

Using fNIRS to Verify Trust in Highly Automated Driving

Jaume R. Perello-March¹, Christopher G. Burns, Roger Woodman², Mark T. Elliott³, and Stewart A. Birrell⁴

Abstract—Trust in automation is crucial for the safe and appropriate adoption of automated driving technology. Current research methods to measure trust mainly rely on subjective scales, with several intrinsic limitations. This empirical experiment proposes a novel method to measure trust objectively, using functional near-infrared spectroscopy (fNIRS). Through manipulating participants' expectations regarding driving automation credibility, we have induced and successfully measured opposing levels of trust in automation. Most notably, our results evidence two separate yet interrelated cortical mechanisms for trust and distrust. Trust is demonstrably linked to decreased monitoring and working memory, whereas distrust is event-related and strongly tied to affective (or emotional) mechanisms. This paper evidence that trust in automation and situation awareness are strongly interrelated during driving automation usage. Our findings are crucial for developing future driver state monitoring technology that mitigates the impact of inappropriate reliance, or over trust, in automated driving systems.

Index Terms—fNIRS, highly automated driving, trust in automation.

I. INTRODUCTION

RESEARCH has identified trust in automation (TiA) as a critical human factor for the acceptance and correct usage of automated driving systems [1]. According to the current state-of-the-art, TiA has several layers. Whereas dispositional trust can be static during an adult lifespan, situational trust fluctuates with experience [2]. This is known as trust calibration [3]. Several subjective scales exist to measure dispositional and situational TiA but these cannot measure TiA objectively and in real-time.

Lee and See described three processes for trust judgements at a contextual level: affective, analogic or analytic [4]. The stronger emotional -in the former- or executive -in the latter- component define their nature. The occurrence of trust judgements with greater emotional content or rationally-calculated outcomes will depend on several factors, such as

Manuscript received 1 April 2022; revised 5 August 2022; accepted 27 September 2022. Date of publication 11 October 2022; date of current version 26 January 2023. This work was supported by WMG, The University of Warwick. The Associate Editor for this article was L. Li. (*Corresponding author: Jaume R. Perello-March.*)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Biomedical and Scientific Research Ethics Committee from the University of Warwick.

Jaume R. Perello-March is with the Centre for Future Transport and Cities, Coventry University, CV1 2TT Coventry, U.K. (e-mail: jaume.perellomarch@coventry.ac.uk).

Christopher G. Burns is with the School of Aerospace, Transport and Manufacturing (SATM), Cranfield University, Cranfield, U.K.

Roger Woodman and Mark T. Elliott are with WMG, The University of Warwick, CV4 7AL Coventry, U.K.

Stewart A. Birrell is with the Centre for Future Transport and Cities, Coventry University, CV1 2TT Coventry, U.K.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TITS.2022.3211089>, provided by the authors.

Digital Object Identifier 10.1109/TITS.2022.3211089

time availability, experience or expertise with the automated system. These processes could conceivably be a way to measure situational TiA using modern wearable neurophysiology equipment in realistic lab setups.

II. BACKGROUND

Previous experimental research in economics has explored the neural correlates of interpersonal trust and distrust using neurophysiology. These studies have explored reciprocal social exchanges [5], seller profile's trustworthiness [6], and trustworthiness evaluations of online offers [7] using fMRI. Overall, these studies evidence that:

- (1) Trust and distrust are distinct yet related constructs regarding the nature and type of neural responses involved and the timescale required for their development [6].
- (2) The neural mechanisms of trust and distrust involve emotional and cognitive structures. Distrust is more dependent upon autonomic emotional processes, whereas trust is more dependent on intentional, calculated decision-making [5], [6], [7].

The overall trust process is similar for humans and automated agents, yet significant differences exist between interpersonal trust and TiA [8].

An emerging perspective that tackles TiA from a neuroergonomics approach [9], [10] offers the potential to apply known neural correlates of decision-making, theory of mind and anticipation of rewards or losses, to the current frameworks of TiA. Research in neuroergonomics has explored the neural correlates of TiA, often with wearable electroencephalogram (EEG) and functional near-infrared spectroscopy (fNIRS) devices.

Work in this domain identified event-related potential (ERP) components from the EEG signal in the anterior cingulate cortex as neural markers for error monitoring [10]. These ERP components were used to infer miscalibrated trust while participants with opposing algorithm credibility expectations (expected performance) monitored the algorithm's reliability (actual performance). 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.

A similar experiment in neuroergonomics used EEG while participants monitored algorithm reliability and rated their trust levels [11]. The authors examined how credibility and reliability affect the causal relationships among different brain regions. That is, the way brain regions are influenced by credibility and reliability in the context of human-automation interaction. Their findings corroborated those from [10] in that initial credibility modulates the formation of initial

trust, with automation reliability derived from the experience being the main factor influencing the calibration of TiA. Furthermore, their results also agree with previous literature evidencing that trust and distrust elicit different connectivity patterns in the brain, and thus, these are two distinct cognitive processes. In particular, distrust elicits a complex quick and episodic top-down response involving several neural networks (i.e., prefrontal cortex, posterior cingulate cortex, and the temporoparietal junction) requiring additional cognitive resources compared to trust, which instead consists of a slower, cumulative and deliberate process based upon the long-term experience [6].

Hopko and Mehta used fNIRS to measure cortical activation and functional connectivity associated to trust during a human-robot collaboration task [12]. Their participants had to complete a surface finishing task in collaboration with a robot with varying reliability conditions. Results indicate increased cortical activation during unreliable robot behaviour within the dorsolateral prefrontal cortex (DLPFC) the anterior prefrontal cortex, primary motor cortex, and primary visual cortex.

In the context of automated driving, to the authors' knowledge very limited research has attempted to explore the neural correlates of trust. Seet and colleagues used EEG power spectral density and functional connectivity analysis, combined with behavioural parameters and self-reports in a driving simulator study [13]. The simulation involved conditionally automated driving (SAE Level 3) and fully automated driving (SAE Level 5). Participants were exposed to several system failures across the study. Results indicated a significant reduction in right-frontal alpha band activity during system malfunction in fully automated driving, supported by decreased self-reported trust in this condition. The authors argued that the reduction of right frontal power derives from a lateralised left-frontal power increase, and thus brain activity in the left-frontal area would have increased as the motivation to re-engage manual control. In other words, during a fully automated driving malfunction, participants reported lower trust and were motivated to take-over control. These findings would only align partly with [11]. While they agree that distrust increased cognitive load due to increasing attention, vigilance and cognitive control, those from [11] observed connectivity increments directed towards the right prefrontal cortex instead.

What has been observed for distrust also seems aligned with studies exploring other constructs with certain similarities to distrust and low trust. For example, suspicion induced from a computer malware manipulation was associated with participants' increased oxygenated haemoglobin (HbO), measured with fNIRS, in Broca's area, the DLPC, the frontopolar region, and the orbitofrontal cortex and the anterior cingulate cortex [14]. These HbO variations were mainly localised in the left hemisphere, aligned with the findings discussed in the previous paragraph from [13] but opposed to those from [11]. Hirshfield and colleagues argued that suspicion is a construct related to low trust and distrust because it also leads to a higher engagement in monitoring the automation [14], as noted by [13]. Indeed, the relationship of state-level suspicion with distrust was extensively described in [15]. According to the authors, distrust in automation increases mental workload and emotional arousal. This finding is important because it shows the main difference between interpersonal and human-

automation trust. Higher interpersonal trust involves social engagement, whereas higher TiA implies disengagement from the driving tasks, and hence, conversely, low trust increases engagement with the driving task [16], [17].

The use of fNIRS in the domain of TiA is still scarce but promising. Work by Palmer *et al.* monitored participants with fNIRS while supervising aerial and ground uncrewed vehicles under varying levels of integrity and control during a military-related experimental task [18]. Uncrewed vehicles were controlled using a visual interface allowing three levels of automated support -i.e. assisted manual, assisted automated and fully automated- and their integrity was manipulated to generate correct and incorrect behaviours. Results indicated that the uncertainty of judging the reliability of the uncrewed vehicle's abilities under assisted manual and assisted automated increased oxygenation in the orbitofrontal cortex, specifically in Brodmann Area (BA) 10, and the right DLPC in BA 46. This finding could be associated with those from [12] who found increased activation in the DLPC and the anterior prefrontal cortex during robot malfunction. As well as to those from [6], who found the orbitofrontal and anterior cingulate cortex to play a crucial role in intentional engagement and hence in the intentional decision of trusting. Both orbitofrontal and cingulate cortex are concomitant areas and could be part of the neural network of active trust judgements.

Similarly, results from Palmer *et al.* also found that the ventrolateral prefrontal cortex (VLPC) (BAs 44, 45, 47) could be implicated in the development of distrust due to poor decision making [18]. The VLPC is very proximal to the insular cortex, an area triggered by intense emotions such as fear, the anticipation of losses and distrust [6]. Palmer and colleagues discuss the application of their findings with trust in automated vehicles [18], but their study used a military task context and should be interpreted cautiously concerning its transferability to highly automated driving (HAD, i.e., SAE levels 3-4).

Sibi *et al.* compared mental workload levels derived from several automated driving modes using fNIRS in a driving simulator study [19]. They observed that the DLPC activation during lane changes performed with partially automated mode was comparable to that during a manual lane change, suggesting that partially automated driving is as cognitively demanding for drivers as manual driving. In addition, they also decided to evaluate self-reported trust from each mode of automated driving control, but the results were inconclusive.

Although EEG has been the preferred technique for exploring the neural correlates of TiA, the use of fNIRS has been increasing over the last few years, as it has become a viable solution for realistic and naturalistic setups. Arguably, they are entirely different measures of different parameters. FNIRS offers a good spatial resolution allowing the localisation of specific functional regions with montages of roughly up to 80 sources, but EEG montages with over 256 electrodes are possible. Additionally, in terms of temporal resolution fNIRS is not comparable to EEG, which is near-instantaneous (i.e. there is no lag in responses) and can have sampling rates over 20,000Hz.

Its use in human factors automotive research has also been growing, e.g., for detecting drivers' braking intentions [20], [21], mental workload [22], [23], [24], [25], attentional levels [26], responses to changing vehicle dynamics [27], inhibitory

control [28], fatigue [29], and drowsiness [30]. However, none of these studies has focused on TiA yet. Hence, it remains unclear whether the findings described in this section relating to TiA would also be transferable to the HAD context and whether fNIRS would be a reliable technique for the real-time measurement of situational TiA.

Different trust levels can be induced through credibility expectations to naïve participants [10], [12], [31], [32], [33]. In other words, the automated system's performance credibility can be manipulated through previous information and induce trust/distrust in naïve participants. Thus, it could be argued that if naïve participants lack expertise with a given automated system, they would not be able to make analytic trust judgements, according to Lee & See's framework [4]. Arguably, they could only use preconceived expectations regarding automated systems, and thus affective or analogic processes would be controlling situational trust judgements in such a context. If so, experimentally inducing such expectations to naïve users should trigger either affective or analogic trust calibration processes when driving across varying scenarios.

III. RESEARCH HYPOTHESES

We designed a HAD simulator experiment to investigate TiA levels based on induced automation credibility. The experimenter verbally provided two opposing vehicle performance descriptions regarding automation credibility, which served as a grouping factor for our two groups of participants, -i.e. low credibility vs high credibility. Automation credibility was expected to induce opposing levels of TiA. However, because vehicle dynamics and performance were equal for both groups, we expected to find a trust calibration similar to that in [10] and [11] during lower traffic complexity scenarios. Traffic complexity refers to the combination of several environmental features such as traffic volume, flow and lane change presence among other road users [34].

This research aims to measure variations of situational TiA during highly automated driving, using fNIRS in a realistic high-fidelity driving simulator setup. We aim to provide a methodological basis for real-time objective measurement of situational trust in highly automated vehicles, from which further research exploring this construct should benefit. The hypotheses proposed are:

- 1) Trust will quickly calibrate for the low credibility group (LC) aligned with vehicle performance, and thus, **no group differences** for brain activity will be observed **during low complexity traffic conditions**.
- 2) As the driving scenarios become more complex and risky, participants will recalibrate their TiA. Because the only information they will have available will be the credibility expectations induced, we expect **group differences during complex traffic conditions**, with the LC group showing greater brain activation across the orbitofrontal, ventrolateral and dorsolateral prefrontal cortex.

IV. METHOD

A. Participants

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 and had no previous

experience with HAD. Seven participants withdrew due to motion sickness and their data was excluded from analysis. A total of twenty-seven participants completed the trials and were included for data analysis (20 male and 7 female). 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. They all received a £10 voucher after the experiment.

Participants were randomly assigned to two groups of HAD credibility expectations. HAD performance was described to the low credibility group (N = 12) as a not-entirely-reliable, early prototype system capable of self-driving and adapting to road conditions still under development. Conversely, the HAD system was described to the HC group (N = 15) as a fully reliable HAD 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, only induced reliability expectations were manipulated.

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). Consistent gender differences have not been described as a modulator of trust during driving conditions yet, and thereby there is no reason to suspect these would influence our results. Participants were mostly aged between 18 and 35 years old (85.19%). 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, and thirteen of them (48%) driving an average of more than 10k miles a year. Both groups were instructed not to attempt to take control of the vehicle under any circumstances to generate the vulnerability required for TiA [2], [4].

B. Apparatus

The trials were conducted using the 3xD driving simulator at the University of Warwick Fig. 1. The 3xD is a fixed-base high-fidelity driving simulator, equipped with a whole-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 (for technical details see [35]). The simulated driving 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 for predefined use cases, and also generated road motion vibration through the seats and environmental sound with the in-vehicle sound system.

Neurophysiological data was obtained from the prefrontal cortex with a NIRSport CW-NIRS device (NIRx Medical Technologies LLC, USA) (Fig. 2). Data were extracted using NIRStar acquisition software (CA, USA; version 15.0). 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, 760 nm and 850 nm, allowing quantification of oxygenated



Fig. 1. The 3x3D driving simulator at the University of Warwick.



Fig. 2. Channels montage and representation of the whole setup device –in our case, the hardware was placed behind the passenger seat.

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).

Plastic spacers located at a distance of 3 cm between each source and detector pair constitute a recording channel, thus resulting in 22 recording channels. Channels were mounted within the Montreal Neurological Institute (MNI) coordinate space for consistency across head size variation [36]. These coordinates allow subsets of fNIRS channels down to those directly measuring particular regions of interest (ROIs) (TABLE I).

Self-reported trust was collected using the Trust in Automated Systems Scale [37]. 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. This is an established scale widely used in research to measure operators' trust in automated systems [38], [39], [40].

C. Automated Driving Scenarios

Drivers during HAD are expected to be engaged in non-driving related tasks (NDRTs) since monitoring the driving task is not required during predefined HAD use cases. Hence, our first experimental condition involved performing a verbal 2-back task for 2 minutes. A 2-back is a working memory task involving speech and is well established for generating mental workload. Previous neuroergonomics studies have used fNIRS to measure mental workload elicited by an N-back task, both in flight and driving simulators [41], [42], [43]; as well as human factors studies measuring mental workload with other physiological devices [44], [45], [46], [47]. This NDRT was carried out while the highly automated vehicle (HAV) was driving across a highway scenario, and we expect

TABLE I
LIST OF CHANNELS AND REGIONS OF INTEREST

Source	Detector	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

this condition to be a control condition for mental workload, as the 2-back task should not affect TiA (Fig. 3).

HAD use cases will entail a wide range of different road layouts and traffic conditions. Each scenario involves different vehicle dynamics which could increase the driver's mental workload when assessing actual vehicle performance against expected automated system reliability. Such compensating behaviours are well known to reallocate cognitive resources depending on road layout changes [27]. Similarly, traffic conditions and road type cause higher stress. It was found that urban scenarios generated higher stress compared to highway-motorway driving [23], [47], [48]. Higher stress in such road layouts is influenced by the increased amount of contextual information and stimuli to process, which require cognitive resources allocation for increased attention and monitoring [23]. Indeed, traffic and road complexity have been found modulators of trust in the context of automated driving [49]. Vehicle users showed higher trust in an automated system interface displaying recognised traffic objects in augmented reality. Participants were more confident when they could ensure the vehicle was fully aware of the situation on their behalf. Since situational TiA may calibrate according to contextual changes [2], [50], it could be expected that high traffic density and urban environments will affect TiA in line with the credibility assigned to each group, and this would be observable with fNIRS (see Fig. 4 for a summary of this process).

HAVs will have to cope with unexpected events resulting from other road users when driving across complex traffic conditions on busy urban roads, prone to hazardous situations. The rationale of including a risky event in our simulated driving scenarios lies in that HAV users also perceive such scenarios as potentially risky [33], [51], and risk perception plays a crucial role in the calibration of TiA [2], [4]. According to [33]:

“A highly reliable automated system or a driver's trust in the system will mitigate the perceived relational risk level

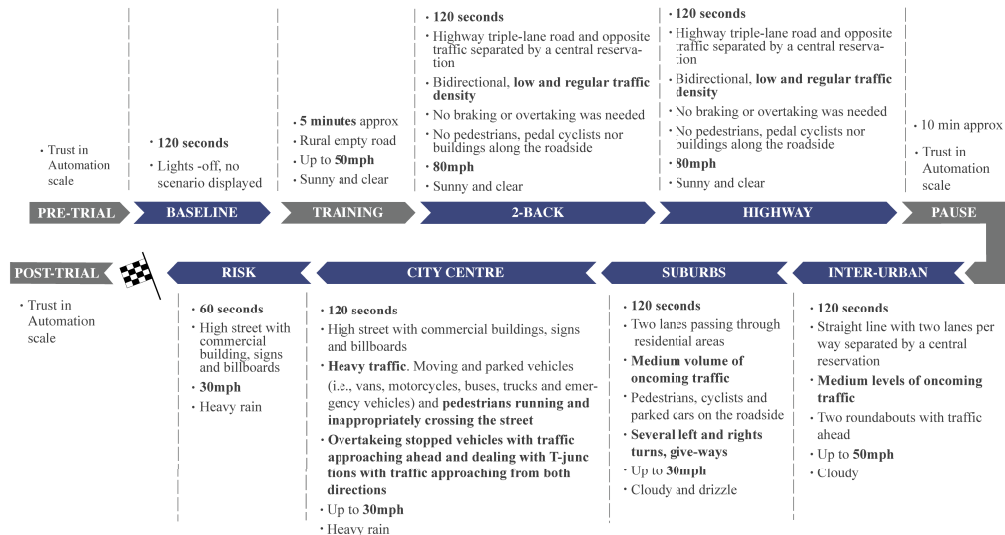


Fig. 3. Experimental conditions in order of occurrence. Blue shaded boxes indicate data analysed. Conditions may have been longer but only 120 seconds were extracted for analysis.



Fig. 4. Relationship between trust, risk, vigilance, mental workload and cortical activation.

even with a high level of situational risk on the road. On the other hand, if a driver distrusts the system, the perceived relational risk level will be high no matter whether situational risk is present (p. 181).” With this in mind, we designed a risky scenario (Fig. 3) where the HAV follows a van, and immediately after a left bend, both encounter a cyclist and proceed to overtake while approaching a junction with the right-of-way. Right after the van passes the junction, and while the HAV overtakes the cyclist, an ambulance with emergency lights and a siren moves into view at high speed from the left side of the junction. The HAV performs an emergency braking and evasive manoeuvre to avoid crashing against the ambulance, and immediately after, a police vehicle follows the ambulance, so the HAV has to brake again.

As discussed in hypothesis 2, if participants trust the automated system’s capabilities, they should be less vigilant and engaged with the driving task than those who distrust. If so, we should expect a reduced cognitive load for the HC group and increased brain oxygenation across the pre-frontal cortex for the LC group during a risky scenario.

D. 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, so participants only had to complete the first TiA scale at the start of the experiment. Participants were instructed on the 2-back task and performed a short practice session. After the 2-back training, they 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. They were

instructed to be particularly careful not to apply any pressure to the sensors or stretch the cables to avoid signal spikes and artefacts. Driving simulator lights were switched off to achieve optimal signal quality during calibration. The signal was calibrated using the NIRStar acquisition software (version 15.0) until achieving excellent quality from all channels. Following, we recorded participants’ baseline for 2 minutes with the lights switched off and without projecting the driving scenario.

Participants started with a 5 minutes familiarisation trial consisting of driving manually across empty rural roads, which minimised motion sickness impact [52]. Participants were instructed to drive cautiously to gain familiarity and 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 instructed to engage in automated driving 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 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 paused as longer exposures to driving simulators tend to increase the risk of simulator sickness [52]. Participants left the vehicle and went into the control room to fill in the second TiA scale. The signal was calibrated once again before resuming the subsequent scenarios.

Upon resuming, the scenario began with HAD engaged from the same stopping point and leading to an interurban drive with low traffic complexity for 2 minutes –i.e. third experimental condition. After this, the vehicle entered the suburbs, where traffic complexity slightly increased throughout

the scenario –fourth experimental condition. Two minutes later, traffic density increased, leading to a 2 minute city centre scenario –fifth experimental condition. The experiment ended with the HAV performing an evasive manoeuvre, the risk scenario. After this, participants left the driving simulator and filled in the third TiA scale.

E. Data Pre-Processing

Raw fNIRS data were pre-processed using HomER 3 [53] scripts running on MATLAB R2019a (Mathworks Inc.) and followed the current recommendations for pre-processing fNIRS data [54] (IV-F). For fNIRS current best practices and publication guidelines, see Yücel *et al.* [55]. Corrected optical density data were then converted to HbO, HbR and HbT concentrations using the modified Beer-Lambert law. Once calculated optical density concentrations, data was block averaged and exported as Hemodynamic Response Function (HRF) means.

F. Data Analysis

Block averaged HbO and HbR values from HomER 3 were exported in excel files containing HRF means for each channel, condition and participant. The underlying ROIs were determined using the NIRS Brain AnalyzIR toolbox [56] to calculate the corresponding anatomical labels for each position. The toolbox creates a variable that lists the channels and 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 complete picture, so we also extracted a ‘depth’ value for each channel and BA. Depth values represent the distance on average between the channel and the BA -i.e., the further the distance, the lower the likelihood that the channel captures that BA. Therefore, we selected up to three channels accounting for at least a combined relative weight of 0.80 (i.e. covering at least 80% of a particular ROI) and for the lowest combined depth value (i.e. the smallest combined distance on average).

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, the response will be negated. Following [57], we established averaging together only up to 3 channels. The most sensitive channels of each ROI were grouped. This led to 10 ROIs: Bilateral BAs 08, 09, 10 and 46, and left BA44 and 45 (Fig. 5). 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 for each ROI and each chromophore (TABLE I). Each single mean concentration value was then transformed into Z-scores ($M = 0$; $SD = 1$) against the mean group baseline value and its standard deviation (i.e., $Z = (X - \text{baseline mean}) / \text{baseline SD}$) to enable inter-individual and intra-individual comparisons (see TABLE III 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 [29], [41], [43], [58], [59], [60], [61].

The General Linear Model is the standard approach for analysing and interpreting hemodynamic responses [54], [62].

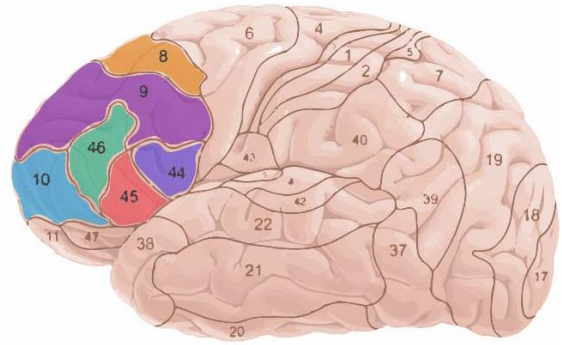


Fig. 5. Brodmann Areas covered by the fNIRS montage coloured.

TABLE II
DATA PRE-PROCESSING

Step	Description	Function	Input values
1	Remove 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	Transform fNIRS raw data into optical density	hmrRIntensity2OD	None
3	Identify and correct motion artefacts	hmrRMotionCorrectPCArecurse	tMotion = 0.5 tMask = 1 STDthresh = 20 AMPthresh = 5 nSV = 0.97 maxIter = 5
4	Eliminate noise from physiological activity, low frequency signal drifts or machine noise.	hmrRBandPassFilt	hpf = 0.01 lpf = 0.5
5	Convert optical density to concentrations	hmrROD2concNew	ppf = 1 (760nm) 1 (850nm)
6	Calculate the block average of the given conditions	hmrR_BlockAvg	None

Among the range of possibilities this approach offers, the well-known 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 [63]. 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 [64], [65], [66], [67]. 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 [68], [69], [70]. Therefore, following these recommendations, we will perform 2 (low credibility/high credibility) by 7 (baseline, 2-back, highway, interurban, suburbs, city centre, and risk) mixed ANOVAs to determine changes in haemoglobin concentrations on each

TABLE III
DATA ANALYSIS

Step	Description	Function	Criteria
1	Determining underlying ROIs for each channel	nirs.util.con verlabels2ro i	
2	Determining sensitivity for each channel using relative weights and depth values	nirs.util.dept hmap	
3	Determining most sensitive channels combined for each ROI		Up to three channels with the highest relative weight and lowest depth values
4	Averaging most sensitive channels for each ROI and for each experimental condition	Mean	
5	Standardising individual mean HRF concentrations	$Z = (X - \text{baseline mean}) / \text{baseline SD}$	
5	Test for normality assumption	Shapiro-Wilk	$p \geq 0.05$
6	Test for sphericity assumption	Mauchly	$p \leq 0.05$
7	Test for main statistical effects and interactions using mixed repeated-measures ANOVA	Analysis of Variance	$p \leq 0.05$
8	Follow-up pairwise comparisons	Bonferroni correction	$p \leq 0.05$

chromophore separately, and report the features in tandem to interpret hemodynamic responses.

HbO, HbR and HbT mean HRF concentrations grouped in ROIs were imported and analysed with IBM SPSS Statistics 26 software. The Shapiro-Wilk's test ($p \geq 0.05$) was used to assess normality assumption violations, and Mauchly's test was used to assess the assumption of sphericity. Thus, mixed 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). Main effects and interactions were followed-up by pair-wise comparisons corrected by Bonferroni.

V. RESULTS

This experiment investigated variations in TiA during highly automated driving by inducing two opposing automation credibility expectations within our groups of participants. Participants sat in the driving simulator and experienced different scenarios during the trial (see Fig. 3 for details). We expected that these credibility expectations would inversely affect their trust calibration when being driven across complex traffic conditions and would trigger different neural responses for each group. The Trust in Automated Systems Scale was rated three times during the experiment (i.e. pre-, mid- and after the driving trial) and was used to explore whether credibility expectations had the hypothesised effect on self-reported TiA. These ratings were then used to infer the neural correlates of TiA from each credibility group using the data collected with fNIRS.

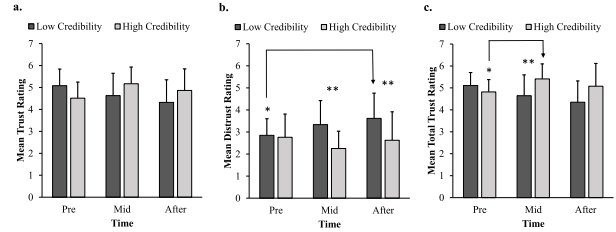


Fig. 6. Mean Trust (a.), Distrust (b.), and Total trust (c.) ratings taken Pre, Mid and After the trials. Double asterisks (**) indicate interaction effects, single asterisks (*) indicate main effects. Error bars indicate standard deviation.

TABLE IV
BETWEEN-SUBJECTS EFFECTS FOR HbO ACROSS ALL SCENARIOS

Brodmann area	$F(1, 22)$	P value	$\eta^2 p$
10 right	9.096	0.006	0.293
44 left	11.243	0.003	0.338
46 right	4.819	0.039	0.180

A. Trust in Automated Systems Scale

Interaction effects for the factor of mean Trust were reported ($F(2, 50) = 4.823$, $p = 0.012$, $\eta_p^2 = 0.162$), but these diminished after pairwise comparisons. Interaction effects were also observed for Distrust ($F(2, 50) = 4.961$, $p = 0.011$, $\eta_p^2 = 0.166$) indicating that the low credibility (LC) group (3.333 ± 1.086) reported more distrust than the high credibility (HC) group (2.253 ± 0.787 , $p = 0.006$) by the middle of the trial (Fig. 6). This trend remained after the trial was completed, supporting the effect of induced LC expectations (3.617 ± 1.146) compared to the HC group (2.627 ± 1.289 , $p = 0.048$) for distrust (Fig. 6). These findings were corroborated by interaction effects for the factor of Total trust ($F(2, 50) = 6.136$, $p = 0.004$, $\eta_p^2 = 0.197$), highlighting the detrimental effect of LC expectations (4.646 ± 0.944) compared to HC expectations (5.411 ± 0.682 , $p = 0.022$) during the mid-study pause (Fig. 6).

A further exploration on each group individually revealed a main effect of time for Distrust within the LC group ($F(2, 50) = 4.961$, $p = 0.011$, $\eta_p^2 = 0.166$). Follow-up pairwise tests revealed an increase in distrust between pre- (2.850 ± 0.749) and post- study (3.617 ± 1.146 , $p = 0.041$) (Fig. 6). A main effect of time for Total trust within the LC group was also observed ($F(2, 50) = 6.136$, $p = 0.004$, $\eta_p^2 = 0.197$), although diminishing after post-hoc corrections.

On the contrary, a main effect of time for Total trust within the HC group ($F(2, 50) = 6.136$, $p = 0.004$, $\eta_p^2 = 0.197$), revealed an increase between pre- (4.817 ± 0.564 , $p = 0.031$) and mid-study (5.411 ± 0.682) (Fig. 6).

These findings indicate that TiA levels were aligned with credibility as expected. Distrust increased for the LC group whilst Total trust increased for the HC group. Henceforth, in the next section, we will discuss how brain activity within the HC group could be associated with trust, whilst that from the LC group could be inferred to distrust.

B. fNIRS Data

Data from three participants were discarded for analysis due to artefacts, resulting in data from 24 participants included for analysis (12 participants per group). BAs 08-L/R, 45-L and 46-L did not report any statistical effects, however other significant and relevant differences will be discussed below.

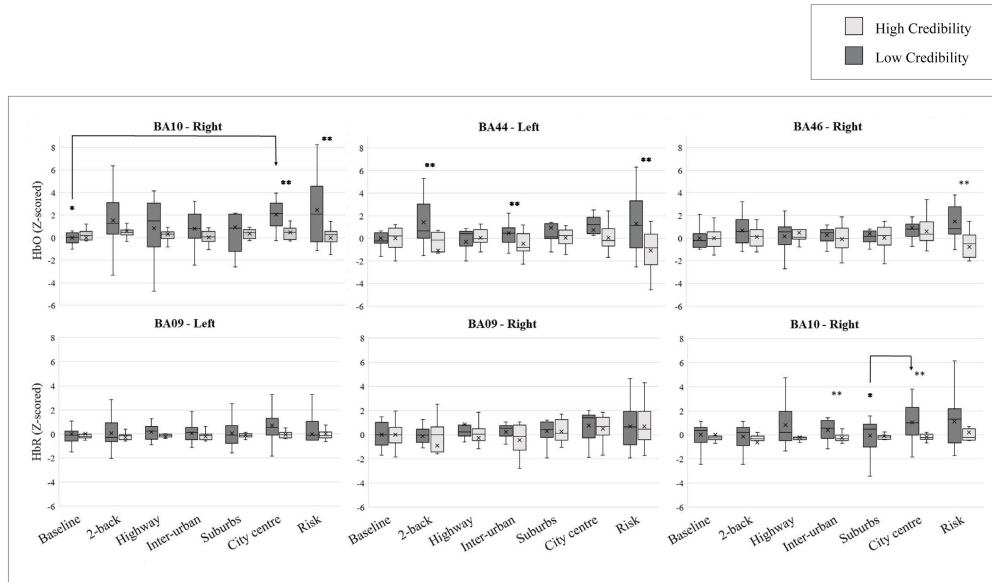


Fig. 7. HbO levels in BA10-right, BA44-left and BA46-right (top), and HbR levels in BA09-left, BA09-right, BA10-right (bottom); across driving conditions for each group. Double asterisks (**) indicate effects between groups, single asterisks (*) indicate main effects for condition. Mean is indicated by (x). Error bars indicate standard error.

TABLE V

INTERACTION EFFECTS FOR HbO ACROSS ALL SCENARIOS

Brodmann area	$F(6, 132)$	P value	$\eta^2 p$
44 left	2.767	0.034	0.112
46 right	2.461	0.027	0.101

TABLE VI

BETWEEN-SUBJECTS EFFECTS FOR HbR ACROSS ALL SCENARIOS

Brodmann area	$F(1, 22)$	P value	$\eta^2 p$
09 left	6.302	0.020	0.223
09 right	4.577	0.044	0.172
10 right	16.322	< 0.001	0.470

1) *Oxygenated Haemoglobin Concentrations (HbO)*: HbO levels varied between participants (TABLE IV). The LC group showed greater concentrations in the orbitofrontal cortex (i.e., BA10 - R) during the city centre (2.049 ± 1.306 , $p = 0.001$) and risk conditions (2.444 ± 2.865 , $p = 0.011$) than the HC group (0.458 ± 0.597 ; and -0.003 ± 1.094 , respectively) (Fig. 7). A further exploration revealed significant HbO variations within the LC group between baseline (0.001 ± 1 , $p < 0.001$) and city centre (Fig. 7) conditions.

Group differences in HbO were also found in the left VLPC (i.e., BA44-L) (TABLE IV). The LC group showed greater concentrations during the 2-back (1.401 ± 2.174 , $p = 0.020$), inter-urban (0.442 ± 1.027 , $p = 0.038$) and risk (1.302 ± 2.636 , $p = 0.019$) conditions compared to the HC group (in order: -1.088 ± 2.655 ; -0.469 ± 0.997 ; -1.082 ± 1.898) (Fig. 7). A group by condition interaction supported these findings (TABLE V).

Similar effects between groups were also observed in the right DLPC (i.e., BA46-R) revealing higher HbO concentrations for the LC group (1.488 ± 1.572 , $p = 0.003$) during the risk event compared to the HC group (-0.765 ± 1.732) (Fig. 7). A group by condition interaction supported these findings (TABLE V).

2) *Deoxygenated Haemoglobin Concentrations (HbR)*: Bilateral HbR levels in the DLPC (i.e. BA09-L/R) and the right orbitofrontal region (i.e., BA10-R) varied between groups

TABLE VII

BETWEEN-SUBJECTS EFFECTS FOR HbT ACROSS ALL SCENARIOS

Brodmann area	$F(1, 22)$	P value	$\eta^2 p$
09 right	4.777	0.040	0.178
46 right	5.489	0.029	0.200
10 right	10.252	0.004	0.318
44 left	12.937	0.002	0.370

(TABLE VI). Group effects in both BA09-L/R diminished after post-hoc pairwise tests corrections.

Higher HbR concentrations were observed in BA10-R for the LC group during inter-urban (0.394 ± 0.870 , $p = 0.026$) and city centre (1.050 ± 1.495 , $p = 0.010$) conditions, than the HC group (-0.247 ± 0.336 ; -0.192 ± 0.279 , respectively) (Fig. 7). HbR variations within the LC group were observed from suburbs (-0.064 ± 1.498 , $p = 0.031$) to city centre (Fig. 7).

3) *Total Haemoglobin Concentrations (HbT)*: HbT concentrations varied between participants (TABLE VII). Greater concentrations in the right DLPC were observed during the risk event for the LC group (BA09-R: 1.412 ± 2.001 ; BA46-R: 1.432 ± 1.469) than the HC group (BA09-R: -0.364 ± 2.133 , $p = 0.047$; BA46-R: -0.371 ± 1.766 , $p = 0.013$) (Fig. 8).

HbT levels in the right orbitofrontal region (i.e., BA10-R) (TABLE VII) also showed greater concentrations for the LC group (2.023 ± 1.147) than the HC group (0.641 ± 1.301 , $p = 0.011$) in the city centre scenario (Fig. 8). Moreover, HbT levels varied within the LC group from baseline (0.001 ± 1.001) to city centre (2.023 ± 1.146 , $p = 0.002$) (Fig. 8) conditions.

Finally, HbT levels in the left VLPC (i.e., BA44-L) revealed group differences (Table VII) in the inter-urban condition (LC: 0.438 ± 0.957 ; HC: -0.443 ± 1.047 , $p = 0.043$) and the risk event (LC: 1.519 ± 2.605 ; HC: 1.019 ± 2.064 , $p = 0.015$) (Fig. 8).

The main results from this experiment can be summarised as follows:

- The Low Credibility group (LC) reported higher Distrust and lower Total trust (Fig. 6).

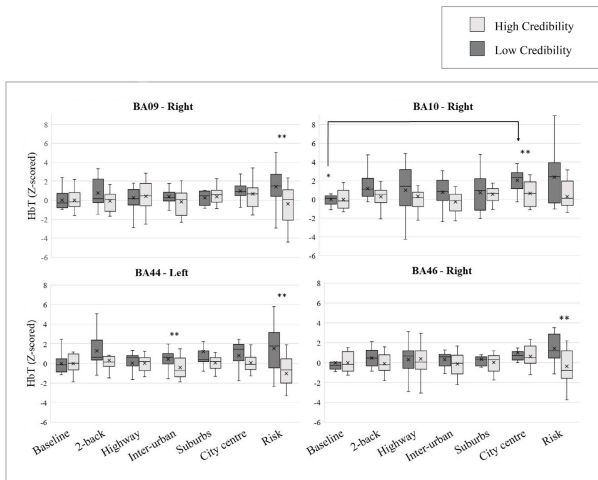


Fig. 8. HbT levels in BA09-right, BA10-right, BA44-left and BA46-right across driving conditions for each group. Double asterisks (**) indicate effects between groups, single asterisks (*) indicate main effects for conditions. Mean is indicated by (x). Error bars indicate standard error.

- The LC group showed a greater brain oxygen metabolism than the HC group towards variations among the driving scenarios (Fig. 7 and Fig. 8).
- The LC group showed increased brain oxygen metabolism during the complex driving scenarios (i.e., inter-urban, city centre and risk) (Fig. 7 and Fig. 8).
- The High Credibility (HC) group reported a higher Total trust (Fig. 6).
- The HC group showed significantly lower brain oxygen metabolism during the risk event (Fig. 7 and Fig. 8).

VI. DISCUSSION

This empirical research measured different levels of trust in highly automated driving (HAD) between two groups of participants with induced opposing automation credibility expectations for simulated driving scenarios with varying traffic complexities. We expected that inducing low automation credibility (LC) would increase participants' distrust whilst high credibility (HC) would increase trust. Assuming that trust and distrust affect drivers' monitoring and engagement with the driving task, we hypothesised that under simpler traffic conditions, trust in automation (TiA) would calibrate equally for both groups, and no main differences between brain activity in the prefrontal cortex would be observed. However, under more complex and challenging traffic conditions, TiA would recalibrate according to the credibility expectations initially induced, thus the LC group would be more engaged with the driving task than the HC group, consequently showing greater brain activity across the prefrontal cortex.

A. Hypothesis 1 – Calibration of TiA

Our first hypothesis predicted that trust would initially calibrate in line with actual vehicle performance (equal for both groups) in low traffic conditions, and thus, brain activity would not differ between groups during the 2-back, highway, interurban and suburban conditions.

Self-reports indicated that initial TiA did not calibrate according to vehicle performance as in [10], and that participants reported trust/distrust levels according to the credibility expectations initially induced as in [31], [32], and [33]. That

is, during the mid-study pause, the LC group was reporting a significant distrust increase. On the contrary, the HC group reported increased Total trust scores by the mid-study pause. These results are informative regarding initial TiA calibration since they were taken after participants had experienced the highway scenario which included performing the 2-back, cognitive workload inducing task. Considering that vehicle performance was equal for both groups, and that highway was a simplified driving layout, we expected the LC group to calibrate their TiA according to the vehicle performance, but they did not. Instead, they reported distrusting even though the vehicle was driving reliably. Perhaps participants judged the vehicle reliability upon their pre-existing expectations instead of the actual driving performance, which could be inferred as an analogic trust judgement [4].

Such self-reported distrust was supported by unique variations in the right orbitofrontal (i.e., BA10-R) for the LC group. This finding within the LC group was consistent among all three chromophores in two different driving scenarios (i.e., suburbs and city centre), thus suggesting the right orbitofrontal might be involved in assessing the driving context to calibrate trust, as suggested in previous work [12], [18]. That being the case, this would indicate incremented monitoring towards the changes in the driving environment for distrusting participants.

Aligned with our hypothesis, no group effects in haemoglobin concentrations were observed for highways and suburbs. However, between-subjects effects were observed in the left VLPC (BA44) during 2-back for HbO levels (Fig. 7). The increment observed for the LC group could be attributed to meeting the task demands, as found in previous research [42]. However, for the HC group, HbO levels decreased unexpectedly. Because we did not compare participants' performance, potentially a difference in performance might explain this phenomenon. Nonetheless, we ensured that participants from both groups were engaged in the 2-back task by verbally encouraging them to continue with the task.

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" [68], [69], [71]. Remarkably, this mirrored trend occurred in Broca's area, mainly known for language processing [69], during a verbal working memory task. Hence, the observed localised deactivation in BA44 during the 2-back task for the HC group might be due to a reallocation of cognitive resources towards other neural regions rather than a signal of poor task performance. A similar reallocation effect was also found in [13] during a highly automated vehicle (HAV) malfunction. The authors reported a reduction of right frontal brain activity deriving from a lateralised left-frontal increase, which would have increased due to the motivation to re-engage manual control and distrust.

In addition, group differences during the interurban driving scenario for HbO and HbT concentrations in BA44 and HbR concentrations in BA10-R were not aligned with our hypothesis either. These findings are likely indicative of the greater engagement in the driving task resulting from the distrust among the LC group. Indeed, [18] also found that increased HbO in the VLPC (including BA44) could be implicated in the development of distrust because of poor decision-making, as earlier noted by [7]. This area has also

been associated with suspicion during computer malfunctions [14], deliberate deception -lying- [72], frustration during automated driving [73], and even a predictor of emotional valence levels in a previous fNIRS study [74]. Furthermore, BA44 is anatomically proximal to the insular cortex, a region triggered by intense negative emotions, fear and anticipation of losses, associated with distrust [6], [7].

Overall, these results suggest that situational TiA did not calibrate according to vehicle performance for the LC group. Hence, the situational TiA calibration process was strongly biased by the initial distrust induced by poor expectations of reliability. Given the strong emotional component of distrust, it could be argued that this calibration was possibly framed on the affective process of trust calibration described in [4].

In contrast, the high credibility (HC) expectations provided to the other group matched with the actual vehicle performance, thus increasing trust and reducing the engagement with the driving task. Similar findings were also observed in [12] with human collaborative robots. Our participants in the HC group based their trust calibration on the heuristics (i.e., the mental model concerning the HAV reliability) generated by the vehicle capabilities provided, thus indicating an analogic process of trust calibration according to [4].

B. Hypothesis 2 – Recalibration of TiA

The second hypothesis predicted a recalibration of TiA as driving scenarios become more complex and hazardous. Thus, group differences were expected, particularly during the city centre and risk conditions.

Substantial evidence in favour of this hypothesis was found in self-reported data. Towards the end of the experiment, distrust had significantly increased for the LC group compared to the pre- and mid-study stages. This agrees with previous studies which have also provided information regarding the driving automation as an independent variable to manipulate TiA [10], [31], [32], [33], [75], [76].

Cortical haemodynamic concentrations were also in favour of our hypothesis. 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 uncrewed vehicle [18]; as well as during unreliable conditions of human-robot collaboration [12]. The orbitofrontal and anterior cingulate cortex have been found to play a critical role in intentional engagement [6]. Hence suggesting that our participants in the LC group were possibly judging the credibility of the driving automation, calibrating their TiA, and maybe even intending to take over manual control during city centre and risk scenarios. This statement would agree not only with self-reported distrust (Fig. 6) but also with the variations observed exclusively within this group from baseline to city centre, indicating an increase in brain activity (\uparrow HbO and \uparrow HbT) in BA10-right, possibly due to the uncertainty generated by increased traffic complexity (Fig. 7 and Fig. 8). HbR results coupled with these trends show aligned variations in deoxygenated haemoglobin between suburbs and city centre scenarios.

Another argument favouring this hypothesis was found in haemodynamic concentrations for the LC group in the VLPC (BA44). This group reported significantly greater HbO and HbT concentrations in this area during the risk scenario (Fig. 7 and Fig. 8). This finding would strongly agree with

the broader literature linking this area with distrust [6], [7], [18] and intense negative emotions [6], [14], [72], [73].

Finally, a lateralised activation (\uparrow HbO and \uparrow HbT) in the right DLPC (BA09-R and BA46-R) was also observed in the LC group, only during the risk scenario (Fig. 7, and Fig. 8). This seems to agree with those findings from [18], who found increased HbO in the right DLPC, and particularly in BA46, when participants were judging the credibility of the vehicle's abilities under assisted manual and assisted automated control. The DLPC was also found more active under low robot reliability conditions along with lower perceptions of trust [12]. Relatedly, [7] 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 [19], [27], [77]. In particular, [27] 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. These findings highlight the active role of the DLPC in situational TiA calibration and judging the contextual reliability of the HAV.

Overall, these findings align with Hypothesis 2 with the LC group showing higher activity in both the DLPC and VLPC during the risk scenario when reported the highest ratings for distrust. Both areas are predictors of emotional valence levels in a previous fNIRS study [74]. This would agree with the broader literature in that distrust is quick and episodic (i.e., event-related) and linked to emotional brain mechanisms [5], [6], [7]. This could result from an affective decision-making process made upon the strong emotional cues generated by distrust. Even though these participants had no reason to distrust the HAV - as it proved reliable across the scenarios - they kept distrusting.

On the contrary, the HC group seemed to follow the same trend reported in Hypothesis 1 –i.e. actively trusting the HAV as the traffic context was becoming more complex, thus suggesting their trust calibration relied on heuristics. Once they observed that the vehicle performance matched their credibility expectations, they disengaged from the driving task. This was particularly evident during the risk scenario where this group showed minimal cognitive workload overall as indicated by fNIRS data.

In summary, this research presents the first contribution to measure situational TiA under HAD in a realistic driving simulator setup. The three mental processes for trust calibration described in [4] acknowledged the importance of emotional cues in distrust judgements. As shown, these cues can bias trust judgements irrespectively of the actual HAV reliability. This finding is significant since affective and analogic judgements are prone to inappropriate behaviours like automation misuse due to overtrust or disuse due to distrust [78]. Because affective and analogic judgements do not rely on the actual knowledge of system limitations, capabilities and driving performance under specific contexts, but instead are led by feelings, beliefs or impressions.

The phenomenon known as “autonowashing” [79] refers to the usage of misleading terminology to describe current automated driving technology, exaggerating the actual capabilities of such systems. With these regards, research must focus on better understanding trust in automation and the

reliance behaviours of automated vehicle users to ensure this technology is safer than manual driving. Driver state monitoring systems supporting safe take-over transitions and actively preventing automation misuse will be necessary. fNIRS stands out for offering specific advantages to become a vital tool for driving research in neuroergonomics [70]. This knowledge will help develop and integrate future AI-based driver state monitoring and infotainment systems [32], [80].

VII. CONCLUSION

We expect further related work will benefit from this knowledge since it provides a considerable research methodology to assess TiA in real-time and objectively. Neurophysiology has the potential to become the longed-for objective measure of TiA. Notwithstanding, accurate self-reported tools are required to interpret neurophysiological data, and in this case, the scale used may have limited this. Although established and widely used, some authors have criticised this scale as not an accurate measure of situational trust but rather a propensity or disposition to trust [1], [81]. In addition, the experimental design with conditions in a fixed order, a relatively small sample and an fNIRS montage covering only the pre-frontal cortex may have limited our findings. Future work should consider counterbalanced or Latin-Square experimental designs where practical in a simulated environment. In addition, a dropout rate of roughly 25% in driving simulator experiments due to motion sickness should be considered when recruiting the sample [52]. Ultimately, we recommend further research should include montage set-ups covering the temporal, parietal and occipital cortices.

These results expand our existing knowledge in the following areas:

- Provide supporting evidence of two separate neural processes for trust and distrust.
- Where distrust is event-related and strongly tied to affective mechanisms, trust seems to decrease monitoring and working memory.
- Thus supporting the view that TiA and situation awareness are strongly related during driving automation usage [82], [83].

Considering these results as a whole, orbitofrontal, ventrolateral and dorsolateral prefrontal cortex structures are the most promising areas responsible for shaping part of the neural network responsible for situational TiA in HAD.

REFERENCES

- [1] B. E. Holthausen, P. Wintersberger, B. N. Walker, and A. Riener, "Situational trust scale for automated driving (STS-AD): Development and initial validation," in *Proc. 12th Int. ACM Conf. Automot. User Interfaces Interact. Vehicular Appl. (AutomotiveUI)*, 2020, pp. 40–47, doi: [10.1145/3409120.3410637](https://doi.org/10.1145/3409120.3410637).
- [2] K. A. Hoff and M. Bashir, "Trust in automation: Integrating empirical evidence on factors that influence trust," *Hum. Factors*, vol. 57, no. 3, pp. 407–434, 2015, doi: [10.1177/0018720814547570](https://doi.org/10.1177/0018720814547570).
- [3] B. M. Muir, "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems," *Ergonomics*, vol. 37, no. 11, pp. 1905–1922, Nov. 1994, doi: [10.1080/00140139408964957](https://doi.org/10.1080/00140139408964957).
- [4] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004, doi: [10.1518/hfes.46.1.50_30392](https://doi.org/10.1518/hfes.46.1.50_30392).
- [5] F. Krueger *et al.*, "Neural correlates of trust," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 50, pp. 20084–20089, Dec. 2007, doi: [10.1073/pnas.0710103104](https://doi.org/10.1073/pnas.0710103104).
- [6] A. Dimoka, "What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study," *Mis Quart.*, vol. 34, no. 2, pp. 373–396, 2010. [Online]. Available: <http://www.jstor.org/stable/20721433>
- [7] M. Hubert, M. Hubert, M. Linzmajer, R. Riedl, and P. Kenning, "Trust me if you can—Neurophysiological insights on the influence of consumer impulsiveness on trustworthiness evaluations in online settings," *Eur. J. Marketing*, vol. 52, no. 1/2, pp. 118–146, Feb. 2018, doi: [10.1108/EJM-12-2016-0870](https://doi.org/10.1108/EJM-12-2016-0870).
- [8] P. Madhavan and D. A. Wiegmann, "Similarities and differences between human–human and human–automation trust: An integrative review," *Theor. Issues Ergonom. Sci.*, vol. 8, no. 4, pp. 277–301, Jul. 2007, doi: [10.1080/14639220500337708](https://doi.org/10.1080/14639220500337708).
- [9] K. Drnec, A. R. Marathe, J. R. Lukos, and J. S. Metcalfe, "From trust in automation to decision neuroscience: Applying cognitive neuroscience methods to understand and improve interaction decisions involved in human automation interaction," *Frontiers Hum. Neurosci.*, vol. 10, pp. 1–14, Jun. 2016, doi: [10.3389/fnhum.2016.00290](https://doi.org/10.3389/fnhum.2016.00290).
- [10] E. J. de Visser *et al.*, "Learning from the slips of others: Neural correlates of trust in automated agents," *Frontiers Hum. Neurosci.*, vol. 12, p. 309, Aug. 2018, doi: [10.3389/fnhum.2018.00309](https://doi.org/10.3389/fnhum.2018.00309).
- [11] J. Huang, S. Choo, Z. H. Pugh, and C. S. Nam, "Evaluating effective connectivity of trust in human–automation interaction: A dynamic causal modeling (DCM) study," *Hum. Factors*, vol. 64, no. 4, Mar. 2021, Art. no. 18720820987443, doi: [10.1177/0018720820987443](https://doi.org/10.1177/0018720820987443).
- [12] S. K. Hopko and R. K. Mehta, "Trust in shared-space collaborative robots: Shedding light on the human brain," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 2022, Jun. 2022, Art. no. 001872082211090, doi: [10.1177/00187208221109039](https://doi.org/10.1177/00187208221109039).
- [13] M. Seet, J. Harvy, R. Bose, A. Dragomir, A. Bezerianos, and N. Thakor, "Differential impact of autonomous vehicle malfunctions on human trust," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 548–557, Jan. 2020, doi: [10.1109/TITS.2020.3013278](https://doi.org/10.1109/TITS.2020.3013278).
- [14] L. M. Hirshfield *et al.*, "Using noninvasive brain measurement to explore the psychological effects of computer malfunctions on users during human–computer interactions," *Adv. Hum.-Comput. Interact.*, vol. 2014, pp. 1–13, Jan. 2014, doi: [10.1155/2014/101038](https://doi.org/10.1155/2014/101038).
- [15] P. Bobko, A. J. Barelka, and L. M. Hirshfield, "The construct of state-level suspicion: A model and research agenda for automated and information technology (IT) contexts," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 56, no. 3, pp. 489–508, May 2014, doi: [10.1177/0018720813497052](https://doi.org/10.1177/0018720813497052).
- [16] N. R. Bailey and M. W. Scerbo, "Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust," *Theor. Issues Ergonom. Sci.*, vol. 8, no. 4, pp. 321–348, Jul. 2007, doi: [10.1080/14639220500535301](https://doi.org/10.1080/14639220500535301).
- [17] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 52, no. 3, pp. 381–410, Jun. 2010, doi: [10.1177/0018720810376055](https://doi.org/10.1177/0018720810376055).
- [18] S. Palmer, D. Richards, G. Shelton-Rayner, K. Izzetoglu, and D. Inch, "Assessing variable levels of delegated control—A novel measure of trust," in *HCI International 2020—Late Breaking Papers: Cognition, Learning and Games*, vol. 12425. Cham, Switzerland: Springer, Jul. 2020, pp. 202–215, doi: [10.1007/978-3-030-60128-7_16](https://doi.org/10.1007/978-3-030-60128-7_16).
- [19] S. Sibi, S. Baiters, B. Mok, M. Steiner, and W. Ju, "Assessing driver cortical activity under varying levels of automation with functional near infrared spectroscopy," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1509–1516, doi: [10.1109/IVS.2017.7995923](https://doi.org/10.1109/IVS.2017.7995923).
- [20] L. Zhu, S. Li, Y. Li, M. Wang, Y. Li, and J. Yao, "Study on driver's braking intention identification based on functional near-infrared spectroscopy," *J. Intell. Connected Vehicles*, vol. 1, no. 3, pp. 107–113, Feb. 2019, doi: [10.1108/jicv-09-2018-0007](https://doi.org/10.1108/jicv-09-2018-0007).
- [21] L. Zhu *et al.*, "Analysis of braking intention based on fNIRS in driving simulation experiments," *IET Intell. Transp. Syst.*, vol. 13, no. 7, pp. 1181–1189, Jul. 2019, doi: [10.1049/iet-its.2018.5304](https://doi.org/10.1049/iet-its.2018.5304).
- [22] A. Unni, K. Ihme, M. Jipp, and J. W. Rieger, "Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study," *Frontiers Hum. Neurosci.*, vol. 11, pp. 1–14, Apr. 2017, doi: [10.3389/fnhum.2017.00167](https://doi.org/10.3389/fnhum.2017.00167).
- [23] H. J. Foy and P. Chapman, "Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation," *Appl. Ergonom.*, vol. 73, pp. 90–99, Nov. 2018, doi: [10.1016/j.apergo.2018.06.006](https://doi.org/10.1016/j.apergo.2018.06.006).
- [24] S. Sibi, H. Ayaz, D. P. Kuhns, D. M. Sirkin, and W. Ju, "Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 419–425.

- [25] M. Johns, S. Sibi, and W. Ju, "Effect of cognitive load in autonomous vehicles on driver performance during transfer of control," in *Proc. 6th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, 2014, pp. 1–4. [Online]. Available: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84910597509&partnerID=tZOtx3y1>
- [26] K. Yamamoto, H. Takahashi, T. Sugimachi, and Y. Suda, "Validation study of driver's attention level during actual driving using fNIRS," in *Proc. 6th Driver Distraction Inattention Conf.*, 2018, pp. 1–9.
- [27] J. L. Bruno *et al.*, "Mind over motor mapping: Driver response to changing vehicle dynamics," *Hum. Brain Mapping*, vol. 39, no. 10, pp. 3915–3927, Oct. 2018, doi: [10.1002/hbm.24220](https://doi.org/10.1002/hbm.24220).
- [28] H. J. Foy, P. Runham, and P. Chapman, "Prefrontal cortex activation and young driver behaviour: A fNIRS study," *PLoS One*, vol. 11, no. 5, pp. 1–19, 2016, doi: [10.1371/journal.pone.0156512](https://doi.org/10.1371/journal.pone.0156512).
- [29] C. T. Lin *et al.*, "Exploring the brain responses to driving fatigue through simultaneous EEG and fNIRS measurements," *Int. J. Neural Syst.*, vol. 30, no. 1, pp. 1–12, 2020, doi: [10.1142/S0129065719500187](https://doi.org/10.1142/S0129065719500187).
- [30] M. A. Tanveer, M. J. Khan, M. J. Qureshi, N. Naseer, and K.-S. Hong, "Enhanced drowsiness detection using deep learning: An fNIRS study," *IEEE Access*, vol. 7, pp. 137920–137929, 2019, doi: [10.1109/ACCESS.2019.2942838](https://doi.org/10.1109/ACCESS.2019.2942838).
- [31] M. Körber, E. Baseler, and K. Bengler, "Introduction matters: Manipulating trust in automation and reliance in automated driving," *Appl. Ergonom.*, vol. 66, pp. 18–31, Jan. 2018, doi: [10.1016/j.apergo.2017.07.006](https://doi.org/10.1016/j.apergo.2017.07.006).
- [32] J. Perello-March, C. Burns, R. Woodman, M. T. Elliott, and S. Birrell, "Driver state monitoring: Manipulating reliability expectations in simulated automated driving scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 5187–5197, Jun. 2021, doi: [10.1109/TITS.2021.3050518](https://doi.org/10.1109/TITS.2021.3050518).
- [33] M. Li, B. E. Holthausen, R. E. Stuck, and B. N. Walker, "No risk no trust: Investigating perceived risk in highly automated driving," in *Proc. 11th Int. Conf. Automot. User Interfaces Interact. Veh. Appl.*, Sep. 2019, pp. 177–185, doi: [10.1145/3342197.3344525](https://doi.org/10.1145/3342197.3344525).
- [34] E. Teh, S. Jamson, O. Carsten, and H. Jamson, "Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 22, pp. 207–217, Jan. 2014, doi: [10.1016/j.trf.2013.12.005](https://doi.org/10.1016/j.trf.2013.12.005).
- [35] S. Khastgir, S. Birrell, G. Dhadyalla, and P. Jennings, "Identifying a gap in existing validation methodologies for intelligent automotive systems: Introducing the 3xD simulator," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 648–653.
- [36] G. A. Z. Morais, J. B. Balardin, and J. R. Sato, "fNIRS optodes' location decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, Dec. 2018, doi: [10.1038/s41598-018-21716-z](https://doi.org/10.1038/s41598-018-21716-z).
- [37] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cognit. Ergonom.*, vol. 4, no. 1, pp. 53–71, Mar. 2000, doi: [10.1207/S15327566IJCE0401_04](https://doi.org/10.1207/S15327566IJCE0401_04).
- [38] K. Satterfield, C. Baldwin, E. J. de Visser, and T. Shaw, "The influence of risky conditions in trust in autonomous systems," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, 2017, vol. 61, no. 1, pp. 324–328, doi: [10.1177/1541931213601562](https://doi.org/10.1177/1541931213601562).
- [39] Q. Zhang, L. P. Robert, N. Du, and X. J. Yang, "Trust in AVs: The impact of expectations and individual differences," in *Proc. Conf. Auto. Vehicles Soc., Building Res. Agenda*, Mar. 2018, pp. 1–6, [Online]. Available: <https://deepblue.lib.umich.edu/handle/2027.42/142567>
- [40] V. A. Banks and N. A. Stanton, "Keep the driver in control: Automating automobiles of the future," *Appl. Ergonom.*, vol. 53, pp. 389–395, Mar. 2016, doi: [10.1016/j.apergo.2015.06.020](https://doi.org/10.1016/j.apergo.2015.06.020).
- [41] K. J. Verdière, R. N. Roy, and F. Dehais, "Detecting Pilot's engagement using fNIRS connectivity features in an automated vs. manual landing scenario," *Frontiers Hum. Neurosci.*, vol. 12, pp. 1–14, Jan. 2018, doi: [10.3389/fnhum.2018.00006](https://doi.org/10.3389/fnhum.2018.00006).
- [42] A. Unni, K. Ihme, M. Jipp, and J. Rieger, "Corrigendum: Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study," *Frontiers Hum. Neurosci.*, vol. 12, p. 498, Dec. 2018, doi: [10.3389/fnhum.2018.00498](https://doi.org/10.3389/fnhum.2018.00498).
- [43] G. Durantin, J.-F. Gagnon, S. Tremblay, and F. Dehais, "Using near infrared spectroscopy and heart rate variability to detect mental overload," *Behav. Brain Res.*, vol. 259, pp. 16–23, Feb. 2014, doi: [10.1016/j.bbr.2013.10.042](https://doi.org/10.1016/j.bbr.2013.10.042).
- [44] T. M. Gable, A. L. Kun, B. N. Walker, and R. J. Winton, "Comparing heart rate and pupil size as objective measures of workload in the driving context: Initial look," in *Proc. Adjunct Proc. 7th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, Sep. 2015, pp. 20–25, doi: [10.1145/2809730.2809745](https://doi.org/10.1145/2809730.2809745).
- [45] M. A. Hogervorst, A.-M. Brouwer, and J. B. F. van Erp, "Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload," *Frontiers Neurosci.*, vol. 8, pp. 1–14, Oct. 2014, doi: [10.3389/fnins.2014.00322](https://doi.org/10.3389/fnins.2014.00322).
- [46] J. K. Lenneman and R. W. Backs, "Cardiac autonomic control during simulated driving with a concurrent verbal working memory task," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 51, no. 3, pp. 404–418, Jun. 2009, doi: [10.1177/0018720809337716](https://doi.org/10.1177/0018720809337716).
- [47] V. Melnicuk, S. Thompson, P. Jennings, and S. Birrell, "Effect of cognitive load on drivers' state and task performance during automated driving: Introducing a novel method for determining stabilisation time following take-over of control," *Accident Anal. Prevention*, vol. 151, pp. 1–14, Mar. 2021, doi: [10.1016/j.aap.2020.105967](https://doi.org/10.1016/j.aap.2020.105967).
- [48] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005, doi: [10.1109/TITS.2005.848368](https://doi.org/10.1109/TITS.2005.848368).
- [49] P. Wintersberger, T. von Sawitzky, A.-K. Frison, and A. Rieger, "Traffic augmentation as a means to increase trust in automated driving systems," in *Proc. 12th Biannual Conf. Italian SIGCHI Chapter*, Sep. 2017, pp. 18–20, doi: [10.1145/3125571.3125600](https://doi.org/10.1145/3125571.3125600).
- [50] J. Kraus, D. Scholz, D. Stiegemeier, and M. Baumann, "The more you know: Trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 62, no. 5, pp. 718–736, Aug. 2020, doi: [10.1177/0018720819853686](https://doi.org/10.1177/0018720819853686).
- [51] J. R. Perello-March, C. G. Burns, S. A. Birrell, R. Woodman, and M. T. Elliott, "Physiological measures of risk perception in highly automated driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4811–4822, May 2022, doi: [10.1109/TITS.2022.3146793](https://doi.org/10.1109/TITS.2022.3146793).
- [52] J. Smyth, P. Jennings, P. Bennett, and S. Birrell, "A novel method for reducing motion sickness susceptibility through training visuospatial ability—A two-part study," *Appl. Ergonom.*, vol. 90, Jan. 2021, Art. no. 103264, doi: [10.1016/j.apergo.2020.103264](https://doi.org/10.1016/j.apergo.2020.103264).
- [53] T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain," *Appl. Opt.*, vol. 48, no. 10, p. D280, Apr. 2009, doi: [10.1364/AO.48.00D280](https://doi.org/10.1364/AO.48.00D280).
- [54] P. Pinti, F. Scholkmann, A. Hamilton, P. Burgess, and I. Tachtsidis, "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 Hum. Neurosci.*, vol. 12, pp. 1–21, Jan. 2019, doi: [10.3389/fnhum.2018.00505](https://doi.org/10.3389/fnhum.2018.00505).
- [55] M. A. Yücel *et al.*, "Best practices for fNIRS publications," *Neurophotonics*, vol. 8, no. 1, pp. 1–34, Jan. 2021, doi: [10.1117/1.NPh.8.1.012101](https://doi.org/10.1117/1.NPh.8.1.012101).
- [56] H. Santosa, X. Zhai, F. Fishburn, and T. Huppert, "The NIRS brain AnalyzIR toolbox," *Algorithms*, vol. 11, no. 5, p. 73, May 2018, doi: [10.3390/A11050073](https://doi.org/10.3390/A11050073).
- [57] I. M. Wiggins, C. A. Anderson, P. T. Kitterick, and D. E. H. Hartley, "Speech-evoked activation in adult temporal cortex measured using functional near-infrared spectroscopy (fNIRS): Are the measurements reliable?" *Hearing Res.*, vol. 339, pp. 142–154, Sep. 2016, doi: [10.1016/j.heares.2016.07.007](https://doi.org/10.1016/j.heares.2016.07.007).
- [58] N. Roche-Labarbe, B. Zaaïmi, P. Berquin, A. Nehlig, R. Grebe, and F. Wallois, "NIRS-measured oxy- and deoxyhemoglobin changes associated with EEG spike-and-wave discharges in children," *Epilepsia*, vol. 49, no. 11, pp. 1871–1880, Nov. 2008, doi: [10.1111/j.1528-1167.2008.01711.x](https://doi.org/10.1111/j.1528-1167.2008.01711.x).
- [59] Y. Minematsu, K. Ueji, and T. Yamamoto, "Activity of frontal pole cortex reflecting hedonic tone of food and drink: fNIRS study in humans," *Sci. Rep.*, vol. 8, no. 1, p. 16197, Dec. 2018, doi: [10.1038/s41598-018-34690-3](https://doi.org/10.1038/s41598-018-34690-3).
- [60] U. Leon-Dominguez *et al.*, "Molecular concentration of deoxyHb in human prefrontal cortex predicts the emergence and suppression of consciousness," *NeuroImage*, vol. 85, pp. 616–625, Jan. 2014, doi: [10.1016/j.neuroimage.2013.07.023](https://doi.org/10.1016/j.neuroimage.2013.07.023).
- [61] M. Tanida, K. Sakatani, R. Takano, and K. Tagai, "Relation between asymmetry of prefrontal cortex activities and the autonomic nervous system during a mental arithmetic task: Near infrared spectroscopy study," *Neurosci. Lett.*, vol. 369, no. 1, pp. 69–74, Oct. 2004, doi: [10.1016/j.neulet.2004.07.076](https://doi.org/10.1016/j.neulet.2004.07.076).
- [62] M. Monti, "Statistical analysis of fMRI time-series: A critical review of the GLM approach," *Frontiers Hum. Neurosci.*, vol. 5, p. 28, Mar. 2011, doi: [10.3389/fnhum.2011.00028](https://doi.org/10.3389/fnhum.2011.00028).
- [63] S. Tak and J. C. Ye, "Statistical analysis of fNIRS data: A comprehensive review," *Neuroimage*, vol. 85, pp. 72–91, Jan. 2014, doi: [10.1016/j.neuroimage.2013.06.016](https://doi.org/10.1016/j.neuroimage.2013.06.016).

- [64] M. L. Schroeter, S. Zysset, T. Kupka, F. Kruggel, and D. Y. von Cramon, "Near-infrared spectroscopy can detect brain activity during a color-word matching stroop task in an event-related design," *Hum. Brain Mapping*, vol. 17, no. 1, pp. 61–71, Sep. 2002, doi: [10.1002/hbm.10052](https://doi.org/10.1002/hbm.10052).
- [65] A. J. Fallgatter and W. K. Strik, "Frontal brain activation during the Wisconsin card sorting test assessed with two-channel near-infrared spectroscopy," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 248, no. 5, pp. 245–249, Oct. 1998.
- [66] D. J. Mehagnoul-Schipper *et al.*, "Simultaneous measurements of cerebral oxygenation changes during brain activation by near-infrared spectroscopy and functional magnetic resonance imaging in healthy young and elderly subjects," *Hum. Brain Mapping*, vol. 16, no. 1, pp. 14–23, May 2002, doi: [10.1002/hbm.10026](https://doi.org/10.1002/hbm.10026).
- [67] G. Taga, K. Asakawa, A. Maki, Y. Konishi, and H. Koizumi, "Brain imaging in awake infants by near-infrared optical topography," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 19, pp. 10722–10727, Sep. 2003, doi: [10.1073/pnas.1932552100](https://doi.org/10.1073/pnas.1932552100).
- [68] H. Obrig, "From acoustic segmentation to language processing: Evidence from optical imaging," *Frontiers Neuroenergetics*, vol. 2, no. 13, pp. 1–12, 2010, doi: [10.3389/fnene.2010.00013](https://doi.org/10.3389/fnene.2010.00013).
- [69] V. Quaresima, S. Bisconti, and M. Ferrari, "A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults," *Brain Lang.*, vol. 121, no. 2, pp. 79–89, 2012, doi: [10.1016/j.bandl.2011.03.009](https://doi.org/10.1016/j.bandl.2011.03.009).
- [70] T. Liu, M. Pelowski, C. Pang, Y. Zhou, and J. Cai, "Near-infrared spectroscopy as a tool for driving research," *Ergonomics*, vol. 59, no. 3, pp. 368–379, Mar. 2016, doi: [10.1080/00140139.2015.1076057](https://doi.org/10.1080/00140139.2015.1076057).
- [71] R. Wenzel *et al.*, "Saccadic suppression induces focal hypooxygenation in the occipital cortex," *J. Cerebral Blood Flow Metabolism*, vol. 20, no. 7, pp. 1103–1110, Jul. 2000.
- [72] S. Bunce, A. Devaraj, M. Izzetoglu, B. Onaral, and K. Pourrezaei, "Detecting deception in the brain: A functional near-infrared spectroscopy study of neural correlates of intentional deception," *Proc. SPIE*, vol. 5769, pp. 24–32, May 2005, doi: [10.1117/12.600601](https://doi.org/10.1117/12.600601).
- [73] W. Damm, M. Franzle, A. Ludtke, J. W. Rieger, A. Trende, and A. Unni, "Integrating neurophysiological sensors and driver models for safe and performant automated vehicle control in mixed traffic," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 82–89, doi: [10.1109/IVS.2019.8814188](https://doi.org/10.1109/IVS.2019.8814188).
- [74] D. Bandara, S. Velipasalar, S. Bratt, and L. Hirshfield, "Building predictive models of emotion with functional near-infrared spectroscopy," *Int. J. Hum.-Comput. Stud.*, vol. 110, pp. 75–85, Feb. 2018, doi: [10.1016/j.ijhcs.2017.10.001](https://doi.org/10.1016/j.ijhcs.2017.10.001).
- [75] S. Khashtgir, S. Birrell, G. Dhadyalla, and P. Jennings, "Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 290–303, Jan. 2018, doi: [10.1016/j.trc.2018.07.001](https://doi.org/10.1016/j.trc.2018.07.001).
- [76] F. Walker, J. Wang, M. H. Martens, and W. B. Verwey, "Gaze behaviour and electrodermal activity: Objective measures of drivers' trust in automated vehicles," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 64, pp. 401–412, Jul. 2019, doi: [10.1016/j.trf.2019.05.021](https://doi.org/10.1016/j.trf.2019.05.021).
- [77] E. J. Anderson *et al.*, "Involvement of prefrontal cortex in visual search," *Exp. Brain Res.*, vol. 180, no. 2, pp. 289–302, Jun. 2007, doi: [10.1007/s00221-007-0860-0](https://doi.org/10.1007/s00221-007-0860-0).
- [78] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Hum. Factors*, vol. 39, no. 2, pp. 230–253, 1997.
- [79] L. Dixon, "Autonowashing: The greenwashing of vehicle automation," *Transp. Res. Interdiscipl. Perspect.*, vol. 5, May 2020, Art. no. 100113, doi: [10.1016/j.trip.2020.100113](https://doi.org/10.1016/j.trip.2020.100113).
- [80] J. Perello-March, C. Burns, M. T. Elliott, and S. Birrell, "Integrating trust in automation into driver state monitoring systems," in *Human Interaction and Emerging Technologies (Advances in Intelligent Systems and Computing)*, vol. 1018. Cham, Switzerland: Springer, 2020, pp. 344–349, doi: [10.1007/978-3-030-25629-6](https://doi.org/10.1007/978-3-030-25629-6).
- [81] B. D. Adams, L. E. Bruyn, S. Houde, and P. Angelopoulos, "Trust in automated systems: Literature review," Defence Res. Develop. Canada, Toronto, ON, Canada, Tech. Rep. CR-2003-09, 2003.
- [82] M. R. Endsley, "From here to autonomy: Lessons learned from human-automation research," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 59, no. 1, pp. 5–27, Feb. 2017, doi: [10.1177/0018720816681350](https://doi.org/10.1177/0018720816681350).
- [83] D. R. Large, G. Burnett, D. Salantri, A. Lawson, and E. Box, "A longitudinal simulator study to explore drivers' behaviour in level 3 automated vehicles," in *Proc. 11th Int. Conf. Automot. User Interface Interact. Veh. Appl.*, Sep. 2019, pp. 222–232, doi: [10.1145/3342197.3344519](https://doi.org/10.1145/3342197.3344519).

Jaume R. Perello-March received the B.Sc. degree in psychology and the M.Sc. degree in human evolution and cognition from the University of the Balearic Islands, Spain, in 2014 and 2016, respectively, and the Ph.D. degree in human factors engineering from the University of Warwick in 2022. He holds previous automotive research and industrial experience in leading European centres, such as the Swedish National Road Transport Research Institute (VTI) and the Galician Automotive Technology Centre (CTAG). He is currently a Research Fellow in human factors for future transport with the National Transport Design Centre (NTDC), Coventry University. He has articles published in his field in Q1/4 journals and has presented his work at international conferences. His research interests include driver state monitoring and behavior.

Christopher G. Burns received the B.Sc., Master by Research (M.Res.), and Ph.D. degrees in psychology from The University of Edinburgh, with a specialist background in quantitative research methods and statistics in individual differences with psychometric and psychophysiological methods. He joined the WMG's Human Factors Section, as a Post-Doctoral Research Fellow on the U.K. Autodrive Project, embedded with a Research Team at Jaguar Land Rover, in 2017. His involvement in U.K. Autodrive, consisted of designing and refining methodologies, then conducting and analyzing practical experiments in passenger and pedestrian experiences of a prototype autonomous low-speed electric vehicle (L-SATS) operating in a controlled arena. These studies included user attitudes and intentions-to-use, trust formation, technology acceptance and internal/external human-machine interface evaluations, using both quantitative and qualitative methods. Previously, he worked on projects involving driving simulation, sustained attention, mental workload, and emotional and attitudinal responses. He previously worked on ultrasonographic simulation and medical training in teaching contexts.

Roger Woodman received the Ph.D. degree from the Bristol Robotics Laboratory in 2013. He is currently an Assistant Professor and the Human Factors Research Lead at WMG. He is also the Deputy Director of the Centre for Doctoral Training, WMG, training doctoral researchers in the areas of intelligent and electrified mobility systems. He lectures on the topics of human factors of future mobility. He has more than 20 years of experience working in industry and academia. He has several scientific articles published in the field of connected and autonomous vehicles. His research interests include self-driving vehicles, shared mobility, and human-machine interaction.

Mark T. Elliott received the M.Eng. degree in electronic systems engineering and the Ph.D. degree in biomedical engineering from Aston University in 2000 and 2007, respectively, developing intelligent systems to discriminate between different walking patterns. He is currently an Associate Professor at the Institute of Digital Healthcare, WMG, University of Warwick. He subsequently spent a number of years as a Research Fellow with the Sensory Motor Neuroscience Laboratory, University of Birmingham, modeling multisensory integration in the context of human movement coordination. His research interests include measuring health, wellbeing, and behavior through data-driven approaches. This primarily involves analyzing and modeling data from wearable and mobile devices that capture movement and physiological responses. Much of his research is highly applied and involves collaborating with commercial and public-sector partners.

Stewart A. Birrell received the B.Sc. degree in sport science and the Ph.D. degree in ergonomics from Loughborough University, U.K., in 2002 and 2007, respectively. He is currently a Professor of human factors for future transport with the National Transport Design Centre (NTDC), Coventry University. He has spent the previous 15 years working within the transportation sector within industry and academia, with expertise ranging from driver behavior and distraction, multimodal warnings, user state monitoring and information requirements—all underpinned by the design of in-vehicle information systems, and their evaluation using driving simulators, virtual reality (VR), and field operational trials. He also applies innovative human factors engineering methodologies to enable real-world and virtual evaluation of user interaction with connected and autonomous vehicle (CAV), electric vehicle (EV), and urban air mobility (UAM) technologies and services. He has over 100 journals and conference papers, book sections and articles published in his field to date. He is an Editor of the internationally renowned, Q1/4* journal IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.