et al., 2018, 2019; Radhakrishnan et al., 2020; Walker et al., 2019; Wintersberger et al., 2017), as well as for emotion recognition in other human-computer interaction domains (Cowley et al., 2016; Egger et al., 2019; López-Gil et al., 2016). Embedding these sensors into smartwatches also offers the possibility of using keyless wristbands such as the Activity Key presented by Jaguar-Land Rover (Jaguar-LandRover, 2018). Although the data quality from these signals is still limited compared to research-grade devices, huge improvements have been made in the last years, suggesting a promising future for wearables as part of a multimodal DSM system fusing eye-trackers and contactless sensors as discussed in the previous section.

If, as it seems, conditional driving automation is to be pervasively integrated into production vehicles in due course, all stakeholders must take into consideration the potential consequences on drivers' readiness of inappropriate distrust among inexpert or novice users too – and note here that an experienced driver may still be inexpert with automated driving technology. TiA plays a crucial role in technology acceptance and correct use, and these challenges derived from distrust cannot be ignored. As earlier discussed in section 6.2.1, this is another argument in favour of the importance of bringing driver's functional state in the loop of human-autonomy driving performance assessment. Automated driving systems with shared control must assess human drivers' functional state to determine their availability to take-over control.

6.2.3 Can functional near-infrared spectroscopy (fNIRS) measures be used as indicators of factors modulating driver availability state?

Extensive research has used brain activity to investigate driver behaviour and human factors derived from the driving context (for a review, see Lohani et al., 2019). Despite the existing evidence, HAD also entailed new human factors considerations, which have received less attention, and only very few studies have used fNIRS to investigate these factors. Using fNIRS in this research has contributed to understanding factors like TiA and perceived risk but has further implications concerning SA. Our findings agree with previous work showing the sensitivity of fNIRS in measuring TiA fluctuations affecting SA, and thus, driver availability states. Regardless of not finding significant effects for mental workload in the driver availability experiment – possibly due to the confluence with TiA –, current literature suggests that fNIRS parameters are also valid indicators of mental workload and task demand (Durantin et al., 2014; Foy & Chapman, 2018; Girouard et al., 2009; Unni et al., 2018).

Results from the driver availability experiment (chapter 4) showed that fNIRS indicators are sensitive to opposing levels of TiA in the HAD context. In addition, by proving that credibility expectations can induce overtrust or distrust, we provide the first experimental evidence of the *autonowashing* phenomenon measured with objective indicators. As earlier discussed in section 1.1, *autonowashing* – i.e., driving automation credibility expectations induced by marketing strategies and mass media – may become a safety-critical issue when consumers' expectations on driving automation capability exceed the actual driving automation system performance (Dixon, 2020). By providing evidence of this phenomenon, policymakers, legislators and relevant stakeholders can benefit from this knowledge to take action accordingly to regulate the naming of driving automation technology for the general public.

Results from the risk perception experiment (chapter 5) indicate that fNIRS features are also sensitive to the perceived risk in the HAD context. Thus, this finding provides insight into one of the questions raised from the driver availability experiment (chapter 4) whether perceived risk could be experimentally induced in driving simulators. Therefore, if driving simulators can induce risk perception, it is likely that the general lack of arousal effects in the driver availability experiment for the within-participants variable of Driving Conditions (i.e., H2 – Traffic Complexity) was attributable to low-risk perception, and especially within the trust-promoted group. According to some authors, lower perception of risk, reflected in lower arousal/brain activity variations, would indicate good SA (McKendrick et al., 2016). Bracken et al. (2021) argued that situationally aware drivers would expect potential hazards, and when these occur, they will not come as a surprise. This would explain why neither trust-lowered nor trust-promoted participants showed variations in peripheral physiology (i.e., ECG and EDA) between the risk event and the other driving conditions in the driver availability experiment, even though the event was perceived as potentially hazardous as demonstrated in the risk perception experiment. A claim supported by two facts:

1- The trust-lowered group (LC) reported higher brain activation during the risk event than the trust-promoted group (HC), indicating a higher perception of risk.
2- Only the trust-lowered group (LC) reported effects within conditions in HbO, HbR and HbT levels, and noticeably all from the city centre scenario, which would indicate a higher perceived risk due to distrust.

Since these participants expected a potential driving hazard because driving conditions become more complex, they were not surprised by the risk event regardless of a higher perception of risk. They were situationally aware, and only the low trustworthiness of the vehicle made a significant difference during the event. Trust promoted participants had a similar situation awareness level, but their risk perception was lower because of the reliable driving automation. Hence, participants from the HC group did not show any arousal or brain activity effects for driving conditions.

The rationale underlying such an intricate relationship between mental workload, trust in automation, situation awareness and risk perception would lie upon the boundaries of the situation awareness spectrum. Whereas under-arousal levels have been associated with driver drowsiness and fatigue (Dong et al., 2011; Li et al., 2009), and thus, low situation awareness; mental overload can lead to complacency states, which Bailey & Scerbo (2007) and Parasuraman & Manzey (2010) also attributed to low SA. Following the *Yerkes-Dodson law* (see 2.1), mental overload and low arousal states are described as the upper and lower boundaries of the SA spectrum, where both are indicative of low SA (Bracken et al., 2021). Adequate SA would therefore lie within these boundaries, modulated by trust in automation and risk perception, and could be determined from the three layers of SA described in 2.2.3. Existing literature has evidenced that fNIRS can be used to measure mental workload and attention levels (Durantin et al., 2014; Foy & Chapman, 2018; Hidalgo-Muñoz et al., 2019; Sibi et al., 2016), drowsiness (Tanveer et al., 2019), and even trust in automation (see Chapter 4). However, only a few studies have been undertaken using fNIRS to measure situation awareness, and none in the automated driving context (e.g., McKendrick et al., 2016).

A recent proposal has attempted to fill this gap by describing how situation awareness could be measured with physiological indicators based on existing frameworks (Bracken et al., 2021). According to this proposal, EEG would be particularly suitable for measuring SA throughout its three layers: (1) perception, (2) comprehension, and (3) projection. Neurophysiological techniques such as EEG and fNIRS (Verdière et al., 2018; Wang et al., 2016) have already been used as measures of attention and visual engagement, both indicators of the perception of relevant changes in the environment (1). Whereas for SA levels (2) comprehension and (3) projection, Bracken and colleges propose measures of expectancy (e.g., P300 components of the ERP from the EEG signal) as indicators of the comprehension of environmental changes (SA level 2) and projection of their state in the near future (SA level 3) of these environmental changes (Bracken et al., 2021). They argue that surprising events or stimuli would indicate poor SA, rationalising that proper SA implies users expect certain critical cues or upcoming events. Indeed, recent work has already attempted to validate this proposal using exclusively EEG indicators and machine learning classifiers with limited success, where classification accuracy was only 67% (Kästle et al., 2021). This lack of classification accuracy may be due to relying only upon EEG parameters. The authors argue that EEG parameters are more accurate than subjective measures and more sensitive than peripheral physiology (e.g., ECG or EDA) but ignore that EEG indicators (or any other physiological measure) in isolation are barely interpretable. It could be argued that SCRs could also indicate unexpected events (Dawson et al., 2016), and by extension, poor SA. Another reason which may explain this result is that the proposal from Bracken et al. assumes measures of expectancy (e.g., ERPs from the EEG signal) as indicators of SA levels 2 and 3. This assumption must be carefully considered because:

1. Measures of expectancy for determining executive components of SA levels 2 and 3 (i.e., comprehension and projection) may be problematic. Unexpected events can occur regardless of the driver engagement and awareness of the driving task (e.g., even the most situationally aware and focussed driver could be surprised by an animal coming across from nowhere which would trigger ERPs or SCRs).

2. Perceiving a hazardous event as potentially risky could be considered good SA in certain situations, e.g., a safety-critical situation in an SAE L3 vehicle (perception, comprehension, and projection). A startle response to this scenario would indicate being situationally aware in terms of perceiving the risk, understanding its meaning, and projecting its potential consequences.

3. Differences in prefrontal activity and mental workload do not necessarily have to be attributed to mental overload, unexpected events, and poor SA. Prefrontal activity is dynamic, heterogeneous and could also be attributed to the active comprehension and projection of the information. For example, the orbitofrontal cortex has been found activated in judging the credibility of the driving automation (Palmer et al., 2020), the ventrolateral prefrontal cortex in suspicion and anticipation of losses (Dimoka, 2010; Hirshfield et al., 2014), and the dorsolateral prefrontal cortex in deliberate evaluations of trustworthiness (Hubert et al., 2018) or in judging abnormal driving behaviours (Bruno et al., 2018). In addition, variations in cortical activity cannot only be considered upon the location of such activation but also the “quantity” of activation each task or situation is inducing.

Considered together, these drawbacks stress the importance of developing multimodal systems. Because signals from single measurements may lead to misinterpretations, they offer a clearer picture of the driver state when combined. Sensor fusion would therefore contribute to mitigating this and other related issues such as artefact compensation, source separation, and coverage rate enhancement by applying complex algorithms and models used to integrate the different sources or modalities (Leonhardt et al., 2018). For example, adaptive filtering with one sensor serving as the noise signal can be used for artefact compensation. Applying algorithms based on statistical dependencies between signals, such as the independent component analysis, can be used for source separation. Coverage rate enhancement involves using multiple sensors measuring the same physiology parameter to handle missing data from one sensor (e.g., if the hand moved from the steering wheel, a wristband could also obtain EDA data). However, it is unlikely that brain activity modalities will ever be integrated into production vehicles for driver availability monitoring unless drivers are keen to wear a headset or helmet because of their invasiveness and intrusiveness. These modalities are likely to be integrated instead in those systems requiring human operators to wear a helmet or a headset. For example, in the aviation domain, air traffic controllers, commercial and military pilots will benefit from a headset using integrated fNIRS and EEG neurophysiology sensors. Either for detecting mental overload and situation awareness (Verdière et al., 2018; Wilson & Russell, 2004), but also even to anticipate intentions (Zhu et al., 2019), emotional states (Bandara et al., 2018) or state-level suspicion (Bobko et al., 2014; Hirshfield et al., 2014), which are indicative of trust in automation. Indeed, some authors are already testing the validity of fNIRS in such contexts where human-agent teaming involves dynamic adaptive automation and intelligent decision support systems (Palmer et al., 2019, 2020). Other contexts where brain physiology might also be a helpful solution is, for example, in monitoring motorcyclists, electric scooters or cyclists' functional states, given that helmets are already a safety feature. However, including eye-tracking or other physiological modalities such as skin conductance and cardiac activity might be extremely challenging.

6.2.4 Considerations for integrating physiology measures into a multimodal driver state monitoring (DSM) system

In the driver availability experiment (chapter 4), we expected significant variations in blood oxygenation levels during the 2-back task. Despite the 2-back condition reported effects of cardiac activity and self-reports indicating an increase in mental workload, no variations within-participants were observed from fNIRS parameters. We argued this lack of effects was possibly due to the "steal effect" observed on the HC group (i.e., trustors), reducing HbO levels unexpectedly and instead reallocating cognitive resources towards other areas implicated in the calibration of TiA. Trust calibration requires mental workload, same as undertaking a non-driving related task. The motivations and implications of each upon driving are remarkably different, although both may manifest similarly on brain activity. This finding is an excellent example of what could be analogous to the term known in medicine as comorbidity – i.e., the co-occurrence of two or more medical conditions in an individual. Henceforth, in the discipline of human factors and ergonomics, this phenomenon could be described as the confluence of several human factors in space and time. Since the burdens between trust, mental workload or situation awareness are often diffuse, their confluence may lead to different results to those observed when investigated in isolation, and possible "synergistic" effects. Further research should investigate the implications of the confluence of these human factors and how this could challenge the monitoring of functional states or the training of machine learning classifiers.

Conversely, when only investigating the effect of a single factor in the risk perception experiment, fNIRS data were aligned with EDA and ECG, indicating levels of perceived risk. These and other results from the broader literature suggest that these measures are usually sensitive to those constructs when investigated in isolation in controlled experiments. For example, experiments generating mental workload from a secondary task often report significant results in brain activity or arousal. In isolation and a controlled experiment, this can be expected. Comparing participants in a resting state with participants performing a demanding task does induce mental activity and body responses. Nevertheless, in real-world situations, these human factors constructs will often concur or overlap, and physiological indices may even show contradictory parameters like those described in the previous paragraph. Combining multimodal sources of data in DSM systems is therefore crucial for driver monitoring because, as mentioned, cardiac activity did capture such mental workload variations from the 2-back task that fNIRS did not; whilst at the same time, fNIRS captured differences in TiA that cardiac activity could not detect.

Understanding human factors confluence will be essential to develop DSM systems capable of disentangling several human factors concurring in time. Especially with growing evidence indicating that from the very first interactions, users seem to show a shift towards riskier driving behaviour with automated driving assistance. Drivers are less attentive (Hidalgo-Muñoz et al., 2019; Sibi et al., 2016), less aroused (Biondi et al., 2018; Carsten et al., 2012; De Winter et al., 2014; Melnicuk et al., 2017) and worsened their driving performance (Hidalgo-Muñoz et al., 2019; Rudin-Brown & Parker, 2004). Indeed, we suspect that our results from the driver availability experiment (chapter 4) indicating lower risk perception rates might be explained by this phenomenon. An assumption grounded upon the driver-in-control model (Hollnagel et al., 2003) predicts that automating certain driving tasks would influence drivers' situation assessment, and thus, SA, which was later evidenced in Carsten et al. (2012). Additional evidence can be found from a more recent model, the Human-Autonomy System Oversight (HASO) (Endsley, 2017), which builds upon previous frameworks of SA, mental workload, and TiA, to provide a more comprehensive view of factors affecting human supervision and intervention in automated systems. According to the HASO, the higher the automation level, the lower the situation awareness of human operators, and the less chance they will be available for taking over manual control when required.

Another aspect of particular relevance is the dynamic nature of these human factors. Trust in automation, situation awareness, or mental workload are not static; they all evolve along with the interactions with the system and contribute to a phenomenon known as behavioural adaptation derived from the repeated usage of driving automation (Jamson et al., 2013; Metz et al., 2021; Rudin-Brown & Parker, 2004). Behavioural adaptation refers to those ''behaviours which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change" (Metz et al., 2021). The current approach aims to understand how repeated experience with automated driving changes user attitudes and trust and identify those factors affecting driver behaviour in terms of engagement in NDRTs, take-over performance and mode awareness. A few studies have explored the long-term effects of the information displayed on a human-machine interface (HMI) on drivers' attitudes. For example, in a longitudinal driving simulator study using a partially automated vehicle (SAE L3), Ulahannan and colleagues measured gaze behaviour to investigate changes in information usage from an adaptive HMI after repeated interactions (Ulahannan et al., 2021). They found the usage of different information sources varied along with the repeated exposure to particular events and familiarisation with the driving automation. In a similar driving simulator study, Large et al. (2019) used a novel HMI to keep drivers in the loop during automated driving and help them regain situation awareness when take-over was required. After five consecutive days of exposure to the system, they found increments in trust and technology acceptance and take-over performance improvement. Behavioural adaptation is an area of growing interest since almost all research up to date has usually focused only on the early (1 or 2 hours) interactions of naïve participants encountering HAD technology for the very first time, and it might be an unforeseen source of human error derived from an inadequate deployment of driving automation. To date, the knowledge concerning the effects derived from the long-term interaction with automated driving technology is scarce.

With these concerns, the relational view of trust (Chiou & Lee, 2021) arises as to the need of understanding the relationship of cooperation and resilience with increasingly capable automated systems, since most of the research up to date has been focusing on understanding trust during the first interactions and its calibration for appropriate reliance in systems where humans have shared – if not complete – control. With technology swiftly evolving and humans becoming used to collaborating with increasingly capable automated systems artificially intelligent agents are likely to soon become co-workers rather than just simple tools. This framework stresses the need for a paradigm shift: from human operators to co-operators. Accordingly, it is important to consider not just human characteristics and how automation influences individuals but also how humans can influence automation and how such inter-independent interactions affect trusting automation (Chiou & Lee, 2021). A view that is also addressed in the HASO model aforementioned (Endsley, 2017). Cooperative behaviours such as collaboration, coordination, task reallocation or planning with increasingly anthropomorphised automated systems which exhibit social behaviours (Forster et al., 2017; Pak et al., 2012; Waytz et al., 2014), are likely to be modulated by relational trust aspects. These relational aspects will eventually determine shared SA and joint team performance. Therefore, anthropomorphised artificially intelligent agents will require human parameters such as behaviour, gestures, voice, gaze or arousal states to determine their behaviour towards their human partners (Chiou & Lee, 2021).

Privacy issues and the willingness to be monitored will be other determinant aspects for trusting highly automated driving technology and its acceptance. Early in the 90s, the researchers involved in the DETER project already discussed the issues related to the public acceptance and voluntary adoption of devices that monitor the driver. The authors argued that because drivers would not be expected to adopt such systems voluntarily, it should be considered to impose DSM through legislation (Brookhuis, 1993). Individuals may feel their privacy is violated with camera-based monitoring technologies and be concerned about protecting their personal data. Nevertheless, other individuals may tolerate sharing their data or customise what information is shared in order to benefit from the possibilities offered by DSM systems (Boeglin, 2015; Brown, 2016; Ulahannan et al., 2019). For example, the adoption of smartphone technology has found little to no market resistance. Recent studies exploring the public acceptance of DSM systems have identified social influence (Melnicuk et al., 2019; Smyth, Chen, et al., 2021), effort expectancy, performance expectancy, and attitudes towards using new technology as factors positively related to the behavioural intention to use DSM systems, whereas anxiety was negatively related (Smyth, Chen, et al., 2021). Social influence seems to be an essential aspect of public acceptance of DSM systems, as it was a factor consistent across both studies. According to Melnicuk and colleagues, this factor could be potentially promoted through endorsement by important others, such as family members or transportation authorities. Effort expectancy seems to be another relevant factor to consider, as it was the most determinant factor of use intention in Smyth and colleagues, highlighting the importance of integrated DSM systems (as opposed to wearable devices). In addition, system transparency and whether the information is only shared with the driver or with a third party can influence the acceptance and, thus, the effectiveness of DSM systems (Ghazizadeh & Lee, 2014).

In any case, legislation to determine liability in those accidents involving HAD technology is being developed, and this will likely include DSM data, raising concerns regarding the data management, storage and other privacy implications (Boeglin, 2015; Collingwood, 2017; Taeihagh & Lim, 2019). For example, will drivers be liable if they neglect a TOR by the system? Will the driver's state before the crash (e.g., sleeping vs awake) be considered mitigating or aggravating circumstances in such a case? On the contrary, will manufacturers be liable if their vehicles transfer the driving control without ensuring drivers are ready to drive manually? Similar to aircraft black boxes, Event Data Recorders (EDR) and Data Storage Systems for Automated Driving (DSSAD) are being considered valuable tools to investigate the causes of an accident more accurately (Taeihagh & Lim, 2019). The EDR is a data memory device that only records data when triggered (i.e., usually by the airbag) and is used for analysing and reconstructing traffic accidents. In contrast, the DSSAD is a driving mode storage device exclusive to HAD (SAE levels 3 and 4). It records continuous data regarding the driving control status (e.g., whether the system or the driver was in control at a specific time or whether a takeover was requested) before and after a safety-related event has occurred (Böhm et al., 2020; Kreutner et al., 2020). Whereas these systems will help determine the liability and clarify the causes of HAD-related accidents, it might be at the expense of drivers' freedom and privacy. However, liability cannot be one-sided at the expense of one party, and both the driver and vehicle manufacturer must be incentivised to behave with care.

Aside from drivers' functional state, other sources of personal data such as vehicle location or movements might be used for V2V (vehicle-to-vehicle), V2I (vehicle-to-infrastructure) or V2X (vehicle-to-anything) communications to deliver transport-related services, posing a severe threat of data privacy breaches (Seetharaman et al., 2021). A myriad of information of potential interest for third parties that might be sold or used for targeted marketing and commercial purposes (Collingwood, 2017), for the insurance industry to calculate the costs of their customers more accurately (Boeglin, 2015; Ghazizadeh & Lee, 2014) or even for logistics, delivery or taxi companies to surveil their professional drivers (Ghazizadeh & Lee, 2014). Cyber-intrusion and privacy issues may increase perceived risk towards the usage of automated driving technology, hampering individuals' propensity to trust (Hoffman et al., 2009; Seetharaman et al., 2021; Zhang et al., 2019). Cyber-security breaches may represent a risk of leaking personal information but also a safety-critical risk if hackers and terrorist organisations take control of these vehicles to commit criminal acts (Collingwood, 2017). All stakeholders must carefully consider privacy and cyber-security aspects for achieving relational trust. As discussed in section 2.2.1, trust is fundamental for technology acceptance and for establishing a relationship of cooperation, resilience and collaboration with automated agents.

To conclude, these models envision a future where the driving task will primarily be performed by the automated driving system without requiring human intervention or supervision, except for a limited range of situations where humans may have to take over control and in which the driver availability for optimal take-over manoeuvre will have to be supervised by the driver state monitoring system. Humans will likely become co-drivers, and thus, methods for supporting shared SA in human–autonomy teams will be required (Endsley, 2017). This thesis provides evidence favouring multimodal DSM systems as strong candidates for such a challenging task, bringing the human state into the joint driving performance loop to support a process of trusting through automation responsivity and shared situation awareness.

The key recommendations for measuring driver availability through physiology measures are summarised below:

- Heart rate variability parameters HF and RMSSD would be recommended for measuring long-term (> 2 min.) variations in vagal modulation (i.e., PNS). These are particularly useful for detecting slowly-evolving variations in mental workload and situation awareness but also would be good candidates for measuring overtrust and complacency.
- Heart rate and skin conductance level would be good candidates for measuring long-term levels of activation (SNS). This can be useful, e.g., for detecting overall levels of alertness, drowsiness, sleepiness, fatigue or complacency-overtrust.
- Skin conductance responses (i.e., count, magnitude or amplitude), oxygenated haemoglobin (HbO) or total haemoglobin (HbT) would be recommended for measuring rapid and short-term (< 2 min.) variations in levels of activation (SNS). This would ideally serve to detect the perception of risks, warnings or alerts, or momentary peaks of high mental workload and distrust.
- Comparably to current smartwatches and consumer-grade wearables, machine learning methods used to classify among such driver states will enormously benefit from repeated exposure to data inputs from daily driving situations, ideally in terms of several days or weeks, to achieve a high classification accuracy.

6.3 Recommendations for future work

6.3.1 Methodological recommendations

One of the aims and objectives of this thesis was to provide future work with a solid methodological grounding from which to work towards developing multimodal DSM systems. Aside from an original contribution to knowledge, we intended to identify limitations, drawbacks, and other relevant methodological aspects to consider in future work. With this aim in mind, some of the lessons learnt relate to the recording and processing of physiological data.

An important aspect to bear in mind is determining sufficient sample size. In both studies from this PhD project, sample sizes were determined based on related research, usually between 50 and 20 participants, so we established a target of roughly thirty participants per study. Although in the second study, the sample was smaller than the first one, a convenience sample included participants from both sexes equally distributed and with several professional roles and ages. As discussed in **Error! Reference source not found.**, this may have limited our results, but we are confident about the robustness of our findings. Another consideration when determining the sample size is the dropout rate of approximately 25% due to motion sickness in driving simulator experiments (Smyth, Jennings, et al., 2021). Ideally, future work should determine the sample size based on a power analysis to estimate the minimum sample size needed to obtain a specific effect size at a preset power level (1 - β, −i.e. 1 – the probability of a type II error, conventionally set to 0.8) and α (the probability of a type I error, conventionally set to 0.05) (Faul et al., 2007; Rossi, 2013).

Another methodological aspect worth considering is the order of experimental conditions. The "rule of thumb" would be counterbalancing the order of conditions to mitigate any order effects. For example, in the first study described in Chapter 4, having the 2-back task as the first condition may have contributed to generate a heart rate during the highway scenario unexpectedly higher than inter-urban and suburbs (Figure 21). However, to make the simulation more immersive, driving scenarios were designed as a single-run. An alternative to mitigate order effects could have been making several single-run driving scenarios with the order of conditions counterbalanced in each (e.g., starting with urban scenarios leading to a highway), but it would not have been practical as it took a number of months to create a single run with seven different scenarios. Future work should consider finding a balance between practicalities and mitigating effects of uncontrolled variables.

As discussed in both chapters 4 and 5, it is important that experiments using physiology measures ensure baseline recordings are taken in a familiar and dull environment, comprising at least five minutes. However, it would be unrealistic to expect future vehicle users to remain in a resting state for five minutes before every drive, so we encourage future research to explore the sensitivity of physiological indicators from several baseline states, not only from resting, simulating daily situations. Drivers' states may vary from day to day, so it is essential to investigate whether physiological measures are equally sensitive to arousal variations from, e.g., low to moderate, from moderate to high, from high to moderate or from low to high. Human factors research must find a balance between sufficient experimental control and sufficient experimental flexibility to replicate realistic situations and guarantee the generalisation and validity of results.

Relatedly, the development of physiological devices is progressing at a pace, and wearables have gone through considerable improvements in recent years (Egger et al., 2019; Kumari et al., 2017). As with the devices we have used in this thesis, current research-grade devices can be wireless and nomadic. While this is promising for its inclusion into connected and automated vehicles, researchers will have to deal with several drawbacks associated with their use until this technology is ready. For example, NIRSport (see 3.3) is a cutting-edge, wearable and research-grade fNIRS device, which allows recording brain metabolic data in a more naturalistic environment since the kit fits into a backpack. However, it must be calibrated beforehand in a dark light environment and is prone to muscular movement artefacts. We found it is broadly incompatible with those eye-tracking systems using infrared illumination directed towards the participant's face because it saturates fNIRS optodes.

Similarly, ECG from the BIOPAC MP-160 kit requires applying three electrodes to the participants' torso and care must be taken in not stretching the cables. HRV data also required minimum 2-minute epoch lengths conditioning the duration of experimental conditions and overall study length (Laborde et al., 2017; Shaffer & Ginsberg, 2017). EDA required two electrodes attached to the participants' non-dominant (left) hand to avoid artefacts when driving (Boucsein, 2012; Dawson et al., 2016), which in the UK happens to be the hand used for interacting with the in-vehicle touchscreen. Hence, this would have limited performing any NDRT involving the left hand with the in-vehicle interface or central console. Other than these, commercially available consumer-grade devices for sports physiology monitoring integrated into smart-watches and wristbands also offer promising sources of physiology data collection and are increasingly being used in research contexts (Beggiato et al., 2018; Melnicuk et al., 2017; Walker et al., 2019). However, the data collected by most of these devices have not been validated for research use yet (Laborde et al., 2017; Lohani et al., 2019). A careful adoption of such devices in research is recommended, as well as cross-validations between studies are required to validate its use in scientific research (Lohani et al., 2019).

6.3.2 Selection of consumer versus research-grade devices

Future work may encounter a decision to use consumer-grade or research-grade devices. Consumer-grade devices are usually more affordable, portable, easier to set up, and data is easier to collect and interpret because these are targeted to the general non-expert public (Beggiato et al., 2019; Melnicuk et al., 2017). However, any data collected is likely to be of worse quality than data from research-grade devices (Laborde et al., 2017; Lohani et al., 2019). This is basically because consumer-grade wearable devices are often crafted in the form of chest belts, wristbands or smart-watches for enhanced comfort and unobtrusiveness, and also they use cheaper sensors and less complex software/algorithms (Chan et al., 2012). However, designing for comfort may work to the detriment of data quality. For example, wristbands that collect skin conductance from the inner part of the wrist (e.g., Empatica E4) neglect the fact that skin conductance in that area is much lower than from the fingertips and palms of the hands, which have different skin properties that make them more suitable locations for placing the sensors (Boucsein, 2012; Dawson et al., 2016). In addition, such devices are prone to several sources of artefacts due to movement, impacts, poor or intermittent contact, or dirty skin and may not allow post-hoc raw data filtering data for artefact removal (Beggiato et al., 2019). Using consumer-grade devices, however, implies a cost reduction and avoids complex setups (Beggiato et al., 2019; Melnicuk et al., 2017) – e.g., instructions and procedures to start recording data, as well as a reduction in time needed to obtain results from the data gathered. Therefore, the decision of using consumer-grade or research-grade devices will essentially depend on the physiological data accuracy required for the study, the comfort of attachment, mobility, data richness and data accessibility (Hänsel et al., 2018), or even the available funding/resources/time. Therefore, the goal would be to have consumer-grade devices for practicality and comfort but more complex signal processing and modelling that deal with low-quality data and still output a robust result.

Recommendations for research investigating subtle – or moderate – arousal variations from changes in a driving context would be to use research-grade equipment. On the contrary, research aiming at measuring behavioural or performance interactions with the automated vehicle (e.g., when taking over control or performing NDRTs) would likely benefit from less obtrusive consumer-grade devices and induce sufficient physiological variations to be captured with such devices. Either way, future work aiming at recording any physiological indices must use several data sources in tandem to the extent possible (Collet & Musicant, 2019; Cowley et al., 2016; Egger et al., 2019; Kumari et al., 2017). For example, ECG and EDA devices in tandem with eye-trackers or using fNIRS in tandem with ECG or EDA. Doing so will not only provide researchers with more variables to interpret and understand their results but will also contribute to the development of future DSM systems in understanding potential drawbacks or challenges derived from the combination of several data modalities. We would recommend exploring any possible combinations in tandem with eye-trackers since the development of this technology is more advanced and even available in current production vehicles (see 2.3.1). Gaze behaviour can provide relevant information of drivers' attention allocation and attentional resources available, as well as being indicative of drowsiness, fatigue, or distractions. Such information could provide a complete picture of the current driver state when combined with arousal indicators or brain activity. It would indicate what the driver is focusing on and the level of engagement or activation. As discussed in chapter 5, this assessment would be key for determining drivers who have perceived a monitoring request and whether they have achieved an optimal state for safe take-over performance.

It is crucial then that physiology monitoring technology providers work closely with academics and researchers in developing enhanced systems allowing more experimental flexibility. It is also recommendable to develop user-friendly interfaces for the novice or inexperienced users, with available manuals and guidelines with basic principles for data pre-processing and analysis, not only for setting up the devices. BIOPAC, for example, offers several guidelines and technical advice online, with valuable practices and recommendations in these regards, although some manuals require an urgent update as the software has changed. Regarding fNIRS devices, massive efforts have been undertaken in recent years to achieve standards, guidelines and recommendations for the publication of fNIRS data (Pinti et al., 2019; Yücel et al., 2021). Nonetheless, analysing this data in the present study was the most time-consuming, complex and demanding process compared to the other data extracted, as one could quickly tell from the methods described in 3.3. There is a frustrating lack of standardisation in terminology (e.g., aside from HbR, deoxygenated haemoglobin is often referred to as *ΔDeoxyHb*, *deoxy-Hb*, *Hb*, or *deoxy*) and several aspects affecting data analysis procedures (e.g., GLM versus block averaging statistical approaches). Non-experts asking for technical advice on what is the best practice will usually find themselves with a response: "it depends, there is no good or bad way, it is under your criteria". Novice users usually lack such criteria, sufficient knowledge and understanding for making such decisions affecting the gross of their data. From our experience, there is currently a huge step between the fNIRS community of experts and inexperienced researchers who want to include this technology in their experiments. A step which is usually smaller and achievable with other devices for driving monitoring like eye-tracking or psychophysiology – although these are also more established and hold more years of research behind than fNIRS. Avoiding obstacles for non-experts in fNIRS should be a priority of its scientific leading community as it may result in published papers with questionable quality or methods, hampering the development of this technology.

6.3.3 Selection of questionnaires and scales

Last but not least, it is also recommendable to use self-report questionnaires updated according to the most current frameworks. An adequate scale or questionnaire choice can make a difference when measuring multidimensional constructs such as trust in automation, mental workload, or situation awareness. Especially for more recent or under-studied concepts, it is pertinent to use scales that have been created with an up-to-date theoretical background. Trust in automation is a relatively new construct receiving increasing attention with the appearance of automated driving technology, uncrewed air vehicles and an assortment of all-purpose assistant robots. Carefully identifying the most suitable scale for measuring trust in automated driving will be necessary for developing DSM systems. The well-known and established TASS (Jian et al., 2000) has its own merits for being used here and in several other studies since it made its appearance. However, after twenty-two years and – at least – two significant theoretical milestones in trust in automation (Hoff & Bashir, 2015; Lee & See, 2004), the TASS might be outdated in certain aspects. A study recently suggested that administering the scale without randomising the items may induce a positive bias. The scale is skewed towards positive ratings (Gutzwiller et al., 2019). It does not mean this scale is not valid for measuring trust and distrust anymore. In fact, according to this study, more than a hundred studies have administered the scale previously without randomising the items. If the scale is to be used, our recommendation is to randomise the items to avoid skew. Otherwise, ratings must be cautiously interpreted if moderate trust results are obtained. Furthermore, research has progressed, and now we know trust has three layers (Hoff & Bashir, 2015). It has been argued that the TASS would measure one of these layers, possibly *dispositional* trust, an individual trait considered to be invariant throughout the lifespan (Holthausen, 2020). Even though trust variations in within-participants experimental manipulations have been found here and in previous work (Banks & Stanton, 2016; Satterfield et al., 2017; Zhang et al., 2018), the items in the TASS are not addressing any specific feature concerning to the external variability in where the driving automation is tested. In other words, the TASS does not address the component of *situational* trust (Hoff & Bashir, 2015). In the driver availability experiment presented in chapter 4, changes in driving conditions that were expected to modulate trust levels would refer to such situational trust. It does not necessarily mean the variations in trust and distrust observed were not valid, but perhaps using a scale considering the situational aspect of trust would have been more accurate. A recently developed scale, the Situational Trust Scale for Automated Driving (STS-AD) (Holthausen, 2020), would have likely been a good choice in that case. Thus, we encourage future work aiming at measuring variations in trust in automated driving derived from situational manipulations to use the STS-AD. Otherwise, a potential and updated alternative to the TASS for measuring dispositional and learned aspects of TiA would be the Trust in Automation Scale from Körber (2018), which is aimed at a similar purpose but is based on a more updated framework.

Mental workload is a construct that has historically received more attention than TiA, and extensive work has been carried out in attempting to understand it, as earlier discussed in section 2.2.2. Therefore, the tools for self-reported mental workload assessment developed around twenty years ago are still valid and used nowadays. One of the most often used scales is the NASA-Task Load Index (Hart, 2006; Hart & Staveland, 1988) and still very current in the measure of mental workload in tandem with other physiological indicators (Du et al., 2019; Foy & Chapman, 2018; Solís-Marcos & Kircher, 2018). The Driving Activity Load Index (DALI) (Pauzié, 2008) is an adaptation of the NASA-TLX to the driving context and is also used in current research (Melnicuk et al., 2021). Regardless of having found effects on mental workload when performing the 2-back task in the driver availability experiment, its usage would not be recommended if the trials involve a full driving automation system (e.g., SAE L5), where drivers are not required to drive at all because the DALI is intended to measure mental workload from the driving task. Other recommendations for mental workload assessment are the Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988) and the Rating Scale Mental Effort (RSME). These four scales are the most used in the driving context (Paxion et al., 2014).

Regardless of not directly measuring SA in this thesis – as instead, we assessed emotions through the Self-Assessment Manikin (chapter 4) and perceived risk (chapter 5) –, we would recommend future related work to do so whenever it is possible since it would provide additional information on such a relevant construct for automated driving. There are currently several methods available for assessing SA, such as the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995a), the Situational Awareness Rating Technique (SART) (Taylor, 2018) or the Situation Present Assessment Method (SPAM) (Durso et al., 1998). Instead, if researchers want to assess self-reported emotions derived from the driving situation as we did in the driver availability experiment, aside from the SAM, another standard tool for measuring self-reported emotions is the Positive and Negative Affect Scale

(PANAS) (Watson et al., 1988). To assess perceived risk, to the best of our knowledge, only a few authors have attempted to do so within the automated driving context (Li et al., 2019; Rajaonah et al., 2008). Both studies used the same scale to measure perceived relational and situational risk. Relational risk refers to the drivers' attitude towards the automated vehicle based on previous and current experiences and would relate to the dispositional and learned dimensions of trust in automation. Whereas situational risk refers to the drivers' attitude towards the ongoing driving task or context, based on the evaluation of the likelihood and consequences of potential accidents, and relates with the situational component of TiA (Li et al., 2019). In the risk perception experiment, we were interested in this second form of perceived risk but, as argued in section 5.4.2, this scale was created to be rated straight after experiencing each driving condition, and due to the nature of our simulator study, pausing the trial for participants to respond these items would have been counter-indicated. Therefore, we strongly encourage future work to develop a perceived risk from the driving context scale suitable to be used post-experiment and including both relational and situational dimensions of risk.

The key recommendations for future related work are listed below:

- Baseline recordings of at least 5 minutes should always be taken when measuring variations in physiology parameters. The baseline recording does not necessarily have to be resting in the supine position. If the experiment implies a very naturalistic setup (i.e., real-world testing), this may just be taken in the drivers' seat before starting the trial.
- Highly naturalistic setups would likely benefit from consumer-grade physiology devices. However, it may compromise the quality of the data gathered, so care must be taken to control other sources of potential signal noise or additional artefacts, e.g., speaking (may induce ECG artefacts because of altered respiration or signal spikes on EDA).
- It would be recommendable for lab and driving simulator setups to use research-grade devices. However, researchers should first familiarise themselves with the current best practices and guidelines to guarantee good data quality.
- We strongly encourage researchers to record several sources of physiology data, especially data that can be correlated or coupled (e.g., HR + HRV, HF + RMSSD, HR + respiration, HR + SCL, SCR magnitude + SCR amplitude). This will allow testing the internal validity of results to make more robust interpretations.

- Regarding the selection of questionnaires or scales, it is recommendable to find a balance between validated, robust and updated tools.

6.4 Conclusions

Developing automated driving assistance systems, which involve human intervention to some extent, without considering the driver's functional state is not supported by this current research, and it is being legislated against (see section 1.1). Findings from air traffic and military domains (discussed previously) show that adaptive driving assistance systems based on operators' functional states enhance overall system performance. Hence, including driver cognitive and physiological states into the driving task assessment loop through DSM systems should enhance overall driver-automated driving system performance and safety.

Vehicle manufacturers are currently offering partially or conditionally automated driving technology without considering actual human operators' limitations and capabilities, which vary within contexts. Vehicle manufacturers and technology providers are racing to commercialise the highest driving automation level, missing important safety-critical aspects (which sooner or later arise as with the Tesla accidents reported in section 1.1); when instead they should focus on developing a safe driving automation technology that adapts to the driver's capability to assume manual control optimally. Hence, if highly automated driving is meant to be available for public use soon, it should be deployed to mitigate human error from road transport and bring about a whole new range of opportunities for elderly, impaired, and disabled individuals.

The results from this thesis indicate that this could only be achieved if humans are monitored and assessed by the DSM system. Unmonitored human drivers with shared control in the driving task (i.e., SAE levels 1 to 4) are still prone to errors. In HAD, they are even the source of errors not occurring within the lower levels of driving automation or non-assisted driving, such as those addressed in this thesis concerning overtrust, out-of-the-loop drivers, and mental workload derived from non-driving related activities.

Erasing human error from the driving task can only be achieved by excluding humans from the equation, and this scenario seems to be far away yet. Achieving SAE L5 driving automation available for the mass public soon is unlikely, regardless of the marketing promises made by the industry. Nearly to the completion of this thesis, relevant stakeholders in the industry have stressed publicly that such technology is not ready to be implemented yet. The most recent example comes from Mr Akido Toyota, president of Toyota Motor Corporation, after the embarrassing scene in the Olympics 2021, where an automated pod collided with a Paralympic athlete. In the meantime, highly automated driving technology using DSM systems relying on hybrid physiological cues and gaze behaviour, capable of providing dynamic driving assistance adapted to the driver functional state, could enhance road-transport safety.

To conclude, this research has highlighted the importance of DSM systems for the upcoming highly automated driving technology. In addition, it has also evidenced that such systems would benefit from multiple sources of driver data. Our findings set the path for future research investigating data fusion techniques and machine learning classifiers to develop real-time profiles of driver functional states. Considering that such systems will become a requirement in most Western countries in the upcoming years (see 1.1), this thesis turns out to be extremely timely. We expect automotive manufacturers and technology providers will enormously benefit from our work and that it would also encourage other research in the development of this technology.

Appendix 1

WARWICK
THE UNIVERSITY OF WARWICK

Participant:

Title of Project: Using Biometrics to better Understand User Engagement with Automated Vehicles

Name of Researcher(s): Jaume Perello-March, Stewart Birrell, Mark Elliott and Christopher Burns

Please tick the appropriate box:

1) Your gender
__ Male __ Female __ Other __ Prefer not to say

2) Your age
__ 25-30 __ 31-35 __ 36-40

__ 41-45 __ 46-50 __ 50+

3) Occupation
__ Professional and managerial __ Student __ Other
__ Clerical and sales __ Skilled or semi-skilled

4) Driving Experience
Years having driving license:
__ 0 – 2 __ 3 - 5 __ 6 - 15 __ 16 - 25 __ 26 +

Average mileage per year:
__ 0 – 10k __ 11k – 20k __ 21k – 30k __ 31k – 40k
__ 41k – 50k __ 51k – 60k __ 61k +

SHORT FIVE PERSONALITY TRAITS

Please read the following statements and indicate to what extent they apply to you. Write the suitable number in the box in front of the descriptive statement. Use the following scale:

(-3) Completely wrong (-2) Mostly wrong (-1) More wrong than right.
(0) Neutral; neither right nor wrong
(1) More right than wrong (2) Mostly right. (3) Completely right.

I am often nervous, fearful, and anxious, and worry that something might go wrong.	
I like people; I am friendly and open talking to strangers.	
I have a vivid imagination. I like to fantasize and let my thoughts run free.	
I trust people and I believe that everyone is honest and has good intentions most of the time.	
I am a serious rather than a cheerful person. I have rarely been overflowing with joy.	
I have quite traditional values; I am considered to be somewhat close-minded when it comes to the values of other cultures and other groups of people.	
I often rush into action without considering the consequences of my actions and decisions.	
I enjoy meeting and associating with a lot of people. I take pleasure in the company of others -the more people, the better.	
I have a deep appreciation for fine arts and beauty. I am much impressed by and interested in music, poetry, and art.	
I am a methodical person and I love cleanliness and order. I want every thing to be in its right place.	
I am not looking for excitement or adventures. I do not like to take risks.	
I am not interested in abstract or theoretical ideas. In my mind, ideas without practical applications are a waste of time.	
I often postpone difficult or unpleasant activities and leave things unfinished. It is difficult for me to pull myself together and do the things that I have to.	
I am a reliable person, who values ethical principles; I keep my promises and work carefully and thoroughly.	
I feel comfortable around other people; Most of the time I am not bothered by teasing or by embarrassing situations.	
I am a stubborn person who often gets into arguments; I openly express my anger or my dislike for someone.	
I am active and I like to keep myself busy; I often feel bursting with energy.	
I like to try different activities, to visit different places, to try out unfamiliar and exotic things from time to time; I love novelty and variety.	
I know for certain what I want to accomplish and I work hard for it.	
I rarely feel hopeless; I do not tend to blame myself needlessly. In general, I am content with myself and my life.	

I prefer to remain unnoticed in the background. I often let other people talk and decide things on my behalf.	
My feelings are not important to me; most of the time I do not pay any attention to them.	
I do not want to deal with other people's problems; I am considered a selfish and egotistical person.	
It is very difficult for me to resist temptation and to keep my desires and feelings in check; I do things that I regret later.	
I do not want to be in the centre of attention. I do not like to talk about myself or my accomplishments.	
I am a well composed person and it is difficult to upset or anger me.	
I believe that honesty does not take one very far in life. When necessary, I try to take advantage of others.	
I often feel helpless and indecisive, especially in complicated situations. I easily lose my nerve when I feel that I cannot cope.	
I believe that every person deserves respect. I feel compassion for those people who have been less lucky in life than I have	
I often feel that I am not competent enough to do something; I am not very productive and effective in my work.	

DRIVER BEHAVIOUR QUESTIONNAIRE

The next two pages of the booklet require you to judge the frequency of your own errors and violations. Twenty-four instances of these behaviours are listed. For each item you are asked to indicate how often, if at all, this kind of thing has happened to you. Base your judgments on what you remember of your driving over, say, the past year.

Please indicate your judgments by ticking one of the columns in the grid next to each item. You will notice that these columns are headed by numbers between 0 and 5. These mean the following: **0 = Never; 1 = Hardly ever; 2 = Occasionally; 3 = Quite often; 4 = Frequently; 5 = Nearly all the time**. Remember we do not expect precise answers, merely your best guesses; so please do not linger too long over anyone item.

	0	1	2	3	4	5
Attempt to drive away from traffic lights in third gear						
Become impatient with a slow driver in the outer lane and overtake on the inside						
Drive especially close to the car in front as a signal to its driver to go faster or get out of the way						
Attempt to overtake someone that you hadn't noticed to be signalling a right tum						
Forget where you left your car in car park						
Switch on one thing, such as the headlights, when you meant to switch on something else, such as the wipers						

Realise that you have no clear recollection of the road along which you have just been travelling						
Cross a junction knowing that the traffic lights have already turned against you						
Fail to notice that pedestrians are crossing when turning into a side street from a main road						
Angered by another driver's behaviour, you give chase with the intention of giving him/her a piece of your mind						
Misread the signs and exit from a roundabout on the wrong road						
Disregard the speed limits late at night or early in the morning						
On turning left, nearly hit a cyclist who has come up on your inside						
Queueing to tum left onto main road, you pay such close attention to the main stream of traffic that you nearly hit the car in front						
Drive even though you realise that you may be over the legal blood-alcohol limit						
Have an aversion to a particular class of road user, and indicate your hostility by whatever means you can						
Underestimate the speed of an oncoming vehicle when overtaking						
Hit something reversing that you had not previously seen						
Intending to drive to destination A, you 'wake up' to find yourself on the road to destination B, perhaps because the latter is your more usual destination						
Get into the wrong lane approaching a roundabout or a junction						
Miss 'Give Way' signs, and narrowly avoid colliding with traffic having the right of way						
Fail to check your rear view mirror before pulling out, changing lanes, etc.						
Get involved in unofficial 'races' with other drivers.						
Brake too quickly on a slippery road, or steer the wrong way into a skid.						

DRIVING INTERNALITY (DI) AND DRIVING EXTERNALITY (DE) SCALE

You will find in the following some opinions stated by various drivers concerning causes of accidents. Please express your degree of agreement or disagreement with each statement, selecting a number from the following scale: **Disagree very much (0), Disagree quite a bit (1), Disagree some (2), Agree a little (3), Agree quite a bit (4), Agree very much (5).**

	0	1	2	3	4	5
Driving with no accidents is mainly a matter of luck						
Accidents happen mainly because of different unpredictable events						
The driver can do nothing more than drive according to traffic regulations						
Accidents happen because of so many reasons we will never know the most important one						
People who drive a lot with no accidents are merely lucky; it is not because they are more careful						
The careful driver can prevent any accident						
When a driver is involved in an accident, it is because he did not drive as he should						
When a driver is involved in an accident it is because he did not pay attention to his driving						
Accidents are only the result of mistakes made by the driver						
The driver is to be blamed almost always when an accident occurs						
It is difficult to prevent accidents in bad conditions such as darkness, rain, narrow roads, curves, and so on						
Most accidents happen because of bad roads, lack of appropriate signs, and so on						
It is very hard to prevent accidents involving pedestrians who come out from between parked cars						
Accidents in which children are involved are hard to prevent because they do not know how to be careful						
It is very hard to prevent accidents in which old people are involved because they cannot hear nor see well						
Accidents happen because drivers have not learned how to drive carefully enough						
It is always possible to predict what is going to happen on the road and so it is possible to prevent almost any accident						
Accidents happen when the first driver does not take into consideration all the possible actions of the second driver						
Accidents happen because the driver does not make enough effort to detect all sources of danger while driving						
Most accidents happen because of lack of knowledge or laziness on the part of the driver						
If you are to be involved in an accident, it is going to happen anyhow, no matter what you do						
Most accidents happen because the second driver does not pay attention to traffic regulations even when the first driver does						
The driver does not have enough control over what happens on the road						
Most accidents happen because of mechanical failures						
There will always be accidents no matter how much drivers try to prevent them						

Accidents happen when the driver does not take into consideration all the possible behaviours of pedestrians						
Accident-free driving is a result of the driver's ability to pay attention to what is happening on the roads and sidewalks						
The driver can always predict what is going happen; that is why there is no room for surprises on the road						
It is possible to prevent accidents even in the most difficult conditions such as narrow roads, darkness, rain, and so on						
Prevention of accidents depends only on the driver and his characteristics rather than on external factors						

TRUST IN AUTOMATED SYSTEMS SCALE

Please, mark an X on each line at the point which best describes your feeling or your impression regarding the driving automation you are about to test. (**Not at all =1; Extremely =7**)

1. The system is deceptive.

1 2 3 4 5 6 7

2. The system behaves in an underhanded manner.

1 2 3 4 5 6 7

3. I am suspicious of the system's intent, action or outputs.

1 2 3 4 5 6 7

4. I am wary of the system.

1 2 3 4 5 6 7

5. The system's actions will have a harmful or injurious outcome.

1 2 3 4 5 6 7

6. I am confident in the system.

1	2	3	4	5	6	7

7. The system provides security.

1	2	3	4	5	6	7

8. The system has integrity.

1	2	3	4	5	6	7

9. The system is dependable.

1	2	3	4	5	6	7

10. The system is reliable.

1	2	3	4	5	6	7

11. I can trust the system.

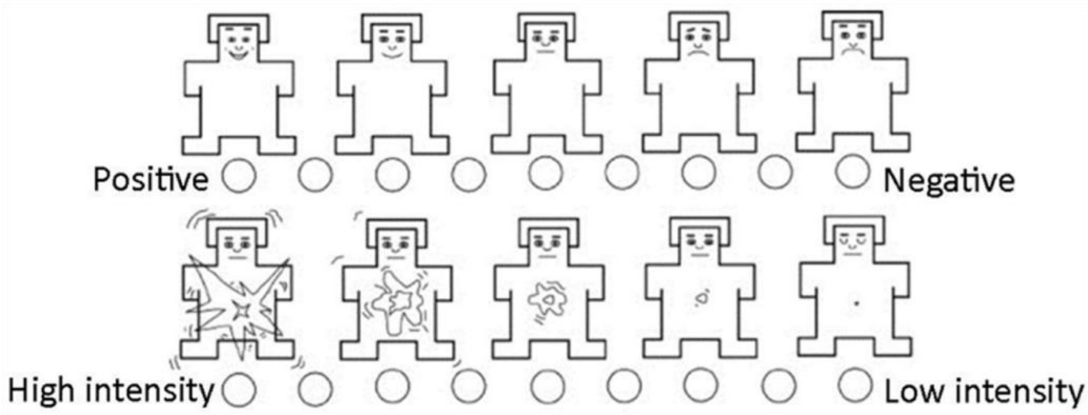
1	2	3	4	5	6	7

12. I am familiar with the system.

1	2	3	4	5	6	7

SELF-ASSESSMENT MANIKIN

Finally, please tick the circle that better represents your feelings after / before your experience with the automated driving vehicle.



SIMULATOR SICKNESS QUESTIONNAIRE

Please, indicate by ticking in the appropriate box whether you have experienced any of the following symptoms:

Symptom	Rating			
	None	Slight	Moderate	Severe
General discomfort	☐	☐	☐	☐
Fatigue	☐	☐	☐	☐
Headache	☐	☐	☐	☐
Eyestrain	☐	☐	☐	☐
Difficulty focusing	☐	☐	☐	☐
Increased salivation	☐	☐	☐	☐
Sweating	☐	☐	☐	☐
Nausea	☐	☐	☐	☐
Difficulty concentrating	☐	☐	☐	☐
Fullness of head	☐	☐	☐	☐
Blurred vision	☐	☐	☐	☐
Dizzy (eyes open)	☐	☐	☐	☐
Dizzy (eyes closed)	☐	☐	☐	☐
Vertigo	☐	☐	☐	☐
Stomach awareness	☐	☐	☐	☐
Burping	☐	☐	☐	☐

DRIVING ACTIVITY LOAD INDEX

Please rate the following items using the scale **1 = very low / 10 = very high.**

1. EFFORT OF ATTENTION. to evaluate the attention required by the activity – to think about, to decide, to choose, to look for and so on

1 2 3 4 5 6 7 8 9 10

2. VISUAL DEMAND. to evaluate the visual demand necessary for the activity.

1 2 3 4 5 6 7 8 9 10

3. AUDITORY DEMAND. to evaluate the auditory demand necessary for the activity.

1 2 3 4 5 6 7 8 9 10

4. TEMPORAL DEMAND. to evaluate the specific constraint owing to timing demand when running the activity

1 2 3 4 5 6 7 8 9 10

5. INTERFERENCE. to evaluate the possible disturbance when running the driving activity simultaneously with any other supplementary task such as phoning, using systems or radio and so on.

1 2 3 4 5 6 7 8 9 10

6. SITUATIONAL STRESS. to evaluate the level of constraints/stress while conducting the activity such as fatigue, insecure feeling, irritation, discouragement and so on.

1 2 3 4 5 6 7 8 9 10

Appendix 2

WARWICK
THE UNIVERSITY OF WARWICK

Participant ID:

Title of Project: Using Biometrics to better Understand User Engagement with Automated Vehicles

Name of Researcher(s): Jaume Perello-March, Stewart Birrell, Mark Elliott and Christopher Burns

Please tick the appropriate box:

1) Your gender

Male___ Female___ Other: ______ Prefer not to say___

2) Your age

___

3) Occupation

Professional and managerial___ Student___ Other___
Clerical and sales___ Skilled or semi-skilled___

4) Driving Experience
Years having driving license:

___

Average mileage per year:

___

Please, circle the number which best describes your feeling or your impression regarding the driving automation you are about to test.
(**Not at all =1; Extremely =7**)

13. The system is deceptive.

1 2 3 4 5 6 7

14. The system behaves in an underhanded manner.

1 2 3 4 5 6 7

15. I am suspicious of the system's intent, action or outputs.

1 2 3 4 5 6 7

16. I am wary of the system.

1 2 3 4 5 6 7

17. The system's actions will have a harmful or injurious outcome.

1 2 3 4 5 6 7

18. I am confident in the system.

1 2 3 4 5 6 7

19. The system provides security.

1 2 3 4 5 6 7

20. The system has integrity.

1 2 3 4 5 6 7

21. The system is dependable.

1 2 3 4 5 6 7

22. The system is reliable.

1 2 3 4 5 6 7

23. I can trust the system.

1 2 3 4 5 6 7

24. I am familiar with the system.

1 2 3 4 5 6 7

Please, indicate by ticking in the appropriate box whether you are experiencing any of the following symptoms:

Symptom	Rating			
	None	Slight	Moderate	Severe
General discomfort	☐	☐	☐	☐
Fatigue	☐	☐	☐	☐
Headache	☐	☐	☐	☐
Eyestrain	☐	☐	☐	☐
Difficulty focusing	☐	☐	☐	☐
Increased salivation	☐	☐	☐	☐
Sweating	☐	☐	☐	☐
Nausea	☐	☐	☐	☐
Difficulty concentrating	☐	☐	☐	☐
Fullness of head	☐	☐	☐	☐
Blurred vision	☐	☐	☐	☐
Dizzy (eyes open)	☐	☐	☐	☐
Dizzy (eyes closed)	☐	☐	☐	☐
Vertigo	☐	☐	☐	☐
Stomach awareness	☐	☐	☐	☐
Burping	☐	☐	☐	☐

Think back across the simulation you have just experienced. Please circle the number which best describes your feelings about:

(**Not at all =1; Extremely =7**)

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

1 2 3 4 5 6 7

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

1 2 3 4 5 6 7

Appendix 3

```
import mainscriptthread


class ScriptBiopac(mainscriptthread.MainScriptThread):
    """
    Script for Biopac
    """
    def __init__(self, connection):
        mainscriptthread.MainScriptThread.__init__(self, connection)


    def run(self):


        print("Script started...\n")
        self.sendKeyForTime(99999999, "key:F1", 5)


import mainscriptthread


class ScriptFNIRS(mainscriptthread.MainScriptThread):
    """
    Script for fNIRS
    """
    def __init__(self, connection):
        mainscriptthread.MainScriptThread.__init__(self, connection)


    def run(self):


        print("Script started...\n")
        self.sendKeyForTime(99999999, "key:F1", 15)
```

Bibliography

Adams, B. D., Bruyn, L. E., Houde, S., & Angelopoulos, P. (2003). *Trust in Automated Systems: Literature Review* (Issue June).

Ajenaghughrure, I. Ben, Sousa, S. D. C., & Lamas, D. (2020). Measuring trust with psychophysiological signals: A systematic mapping study of approaches used. In *Multimodal Technologies and Interaction* (Vol. 4, Issue 3, pp. 1–29). https://doi.org/10.3390/mti4030063

Alrefaie, M. T., Summerskill, S., & Jackon, T. W. (2019). In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles. *Accident Analysis and Prevention*, *131*, 180–190. https://doi.org/10.1016/j.aap.2019.06.011

Anderson, E. J., Mannan, S. K., Husain, M., Rees, G., Sumner, P., Mort, D. J., McRobbie, D., & Kennard, C. (2007). Involvement of prefrontal cortex in visual search. *Experimental Brain Research*, *180*(2), 289–302. https://doi.org/10.1007/s00221-007-0860-0

Backs, R. W., Lenneman, J. K., Wetzel, J. M., & Green, P. (2003). Cardiac Measures of Driver Workload during Simulated Driving with and without Visual Occlusion. *Human Factors*, *45*(4), 525–538. https://doi.org/10.1518/hfes.45.4.525.27089

Backs, R. W., Rohdy, J., & Barnard, J. (2005). Cardiac control during dual-task performance of visual or auditory monitoring with visual-manual tracking. *Psychologia*, *48*(2), 66–83. https://doi.org/10.2117/psysoc.2005.66

Baek, H. J., Lee, H. B., Kim, J. S., Choi, J. M., Kim, K. K., & Park, K. S. (2009). Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. *Telemedicine and E-Health*, *15*(2), 182–189. https://doi.org/10.1089/tmj.2008.0090

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321–348. https://doi.org/10.1080/14639220500535301

Bandara, D., Velipasalar, S., Bratt, S., & Hirshfield, L. (2018). Building predictive models of emotion with functional near-infrared spectroscopy. *International Journal of Human Computer Studies*, *110*(August 2017), 75–85. https://doi.org/10.1016/j.ijhcs.2017.10.001

Banks, V. A., & Stanton, N. A. (2016). Keep the driver in control: Automating automobiles of the future. *Applied Ergonomics*, *53*, 389–395. https://doi.org/10.1016/j.apergo.2015.06.020

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving. *Frontiers in Human Neuroscience*, *12*, 338. https://doi.org/10.3389/fnhum.2018.00338

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

Begum, S. (2013). Intelligent driver monitoring systems based on physiological sensor signals: A review. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 282–289. https://doi.org/10.1109/ITSC.2013.6728246

Berntson, G. G., & Cacioppo, J. T. (2004). Heart Rate Variability: Stress and Psychiatric Conditions. In *Dynamic Electrocardiography* (pp. 57–64). https://doi.org/10.1002/9780470987483.ch7

Bhardwaj, R., & Balasubramanian, V. (2019). Driver's Cardiac Activity Performance Evaluation Based on Non-contact ECG System Placed at Different Seat Locations. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & F. Y. (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). IEA 2018. Advances in Intelligent Systems and Computing* (Vol. 823, pp. 278–285). Springer, Cham. https://doi.org/10.1007/978-3-319-96074-6_30

Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., & Strayer, D. L. (2018). 80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal. *Proceedings of the Human Factors and Ergonomics Society*, *3*(1), 1878–1882. https://doi.org/10.1177/1541931218621427

Birrell, S., Fowkes, M., & Stubbs, D. (2012). Evaluation of smart driving advisors: smartphone apps or value added services. *Driver Behaviour and Training*, 451–470.

Bobko, P., Barelka, A. J., & Hirshfield, L. M. (2014). The construct of state-level suspicion: A model and research agenda for automated and information technology (IT) contexts. *Human Factors*, *56*(3), 489–508. https://doi.org/10.1177/0018720813497052

Boeglin, J. (2015). The Cost of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation. *Yale Journal of Law & Technology*, *17*(171), 30. https://heinonline.org/HOL/Page?handle=hein.journals/yjolt17&id=171&div=&collection=

Böhm, K., Kubjatko, T., Paula, D., & Schweiger, H. G. (2020). New developments on EDR (Event Data Recorder) for automated vehicles. *Open Engineering*, *10*(1), 140–146. https://doi.org/10.1515/eng-2020-0007

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. In *Neuroscience and Biobehavioral Reviews* (Vol. 44, pp. 58–75). https://doi.org/10.1016/j.neubiorev.2012.10.003

Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4614-1126-0

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bracken, B., Tobyne, S., Winder, A., Shamsi, N., & Endsley, M. R. (2021). Can Situation Awareness Be Measured Physiologically? In A. H., A. U., & P. L. (Eds.), *Advances in Neuroergonomics and Cognitive Engineering. AHFE 2021. Lecture Notes in Networks and Systems.* (Vol. 259, pp. 31–38). Springer, Cham. https://doi.org/10.1007/978-3-030-80285-1_4

Bradley, M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Semantic Differential Manikin and the. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(I), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Braithwaite, J. J., & Watson, D. G. (2015). *Issues Surrounding the Normalization and Standardisation of Skin Conductance Responses (SCRs)*.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). A Guide for Analysing Electrodermal Activity (EDA) and Skin Conductance Responses (SCRs) for Psychological Experiments. In *Technical Report, 2nd version*. https://www.biopac.com/wp-content/uploads/EDA-SCR-Analysis.pdf

Brookhuis, K. A. (1993). Detection, tutoring and enforcement of traffic rules violations - the DETER project. *Proceedings of the IEEE-IEE Vehicle Navigation and Informations Systems Conference*, 698–701. https://doi.org/10.1109/vnis.1993.585723

Brookhuis, K. A., van Driel, C. J. G., Hof, T., van Arem, B., & Hoedemaeker, M. (2009). Driving with a congestion assistant; mental workload and acceptance. *Applied Ergonomics*, *40*(6), 1019–1025. https://doi.org/10.1016/j.apergo.2008.06.010

Brown, D. (2016). *Responsibility for Vehicle Security and Driver Privacy in the Age of the Connected Car* (Issue EMEA41026016). https://media.scmagazine.com/documents/213/veracode_idc_connected_car_stu_53120.pdf

Bruno, J. L., Baker, J. M., Gundran, A., Harbott, L. K., Stuart, Z., Piccirilli, A. M., Hosseini, S. M. H., Gerdes, J. C., & Reiss, A. L. (2018). Mind over motor mapping: Driver response to changing vehicle dynamics. *Human Brain Mapping*, *39*(10), 3915–3927. https://doi.org/10.1002/hbm.24220

Bunce, S., Devaraj, a, Izzetoglu, M., Onaral, B., & Pourrezaei, K. (2005). Detecting deception in the brain: a functional near-infrared spectroscopy study of neural correlates of intentional deception. *Proceedings of SPIE*, *5769*(May 2005), 24–32. https://doi.org/10.1117/12.600601

Carsten, O., Lai, F., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control Task Substitution in Semiautomated Driving. *Human Factors*, *54*(5), 747–761. https://doi.org/10.1177/0018720812460246

Chan, M., Estève, D., Fourniols, J. Y., Escriba, C., & Campo, E. (2012). Smart wearable systems: Current status and future challenges. *Artificial Intelligence in Medicine*, *56*(3), 137–156. https://doi.org/10.1016/j.artmed.2012.09.003

Chen, L. L., Zhao, Y., Zhang, J., & Zou, J. Z. (2015). Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems with Applications*, *42*(21), 7344–7355. https://doi.org/10.1016/j.eswa.2015.05.028

Chiou, E. K., & Lee, J. D. (2021). Trusting Automation: Designing for Responsivity and Resilience. *Human Factors*, *00*(0), 1–29. https://doi.org/10.1177/00187208211009995

Cho, H. S., & Park, Y. J. (2018). Detection of heart rate through a wall using UWB impulse radar. *Journal of Healthcare Engineering*, *2018*, 1–8. https://doi.org/10.1155/2018/4832605

Coeugnet, S., Mars, F., Bueno, M., Sentouh, C., Popieul, J.-C., Koustanaï, A., Pauzié, A., & Tattegrain, H. (2021). A User-Centered Approach to Adapt the Human-Machine Cooperation Strategy in Autonomous Driving. In N. L. Black, W. P. Neumann, & I. Noy (Eds.), *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)* (pp. 600–606). Springer International Publishing.

https://doi.org/10.1007/978-3-030-74608-7_73

Collet, C., & Musicant, O. (2019). Associating vehicles automation with drivers functional state assessment systems: A challenge for road safety in the future. *Frontiers in Human Neuroscience*, *13*(131), 1–12. https://doi.org/10.3389/fnhum.2019.00131

Collet, C., Petit, C., Priez, A., & Dittmar, A. (2005). Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation. *Biological Psychology*, *69*(2), 195–203. https://doi.org/10.1016/j.biopsycho.2004.07.003

Collingwood, L. (2017). Privacy implications and liability issues of autonomous vehicles. *Information and Communications Technology Law*, *26*(1), 32–45. https://doi.org/10.1080/13600834.2017.1269871

Costa Jr., P. T., & McCrae, R. R. (2008). *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing*. SAGE Publications Ltd. https://doi.org/10.4135/9781849200479

Cowley, B., Filetti, M., Lukander, K., Torniainen, J., Henelius, A., Ahonen, L., Barral, O., Kosunen, I., Valtonen, T., Huotilainen, M., Ravaja, N., & Jacucci, G. (2016). The Psychophysiology Primer: A Guide to Methods and a Broad Review with a Focus on Human–Computer Interaction. *Foundations and Trends® in Human–Computer Interaction*, *9*(3–4), 151–308. https://doi.org/10.1561/1100000065

Daimler AG. (2009). *ATTENTION ASSIST: Drowsiness-detection system warns drivers to prevent them falling asleep momentarily - Daimler Global Media Site*. https://media.daimler.com/marsMediaSite/en/instance/ko.xhtml?oid=9361586

Damiani, S., Deregibus, E., & Andreone, L. (2009). Driver-vehicle interfaces and interaction: Where are they going? *European Transport Research Review*, *1*(2), 87–96. https://doi.org/10.1007/s12544-009-0009-2

Damm, W., Fränzle, M., Lüdtke, A., Rieger, J. W., Trende, A., & Unni, A. (2019). Integrating neurophysiological sensors and driver models for safe and performant automated vehicle control in mixed traffic∗. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2019-June*, 82–89. https://doi.org/10.1109/IVS.2019.8814188

Dawson, M. E., Schell, A. M., & Filion, D. L. (2016). The electrodermal system. In *Handbook of Psychophysiology, Fourth Edition* (pp. 217–243). Cambridge University Press. https://doi.org/10.1017/9781107415782.010

de Visser, E. J. (2012). *The World Is Not Enough: Trust in Cognitive Agents* [George Mason University Fairfax, VA]. http://www.cs.cmu.edu/~softagents/theses/GMU_deVisser_PhD_Thesis_2012.pdf

de Visser, E. J., Beatty, P. J., Estepp, J. R., Kohn, S., Abubshait, A., Fedota, J. R., & McDonald, C. G. (2018). Learning From the Slips of Others: Neural Correlates of Trust in Automated Agents. *Frontiers in Human Neuroscience*, *12*(309). https://doi.org/10.3389/fnhum.2018.00309

De Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. Universitiy of Groningen.

de Winter, J. C. F., & Dodou, D. (2010). The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *Journal of Safety Research*, *41*, 463–470. https://doi.org/10.1016/j.jsr.2010.10.007

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. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 196–217. https://doi.org/10.1016/j.trf.2014.06.016

Dimoka, A. (2010). What Does the Brain Tell Us About Trust and Distrust? Evidence from a Functional Neuroimaging Study. *MIS Quarterly*, *34*(2), 373–396. http://www.jstor.org/stable/20721433 Accessed:

Dixon, L. (2020). Autonowashing: The Greenwashing of Vehicle Automation. In *Transportation Research Interdisciplinary Perspectives* (Vol. 5). Elsevier Ltd. https://doi.org/10.1016/j.trip.2020.100113

Dixon, L., Megill, W. M., & Nebe, K. (2019). Trust in Automation : An On-Road Study of Trust in Advanced Driver Assistance Systems. *VEHICULAR 2019 : The Eighth International Conference on Advances in Vehicular Systems, Technologies and Applications*, *c*, 85–93. https://www.thinkmind.org/index.php?view=article&articleid=vehicular_2019_6_10_30041

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 596–614. https://doi.org/10.1109/TITS.2010.2092770

Drnec, K., Marathe, A. R., Lukos, J. R., & Metcalfe, J. S. (2016). From Trust in Automation to Decision Neuroscience: Applying Cognitive Neuroscience Methods to Understand and Improve Interaction Decisions Involved in Human Automation Interaction. *Frontiers in Human Neuroscience*, *10*(June), 1–14. https://doi.org/10.3389/fnhum.2016.00290

Drnec, K., & Metcalfe, J. S. (2016). Paradigm Development for Identifying and Validating Indicators of Trust in Automation in the Operational Environment of Human Automation Integration. In D. Schmorrow & C. Fidopiastis (Eds.), *Lecture Notes in Computer Science (Foundations of Augmented Cognition)* (Vol. 9744, pp. 157–167). Springer International Publishing AG. https://doi.org/10.1007/978-3-642-02812-0

Du, N., Haspiel, J., Zhang, Q., Tilbury, D., Pradhan, A. K., Yang, X. J., & Robert, L. P. (2019). Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies*, *104*, 428–442. https://doi.org/10.1016/j.trc.2019.05.025

Durantin, G., Gagnon, J. F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16–23. https://doi.org/10.1016/j.bbr.2013.10.042

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation Awareness as a Predictor of Performance for En Route Air Traffic Controllers. *Air Traffic Control Quarterly*, *6*(1), 1–20. https://doi.org/10.2514/atcq.6.1.1

Egger, M., Ley, M., & Hanke, S. (2019). Emotion Recognition from Physiological Signal Analysis: A Review. *Electronic Notes in Theoretical Computer Science*, *343*, 35–55. https://doi.org/10.1016/j.entcs.2019.04.009

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (1995b). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned from Human-Automation Research. *Human Factors*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

ETSC. (2020). *Road deaths in the European Union*. Road Deaths in the European Union – Latest Data. https://etsc.eu/euroadsafetydata/

EuroNCAP. (2017). *Euro NCAP 2025 Roadmap: In Pursuit of Vision Zero*. https://cdn.euroncap.com/media/30700/euroncap-roadmap-2025-v4.pdf

EuroNCAP. (2018). *For the first time, Euro NCAP puts automated driving technology to the test*. https://www.euroncap.com/en/press-media/press-releases/testingautomation/#nbp

EuropeanComission. (2018). *CARE Reports*. Statistics – Accidents Data | Mobility and Transport. https://ec.europa.eu/transport/road_safety/specialist/statistics_en#

EuropeanCommission. (2019). *Autoconduct: Adaptation of the automation strategy of autonomous vehicles (levels 3-4) to driver needs and driver state under real conditions*. https://trimis.ec.europa.eu/project/adaptation-automation-strategy-autonomous-vehicles-levels-3-4-driver-needs-and-driver-state#tab-outline

EuropeanCouncil. (2019). *Safer cars in the EU*. https://www.consilium.europa.eu/en/press/press-releases/2019/11/08/safer-cars-in-the-eu/

Fairclough, S. H. (1993). Psychophysiological measures of workload and stress. In *DRIVING FUTURE VEHICLES* (pp. 377–390). Taylor & Francis Group. http://worldcat.org/isbn/0748400427

Fairclough, S. H., Brookhuis, K. a, & Vallet, M. (1993). Driver state monitoring system DETER (V2009). *Advanced Transport Telematics: Proceedings of Technical Days Brussels, 8- 9-10 March 1993. Vol. 2*, 330–335. https://cir.nii.ac.jp/crid/1573668924042308864.bib?lang=en

Fairclough, S. H., & Hirst, S. J. (1993). Driver status monitoring: Can In-vehicle IT detect the impaired driver? In E. J. Lovesey (Ed.), *Contemporary Ergonomics: Proceedings of the Ergonomics Society's 1993 Annual Conference* (pp. 476–481). Taylor & Francis Group. https://shora.tabriz.ir/Uploads/83/cms/user/File/657/E_Book/Economics/Contemporary Ergonomics 1993.pdf#page=486

Fallgatter, A. J., & Strik, W. K. (1998). Frontal brain activation during the Wisconsin Card Sorting Test assessed with two-channel near-infrared spectroscopy. *Eur Arch Psychiatry Clin Neurosci*, *248*, 245–249.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferrara, M., & De Gennaro, L. (2000). The sleep inertia phenomenon during the sleep-wake transition: Theoretical and operational issues. *Aviation Space and Environmental Medicine*, *71*(8), 843–848. https://www.researchgate.net/publication/12367344

Ford. (2012). *Ford Biometric System Can Track Your Breathing, Heartbeat And Sweaty Palms*. Huffington Post. goo.gl/uIv99N

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. *IEEE Intelligent Vehicles Symposium, Proceedings*, *2*(Iv), 365–372. https://doi.org/10.1109/IVS.2017.7995746

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

Foy, H. J., Runham, P., & Chapman, P. (2016). Prefrontal cortex activation and young driver behaviour: A fNIRS study. *PLoS ONE*, *11*(5), 1–19. https://doi.org/10.1371/journal.pone.0156512

Gable, T. M., Kun, A. L., Walker, B. N., & Winton, R. J. (2015). Comparing Heart Rate and Pupil Size as Objective Measures of Workload in the Driving Context: Initial Look. *AutomotiveUI 2015: Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive VehicularApplications,* 20–25. https://doi.org/10.1145/2809730.2809745

Gambi, E., Agostinelli, A., Belli, A., Burattini, L., Cippitelli, E., Fioretti, S., Pierleoni, P., Ricciuti, M., Sbrollini, A., & Spinsante, S. (2017). Heart rate detection using microsoft kinect: Validation and comparison to wearable devices. *Sensors*, *17*(8), 1–18. https://doi.org/10.3390/s17081776

Ghazizadeh, M., & Lee, J. D. (2014). Modelling driver acceptance: From feedback to monitoring and mentoring systems. In *Driver Acceptance of New Technology: Theory, Measurement and Optimisation* (pp. 51–70). Ashgate Publishing Ltd. https://doi.org/10.1201/9781315578132-5

Girouard, A., Solovey, E. T., Hirshfield, L. M., Chauncey, K., Sassaroli, A., Fantini, S., & Jacob, R. J. K. (2009). Distinguishing difficulty levels with non-invasive brain activity measurements. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5726 LNCS*(PART 1), 440–452. https://doi.org/10.1007/978-3-642-03655-2_50

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 Manufacturing*, *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. *Human Factors*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Gold, C., Lorenz, L., Damböck, D., & Bengler, K. J. (2013). Partially Automated Driving as a Fallback Level of High Automation. In *6. Tagung Fahrerassistenzsysteme. Der Weg zum automatischen Fahren.* (pp. 1–5). https://mediatum.ub.tum.de/doc/1187198/file.pdf

Gugerty, L. (2011). Situation awareness in driving. *Handbook for Driving Simulation in Engineering, …*, 265–272. http://business.clemson.edu/Psych/gugerty/Gugerty2011SAdriving.pdf

Gutzwiller, R. S., Chiou, E. K., Craig, S. D., Lewis, C. M., Lematta, G. J., & Hsiung, C. (2019). Positive bias in the ‘ Trust in Automated Systems Survey ’? An examination of the Jian et al . (2000) scale. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *June*.

Hänsel, K., Poguntke, R., Haddadi, H., Alomainy, A., & Schmidt, A. (2018). What to put on the user: Sensing technologies for studies and physiology aware systems. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*. https://doi.org/10.1145/3173574.3173719

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meting*, 904–908. https://doi.org/10.1037/e577632012-009

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Healey, J. A., & Picard, R. W. (2005). Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 156–166. https://doi.org/10.1109/TITS.2005.848368

Hecht, T., Feldh, A., Radlmayr, J., Nakano, Y., Miki, Y., & Henle, C. (2019). A Review of Driver State Monitoring Systems in the Context of Automated Driving. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)* (Vol. 1, pp. 398–408). Springer. https://doi.org/10.1007/978-3-319-96098-2

Hidalgo-Muñoz, A. R., Jallais, C., Evennou, M., Ndiaye, D., Moreau, F., Ranchet, M., Derollepot, R., & Fort, A. (2019). Hemodynamic responses to visual cues during attentive listening in autonomous versus manual simulated driving: A pilot study. *Brain and Cognition*, *135*, 1–8. https://doi.org/10.1016/j.bandc.2019.103583

Hirshfield, L., Bobko, P., Barelka, A., Hirshfield, S. H., Farrington, M. T., Gulbronson, S., & Paverman, D. (2014). Using noninvasive brain measurement to explore the psychological effects of computer malfunctions on users during human-computer interactions. *Advances in Human-Computer Interaction*, *2014*, 1–13. https://doi.org/10.1155/2014/101038

Hirshfield, L., Bobko, P., Barelka, A., Sommer, N., & Velipasalar, S. (2019). Toward Interfaces that Help Users Identify Misinformation Online: Using fNIRS to Measure Suspicion. *Augmented Human Research*, *4*(1), 1–13. https://doi.org/10.1007/s41133-019-0011-8

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoffman, R. R., Lee, J. D., Woods, D. D., Shadbolt, N., Miller, J., & Bradshaw, J. M. (2009). The Dynamics of Trust in Cyberdomains. *IEEE Intelligent Systems*, *24*(6), 5–11. https://doi.org/10.1109/MIS.2009.124

Hogervorst, M. A., Brouwer, A. M., & van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in Neuroscience*, *8*(OCT), 1–14. https://doi.org/10.3389/fnins.2014.00322

Hollnagel, E., Nåbo, A., & Lau, I. V. (2003). A systemic model for driver-in-control. *PROCEEDINGS of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 86–91. https://doi.org/Export Date 14 October 2011

Holthausen, B. E. (2020). *Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation* [Georgia Institute of Technology]. https://doi.org/10.1145/3409120.3410637

Hoshi, Y., Huang, J., Kohri, S., Iguchi, Y., Naya, M., Okamoto, T., & Ono, S. (2011). Recognition of Human Emotions from Cerebral Blood Flow Changes in the Frontal Region: A Study with Event-Related Near-Infrared Spectroscopy. *Journal of Neuroimaging*, *21*(2), 94–101. https://doi.org/10.1111/j.1552-6569.2009.00454.x

Hubert, M., Hubert, M., Linzmajer, M., Riedl, R., & Kenning, P. (2018). Trust me if you can – neurophysiological insights on the influence of consumer impulsiveness on trustworthiness evaluations in online settings. *European Journal of Marketing*, EJM-12-2016-0870. https://doi.org/10.1108/EJM-12-2016-0870

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280. https://doi.org/10.1364/AO.48.00D280

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., & Chance, B. (2004). Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. *International Journal of Human-Computer Interaction*, *17*(2), 211–227. https://doi.org/10.1207/s15327590ijhc1702_6

Jaguar-LandRover. (2018). *Activity Key*. https://accessories.jaguar.com/gb/en/f-pace/interior/function-technology/t4a12643-activity-key-vehicle-preparation-kit-pre-21my/#tertiary-tabs

Jamson, A. H., Merat, N., Carsten, O. M. J., & Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, *30*, 116–125. https://doi.org/10.1016/j.trc.2013.02.008

Jian, J.-Y., Bisantz, A. M., & Drury, C. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71. https://doi.org/https://doi.org/10.1207/S15327566IJCE0401_04

Jonah, B. A. (1997). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis and Prevention*, *29*(5), 651–665. https://doi.org/10.1016/S0001-4575(97)00017-1

Kageyama, Y., Odagaki, M., & Hosaka, H. (2007). Wavelet analysis for quantification of mental stress stage by finger-tip photo-plethysmography. *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1846–1849. https://doi.org/10.1109/IEMBS.2007.4352674

Kahneman, D. (1973). Attention and effort. In *The American Journal of Psychology* (Vol. 88, Issue 2). https://doi.org/10.2307/1421603

Kahneman, D., & Tversky, A. (1983). Choices, Values, and Frames. In L. C. MacLean & W. T. Ziemba (Eds.), *Handbook of the Fundamentals of Financial Decision Making* (Vol. 4, pp. 269–278). World Scientific Publishing. https://doi.org/10.1142/9789814417358_0016

Kahneman, D., & Tversky, A. (2019). Prospect theory: An analysis of decision under risk. In *Choices, Values, and Frames* (Vol. 47, Issue 2, pp. 17–43). https://doi.org/10.1017/CBO9780511803475.003

Kang, H. B. (2013). Various approaches for driver and driving behavior monitoring: A review. *Proceedings of the IEEE International Conference on Computer Vision*, 616–623. https://doi.org/10.1109/ICCVW.2013.85

Kästle, J. L., Anvari, B., Krol, J., & Wurdemann, H. A. (2021). Correlation between Situational Awareness and EEG signals. *Neurocomputing*, *432*, 70–79. https://doi.org/10.1016/j.neucom.2020.12.026

Khastgir, S., Birrell, S., Dhadyalla, G., & Jennings, P. (2017). Calibrating Trust to Increase the Use of Automated Systems in a Vehicle. *Advances in Intelligent Systems and Computing*, *484*, 535–546. https://doi.org/10.1007/978-3-319-41682-3_45

Khastgir, S., Birrell, S., Dhadyalla, G., & Jennings, P. (2018). Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation Research Part C: Emerging Technologies*, *96*(July), 290–303. https://doi.org/10.1016/j.trc.2018.07.001

Khastgir, S., Birrell, S., Dhadyalla, G., & Jennings, P. (2019). Effect of Knowledge of Automation Capability on Trust and Workload in an Automated Vehicle: A Driving Simulator Study. *Advances in Intelligent Systems and Computing*, *786*, 410–420. https://doi.org/10.1007/978-3-319-93885-1_37

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kircher, K. (2007). Driver distraction - A review of the literature. *VTI Rapport 594A. VTI Linköping, Sweden*. http://www.erskine-mcmahon.com/files/cell.phone.driving.distraction.article.10.pdf

Kircher, K., Larsson, A., & Hultgren, J. A. (2014). Tactical driving behavior with different levels of automation. *IEEE Transactions on Intelligent Transportation Systems*, *15*(1), 158–167. https://doi.org/10.1109/TITS.2013.2277725

Konstabel, K., Lönnqvist, J. E., Leikas, S., Velázquez, R. G., Qin, H., Verkasalo, M., & Walkowitz, G. (2017). Measuring single constructs by single items: Constructing an even shorter version of the “Short Five” personality inventory. *PLoS ONE*, *12*(8), 1–18. https://doi.org/10.1371/journal.pone.0182714

Konstabel, K., Lönnqvist, J. E., Walkowitz, G., Konstabel, K., & Verkasalo, M. (2012). The ‘Short Five’ (S5): Measuring personality traits using comprehensive single items. *European Journal of Personality*, *26*(1), 13–29. https://doi.org/10.1002/per.813

Körber, M. (2018). Theoretical considerations and development of a questionnaire to measure trust in automation. *20th Triennial Congress of the IEA*, *March*, 1–20. https://doi.org/10.17605/OSF.IO/NFC45

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/j.apergo.2017.07.006

Kraus, J., Scholz, D., Stiegemeier, D., & Baumann, M. (2020). The More You Know: Trust Dynamics and Calibration in Highly Automated Driving and the Effects of Take-Overs, System Malfunction, and System Transparency. *Human Factors*, *62*(5), 718–736. https://doi.org/10.1177/0018720819853686

Kreutner, M. A., Gwehenberger, J., Braxmeier, O., Lauterwasser, C., Borrack, M., &

Reinkemeyer, C. (2020). Needs and Requirements of EDR/DSSAD - Analysis based on insurance claims reported to Allianz Germany. *29th Annual Congress of the European Association for Accident Research (EVU)*, 1–13. https://evuonline.org/attachments/article/3078/EVU2020-05-64 Kreutner EDR-DSSAD.pdf

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., & Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, *104*(50), 20084–20089. https://doi.org/10.1073/pnas.0710103104

Kumari, P., Mathew, L., & Syal, P. (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. *Biosensors and Bioelectronics*, *90*(December 2016), 298–307. https://doi.org/10.1016/j.bios.2016.12.001

Kundinger, T., Riener, A., & Wintersberger, P. (2019, May 2). (Over)trust in automated driving: The sleeping pill of tomorrow? *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312869

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting. *Frontiers in Psychology*, *8*(213), 1–18. https://doi.org/10.3389/fpsyg.2017.00213

Large, D. R., Burnett, G., Salanitri, D., Lawson, A., & Box, E. (2019). A longitudinal simulator study to explore drivers' behaviour in level 3 automated vehicles. *Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019*, 222–232. https://doi.org/10.1145/3342197.3344519

Lazaro, A., Girbau, D., & Villarino, R. (2010). Analysis of vital signs monitoring using an IR-UWB radar. *Progress in Electromagnetics Research*, *100*, 265–284. https://doi.org/10.2528/PIER09120302

Lee, J., Abe, G., Sato, K., & Itoh, M. (2021). Developing human-machine trust: Impacts of prior instruction and automation failure on driver trust in partially automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *81*, 384–395. https://doi.org/10.1016/j.trf.2021.06.013

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & Moray, N. (1994). Trust, self-Confidence, and operators' adaptation to automation. In *International Journal of Human - Computer Studies* (Vol. 40, Issue 1, pp. 153–184). https://doi.org/10.1006/ijhc.1994.1007

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lenneman, J. K., & Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors*, *51*(3), 404–418. https://doi.org/10.1177/0018720809337716

Leon-Dominguez, U., Izzetoglu, M., Leon-Carrion, J., Solís-Marcos, I., Garcia-Torrado, F. J., Forastero-Rodríguez, A., Mellado-Miras, P., Villegas-Duque, D., Lopez-Romero, J. L., Onaral, B., & Izzetoglu, K. (2014). Molecular concentration of deoxyHb in human prefrontal cortex predicts the emergence and suppression of consciousness. *NeuroImage*, *85*, 616–625. https://doi.org/10.1016/j.neuroimage.2013.07.023

Leonhardt, S., Leicht, L., & Teichmann, D. (2018). Unobtrusive vital sign monitoring in automotive environments—A review. *Sensors (Switzerland)*, *18*(9), 1–38. https://doi.org/10.3390/s18093080

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

Li, Z., Zhang, M., Zhang, X., Dai, S., Yu, X., & Wang, Y. (2009). Assessment of cerebral oxygenation during prolonged simulated driving using near infrared spectroscopy: Its implications for fatigue development. *European Journal of Applied Physiology*, *107*(3), 281–287. https://doi.org/10.1007/s00421-009-1122-6

Lin, C. T., King, J. T., Chuang, C. H., Ding, W., Chuang, W. Y., Liao, L. De, & Wang, Y. K. (2020). Exploring the Brain Responses to Driving Fatigue through Simultaneous EEG and fNIRS Measurements. *International Journal of Neural Systems*, *30*(1), 1–12. https://doi.org/10.1142/S0129065719500187

Liu, T., Pelowski, M., Pang, C., Zhou, Y., & Cai, J. (2016). Near-infrared spectroscopy as a tool for driving research. *Ergonomics*, *59*(3), 368–379. https://doi.org/10.1080/00140139.2015.1076057

Lohani, M., Payne, B. R., & Strayer, D. L. (2019). A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving. *Frontiers in Human Neuroscience*, *13*(57), 1–27. https://doi.org/10.3389/fnhum.2019.00057

López-Gil, J.-M., Virgili-Gomá, J., Gil, R., & García, R. (2016). Method for Improving EEG Based Emotion Recognition by Combining It with Synchronized Biometric and Eye Tracking Technologies in a Non-invasive and Low Cost Way. *Frontiers in Computational Neuroscience*, *10*(August), 9–10. https://doi.org/10.3389/fncom.2016.00085

Lu, Z., Zhang, B., Feldhütter, A., Happee, R., Martens, M., & De Winter, J. C. F. (2019). Beyond mere take-over requests: The effects of monitoring requests on driver attention, take-over performance, and acceptance. *Transportation Research Part F: Traffic Psychology and Behaviour*, *63*, 22–37. https://doi.org/10.1016/j.trf.2019.03.018

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Marberger, C., Mielenz, H., Naujoks, F., Radlmayr, J., Bengler, K., & Wandtner, B. (2018). Understanding and applying the concept of "Driver availability" in automated driving. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing* (Vol. 597, pp. 595–605). Springer, Cham. https://doi.org/10.1007/978-3-319-60441-1_58

Martens, M. H., & Van Den Beukel, A. P. (2013). The road to automated driving: Dual mode and human factors considerations. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2262–2267. https://doi.org/10.1109/ITSC.2013.6728564

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management*, *20*(3), 709–734. https://doi.org/10.5465/AMR.1995.9508080335

McKendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., & Ayaz, H. (2016). Into the wild: Neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. *Frontiers in Human Neuroscience*, *10*(MAY2016). https://doi.org/10.3389/fnhum.2016.00216

Mehagnoul-Schipper, D. J., Van Der Kallen, B. F. W., Colier, W. N. J. M., Van Der Sluijs, M. C., Van Erning, L. J. T. O., Thijssen, H. O. M., Oeseburg, B., Hoefnagels, W. H. L., & Jansen, R. W. M. M. (2002). Simultaneous measurements of cerebral oxygenation changes during brain activation by near-infrared spectroscopy and functional magnetic resonance imaging in healthy young and elderly subjects. *Human Brain Mapping*, *16*(1), 14–23. https://doi.org/10.1002/hbm.10026

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

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

Melnicuk, V., Birrell, S., & Thompson, S. (2019). How Acceptable Is It to Monitor Driver State? Using the UTAUT Model to Analyse Drivers' Perceptions Towards the System. *Advances in Human Aspects of Transportation*, *786*, 579–590. https://doi.org/10.1007/978-3-319-93885-1

Melnicuk, V., Thompson, S., Jennings, P., & Birrell, S. (2021). 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. *Accident Analysis and Prevention*, *151*, 1–14. https://doi.org/10.1016/j.aap.2020.105967

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

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2019). The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. *Cognition, Technology and Work*, *21*(1), 87–98. https://doi.org/10.1007/s10111-018-0525-8

Metcalfe, J. S., Marathe, A. R., Haynes, B., Paul, V. J., Gremillion, G. M., Drnec, K., Atwater, C., Estepp, J. R., Lukos, J. R., Carter, E. C., & Nothwang, W. D. (2017). Building a framework to manage trust in automation. *Proc. SPIE 10194, Micro- and Nanotechnology Sensors, Systems, and Applications IX*, 1–12. https://doi.org/10.1117/12.2264245

Metz, B., Wörle, J., Hanig, M., Schmitt, M., Lutz, A., & Neukum, A. (2021). Repeated usage of a motorway automated driving function: Automation level and behavioural adaption. *Transportation Research Part F: Traffic Psychology and Behaviour*, *81*, 82–

100. https://doi.org/10.1016/j.trf.2021.05.017

Miller, D., Johns, M., Mok, B., Gowda, N., Sirkin, D., Lee, K., & Ju, W. (2016). Behavioral measurement of trust in automation: The trust fall. *Proceedings of the Human Factors and Ergonomics Society*, 1842–1846. https://doi.org/10.1177/1541931213601422

Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction becomes engagement in automated driving. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*, 1676–1680. https://doi.org/10.1177/1541931215591362

Minematsu, Y., Ueji, K., & Yamamoto, T. (2018). Activity of frontal pole cortex reflecting hedonic tone of food and drink: fNIRS study in humans. *Scientific Reports*, *8*(1), 16197. https://doi.org/10.1038/s41598-018-34690-3

Montag, I., & Comrey, A. (1987). Internality and Externality as Correlates of Involvement in Fatal Driving Accidents. *Journal of Applied Psychology*, *72*(3), 339–343. https://doi.org/10.1037/0021-9010.72.3.339

Monti, M. M. (2011). Statistical analysis of fMRI time-series: A critical review of the GLM approach. *Frontiers in Human Neuroscience*, *MARCH*. https://doi.org/10.3389/fnhum.2011.00028

Morris, D. M., Erno, J. M., & Pilcher, J. J. (2017). Electrodermal Response and Automation Trust during Simulated Self-Driving Car Use. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1759–1762. https://doi.org/10.1177/1541931213601921

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *Int. J. Man-Machine Studies*, *27*, 527–539. https://doi.org/10.1016/S0020-7373(87)80013-5

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), 1905–1922. https://doi.org/10.1080/00140139408964957

Muir, B. M., & Moray, N. (1996). Trust in automation. part ii. experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

Naujoks, F., Purucker, C., & Neukum, A. (2016). Secondary task engagement and vehicle automation - Comparing the effects of different automation levels in an on-road experiment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *38*, 67–82. https://doi.org/10.1016/j.trf.2016.01.011

Navarro, J. (2018). A State of science on Highly Automated Driving. *Theoretical Issues in Ergonomics Science*. https://doi.org/10.1080/1463922X.2018.1439544

Nemcova, A., Seitl, M., Dominik, T., Semela, M., Sucha, M., Kolar, R., Svozilova, V., Bucsuhazy, K., Smisek, R., Mezl, M., Hesko, B., Belak, M., Bilik, M., & Maxera, P. (2020). Multimodal Features for Detection of Driver Stress and Fatigue: Review. *IEEE Transactions on Intelligent Transportation Systems*, 1–20. https://doi.org/10.1109/tits.2020.2977762

NHTSA. (2022). *NHTSA Proposes Significant Updates to Five-Star Safety Ratings Program*. https://www.nhtsa.gov/press-releases/five-star-safety-ratings-program-updates-proposed

Nothwang, W. D., Gremillion, G. M., Donavanik, D., Haynes, B. A., Atwater, C. S., Canady, J. D., Metcalfe, J. S., & Marathe, A. R. (2016). Multi-sensor fusion architecture for human-autonomy teaming. *Proceedings - 2016 Resilience Week, RWS 2016*, 166–171. https://doi.org/10.1109/RWEEK.2016.7573327

NTSB. (2017). *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*.

NTSB. (2019a). *Collision between vehicle controlled by developmental automated driving system and pedestrian. NTSB/HAR-19/03 PB2019-101402*. https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1903.pdf

NTSB. (2019b). *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*.

NTSB. (2020a). *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*. https://data.ntsb.gov/Docket?ProjectID=96932

NTSB. (2020b). *Collision Between Car Operating with Partial Driving Automation and Truck-Tractor Semitrailer, Delray Beach, Florida, March 1, 2019, NTSB/HAB-20/01*.

Obrig, H. (2010). From acoustic segmentation to language processing: evidence from optical imaging. *Frontiers in Neuroenergetics*, *2*(13), 1–12. https://doi.org/10.3389/fnene.2010.00013

Oliveira, L., Burns, C., Luton, J., Iyer, S., & Birrell, S. (2020). The influence of system transparency on trust: Evaluating interfaces in a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *72*, 280–296. https://doi.org/10.1016/j.trf.2020.06.001

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, *55*(9), 1059–1072. https://doi.org/10.1080/00140139.2012.691554

Pakdamanian, E., Sheng, S., & Baee, S. (2021, May 6). Deeptake: Prediction of driver takeover behavior using multimodal data. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3445563

Palmer, S., Richards, D., Shelton-Rayner, G., Inch, D., & Izzetoglu, K. (2019). Human-Agent Teaming - an Evolving Interaction Paradigm: An Innovative Measure of Trust. *20th International Symposium on Aviation Psychology*, 438–443. https://corescholar.libraries.wright.edu/isap_2019/74

Palmer, S., Richards, D., Shelton-Rayner, G., Izzetoglu, K., & Inch, D. (2020). Assessing Variable Levels of Delegated Control – A Novel Measure of Trust. In C. Stephanidis (Ed.), *HCI International 2020 – Late Breaking Papers: Cognition, Learning and Games.* (Vol. 12425, pp. 202–215). Springer, Cham. https://doi.org/10.1007/978-3-030-60128-7_16

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and Automation : Use, Misuse, Disuse, Abuse. *Human Factors*, *39*(2), 230–253.

Parker, D., Reason, J. T., Manstead, A. S. R., & Stradling, S. G. (1995). Driving errors, driving violations and accident involvement. *Ergonomics*, *38*(5), 1036–1048. https://doi.org/10.1080/00140139508925170

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems*, *2*(4), 315–322. https://doi.org/10.1049/iet-its:20080023

Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. In *Frontiers in Psychology* (Vol. 5, Issue DEC). Frontiers Research Foundation. https://doi.org/10.3389/fpsyg.2014.01344

Paxion, J., Galy, E., & Berthelon, C. (2015). Overload depending on driving experience and situation complexity: Which strategies faced with a pedestrian crossing? *Applied Ergonomics*, *51*, 343–349. https://doi.org/10.1016/j.apergo.2015.06.014

Payre, W., Birrell, S., & Parkes, A. M. (2020). Although autonomous cars are not yet manufactured, their acceptance already is. *Theoretical Issues in Ergonomics Science*, *October*, 1–14. https://doi.org/10.1080/1463922x.2020.1836284

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 252–263. https://doi.org/10.1016/j.trf.2014.04.009

Payre, W., Cestac, J., & Delhomme, P. (2016). Fully Automated Driving: Impact of Trust and Practice on Manual Control Recovery. *Human Factors*, *58*(2), 229–241. https://doi.org/10.1177/0018720815612319

Perello-March, J., Burns, C., Elliott, M. T., & Birrell, S. (2020). Integrating Trust in Automation into Driver State Monitoring Systems. In T. Ahram, R. Taiar, C. S., & C. A. (Eds.), *Human Interaction and Emerging Technologies. IHIET 2019. Advances in Intelligent Systems and Computing* (Vol. 1018, pp. 344–349). Springer International Publishing. https://doi.org/10.1007/978-3-030-25629-6

Perello-March, J., Burns, C., Woodman, R., Elliott, M. T., & Birrell, S. (2021). Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 1–11. https://doi.org/10.1109/TITS.2021.3050518

Perkins, L., Miller, J. E., Hashemi, A., & Burns, G. (2010). Designing for Human-Centered Systems: Situational Risk as a Factor of Trust in Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(25), 2130–2134. https://doi.org/10.1177/154193121005402502

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2019). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience*, *12*(January), 1–21. https://doi.org/10.3389/fnhum.2018.00505

Quaresima, V., Bisconti, S., & Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain and Language*, *121*(2), 79–89.

https://doi.org/10.1016/j.bandl.2011.03.009

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

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society*, *2014-Janua*(1), 2063–2067. https://doi.org/10.1177/1541931214581434

Rajaonah, B., Tricot, N., Anceaux, F., & Millot, P. (2008). The role of intervening variables in driver-ACC cooperation. *International Journal of Human Computer Studies*, *66*(3), 185–197. https://doi.org/10.1016/j.ijhcs.2007.09.002

Rangel, A., Camerer, C., & Montague, P. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*, 545–556. https://doi.org/10.1038/nrn2357

Rani, P., Liu, C., Sarkar, N., & Vanman, E. (2006). An empirical study of machine learning techniques for affect recognition in human-robot interaction. *Pattern Analysis and Applications*, *9*, 58–69. https://doi.org/10.1007/s10044-006-0025-y

Reason, J., Manstead, A., Stephen, S., Baxter, J., & Campbell, K. (1990). Errors and violations on the roads: A real distinction? *Ergonomics*, *33*(10–11), 1315–1332. https://doi.org/10.1080/00140139008925335

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scalling procedure for measureing mental workload. In *In P. A. Hancock and N. Meshkati (eds), Human Mental Workload (Amsterdam: North-Holland)*.

Reuters. (2016). *Germany says Tesla should not use 'Autopilot' in advertising*. https://www.reuters.com/article/us-tesla-germany/germany-says-tesla-should-not-use-autopilot-in-advertising-idUSKBN12G0KS

Riley, V. (1996). Operator Reliance on Automation: Theory and Data. In Mustapha Mouloua & Raja Parasuraman (Eds.), *Automation and Human Performance Theory and Applications* (pp. 19–35). Taylor & Francis Group. https://doi.org/10.1201/9781315137957-2

Roche-Labarbe, N., Zaaimi, B., Berquin, P., Nehlig, A., Grebe, R., & Wallois, F. (2008). NIRS-measured oxy- and deoxyhemoglobin changes associated with EEG spike-and-wave discharges in children. *Epilepsia*, *49*(11), 1871–1880. https://doi.org/10.1111/j.1528-1167.2008.01711.x

Rossi, J. S. (2013). Statistical power analysis. In *Handbook of psychology: Research methods in psychology, Vol. 2, 2nd ed.* (pp. 71–108). John Wiley & Sons, Inc.

Rudin-Brown, C. M., & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): Implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *7*(2), 59–76. https://doi.org/10.1016/j.trf.2004.02.001

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

https://doi.org/https://doi.org/10.4271/J3016_201806

Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS Brain AnalyzIR toolbox. *Algorithms*, *11*(5). https://doi.org/10.3390/A11050073

Satterfield, K., Baldwin, C., de Visser, E. J., & Shaw, T. (2017). The influence of risky conditions in trust in autonomous systems. *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting*, *61*(1), 324–328. https://doi.org/10.1177/1541931213601562

Schroeter, M. L., Zysset, S., Kupka, T., Kruggel, F., & Von Cramon, D. Y. (2002). Near-infrared spectroscopy can detect brain activity during a color-word matching stroop task in an event-related design. *Human Brain Mapping*, *17*(1), 61–71. https://doi.org/10.1002/hbm.10052

Seet, M., Harvy, J., Bose, R., Dragomir, A., Bezerianos, A., & Thakor, N. (2020). Differential Impact of Autonomous Vehicle Malfunctions on Human Trust. *IEEE Transactions on Intelligent Transportation Systems*, 1–10. https://doi.org/10.1109/tits.2020.3013278

Seetharaman, A., Patwa, N., Jadhav, V., Saravanan, A. S., & Sangeeth, D. (2021). Impact of Factors Influencing Cyber Threats on Autonomous Vehicles. *Applied Artificial Intelligence*, *35*(2), 105–132. https://doi.org/10.1080/08839514.2020.1799149

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*(September), 1–17. https://doi.org/10.3389/fpubh.2017.00258

Sheridan, T. B. (1989). Trustworthiness of Command and Control Systems. *Analysis, Design and Evaluation of Man–Machine Systems 1988*, *21*(1965), 427–431. https://doi.org/10.1016/B978-0-08-036226-7.50076-4

Sibi, S., Ayaz, H., Kuhns, D. P., Sirkin, D. M., & Ju, W. (2016). Monitoring Driver Cognitive Load Using Functional Near Infrared Spectroscopy in Partially Autonomous Cars. *IEEE Intelligent Vehicles Symposium*, *Iv*.

Sibi, S., Baiters, S., Mok, B., Steiner, M., & Ju, W. (2017). Assessing driver cortical activity under varying levels of automation with functional near infrared spectroscopy. *IEEE Intelligent Vehicles Symposium, Proceedings*, *Iv*, 1509–1516. https://doi.org/10.1109/IVS.2017.7995923

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*(8), 334–340. https://doi.org/10.1016/j.tics.2009.05.001

Singh, M., & Queyam, A. (2013). Stress Detection in Automobile Drivers using Physiological Parameters: A Review. *International Journal of Electronics Engineering*, *5*(2), 1–5. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Stress+Detection+in+Automobile+Drivers+using+Physiological+Parameters:+A+Review#0

Singh, S. (2015). Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey. *National Highway Traffic Safety Administration*, *February*, 1–2. https://doi.org/https://doi.org/10.1016/j.trf.2015.04.014

Smyth, J., Birrell, S., Mouzakitis, A., & Jennings, P. (2019). Motion Sickness and Human Performance – Exploring the Impact of Driving Simulator User Trials. In N. A. Stanton (Ed.), *Advances in Human Aspects of Transportation. AHFE 2018. Advances in Intelligent Systems and Computing* (Vol. 786, pp. 445–457). Springer Verlag.

https://doi.org/10.1007/978-3-319-93885-1_40

Smyth, J., Chen, H., Donzella, V., & Woodman, R. (2021). Public acceptance of driver state monitoring for automated vehicles: Applying the UTAUT framework. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*, 179–191. https://doi.org/10.1016/j.trf.2021.10.003

Smyth, J., Jennings, P., Bennett, P., & Birrell, S. (2021). A novel method for reducing motion sickness susceptibility through training visuospatial ability – A two-part study. *Applied Ergonomics*, *90*. https://doi.org/10.1016/j.apergo.2020.103264

Solís-Marcos, I., & Kircher, K. (2018). Event-related potentials as indices of mental workload while using an in-vehicle information system. *Cognition, Technology & Work*, *0*(0), 1–13. https://doi.org/10.1007/s10111-018-0485-z

Spiessl, W., & Hussmann, H. (2011). Assessing error recognition in automated driving. *IET Intelligent Transport Systems*, *5*(2), 103. https://doi.org/10.1049/iet-its.2010.0102

Stanton, N. A., Walker, G. H., Young, M. S., Kazi, T., & Salmon, P. M. (2007). Changing drivers' minds: The evaluation of an advanced driver coaching system. *Ergonomics*, *50*(8), 1209–1234. https://doi.org/10.1080/00140130701322592

Strayer, D. L., Watson, J. M., & Drews, F. a. (2011). Cognitive Distraction While Multitasking in the Automobile. In *Psychology of Learning and Motivation - Advances in Research and Theory* (Vol. 54). https://doi.org/10.1016/B978-0-12-385527-5.00002-4

Taeihagh, A., & Lim, H. S. M. (2019). Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport Reviews*, *39*(1), 103–128. https://doi.org/10.1080/01441647.2018.1494640

Taga, G., Asakawa, K., Maki, A., Konishi, Y., & Koizumi, H. (2003). Brain imaging in awake infants by near-infrared optical topography. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(19), 10722–10727. https://doi.org/10.1073/pnas.1932552100

Tak, S., & Ye, J. C. (2014). Statistical analysis of fNIRS data: A comprehensive review. In *NeuroImage* (Vol. 85, pp. 72–91). https://doi.org/10.1016/j.neuroimage.2013.06.016

Tanida, M., Sakatani, K., Takano, R., & Tagai, K. (2004). Relation between asymmetry of prefrontal cortex activities and the autonomic nervous system during a mental arithmetic task: Near infrared spectroscopy study. *Neuroscience Letters*, *369*(1), 69–74. https://doi.org/10.1016/j.neulet.2004.07.076

Tanveer, M. A., Khan, M. J., Qureshi, M. J., Naseer, N., & Hong, K. S. (2019). Enhanced drowsiness detection using deep learning: An fNIRS Study. *IEEE Access*, *7*, 137920–137929. https://doi.org/10.1109/ACCESS.2019.2942838

Taylor, R. M. (2018). Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design. In *Situational Awareness* (1st Editio, pp. 111–128). Routledge. https://doi.org/10.4324/9781315087924-8

Techer, F., Ojeda, L., Barat, D., Marteau, J. Y., Rampillon, F., Feron, S., & Dogan, E. (2019). Anger and highly automated driving in urban areas: The role of time pressure. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 353–360. https://doi.org/10.1016/j.trf.2019.05.016

Teh, E., Jamson, S., Carsten, O., & Jamson, H. (2014). Temporal fluctuations in driving demand: the effect of traffic complexity on subjective measures of workload and driving performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, *22*, 207–217. https://doi.org/10.1016/j.trf.2013.12.005

Teigen, K. H. (1994). Yerkes-Dodson: A Law for all Seasons. In *Theory & Psychology* (Vol. 4, Issue 4, pp. 525–547). https://doi.org/10.1177/0959354394044004

Toyota Motor Corporation. (2008). *Toyota Enhances Pre-crash Safety System with Eye Monitor | Toyota Motor Corporation Official Global Website*. https://global.toyota/en/detail/281564

TRL. (2020, July 17). *EU General Safety Regulation*. https://trl.co.uk/projects/eu-general-safety-regulation

Ulahannan, A., Cain, R., & Dhadyalla, G. (2019). Advances in Human Factors in Communication of Design. In A. G. Ho (Ed.), *Advances in Human Factors in Communication of Design. AHFE 2018. Advances in Intelligent Systems and Computing* (Vol. 796, pp. 3–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-93888-2

Ulahannan, A., Thompson, S., Jennings, P., & Birrell, S. (2021). Using Glance Behaviour to Inform the Design of Adaptive HMI for Partially Automated Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *PP*. https://doi.org/10.1109/TITS.2021.3086882

Unni, A., Ihme, K., Jipp, M., & Rieger, J. (2018). Corrigendum: Assessing the Driver's Current Level of Working Memory Load With High Density Functional Near-Infrared Spectroscopy: A Realistic Driving Simulator Study. *Frontiers in Human Neuroscience*, *12*(498). https://doi.org/10.3389/fnhum.2018.00498

Unni, A., Ihme, K., Jipp, M., & Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in Human Neuroscience*, *11*(April), 1–14. https://doi.org/10.3389/fnhum.2017.00167

Verdière, K. J., Roy, R. N., & Dehais, F. (2018). Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual Landing Scenario. *Frontiers in Human Neuroscience*, *12*(January), 1–14. https://doi.org/10.3389/fnhum.2018.00006

Verwey, W. B., & Veltman, H. a. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied*, *2*(3), 270–285. https://doi.org/10.1037/1076-898X.2.3.270

Vitale, J. J., Giffi, C. A., Pingitore, G., Robinson, R., Schmith, S., & Gangula, B. (2017). *What's ahead for fully autonomous driving: Consumer opinions on advanced vehicle technology Perspectives from Deloitte's Global Automotive Consumer Study*. file:///C:/Users/march_j/Downloads/manufacturing-consumer-opinions-on-advanced-vehicle-technology.pdf

Walker, F., Wang, J., Martens, M. H., & Verwey, W. B. (2019). Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 401–412. https://doi.org/10.1016/j.trf.2019.05.021

Wang, W., Wang, B., Bu, L., Xu, L., Li, Z., & Fan, Y. (2016). Vigilance task-related change in brain functional connectivity as revealed by wavelet phase coherence analysis of near-infrared spectroscopy signals. *Frontiers in Human Neuroscience*, *10*, 11. https://doi.org/10.3389/fnhum.2016.00400

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Waymo. (2018). *The very human challenge of safe driving*. Nov 5. https://medium.com/waymo/the-very-human-challenge-of-safe-driving-58c4d2b4e8ee

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Wenzel, T., Wobst, P., Heekeren, H. H., Kwong, K., Brandt, T. A., Kohl, M., Obrig, H., Dirnagl, U., & Villringer, A. (2000). Saccadic Suppression Induces Focal Hypooxygenation In the Occipital Cortex. In *Journal of Cerebral Blood Flow and Metabolism* (Vol. 20).

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wiggins, I. M., Anderson, C. A., Kitterick, P. T., & Hartley, D. E. H. (2016). Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): Are the measurements reliable? *Hearing Research*, *339*, 142–154. https://doi.org/10.1016/j.heares.2016.07.007

Wilson, G. F., & Russell, C. A. (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(4), 635–644. https://doi.org/10.1518/hfes.45.4.635.27088

Wilson, G. F., & Russell, C. A. (2004). Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 381–389. https://doi.org/10.1518/hfes.45.3.381.27252

Wilson, G. F., & Russell, C. A. (2007). Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. *Human Factors*, *49*(6), 1005–1018. https://doi.org/10.1518/001872007X249875.1006

Wintersberger, P., Riener, A., Schartmüller, C., Frison, A. K., & Weigl, K. (2018). Let me finish before I take over: Towards attention aware device integration in highly automated vehicles. *Proceedings - 10th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2018*, 53–65. https://doi.org/10.1145/3239060.3239085

Wintersberger, P., Von Sawitzky, T., Frison, A. K., & Riener, A. (2017). Traffic augmentation as a means to increase trust in automated driving systems. *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter (CHItaly '17)*, *17*, 18–20. https://doi.org/10.1145/3125571.3125600

Wörle, J., Metz, B., Thiele, C., & Weller, G. (2019). Detecting sleep in drivers during highly automated driving: The potential of physiological parameters. *IET Intelligent Transport Systems*, *13*(8), 1241–1248. https://doi.org/10.1049/iet-its.2018.5529

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strenght of Stimulus to Rapidity of Habit Formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. https://doi.org/10.5860/choice.47-0570

Young, M. S., Stanton, N. A., & Harris, D. (2007). Driving automation: Learning from aviation about design philosophies. *International Journal of Vehicle Design*, *45*(3), 323–338. https://doi.org/10.1504/IJVD.2007.014908

Yücel, M. A., von Lühmann, A., Scholkmann, F., Gervain, J., Dan, I., Ayaz, H., Boas, D., Cooper, R. J., Culver, J., Elwell, C. E., Eggebrecht, A., Franceschini, M. A., Grova, C., Homae, F., Lesage, F., Obrig, H., Tachtsidis, I., Tak, S., Tong, Y., … Wolf, M. (2021). Best Practices for fNIRS publications. *Neurophotonics*, *8*(1), 1–34. https://doi.org/10.1117/1.NPh.8.1.012101

Zhang, Q., Robert, L. P., Du, N., & Yang, X. J. (2018). Trust in AVs: The Impact of Expectations and Individual Differences. *Conference on Autonomous Vehicles in Society: Building a Research Agenda*, *March*, 1–6. https://deepblue.lib.umich.edu/handle/2027.42/142567

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 207–220. https://doi.org/10.1016/j.trc.2018.11.018

Zhou, F., Yang, X. J., & de Winter, J. C. F. (2021). Using Eye-Tracking Data to Predict Situation Awareness in Real Time During Takeover Transitions in Conditionally Automated Driving. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2021.3069776

Zhu, L., Li, S., Li, Y., Wang, M., Zhang, C., Li, Y., Yao, J., & Ji, H. (2019). Analysis of braking intention based on fNIRS in driving simulation experiments. *IET Intelligent Transport Systems*, *13*(7), 1181–1189. https://doi.org/10.1049/iet-its.2018.5304

Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). FNIRS Optodes' Location Decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-21716-z

Zontone, P., Affanni, A., Bernardini, R., Piras, A., Rinaldo, R., Formaggia, F., Minen, D., Minen, M., & Savorgnan, C. (2020). Car Driver's Sympathetic Reaction Detection through Electrodermal Activity and Electrocardiogram Measurements. *IEEE Transactions on Biomedical Engineering*, *67*(12), 3413–3424. https://doi.org/10.1109/TBME.2020.2987168