Keep Calm and Ride Along: Passenger Comfort and Anxiety as Physiological Responses to Autonomous Driving Styles

Nicole Dillen*, **Marko Ilievski***, **Edith Law, Lennart Nacke, Krzysztof Czarnecki, Oliver Schneider**
University of Waterloo, Waterloo, Canada
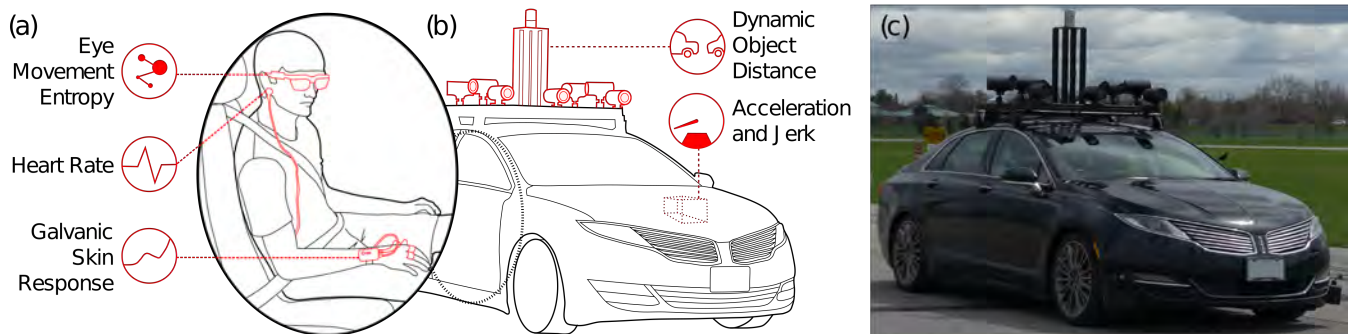{nicole.dillen, marko.ilievski, edith.law, len, kczarnec, oliver.schneider}@uwaterloo.ca

Figure 1: **(a) We measured physiological signals (heart rate, eye-movement entropy, galvanic skin response) and self-reported comfort and anxiety from passengers in an autonomous vehicle. (b) We correlated passenger data with driving style parameters, including acceleration, jerk, and dynamic object distance, as well as four events: following a lead vehicle, stopping at a sign, passing a vehicle, and a tight turn. (c) This study took place on a closed track in an autonomous vehicle.**

ABSTRACT
Autonomous vehicles have been rapidly progressing towards full autonomy using fixed driving styles, which may differ from individual passenger preferences. Violating these preferences may lead to passenger discomfort or anxiety. We studied passenger responses to different driving style parameters in a physical autonomous vehicle. We collected galvanic skin response, heart rate, and eye-movement patterns from 20 participants, along with self-reported comfort and anxiety scores. Our results show that the presence and proximity of a lead vehicle not only raised the level of all measured physiological responses, but also exaggerated the existing effect of the longitudinal acceleration and jerk parameters. Skin response was also found to be a significant predictor of passenger comfort and anxiety. By using multiple independent events to isolate different driving style parameters, we demonstrate a method to control and analyze such parameters in future studies.

---

*Both authors contributed equally to this work.

Author Keywords
autonomous vehicles, passengers, driving style, physiological sensing, comfort, affective computing, empirical study

CCS Concepts
•**Human-centered computing** → **Interaction techniques;**

INTRODUCTION
Despite significant and rapid technological advancements, autonomous vehicles (AVs) face a crisis of trust. Vehicles are quickly progressing towards SAE autonomy level 4, in which manual driving intervention is unnecessary within the operational design domain [12]. These fully-autonomous vehicles could lead to safer, faster, and more accessible transportation [30, 13]. Yet, people are hesitant to ride in them: according to a 2019 AAA survey, 60–80% of drivers interviewed expressed their reluctance to ride in a fully AV [16]. For passengers, trust, safety, and control issues are likely the underlying concerns behind their reluctance to ride in these vehicles [38]. Promoting a sense of comfort in passengers may be the next crucial challenge for autonomous driving.

One opportunity to promote a sense of comfort is by adapting the vehicle's *driving style*, the set of driving parameters that determine, for example, how quickly the vehicle accelerates or how closely it follows another vehicle [5]. Driving styles can either be fixed or adaptive. Currently, fixed driving styles are presets that target physical safety and not comfort; a fixed style also fails to account for how passengers may have varying

individual preferences towards driving styles [5]. Efforts to develop adaptive styles may use techniques like learning a driving style from user demonstration [28], but these do not directly target passenger comfort: passenger preferences may be a direct consequence of their *perceived* driving style, which might significantly differ from that which they employ in practice [5].

Researchers who have studied passenger reactions have typically conducted studies using simulators [6, 7, 26, 40] or Wizard-of-Oz studies [17, 53]. While these have yielded useful insights for other factors, they lack external validity for realistic motion with an AV. Passenger studies in physical AVs have looked at related topics like lane-changing behavior [18], self-reported trust in a controlled lab setting [37], or physiological measures of trust in a naturalistic setting [34].

In our work, we bridge the gap between lab and naturalistic studies and between self-report and physiological measures. We 1) connect passenger reactions of anxiety and comfort to physiological signals, 2) manipulate driving style parameters in a controlled study on a closed track, and 3) extend analysis to events, providing insight for adaptive driving styles. The closest study in this space is Mühl et al. [34], who studied overall trust rather than comfort and anxiety, in a less controlled setting (traffic). They did not, however, find a significant correlation between trust and the physiological device they employed.

Specifically, we collected physiological sensor data, such as galvanic skin response (GSR), heart rate (HR), and eye-movement patterns from 20 participants, along with self-reported comfort and anxiety scores on a 10-point Likert scale. We varied driving style parameters across four trials, each including four events: stopping behind another vehicle, passing a stationary vehicle, stopping at a stop sign, and taking a sharp turn. We isolated the individual parameters (lateral and longitudinal components of acceleration and distance to an agent vehicle) as much as realistically possible, because of the unavoidable confounds that occur as a result of naturalistic driving (for example, slight lateral translations or wavering may still occur on straight road driving); our methods can inform future in-vehicle studies.

We contribute: 1) evidence that the magnitude of certain physiological responses (e.g., GSR and HR) is significantly affected by driving parameters such as acceleration and jerk—more so in the presence and proximity of a lead vehicle; 2) insight into how driving conditions affect passengers; events involving a stop—particularly behind a lead vehicle—led to a higher level of response independent of driving style parameters. 3) In addition—because GSR was found to be a significant predictor of self-reported passenger comfort and anxiety—we are able to use our analysis results to understand the effect of driving style parameters on not just the physiological response, but also on the perceived comfort and anxiety itself. 4) Finally, we have released our generated dataset on an open source platform[1].

---

[1]Data hosted on dataverse: **https://doi.org/10.5683/SP2/FJXBRX**

RELATED WORK
We discuss prior work on physiological responses to stress in various contexts and position our work in the context of state-of-the-art passenger-vehicle interaction research.

Measuring physiological response to stress
Our work aims to understand the effect of autonomous driving style on the comfort and anxiety level of passengers in AVs. Specifically, we study *state anxiety*, defined by Spielberger [47] as a complex emotional response to a perceived threat, characterized by "feelings of tension and heightened autonomic nervous system activity." This leads us to investigate the physiological response of the passenger by measuring autonomous nervous system variables such as GSR, HR and heart rate variability (HRV), as well as eye movement patterns.

GSR
Commonly used GSR variables include the electrodermal activity (EDA) or skin conductance level (SCL), a slowly changing measure of tonic physiological activity, and skin conductance response (SCR), quick bursts of elevated conductance levels resembling peaks. Increased sweat production in the eccrine glands, which is the primary basis for SCL, indicates psychological arousal [32].

In the context of driving, Mühl et al. used SCL to measure passenger trust [34] in an AV, while Alpers et al. [3] found significant effect sizes in SCL among driving-phobic participants before, during, as well as after their exposure to anxiety-inducing, manual driving tasks. Elevations in SCL were also noticed among the non-phobic participants albeit at a lower level.

In a study by Blechert et al. [9], the largest effect sizes in relation to anxiety were demonstrated by electrodermal measures which included both SCL and SCR variables. Liu et al. [31] also found a positive association between SCR rate and amplitude and anxiety.

HR and HRV
Zheng et al. [55] investigated the use of a photoplethysmogram (PPG, a method of measuring cardiac activity) features to predict anxiety and found that the mean pulse rate could be a potential marker for anxiety. HR was similarly used by Katcher, Segal, and Beck [25] to indicate the onset of discomfort and anxiety in patients undergoing dental surgery. Taelman et al. [49] linked mental stress to an increase in the mean heart beat period and HRV, while an elevated heart was also found to be a consequence of discomfort (in relation to environmental noise) in a study conducted by Baker [4].

With frequency domain analysis of HRV, the low frequency (LF) component is generally found to be higher in a state of anxiety as compared to baseline conditions [36]. Rani et al. [41] exploited this phenomenon (with other physiological responses) to produce an index for measuring people's anxiety level when interacting with a robot. Meanwhile, Liu et al. [31] used the LF, along with the peak PPG amplitude and the mean and standard deviation of the interbeat interval, to train a classifier to dynamically adjust difficulty in video games.

Eye tracking
Eye tracking research is primarily based on fixations: clusters of gaze points that occur in quick succession, typically within 100-300 milliseconds, with a small spatial range [14]. Anxiety has been linked to eye movement patterns; in particular, loss of focus on the areas of interest seems to be a trend among individuals experiencing stress or anxiety. In the context of driving, vision-based gaze estimation has been used for driver monitoring to predict driver state and behaviour, such as driver fatigue and failure to signal [19, 29].

In the context of attention and performance during penalty kicks, Wilson et al. [51] found that players made significantly more fixations towards target areas in situations involving high threat but spent far more time focusing on the goal-keeper instead of the goal-target area, resulting in sub-optimal shooting strategies. Similarly, Causer et al. [10] showed that high anxiety caused competitive shotgun shooters to experience higher levels of distraction as evidenced by the decreased duration of final goal-oriented fixations. This was further corroborated by Allsop et al. [2], who found that anxiety increased scanning entropy—the randomness in scan behaviour—in the presence of cognitive load. In addition, they found that high anxiety significantly decreased the dwell time on the area of interest, while increasing it on irrelevant regions.

Positioning this work in the context of related studies
Existing studies on passenger-vehicle interaction for AVs can be grouped into three broad categories: physical AV studies, simulator studies, and Wizard-of-Oz studies.

Several on-road studies have been conducted inside *physical AVs* [18, 34], each of which differ from our research in one or more ways. Unlike our study, which was run entirely in a physical, autonomously driven vehicle and in a controlled real-world environment, Mühl et al. [34] conducted the field study portion of their experiment in an uncontrolled traffic environment and only investigated driving styles in a simulator. Festner et al. [18], meanwhile, studied only lane-changing behavior and did not measure physiological signals. Our work used a wide range of physiological signals to study comfort (and anxiety) complementing Mühl et al.'s focus on trust. Jaguar has also investigated physiological sensing for motion sickness in their vehicles, but did not study the effect of driving style on overall comfort and anxiety [45]. Besides these on-road studies, Oliveira et al. also conducted research inside a physical AV but did so in a controlled lab environment and relied only on self-reported scores and interviews [37].

Simulator studies, while highly controlled, are less ecologically valid [35] as they generally fail to account for realistic physical forces and motion components and provide an unavoidable safety bias. To avoid simulator sickness, these studies are also limited to mostly straight road driving and do not study turn-taking maneuvers. Most related works on passenger-vehicle interaction have been conducted in simulators and either study physiological response [6, 7, 26, 40] or do not measure them at all [5, 8, 21, 42, 43]. Kia did investigate physiological sensing in a prototype vehicle but their work focused on adaptive vehicle interiors as opposed to driving style [15]. Other simulator studies tend to focus mainly

on investigating take-over requests for safety drivers [23, 46, 50, 52].

The remaining *Wizard-of-Oz* studies were conducted with manually driven vehicles to study passenger [17, 53] and pedestrian [44] responses but lack in providing the full experience of autonomous driving to participants (e.g., automatic actuation of the steering wheel). In addition, these studies do not measure physiological responses and instead rely only on self-reported data.

STUDY DESCRIPTION
The study was conducted on a closed test track in Waterloo, Ontario, Canada, depicted in Figure 2, and involved the use of an AV and a human controlled vehicle agent.

The AV, i.e., the **ego vehicle**, used in this study was the "Autonomoose", a Lincoln MKZ hybrid research platform [11] developed at the University of Waterloo to reach Level 3 autonomy. The "Autonomoose", as seen in Figure 1, was fitted with an array of sensors including Novatel inertial measurement unit (IMU), Novatel Global Positioning System (GPS), Velodyne Light-Detection-and-Ranging (LIDAR), and a range of vision sensors. These sensors enable the AV to localize and understand the environment around it. The motion planning algorithm on board the vehicle uses the sensor information to select a trajectory in accordance with the intended driving style and the current environmental constraints [54]. Although the study was entirely run with autonomy engaged, a safety driver was present at all times to take over control in an emergency.

In addition to the ego, one other **agent vehicle**, a manually driven Lexus 450 Rx, was present on the track. This additional vehicle allowed us to investigate how passengers would react to autonomous driving in the presence of other vehicles. Besides this, no other pedestrian or vehicle was present in the study design.

Driving styles
Driving style is comprised of a large number of driving parameters [5] such as acceleration thresholds, or lane-change behaviors. In this study, we manipulated only a subset of the driving style parameters: the longitudinal and lateral thresholds of acceleration and distance. Longitudinal acceleration and distance represented the maximum forward acceleration and minimum distance from an object in front of the AV,

Table 1: We varied the thresholds for the lateral and longitudinal components of two parameters: acceleration and distance. Both components of each parameter were linked for a total of four different driving styles: Low Acceleration Low Distance, Low Acceleration High Distance, High Acceleration Low Distance, and High Acceleration High Distance.

Parameter	Less Aggressive		More Aggressive	
	Long.	*Lat.*	*Long.*	*Lat.*
Acceleration	2.5 m/s^2	2 m/s^2	4 m/s^2	4 m/s^2
Distance	10 m	4.5 m	7.5 m	2 m

respectively. Likewise, lateral acceleration and distance represented the maximum acceleration perpendicular to the track and minimum distance to objects beside the AV.

The parameter thresholds were evaluated in a pilot (N=12) study based on interview feedback and physiological responses and can be found in Table 1. These thresholds reached the maximum safety limit for the vehicle. Speed was also generally constant at the maximum safety threshold of 9.72 m/s.

We linked acceleration and distance thresholds to keep the experiment feasible. Manipulating four different thresholds would have resulted in 16 different driving styles for each of their combinations. To avoid having 16 trials for each participant, the longitudinal and lateral components of acceleration (and distance) were either both set at their more aggressive threshold, or both set at the less aggressive threshold. This resulted in a combination of only four driving styles (Low Acceleration Low Distance, Low Acceleration High Distance, High Acceleration Low Distance, and High Acceleration High Distance), each requiring its own trial.

The order of varying the thresholds in each trial was randomized for each participant in a within-subjects study design.

Although the *thresholds* were manipulated at the trial level, the individual variables themselves continuously vary in a realistic driving scenario. As a result, the analysis was performed not on the thresholds but on samples from the entire signal obtained for each of these variables. In addition, the derivatives of both components of acceleration–longitudinal and lateral jerk signals—were also considered in the analysis.

Manipulation checks
The interview feedback and physiological responses from the pilot demonstrated that the manipulations were indeed perceived by the participants. All participants were included in the final study irrespective of whether the manipulation checks were satisfied; this was done to increase the generalizability of the analysis models to unexplained variance [48].

Teasing apart parameters with events
With realistic driving, it is difficult to completely separate the effects of each driving style parameter. For example, there may always be some very minor contribution of lateral acceleration even when driving along a straight path because of slight, unintended wavering in the lateral translation. This could present confounds when analyzing entire trials.

To tease apart the influence of each individual parameter, four different events were considered, each testing a different parameter:

1. **Passing**: involved passing a parked car oriented perpendicular to the ego. This event tested *lateral distance*.

2. **Intersection-stop**: involved stopping at a clear intersection. This event tested *longitudinal acceleration*.

3. **Car-stop**: involved stopping behind a lead agent vehicle. This event was mainly used to test *longitudinal distance*, although longitudinal deceleration also played a role in the stop itself.
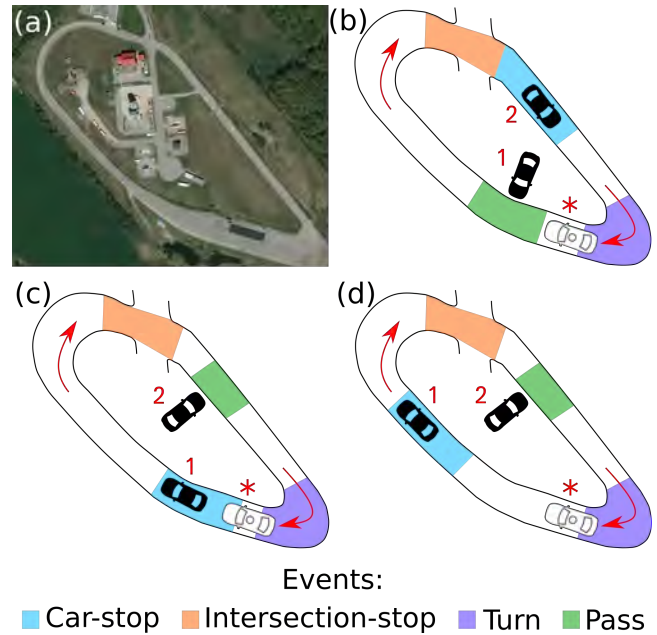

Events:
■ Car-stop ■ Intersection-stop ■ Turn ■ Pass

Figure 2: (a) Satellite image of the test track. (b-d) Order of events for each trial [trial 1 – b, trial 2 – c, trial 3 – d] indicated by the number beside the vehicle; the star represents the start and end location of the ego vehicles. Colored segments of the track represent the locations for the different events: green for passing, orange for intersection-stop, blue for car-stop, and purple for turning events. Trial 4 was a repetition of trial 1.

4. **Turning**: involved making a sharp turn at the end of the track. This event tested *lateral acceleration*.

By analyzing events, we can more precisely isolate the different driving style parameters, while introducing typical driving scenarios as additional effect.

Varying the order of events
It was expected that the same order of events in each trial would result in boredom and a strong order effect. To mitigate this, the order of the passing and car-stop events were varied for each trial (intersection-stop and turning events remained constant because only one intersection and one sharp turn were present in the test track, see Figure 2). Each participant experienced the same sequence of event ordering—it was only the threshold order that was randomized.

Participant task: watching a video
We needed an area of interest to measure eye movement entropy. We thus tasked participants to watch a video played on a 5-inch display smartphone fixed to the dashboard (participants sat in the passenger seat as seen in Figure 3). This task was regarded as one of the most plausible activities for a passenger in an autonomous vehicle while providing a trackable area of interest (seen in the bottom panel of Figure 3). A neutral topic – a documentary on the lost city of Atlantis – was chosen to minimize any associated emotional responses.
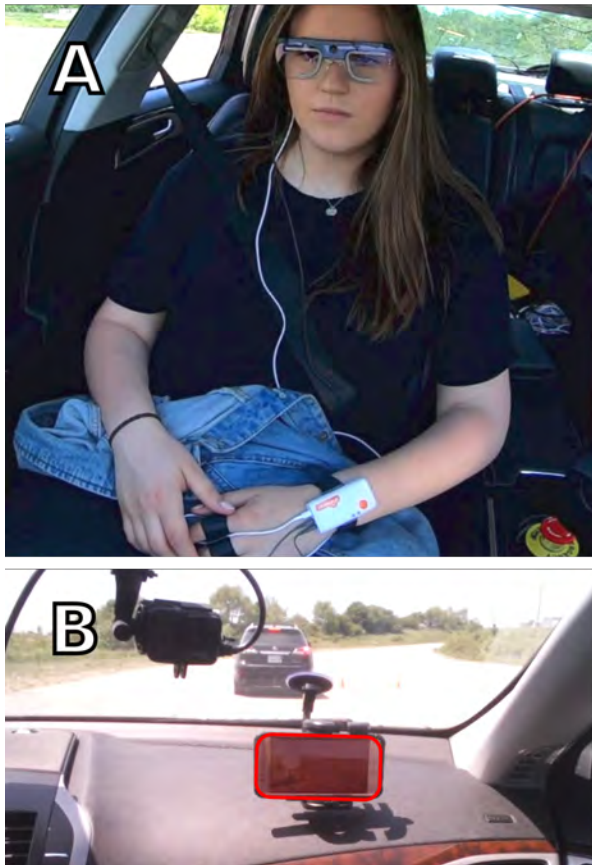
Figure 3: (a) A participant fitted with the physiological sensors. (b) Their view from the passenger seat, which includes the video displayed on the phone screen to watch as their task. The red bounding box indicates the AOI.

Participant data: questionnaires, self-reported scores, and physiological responses

We collected a pre-study questionnaire, self-reported scores, as well as physiological responses from participants.

Pre-study demographics questionnaire

Participants were asked to fill out a pre-study demographics questionnaire which had questions for self-reported age and gender, perceived driving style (very defensive, defensive, aggressive, very aggressive), and perceived knowledge and trust in autonomous vehicles (on Likert scales from 1–10). This questionnaire can be found in the supplementary material.

Self-reported scores

After each event, we asked participants for on-the-fly self-reported scores of comfort and anxiety. Each participant was asked to rate separately their comfort and anxiety on an integer scale from 1 to 10, 1 meaning "Not comfortable (anxious) at all" and 10 meaning "Very comfortable (anxious)".

After each trial, we asked participants to complete the somatic sub-scale of the Competitive State Anxiety Inventory - 2 (CSAI-2) questionnaire [33].

Physiological response

Three sources of physiological response were measured: GSR, HR (and HRV), and eye movement patterns. A Shimmer3+ device was used to measure GSR and obtain a PPG signal for HR and HRV, while a Tobii Glasses 2 Pro device was used to track gaze patterns. The sensor set-up is shown in Figure 3.

Participant recruitment

We recruited 20 participants (10F, 10M) with ages between 19 and 64 years inclusive ($M = 33.5$, $SD = 3.52$) using posters posted at the University of Waterloo. Participants had diverse educational backgrounds, occupations, and familiarity with technology and self-driving vehicles. Participants each received remuneration of \$15 CAD, in addition to reimbursement for transportation to and from the test track, if requested.

Procedure

Participants were first briefed about the experiment and given an information letter. After signing the consent form, they were fitted with GSR, PPG, and eye tracking sensors. In order to minimize movement artifacts in the GSR signal, the sensor device was fitted to the non-dominant hand of the participant using Velcro straps. Velcro straps were also used to snugly wrap the electrodes around the base of the middle and index fingers. To further reduce signal noise, participants were instructed to position their hand on their leg and keep as still as possible during the drive. In addition participants where asked to avoid talking. All electrodes were sanitized with alcohol wipes after each experimental run.

After the sensors were fitted, participants filled out the pre-study questionnaire, and were directed to focus on the video that was playing but to feel free to look up if they felt the need to. The experiment began after eye-tracker calibration. After completing all trials, participants were informally interviewed for their feedback, and were asked to rank each event in order of decreasing discomfort or anxiety.

SIGNAL PROCESSING

In this section, we describe the method of collection and processing of vehicle state and physiological signals used in the experimental analyses.

Ego vehicle driving data

Vehicle state signals were sampled at 20 Hz and can be seen in the middle panel of Figure 4.

Acceleration

Acceleration values were measured using a car-mounted IMU unit. To reduce signal noise, the raw acceleration data was filtered using a second-order Butterworth filter. The acceleration was then broken down into its lateral and longitudinal components which were used in turning and stopping events respectively.

Jerk

Jerk was calculated from the first-order derivatives of all acceleration components. To reduce the noise of these derivatives over discrete values, a mean filter with a window size of 2 seconds was applied on the jerk values.

Distance

The vehicle used the AVOD [27] detection algorithm to detect all surrounding dynamic objects, which in this case consisted of the other agent vehicle. This algorithm relies on input from multiple cameras along with the LIDAR data to produce bounding boxes in KITTI format [20]. After initial detection, each dynamic object bounding box is tracked using a Kalman filter [24]. While the algorithm is sufficiently robust to accomplish autonomous driving, we found that, for our analysis, further manual processing was sometimes required to guarantee the most accurate positioning of the agent vehicle, particularly when the detection algorithm failed to detect it. This additional processing involved forcefully constructing a bounding box in the region where the agent vehicle should have been present, based on the recorded image and LIDAR data.

We calculated the distance metric "bumper-to-bumper" based on the Euclidean distance between the ego vehicle and dynamic object's bounding boxes, using the closest pair of points between the two boxes; this method of distance accurately depicts how close two vehicles come to a collision at any point in time. We resolved the distance into its lateral and longitudinal components, which were used for the passing and car-stop events, respectively.

Participant Physiological Response Signals

The raw physiological response signals can be seen in Figure 4. Each individual response signal was processed further to extract the actual signals of interest.

GSR

The GSR variables used in the analysis included the SCL, number of peaks, and the maximum peak amplitude. Peak-related variables were applicable only to analyses involving aggregation over an interval, discussed later in more detail.

To remove an observed upwards linear trend (and, hence, the baseline), we detrended the raw GSR signal by first fitting an ordinary least-squares regression line to it and then subtracting this fitted line from the signal. The phasic component of the SCL signal was then extracted via a moving median filter with an 8 second window, and peaks were detected as the maximum amplitude between onset (SCL $> 0.01\mu$S) and offset (SCL $< 0\mu$S) pairs [22]. The median filtering also helped in the removal of noise artifacts in the signal.

HR and HRV

Peaks were detected in the PPG signal by finding local maxima in a 10 second sliding window. The number of peaks detected in each window were multiplied by six to obtain the HR in beats per minute (BPM). To measure HRV, peaks were detected over the entire trial and an interbeat interval (IBI) series was constructed from the differences in timing between each pair of consecutive peaks. The IBI series was then interpolated at regularly-spaced intervals and transformed to the frequency domain using a Fourier Transform. The normalized LF ($0.04 - 0.15$ Hz) and HF ($0.15 - 0.4$ Hz) components were then extracted and the ratio RF between them was computed. The HRV could only be extracted over the entire trial as a
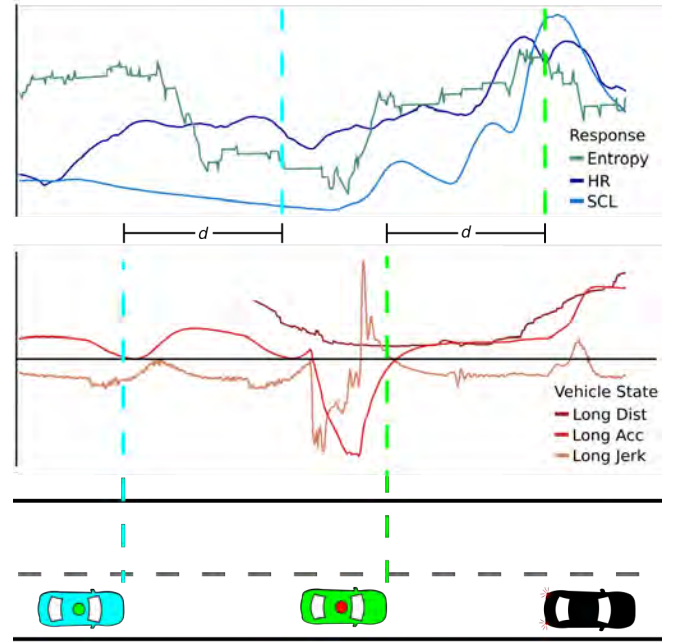

Figure 4: A car-stop event with physiological (above) and ego vehicle state (below) signals. As the ego vehicle approaches the agent vehicle, P19's physiological signals respond. Note that there is an inherent delay, *d*, associated with physiological signals.

smaller interval would not be adequate to capture the required frequency resolution.

Eye movement patterns

There were two areas of interest (AOIs) for studying eye movement patterns: the phone screen and the rest of the environment. Gaze patterns were converted to a sequence of fixations using the iMotions software [39]. Following the approach outlined by Allsop et al. [1], the fixation sequence between the two AOIs was converted to a transition matrix used to calculate the eye movement entropy over 10 second sliding windows, using the following formula:

$$Entropy = \sum_{i=1}^{n} p(i) \sum_{j=1}^{n} p(j|i) \log_2 p(j|i), i \neq j$$

where n is the number of AOIs, $p(i)$ is the zero-order probability of fixating on AOI i, and $p(j|i)$ is the probability of fixating on AOI j given a current dwell on AOI i.

Synchronization

The raw GSR, PPG, and gaze patterns were sampled at 512 Hz, 512 Hz, and 50 Hz respectively. The extracted SCL, HR, and entropy signals were sampled collectively at 20 Hz using linear interpolation and then synchronized with the vehicle state signals (acceleration, jerk, and distance, all sampled at 20 Hz). Physiological and vehicle state signals were measured using different clocks: video data on each system were used to manually label common events in both groups of signals
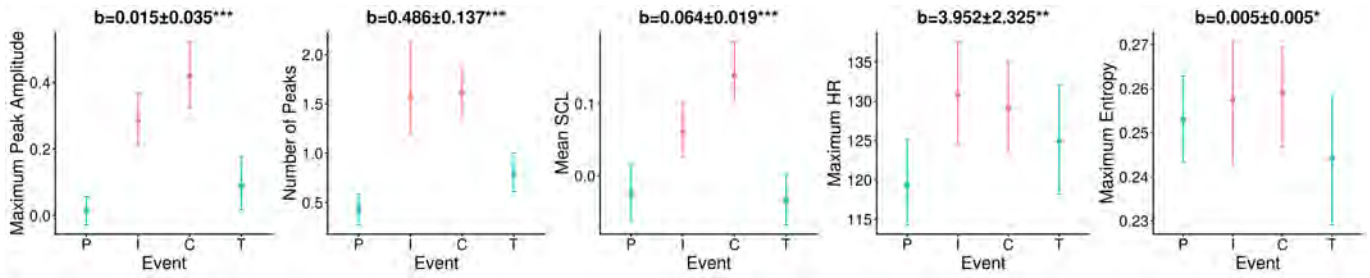
Figure 5: Plots highlighting the significant differences in physiological responses for different events (P=pass, I=intersection-stop, C=car-stop, T=turn). Stopping events (red) had a larger effect than non-stopping events (blue). All *b* values are reported at the top of each plot along with their 95% CIs; *, **, and *** indicate p-values less than 0.05, 0.01, and 0.001, respectively.

and align them in time. Nearest-neighbour interpolation was then used to synchronize and combine all signals.

ANALYSIS I: UNDERSTANDING THE ROLE OF EVENTS

Although the initial purpose of the events was to isolate the different driving style parameters, participants expressed that certain events affected them more than others. We thus investigated effects associated with event type, independent of the driving style parameters studied.

We regressed each physiological response variable against a categorical variable representing the event. We used linear mixed-effects models (LMEs) to perform the regression, controlling for random effects in the within-subjects grouping factor (the participant IDs) and the trial number (used to model learning effects across trials)[2]. To capture both average and extreme effects, each response variable was aggregated into maximum and mean values over the event intervals and included the mean SCL, maximum peak amplitude, number of peaks, maximum and mean HR, and maximum and mean entropy. Planned contrasts grouped the four events into stopping (car-stops and intersection-stops) and non-stopping (turns and passing) events. We discuss the significant findings (Figure 5):

We found that ***stopping events raised GSR, HR, and entropy.*** Most participants also directly stated their discomfort or apprehension with stopping (16 out of 20), particularly stopping behind the lead vehicle (14 out of these 16), which had a significantly higher effect on GSR compared to stopping at an intersection.

ANALYSIS II: DRIVING STYLE PARAMETERS

We used LMEs to regress each response variable on vehicle state variables, controlling for generally significant random effects in the within-subjects grouping factor (the participant IDs) and the trial number (used to model learning effects across trials). Event, window, and time-series models also used sequence number as a predictor along with a first order autoregressive co-variance structure to model the sequential nature of the data points.

---

[2]The source code and LME models used for all analyses can be found in our online Github repository: https://github.com/nicolebd/KeepCalm.

Levels of aggregation

To explore effects at different interval ranges, we conducted regressions at four different levels of aggregation:

1. *Trial level:* aggregations were performed over the entire trial. Additional variables included the normalized HRV variables—LF, HF and RF.

2. *Event level:* aggregations were performed over each event.

3. *Window level:* aggregations were performed over non-overlapping sliding windows, calculated based on position to spatially align them across all participants. A window width of 50 m and an average ego speed of about 10 m/s resulted in approximately 5 s intervals, designed to incorporate the 1-5 s lag in GSR, the slowest response.

4. *Time-series level:* Here, response variables included the SCL, HR, and eye movement entropy at each timestamp, devoid of any aggregation.

As in Analysis I, aggregations were done using the mean and maximum values for variables, where maximum-valued responses were regressed on maximum-valued predictors and mean-valued responses were regressed on mean-valued predictors. Furthermore, while maximum-valued predictors took into account directionality, mean values were calculated on the basis of the absolute values.

Trial level and time-series results were inconclusive: trial-sized intervals may have been too wide for the vehicle state predictors, while the actual time-series samples might have been insufficient to capture delays in physiological response. By extension, we did not find any significant effects on the HRV variables that were analyzed only at the trial level.

For brevity, we discuss only the results from the window level analyses: event-level results, although not reported, were similar and serve as a confirmation of our finer-grained findings. We also acknowledge the possibility of confounds between driving style parameters and events but argue that it would be impossible to study certain parameters in the absence of events (e.g., lateral acceleration and longitudinal distance cannot be studied without turning and car-stop events, respectively). We did, however, analyze driving parameters separately over each event and found similar significant or nearly significant ($p < 0.1$) results.

Table 2: Regression coefficients and confidence intervals for all significant predictors at the window level. No significant effects were found for the mean HR and mean entropy, so these responses were not included; *, **, and *** indicate p-values less than 0.05, 0.01, and 0.001 respectively.

Predictor (max)	Max. Pk. Amp. ($b\pm95\%CI$)	Num. Pks. ($b\pm95\%CI$)	Max. HR ($b\pm95\%CI$)	Max. Entropy ($b\pm95\%CI$)	Predictor (mean)	Mean SCL ($b\pm95\%CI$)
$a_{long}-$	$-0.025\pm0.037**$	$-0.081\pm0.085***$	$-1.949\pm2.964**$			
$a_{long}+$	$0.087\pm0.042**$	$0.246\pm0.094***$	$3.596\pm2.645**$	$-$	a_{long}	$-$
$j_{long}-$	$-0.026\pm0.043**$	$-0.068\pm0.095***$				
$j_{long}+$	$0.088\pm0.039**$	$0.27\pm0.09***$	$-$	$-$	j_{long}	$0.143\pm0.115*$
$pres_{lead}$	$0.12\pm0.072**$	$0.432\pm0.161***$	$6.546\pm4.223**$	$0.014\pm0.012*$	$pres_{lead}$	$0.067\pm0.042**$
d_{long}	$-$	$-0.014\pm0.013*$	$-$	$-$	d_{long}	$-$

Predictors

Vehicle state predictor variables included the lateral and longitudinal components of acceleration and jerk (a_{long}, a_{lat} and j_{long}, j_{lat}, respectively)—the mean and maximum were considered—as well a the presence of a lead vehicle and parked vehicle ($pres_{lead}$ and $pres_{pass}$, respectively).

The minimum lateral and longitudinal distances d_{lat} and d_{long} were considered in separate analyses when $pres_{pass}$ and $pres_{lead}$ were significant, respectively; these analyses were run only on the set of samples that had the parked or lead vehicle present.

Findings

Table 2 provides an overview of all **vehicle state** predictors that were found to be significant at the window level.

GSR and HR were directly proportional to the magnitude of longitudinal acceleration and jerk. Qualitative interview feedback directly supports these observations: participants explicitly referred to the "jerkiness" in braking and "jittery" driving, with one participant likening the motion to "when my kids were learning how to drive".

While the main effect of acceleration and jerk was largely obvious and expected, the more interesting observation was its directionality, as seen in the top panel of Figure 6: typically, the positive direction had a greater effect. For example, the maximum HR went up, on average, by about 3.6 BPM for every m/s^2 increase in positive acceleration, while it decreased by nearly 2 BPM per every m/s^2 increase in magnitude of deceleration. This might be a consequence of relatively high and jerky accelerations during the start of each trial, as well as at the end of each stopping event.

Lead vehicle presence (and proximity) raised GSR, HR, and entropy. This result suggests that passengers may be more apprehensive about their safety when another actor—in this case, the agent vehicle—is added to the picture. Indeed, 14 out of 20 participants considered the car-stop event to be the most uncomfortable and anxiety-inducing: on average, even the maximum HR went up by about 6.5 BPM when the lead vehicle was present, while every 100 m increase in *proximity* generated an additional 1.4 GSR peaks.
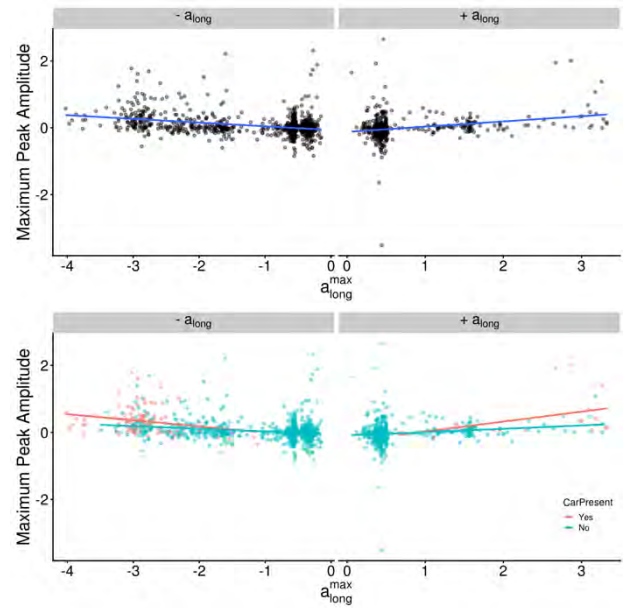

Figure 6: The main effect of acceleration on the GSR variable, maximum peak amplitude (top), and the magnified effect when a lead vehicle is present (bottom).

Lead vehicle presence (and proximity) magnified the effect of longitudinal acceleration and jerk for GSR and HR. The bottom panel of Figure 6 depicts how higher magnitudes of acceleration have an even higher effect on the maximum peak amplitude in the presence of a lead vehicle. Specifically, the interaction between lead vehicle presence and longitudinal acceleration was significant for the mean SCL ($b = 0.207\pm0.154, p < .01$), maximum peak amplitude ($b = 0.247\pm0.188, p < .05$) and number of peaks ($b = 1.130\pm0.417, p < .001$), while that between presence and longitudinal jerk was significant for the number of peaks ($b = 0.504\pm0.309, p < .01$) and maximum HR ($b = 9.115\pm8.029, p < .05$). The effect did not exist for entropy.

Weak ($p < .1$) or nearly significant ($p \approx .05$) interaction effects between lead vehicle *proximity* and longitudinal acceler-

ation or jerk were found for the number of peaks, mean and maximum HR, and mean entropy.

ANALYSIS III: SELF-REPORTED SCORES

Using the same approach as described in the previous analyses, we regressed the CSAI-2 and on-the-fly scores—over entire trial and event level intervals, respectively—on the number of peaks, mean SCL, mean HR, and mean entropy in one case, and the number of peaks, maximum peak amplitude, maximum HR, and maximum entropy in another.

Findings
We report significant results of this analysis in Table 3.

GSR predicted comfort and anxiety. For every unit increase in the number of peaks, we see that, on average, the predicted comfort decreases by about a quarter of a point on the 10-point Likert scale, while the CSAI-2 score increases by about 30%. The fact that these results were obtained at two levels of analysis for distinct methods of self-reporting (the standardized CSAI-2 questionnaire, versus the 10-point Likert scales) indicates that the physiological responses studied are indeed measures of comfort and anxiety, nicely tying together all of our prior results.

DISCUSSION

The use of a physical AV, instead of simulated environments, allows us to draw informed conclusions about the actual effect of each driving style parameter. By extension, the study led to the generation of a naturalistic dataset of passenger responses and the corresponding driving style parameters, a contribution that—along with the methods and findings of our study—can benefit both researchers who do not have access to a physical AV, as well as industrial practitioners investigating methods to make such vehicles more accessible to the public. We present three major findings:

1. **The type of driving event determines the magnitude of physiological responses.** While the initial purpose of modelling events was a methodological contribution towards isolating and testing different driving style parameters, it turned out that the type of event itself, independent of the driving parameters, determined the magnitude of the physiological response (GSR, HR, and entropy). Stopping events, particularly car-stop events, had the largest impact on response. This was corroborated by qualitative interview feedback, which also indicated the relative comfort of the turning and passing events.

2. **The magnitude of physiological responses is affected by external interacting agents.** The positive relationship between longitudinal acceleration and jerk and physiological responses was further positively magnified by the presence and proximity of a lead vehicle. This tells us that the magnitude of physiological responses is context-dependent, influenced not just by the type of event but the mere presence of another interacting agent, and its distance to the ego vehicle. *Passengers may, therefore, not only have individual driving style preferences, but these preferences may manifest differently depending on the scenario in question.* Moreover, participants talked about the "jerkiness" in driving as

Table 3: Regression coefficients and confidence intervals for significant physiological response predictors for self-reported comfort and anxiety; *, **, and *** indicate p-values less than 0.05, 0.01, and 0.001 respectively.

Predictors	Comfort (b ± 95%CI)	Anxiety (b ± 95%CI)	CSAI-2 (b ± 95%CI)
Num. Pks.	−0.245*** ±0.113	0.131* ±0.115	0.305*** ±0.165
Mean SCL	−0.934* ±0.837	–	–

a source of discomfort, an occurrence that would not have manifested in a low or medium fidelity simulator study, confirming the usefulness of our in-vehicle experiment for realistic results.

3. **Physiological responses indicate passenger comfort and anxiety.** Third, we were able to relate passenger skin conductance back to the self-reported comfort and anxiety scores. While similar findings can readily be found in the literature, its usefulness in this study lies in its confirmation of the fact that we were indeed studying the correct problem: the perceived comfort and anxiety experienced by passengers in an autonomous vehicle. Furthermore, a similar study on passenger trust did not find any significant correlations between trust and skin conductance [34], indicating that comfort and anxiety might be better self-report items to use in future research.

Besides our major findings, we also found (from analysis II) generally significant random effects for participant ID across all response variables; the significance of these effects indicates individual differences in baseline comfort levels in an AV suggesting the need to adapt the driving style to passengers at an individual level.

Applications in the real-world
To satisfy individual driving style preferences of passengers in future autonomous vehicles, the driving style parameters could be adapted at the individual level, either in an online or offline fashion to create personalized driving styles. Our findings tell us **which of these parameters** to consider.

Furthermore, the ability to relate physiological responses back to self-reported comfort and anxiety scores advocates for their use in an **implicit passenger-vehicle interface**. Given the vehicle state as context, the onset of discomfort or anxiety could be sensed from the physiological responses and implicitly used to make adjustments to the driving style parameters. Even in the absence of such an implicit, adaptive system, a set of preset driving styles *informed by actual passenger feedback and responses* could still improve the safety-focused presets that are currently employed.

Finally, the responses considered in this study have the advantage of being **amenable to unobtrusive, real-world implementations**, as opposed to other more obtrusive and error-

prone methods such as electroencephalogram [32]. For example, smartwatches and rings could be used to measure GSR and HR, and a stationary eye-tracker could be set up to measure eye movement patterns. Moreover, this has direct applications outside the context of AVs: smartwatches could be used to provide live feedback to Uber drivers (as we have all had Uber drivers that have driven a little too fast).

Limitations in Study Design
There were several limitations associated with the study:

1. Confounds were present in the study. Despite our efforts to isolate, as much as possible, separate parameters by means of events, there is still some amount of confounding as the nature of driving itself involves the combined effect of multiple variables. Confounds could also exist between driving style parameters and events but, as discussed in the section on aggregation levels, this was unavoidable in order to appropriately study each driving parameter.

2. External factors could have influenced the physiological responses and self-report scores. Events such as manual takeovers, and the presence of unexpected entities (e.g., a wild turkey) on the test track occurred during some trials. While the takeovers usually went unnoticed, two participants reported a spike in their anxiety after the takeover occurred. Other participants were noticeably affected by the presence of unexpected parked vehicles, while another was surprisingly calmed by an unexpected beeping alert (ironically a spurious crash warning). The weather also ranged from warm and sunny, to cloudy—and even rainy—throughout the study week. Such external factors, while contributing even more towards a realistic dataset, are hard to control for and were not taken into account during the analysis.

3. There were some threats to external validity. Safety concerns prevented us from running the study on a real road. Also, while the research is aimed at passengers in higher level vehicles, we were only able to run the study in a level 3 vehicle due to state-of-the-art limitations in automated driving. The presence of a safety driver is likely to be another major factor that downplayed passenger responses and feedback.

4. Diversity in perceived driving style would have been preferred. The pre-study demographics survey indicated that most participants preferred to employ a more defensive style of driving. Furthermore, a few participants explicitly expressed their boredom by the end of the third and fourth trials indicating the need for shorter or more interesting trials.

5. There were possible shortcomings in the video task. Although we tried to keep the task as natural as possible without causing any additional apprehension, the fact that there were no follow up questions on the video made it difficult to ensure that the participants actually watched it, possibly affecting the magnitude of the entropy variables.

CONCLUSION
In this work, we studied passenger comfort and anxiety as physiological responses to autonomous driving style. We ran an experiment in a physical AV, contributing a method to control and analyze different driving style parameters, and releasing a naturalistic driving dataset as a by-product.

We measured GSR, HR, and eye movement patterns, and collected self-reported comfort and anxiety scores from participants. By analyzing our data at multiple levels of aggregation, we found that parameters such as longitudinal acceleration and jerk affect all physiological responses, more so in the presence (and, to a lesser extent, proximity) of a lead vehicle. We also found that the type of event, independent of the driving parameters, determined the magnitude of GSR, HR, and entropy; stopping events—particularly stopping behind a lead vehicle—had the most impact.

Finally, we were able to relate the physiological responses back to the self-reported comfort and anxiety scores, indicating that the responses could potentially be used to sense (dis)comfort or anxiety in future passenger-vehicle interfaces.

Future work for this research should primarily lie in addressing the outlined limitations, exploring a more diverse set of events, and comparing passenger responses to automated driving with those achieved during manual driving. Systems that warn the passenger about upcoming stressful events while adapting the driving style to their physiological response should be investigated.

REFERENCES
[1] Jonathan Allsop and Rob Gray. 2014. Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition* 3, 2 (2014), 63–71.

[2] Jonathan Allsop, Rob Gray, Heinrich H Bülthoff, and Lewis Chuang. 2017. Eye movement planning on Single-Sensor-Single-Indicator displays is vulnerable to user anxiety and cognitive load. *Journal of Eye Movement Research* 10, 5 (2017), 8–1.

[3] Georg W Alpers, Frank H Wilhelm, and Walton T Roth. 2005. Psychophysiological assessment during exposure in driving phobic patients. *Journal of Abnormal Psychology* 114, 1 (2005), 126.

[4] Carol F Baker. 1992. Discomfort to environmental noise: Heart rate responses of SICU patients. *Critical Care Nursing Quarterly* 15, 2 (1992), 75.

[5] Chandrayee Basu, Qian Yang, David Hungerman, Mukesh Singhal, and Anca D. Dragan. 2017. Do You Want Your Autonomous Car To Drive Like You?. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17).* ACM, New York, NY, USA, 417–425. DOI: http://dx.doi.org/10.1145/2909824.3020250

[6] Matthias Beggiato, Franziska Hartwich, and Josef Krems. 2018. Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving. *Frontiers in human neuroscience* 12 (2018).

[7] Matthias Beggiato, Franziska Hartwich, and Josef Krems. 2019. Physiological correlates of discomfort in automated driving. *Transportation research part F: traffic psychology and behaviour* 66 (2019), 445–458.

[8] Hanna Bellem, Barbara Thiel, Michael Schrauf, and Josef F Krems. 2018. Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation research part F: traffic psychology and behaviour* 55 (2018), 90–100.

[9] Jens Blechert, Marta Lajtman, Tanja Michael, Jürgen Margraf, and Frank H Wilhelm. 2006. Identifying anxiety states using broad sampling and advanced processing of peripheral physiological information. *Biomedical sciences instrumentation* 42 (2006), 136–141.

[10] Joe Causer, Paul S Holmes, Nickolas C Smith, and A Mark Williams. 2011. Anxiety, movement kinematics, and visual attention in elite-level performers. *Emotion* 11, 3 (2011), 595.

[11] Colwell, Ian. 2018. Runtime Restriction of the Operational Design Domain: A Safety Concept for Automated Vehicles. (2018). http://hdl.handle.net/10012/13398

[12] SAE On-Road Automated Driving Committee and others. *SAE J3016. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Technical Report. tech. rep., SAE International, 2016. Cited on.

[13] Peter Davidson and Anabelle Spinoulas. 2015. Autonomous vehicles: what could this mean for the future of transport. In *Australian Institute of Traffic Planning and Management (AITPM) National Conference, Brisbane, Queensland*.

[14] Andrew T Duchowski. 2007. Eye tracking methodology. *Theory and practice* 328, 614 (2007), 2–3.

[15] Stephen Edelstein. 2018. Kia wants future autonomous cars to be able to read passengers' emotions. https://www.digitaltrends.com/cars/kia-emotion-recognition-ai-tech-at-ces-2019/. (2018). Accessed: 2019-09-15.

[16] Ellen Edmonds. 2019. Three in Four Americans Remain Afraid of Fully Self-Driving Vehicles. https://newsroom.aaa.com/2019/03/americans-fear-self-driving-cars-survey/. (2019). Accessed: 2019-09-15.

[17] Mohamed Elbanhawi, Milan Simic, and Reza Jazar. 2015. In the passenger seat: investigating ride comfort measures in autonomous cars. *IEEE Intelligent Transportation Systems Magazine* 7, 3 (2015), 4–17.

[18] Michael Festner, Alexandra Eicher, and D Schramm. 2017. Beeinflussung der Komfort-und Sicherheitswahrnehmung beim hochautomatisierten Fahren durch fahrfremde Tätigkeiten und Spurwechseldynamik. *von* 11 (2017).

[19] Lex Fridman, Heishiro Toyoda, Sean Seaman, Bobbie Seppelt, Linda Angell, Joonbum Lee, Bruce Mehler, and Bryan Reimer. 2017. What Can Be Predicted from Six Seconds of Driver Glances?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 2805–2813.

[20] A. Geiger, P. Lenz, and R. Urtasun. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 3354–3361. DOI: http://dx.doi.org/10.1109/CVPR.2012.6248074

[21] Franziska Hartwich, Matthias Beggiato, and Josef F Krems. 2018. Driving comfort, enjoyment and acceptance of automated driving–effects of driversâĂŹ age and driving style familiarity. *Ergonomics* 61, 8 (2018), 1017–1032.

[22] iMotions. 2016. Galvanic Skin Response (GSR): The Complete Pocket Guide. https://imotions.com/blog/galvanic-skin-response/. (2016). Accessed: 2019-09-11.

[23] Myounghoon Jeon. 2019. Multimodal Displays for Take-over in Level 3 Automated Vehicles while Playing a Game. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, LBW0229.

[24] Rudolph E Kalman and Richard S Bucy. 1961. New results in linear filtering and prediction theory. *Journal of basic engineering* 83, 1 (1961), 95–108.

[25] Aaron Katcher, Herman Segal, and Alan Beck. 1984. Comparison of contemplation and hypnosis for the reduction of anxiety and discomfort during dental surgery. *American Journal of Clinical Hypnosis* 27, 1 (1984), 14–21.

[26] Toru Kobayashi, Tetsushi Ikeda, Yumiko O Kato, Akira Utsumi, Isamu Nagasawa, and Satoshi Iwaki. 2018. Evaluation of Mental Stress in Automated Following Driving. In *2018 3rd International Conference on Robotics and Automation Engineering (ICRAE)*. IEEE, 131–135.

[27] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven L Waslander. 2018. Joint 3d proposal generation and object detection from view aggregation. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1–8.

[28] Markus Kuderer, Shilpa Gulati, and Wolfram Burgard. 2015. Learning driving styles for autonomous vehicles from demonstration. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2641–2646.

[29] Michael Glazer William Angell Spencer Dodd Benedikt Jenik Jack Terwilliger Aleksandr Patsekin Julia Kindelsberger Li Ding Sean Seaman Alea Mehler Andrew Sipperley Anthony Pettinato Bobbie Seppelt Linda Angell Bruce Mehler Bryan Reimer Lex Fridman, Daniel E. Brown. 2019. MIT Advanced Vehicle Technology Study: Large-Scale Naturalistic Driving Study of Driver Behavior and Interaction With Automation. *IEEE Access* 7 (2019), 102021–102038. DOI:http://dx.doi.org/10.1109/ACCESS.2019.2926040

[30] Todd Litman. 2017. *Autonomous vehicle implementation predictions*. Victoria Transport Policy Institute Victoria, Canada.

[31] Changchun Liu, Pramila Agrawal, Nilanjan Sarkar, and Shuo Chen. 2009. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction* 25, 6 (2009), 506–529.

[32] Regan L Mandryk and Lennart E Nacke. 2016. Biometrics in Gaming and Entertainment Technologies. In *Biometrics in a Data Driven World*. Chapman and Hall/CRC, 215–248.

[33] Rainer Martens, Damon Burton, Robin S Vealey, Linda A Bump, and Daniel E Smith. 1990. Development and validation of the competitive state anxiety inventory-2. *Competitive anxiety in sport* (1990), 117–190.

[34] Kristin Mühl, Christoph Strauch, Christoph Grabmaier, Susanne Reithinger, Anke Huckauf, and Martin Baumann. 2019. Get Ready for Being Chauffeured: Passenger's Preferences and Trust While Being Driven by Human and Automation. *Human factors* (2019), 0018720819872893.

[35] Nadia Mullen, Judith Charlton, Anna Devlin, and Michel Bedard. 2011. *Simulator validity: behaviours observed on the simulator and on the road* (1st ed.). CRC Press, Australia, 1 – 18.

[36] Hiroki Murakami and Hideki Ohira. 2007. Influence of attention manipulation on emotion and autonomic responses. *Perceptual and Motor Skills* 105, 1 (2007), 299–308.

[37] Luis Oliveira, Karl Proctor, Christopher G Burns, and Stewart Birrell. 2019. Driving Style: How Should an Automated Vehicle Behave? *Information* 10, 6 (2019), 219.

[38] Simone Pettigrew, Caitlin Worrall, Zenobia Talati, Lin Fritschi, and Richard Norman. 2019. Dimensions of attitudes to autonomous vehicles. *Urban, Planning and Transport Research* 7, 1 (2019), 19–33.

[39] Imotions: Biometric Research Platform. 2018. Eye Tracking: The Complete Pocket Guide. https://imotions.com/blog/eye-tracking. (2018). Accessed: 2019-07-17.

[40] Kathrin Pollmann, Oilver Stefani, Amelie Bengsch, Matthias Peissner, and Mathias Vukelić. 2019. How to Work in the Car of the Future?: A Neuroergonomical Study Assessing Concentration, Performance and Workload Based on Subjective, Behavioral and Neurophysiological Insights. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 54, 14 pages. DOI:http://dx.doi.org/10.1145/3290605.3300284

[41] Pramila Rani, Nilanjan Sarkar, Craig A Smith, and Leslie D Kirby. 2004. Anxiety detecting robotic system–towards implicit human-robot collaboration. *Robotica* 22, 1 (2004), 85–95.

[42] Patrick Rossner and Angelika C Bullinger. 2019a. Do You Shift or Not? Influence of Trajectory Behaviour on Perceived Safety During Automated Driving on Rural Roads. In *International Conference on Human-Computer Interaction*. Springer, 245–254.

[43] Patrick Rossner and Angelika C Bullinger. 2019b. How Do You Want to be Driven? Investigation of Different Highly-Automated Driving Styles on a Highway Scenario. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 36–43.

[44] Dirk Rothenbücher, Jamy Li, David Sirkin, Brian Mok, and Wendy Ju. 2016. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*. IEEE, 795–802.

[45] Jaguar Land Rover. 2018. Preventing Motion Sickness. https://www.jaguarlandrover.com/2018/preventing-motion-sickness. (2018). Accessed: 2019-09-15.

[46] Shadan Sadeghian Borojeni, Susanne CJ Boll, Wilko Heuten, Heinrich H Bülthoff, and Lewis Chuang. 2018. Feel the movement: Real motion influences responses to take-over requests in highly automated vehicles. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 246.

[47] Charles D Spielberger. 1966. Theory and Research on Anxiety. New York: Academic Press, 3–19.

[48] Detmar Straub, Marie-Claude Boudreau, and David Gefen. 2004. Validation guidelines for IS positivist

research. *Communications of the Association for Information systems* 13, 1 (2004), 24.

[49] Joachim Taelman, Steven Vandeput, Arthur Spaepen, and Sabine Van Huffel. 2009. Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*. Springer, 1366–1369.

[50] Remo van der Heiden, Shamsi T Iqbal, and Christian P Janssen. 2017. Priming Drivers before Handover in Semi-Autonomous Cars. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 392–404.

[51] Mark R Wilson, Greg Wood, and Samuel J Vine. 2009. Anxiety, attentional control, and performance impairment in penalty kicks. *Journal of Sport and Exercise Psychology* 31, 6 (2009), 761–775.

[52] Priscilla NY Wong, Duncan P Brumby, Harsha Vardhan Ramesh Babu, and Kota Kobayashi. 2019. " Watch Out!": Semi-Autonomous Vehicles Using Assertive Voices to Grab Distracted Drivers' Attention. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, LBW2312.

[53] Nidzamuddin Md Yusof, Juffrizal Karjanto, Jacques Terken, Frank Delbressine, Muhammad Zahir Hassan, and Matthias Rauterberg. 2016. The exploration of autonomous vehicle driving styles: preferred longitudinal, lateral, and vertical accelerations. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 245–252.

[54] Yu Zhang, Huiyan Chen, Steven L. Waslander, Jianwei Gong, Guangming Xiong, Tian Yang, and Kai Liu. 2018. Hybrid Trajectory Planning for Autonomous Driving in Highly Constrained Environments. *IEEE Access* 6 (2018), 32800–32819. DOI: http://dx.doi.org/10.1109/ACCESS.2018.2845448

[55] Yali Zheng, Tracy CH Wong, Billy HK Leung, and Carmen CY Poon. 2016. Unobtrusive and multimodal wearable sensing to quantify anxiety. *IEEE Sensors Journal* 16, 10 (2016), 3689–3696.