UNIVERSITÄT OSNABRÜCK

Institute of Cognitive Science

Master thesis

A psychophysiological analysis of a person's need for information to improve mobility concepts

by
Charlotte Brandebusemeyer

First supervisor: Dr. Klas Ihme,
Deutsches Zentrum für Luft- und Raumfahrt

Second supervisor: Prof. Dr. Peter König
IKW, University of Osnabrück

Submission date: 11th July 2022

Declaration of Authorship

I hereby certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other university.

Osnabrück, 11th July 2022

C. Brandebusemey
____
Charlotte Brandebusemeyer

Abstract

Autonomous driving enables new mobility concepts such as on-demand shared autonomous mobility services. With increasing technological progress in the field of autonomous driving, more attention should be given to the passenger-autonomous vehicle interaction. In the current study, the passenger-autonomous vehicle interaction during such an on-demand drive was examined. The study was conducted as a real-world Wizard-of-Oz study in which the passengers were unaware that a driver was controlling the vehicle instead of the vehicle driving autonomously. Each participant was collected from a virtual stop, was driven to a specific destination and driven back to the starting location. On the way back in the vehicle, two unexpected events occurred – an abrupt braking event and an unexpected detour. 19 out of the 37 participants received an error notice followed by content information during or shortly after an event occurred, while the remaining 18 participants only received an error notice during the events. The focus of my thesis is to examine the communication of the automated vehicle to the passenger when unexpected events occur. The psychophysiological data that was gathered during and after the drive in the autonomous vehicle, served to determine a passenger's information need during the drive. Analyses of the physiological data show a cardiological and electrodermal reaction to event situations – especially during the abrupt braking event. The subjective evaluations served as the main source of insight on a passenger's wish for information during events and during the whole drive. The results indicate that explanatory information is highly desirable. Furthermore, information has a positive impact on the passengers' feeling of pleasure, safety and understandability of the vehicle during the drive.

Content

List of Figures

List of Tables

1 Introduction

1.1 Autonomous driving and concepts of shared mobility

At present, transportation research promotes a transition from traditional human-driven vehicles to autonomously driving vehicles. According to the Society of Automotive Engineers (SAE), vehicles that drive autonomously can be subdivided into six levels of driving automation ranging from no driving automation (level 0) to full driving automation (level 5) (On-Road Automated Driving (ORAD) Committee, 2021). There a several reasons why autonomously driving vehicles have become a major topic in transportation research. One idea is the prospect of eliminating accidents caused by human errors. In 2014 the International Transport Forum (ITF) reported that over 90% of road crashes were related to human error and already in 2010, they stated that the transport sector consumes a large portion of energy from non-renewable fossil fuels and that road traffic is responsible for about 80% of transport CO2 emissions (International Transport Forum, 2010, 2014). Transferring driving responsibilities from a human driver to an automated vehicle may increase driving safety (Howard & Dai, 2013; International Transport Forum, 2014) and reduce the use of fossil fuels and emissions (International Transport Forum, 2010).

In the future, more people will live in urban surroundings and cities will grow (Dobbs et al., 2011). This development poses problems for the current infrastructure since, with a growing population, traffic also increases, which leads to more traffic congestion. Traffic congestion again has an impact on people’s health and on the economy (Requia et al., 2018). Additionally, private vehicles are parked 95% of the time (UN HABITAT III, 2016), taking up parking space (R. Zhang et al., 2015). Smart technology and innovative mobility concepts are needed to improve the efficiency of the existing infrastructure (Dia & Javanshour, 2017). According to Dia & Javanshour (2017), there are six key forces that drive mobility transformation: vehicle electrification, automated self-driving, mobile computing, on-demand shared mobility services, Big Data and predictive analytics with the help of Deep Learning / Artificial Intelligence. Autonomous shared mobility has the potential to improve traffic efficiency with regards to less congestion, due to fewer vehicles needed, and less parking space required such that the space can be used for other public purposes (Dia & Javanshour, 2017; Machado et al., 2018; Rode et al., 2017; R. Zhang et al., 2015). Additionally, competitiveness may be reduced while social equality and the quality of life in cities could increase, according to Rode et al. (2017).

Several concepts of shared mobility currently exist and each of them have positive as well as negative aspects. Shared mobility is defined by Machado et al. (2018) as “trip alternatives that aim at maximizing the utilization of the mobility resources that a society can pragmatically afford, disconnecting their usage from ownership”. Public transport is one type of shared mobility. Negative aspects of public transportation are that fixed routes and time schedules as well as low frequency during some times of the day or in some regions limit the flexibility of a passenger. Taxis on the other hand provide more flexibility but are less affordable (Atasoy et al., 2015). Further vehicle- and/or route-sharing options are carsharing (Machado et al., 2018; Shaheen et al., 2016), personal vehicle sharing (Correia et al., 2014) and ridesharing (Fagnant & Kockelman, 2018). These mobility concepts have infrastructural, environmental and monetary benefits (Machado et al., 2018; Shaheen et al., 2016) but flexibility time-wise and the departure point of the trip are somewhat restricted. On-demand ride services can be seen as an innovative door-to-door service for people who use their phones readily and whose focus it is to minimize time and cost (Gupta et al., 2019). Two types of on-demand services are currently in use: ridesourcing (Machado et al., 2018) and ridesplitting (Shaheen et al., 2016). To further adapt to the passenger’s wish for a specific departure place, on-demand transportation can additionally introduce virtual stops (Brost et al., 2019; Hahn et al., 2020; Rehme et al., 2021). A virtual stop is comparable to a bus stop with the difference that there is no indication of the stop in the real world. The virtual stop is only displayed on technical devices. Predictive positioning of autonomous vehicles can additionally reduce the passenger's waiting time (Miller & How, 2016). An increase in flexibility and a decrease in the cost of a trip are the current challenges of shared mobility.

Automated vehicles enable new innovative mobility concepts. Shared mobility concepts which adapt to the passenger's needs are also currently considered for automated vehicles. Besides technological benefits of innovative mobility concepts, challenges regarding acceptance problems and rebound effects need to be considered (Pakusch et al., 2018). According to Pakusch et al. (2018), people do not evaluate all information present to make a rational decision on their mobility choice but compare the advantage of one travel mode over another. Factors that influence the decision are travel time, travel cost, comfort, flexibility, availability, reliability, safety (Pakusch et al., 2018) and environmental concerns (Machado et al., 2018).

Socio-demographically, elderly people, people too young to drive, people living in urban areas with high congestion and parking issues and people under the influence of alcohol could profit most from on-demand and ridesharing and see the services as attractive concepts (Machado et al., 2018). Currently, people aged 30-39 years, with higher income and education than the average population are the main users (Fleury et al., 2017; le Vine & Polak, 2019). A prerequisite to be able to use the (automated) shared mobility services is that people have to be familiar with technological devices (Gupta et al., 2019; Machado et al., 2018). For some people, such as the elderly, this is a limitation. Social questions may also arise when algorithms decide which passengers to pick up and at what price (Meurer et al., 2020). Such socio-demographic differences and issues need to be considered in future research on automated shared mobility services.

Besides ecological benefits, also possible negative ecological effects of autonomous (shared) mobility need to be considered. Simulation-based studies showed that shared self-driving fleets can achieve similar or even the same mobility as today but with fewer cars, which would lead to reduced congestion, fewer parking spaces and reduced emission by replacing heavier cars (Fagnant & Kockelman, 2015; Martinez et al., 2015; Rigole, 2014). The results of a study in which people's time management was examined indicate that driverless mobility will affect people's use of travel time as well as how they manage their time in general (Stevens et al., 2019). People have the desire to complete tasks during the drive and then to use the saved time for normally neglected activities in everyday life such as social and leisure activities (Stevens et al., 2019). These positive aspects can, however, also turn into negative environmental effects due to resource saving: Trip costs might be lower than a taxi drive (Fagnant & Kockelman, 2015) and time efficiency might rise due to being able to complete tasks during the drive (Stevens et al., 2019) leading to a higher user demand and more booked trips per person and therefore higher mileage. An indirect rebound effect can also be that, due to financial savings by travelling in energy-efficient vehicles, people could have more money to spend on overseas holidays (Chitnis et al., 2013) with the effect of higher pollution than can be saved through shared mobility concepts. Empty runs, vehicles that drive around the blocks until they receive the next request or vehicles that park outside the city in cheaper car parks and have to drive back into the city because of a request can also be side-effects that have a negative impact on the environment (Pakusch et al., 2016). Krueger et al. (2016) point out that previous studies have neglected the impact of self-driving autonomous cars on current non-car users. These could prefer on-demand services to walking, cycling or public transport. Pakusch et al. (2018) conducted a study with 302 participants from Germany who filled in surveys on the influence of autonomous driving on mobility behaviour and the user preferences in the future. Their results showed that public transportation becomes less attractive with the growth of automated carsharing. According to Pakusch et al. (2018), sustainable mobility is increased by making public transport more attractive rather than increasing the attractiveness of carsharing concepts. However, carsharing can also be seen as an interim solution to motivate people to use public transport.

Shared mobility requires rethinking the concept of owning a vehicle. It requires a shift from vehicle "ownership" to "usership" (Machado et al., 2018). However, a private car has a clear advantage over other transportation options in three ways, according to Martinez & Viegas (2017): flexibility, comfort and availability. Santos (2018) supports the findings by reporting that on-demand shared mobility services are less attractive than private cars because of waiting time, travel time, comfort and convenience. Maghraoui et al. (2020) showed in their study that car users had the largest potential to change to an automated vehicle followed, however, directly by public transport users. Pakusch et al. (2018) could also not confirm their hypothesis that ownership will become outdated. Private cars – with automation or not – will remain the preferred mode (Pakusch et al., 2018).

In general, the willingness to use autonomous vehicles seems to be present but the attractiveness of shared travelling needs to improve. In the future, autonomous shared mobility concepts have to continue to focus on being easily accessible, efficient, sustainable, inclusive by considering different socio-demographic backgrounds, safe and have advantages in comfort over other transportation options, such as that travel time can be used for tasks apart from driving. It is of importance that governments encourage the concept of shared mobility (Santos, 2018).

1.2 Passenger-autonomous vehicle interaction

Passenger-autonomous vehicle interaction has an impact on a passenger's expectations and attitude towards automated vehicles. Nowadays, people have idealized expectations about technology capabilities of automated shuttles, which might be nourished by todays media (Nordhoff et al., 2019). Tussyadiah et al. (2017) found that people have a low-level negative attitude towards technology but a high level of trust in automated vehicles. At the same time, they prefer some sort of supervision over the autonomously driving shuttles than having no supervision (Nordhoff et al., 2019). The likelihood of using self-driving taxis is, on the one hand, negatively influenced by the view that technology dehumanizes (Tussyadiah et al., 2017). Passengers in autonomous vehicles fear a loss of control and of fun in driving (Howard & Dai, 2013). On the other hand, the likelihood is positively influenced by people's expectations of the reliability, functionality and helpfulness of an autonomous vehicle (Tussyadiah et al., 2017).

Individual and socio-demographic differences in people's expectations in automated vehicles exist. In a survey study in the US, Zhang et al. (2021b) revealed that drivers' expectations in automated vehicles differ significantly between age, gender, ethnicity, education levels, marital status, driving frequency, driving experience and personality. High expectations were found in younger, male, White non-Hispanic, more highly educated, never married, high frequency drivers but with less driving experience and who are highly extravert, agreeable, conscientious and emotionally stable. In the future, the cause for the different individual traits leading to different expectations of automated vehicles should be examined, according to Zhang et al. (2021b). Whilst the attitude towards and the trust in autonomous vehicles seem to be in general high, a certain level of distrust in technology remains. This is an issue that needs to be addressed while considering Socio-demographic differences in people's expectations of automated vehicles.

Three general research directions in human-computer interaction in the area of (shared) autonomous vehicles can be observed: passenger-centred studies, road-interaction studies and interactions of passengers with the shared mobility services (Meurer et al., 2020). Passenger-centred research mostly emphasizes the positive effect for passengers of not having to drive and being able to improve their time management (Meurer et al., 2020; Stevens et al., 2019). Studies on road-interactions examine the interaction between an autonomous vehicle and road users (Meurer et al., 2020; Rothenbücher et al., 2016). The third research direction examines the acceptance and interaction of passengers with shared mobility shuttles (Detjen et al., 2020; Meurer et al., 2020; Nordhoff et al., 2019). A fourth currently rising research area focuses on the internal passenger-automated vehicle interaction. The preferred means of communication between the passenger and the automated vehicle is investigated. While different aspects of human-autonomous vehicle interaction can be studied, the internal passenger-automated vehicle communication is the focus of this study.

In general, passengers travelling in autonomous vehicles wish for interaction with the vehicle, mostly in the form of receiving information (Meurer et al., 2020; Schraagen et al., 2021; Zhang et al., 2021c). According to Schraagen et al. (2021), a system that is as much self-explanatory as possible is seen as most satisfactory, trustworthy and achieves the highest situational awareness. In a real-world Wizard-of-Oz study – a study in which the participants were made believe that they were travelling in an autonomously driving vehicle while actually a person was driving the vehicle – Meurer et al. (2020) found that, during the drive, awareness about the actual route was important for the passengers to relax and feel comfortable. Installing screens to display the route, approximate arrival time, the traffic situation and for navigation and entertainment or work-related activities is proposed by Meurer et al. (2020). The participants also wished for a clear start and end signal of the drive, together with information about the trip. When incidents occurred, passengers were not sure whether they were onlooking bystanders or somewhat involved and responsible for the actions of the vehicle and wished for information in unsafe situations (Meurer et al., 2020). In general, the passengers in the study were quite actively checking the performance of the vehicle and the route that the vehicle was taking. According to the concept of "locus of control", the passenger feels that either they or the automated system are mainly responsible for the behaviour of the vehicle (Stanton & Young, 1998). Insufficient information may lead to an experienced passive role and a feeling of failing to maintain a sense of control (Zhang et al., 2021c). For more people to adopt automated on-demand mobility systems, the altered role of humans in the vehicle needs to be communicated. Human drivers would not become obsolete but receive new roles i.e. new types of employment (Tussyadiah et al., 2017). Detjen et al. (2020) also found that passengers in the autonomously driving vehicle have a relevant role. Passengers might want to communicate requests to the vehicle on a manoeuvre-based level instead of on a control level, for example when they wish to stop for a break. Intervention by the passenger will not be possible in every situation, so available options for the passenger could be communicated by the vehicle. Voice commands or gestures are possible ways to interact with the vehicle. However, also the communication of the vehicle to the passenger about current actions can induce a feeling of trust (Walker et al., 2016).

Information need can be understood as the desire of a person to receive information to satisfy the subjectively experienced lack of information. What type of information to display and the timing of giving certain information to satisfy this need are still ongoing research questions. Providing explanations for actions can promote trust and acceptance in the automated vehicle (Zhang et al., 2021c). Their results indicated that actions of the vehicle need to be explained the right way. For semi-automated driving, Koo et al. (2015) studied how verbalized messages affect the driver's attitude and safety performance. Explanation content can be subdivided into three categories: what-information, why-information and what-and-why information (Zhang et al., 2021c). What-information refers to the action of the automated vehicle, why-information gives the reason for the action of the automated vehicle and what-and-why information provides both what action the vehicle took and why it reacted so. Only giving what-information to the passengers led to the worst driving outcomes in the case of semi-automated driving and lowered the acceptance of the automated vehicle (Koo et al., 2015). Why-information promoted acceptance and trust, a decrease in anxiety (Koo et al., 2015), a sense of control

(Koo et al., 2016) and perceived understandability (Wiegand et al., 2019). What-and-why information can promote trust by making driving automation more accessible because the drivers/passengers do not have to monitor the driving environment to understand the intentions of the system (Forster et al., 2017; Zhang et al., 2021c). However, what-and-why information can also increase anxiety and annoyance compared to the other two categories (Koo et al., 2015, 2016). The need for an explanation is correlated to driving types and driving scenarios (Shen et al., 2020). More aggressive drivers want fewer explanations and near-crash situations lead to the wish for explanations. Shen et al. (2020) did not find a preferred explanation format, i.e. what- vs. what-and-why information. These results stand in contrast to the above-mentioned studies that revealed a preference for why- and what-and-why information display.

Explanation timing is crucial for the effectiveness of the explanation given by the automated vehicle (Zhang et al., 2021c). Koo et al. (2015) suggested the idea to provide information ahead instead of after an event, so the drivers trust in that the vehicle takes over for a reason. The authors suppose that the idea can also be transferred to fully automated vehicles. Higher positive attitudes and lower negative feelings are the consequence of explanations prior to an event (Koo et al., 2016). Du et al. (2019) and Ruijten et al. (2018) conducted studies in driving simulators to investigate the participants' reactions to events and the timing of the information display. Information prior to an event induced the highest trust in both studies compared to receiving no explanation. An explanation display 7 seconds prior to the event additionally led to the least anxiety and workload and was preferred by the participants compared to an information display within one second after the event or no information at all. Information after the event induced the lowest trust (Du et al., 2019). Participants considered a system that gave information prior to an event to be more intelligent and human-like and was liked more compared to a system that did not give information (Ruijten et al., 2018). Ruijten et al. (2018) suppose that interfaces that mimic human behaviour may help increase trust and therefore acceptance. When information was given after an event, trust in and acceptance of automated vehicles did not differ to when no information was given (Körber et al., 2018). Schraagen et al. (2021) even showed that post-hoc explanations lower trust and satisfaction. Koo et al. (2015) noted that post-hoc explanations can be seen as redundant in some cases such that the driver has a higher workload to process the information. However, post-hoc information increases a driver's understanding (Körber et al., 2018), especially for less aggressive drivers and after accidents (Shen et al., 2020).

Schneider et al. (2021) proposed a method of live explanations during a drive in an automated vehicle. The explanations were not based on text displays. In a simulation study they showed a group of participants a head-up display in which a problem was marked in a yellow box. After the simulated drive, the participants additionally received explanatory information on a smartphone. The results showed that both live explanations and retrospective explanations after a drive have a positive impact on the perceived feeling of control. Live explanations increased the understanding and ease of use of the simulated automated vehicle but so did the retrospective explanations if no live explanations were given during the trip.

Demographic differences exist concerning the preferred way in which information should be conveyed. Results of Zhang et al.'s (2021a) study indicated that the explanations given by automated vehicles should in part be designed based on the passenger's age. While before-action explanations produced the highest trust in all age groups, explanation after an action produced high trust in younger drivers. When the system requested permission for an action, older people had the highest trust in the automated vehicle while at the same time younger and middle-aged people felt the highest distrust and anxiety. The different perceptions may be due to older people feeling less comfortable in giving up driving control (Li et al., 2019).

In sum, the type and timing of information display influences the passenger-autonomous vehicle interaction. Explanatory why and what-and-why information is preferred. While the timing of the information display is favoured before/during the event, explanations after events or after a drive can also have a positive impact on the passengers. The type and timing of information displays should be somewhat adaptable to the passenger's age (Zhang et al., 2021a).

1.3 Psychophysiological measures as indicators for emotions and information need

The cognitive and emotional systems of a human have an influence on a person's physiology (McCraty et al., 2009). In the field of psychophysiology, interactions between the physiological, cognitive and emotional system with human behaviour is examined (McCraty et al., 2009). Physiological activity can be associated with metabolic, physical activity and non-metabolic, emotional activity (Brouwer et al., 2018). McCraty et al. (2009) found in multiple self-conducted studies that emotions and physiological activities interact with each other: Emotions have a physiological consequence and patterns of physiological activity continually influence the emotional experience. Cardiac and electrodermal activity can be measured via electrocardiographs (ECG) and electrodermal activity (EDA) sensors respectively to examine a person's physiological arousal. Differentiating physiological arousal into concrete emotions such as anger, anxiety and fear is, however, challenging.

Cardiovascular parameters such as the heart rate (HR) and the heart rate variability (HRV) can be seen as indicators of psychological stress (Boucsein & Backs, 2000; Brouwer et al., 2018). Stress was defined by Selye (1936) as a non-specific adaptation phenomenon of the organism which allows living beings to react to events. Cardiological measures can directly reflect the activity of the autonomous nervous system (Sztajzel, 2004) which underlies the physiological reaction to stress (Lanatà et al., 2014). According to McCraty & Shaffer (2015), fluctuations in the HR result from complex, non-linear interactions between different physiological systems. More specifically, the HR reflects the balance between the sympathetic and parasympathetic activity of the autonomous nervous system: high activity of the parasympathetic (vagus) nerves slows the HR while activity of the sympathetic nerves accelerates it (McCraty & Shaffer, 2015). Sudden changes in the HR is primarily parasympathetically mediated via the vagus nerve (McCraty & Shaffer, 2015). The flexibility of the autonomous nervous system to modify bodily states, which is measured by the HRV, is crucial for adjusting between high and low arousal states and has thus been related to emotional regulation (Pham et al., 2021). The HRV is a measure that reflects heart-brain interactions and autonomic nervous system dynamics (McCraty et al., 2009) and is, like the HR, also associated with psychological stress (Brouwer et al., 2018). HR and HRV are inversely correlated in terms of the so-called cycle length dependence: If the HR increases, there is less time between the heart beats for variation to occur, therefore the HRV decreases (McCraty & Shaffer, 2015). Increased HR and decreased HRV are commonly observed as reactions to (psychologically) stressful events (Beggiato et al., 2018; Brouwer et al., 2018). In sum, HR and HRV are indicators of cardiovascular activity, which reflects the activity of the autonomous nervous system and which in turn is associated with emotion regulation.

EDA describes the perspiration of a person. Besides thermoregulatory sweating, there is also "emotional" perspiration (Kuno, 1965). The peripheral automatic response of emotional perspiration is induced by external or internal stimuli that generate psychological arousal (Christopoulos et al., 2019). EDA is considered an indicator for the activity of the sympathetic nervous system (Christopoulos et al., 2019; Kyriakou et al., 2019), which, when highly active, is associated with the "fight-or-flight" fear reaction (Roelofs, 2017). Since the sympathetic system reacts more slowly to an event than the parasympathetic system (McCraty & Shaffer, 2015), a slight delay in electrodermal reaction to an event is to be expected. Emotional sweating is most profound in the palms of the hands (Christopoulos et al., 2019). It is an ongoing research discussion on which hand to perform the perspiration measurements or whether both hands should be measured simultaneously. The underlying disagreement and questions are whether emotions are hemispherically lateralized and if they are, which hemisphere is associated with which emotions and whether the projections from the hemispheres project ipsi- or contralateral to the hands (Banks et al., 2012; Gainotti, 2019; Kasos et al., 2018; Picard et al., 2016; Ross, 2021; Wyczesany et al., 2018). Thammasan et al. (2019) put forward the general idea that fast EDA fluctuations measured in both hands are caused by emotional processes while dissimilarities are caused by artefacts. For the analysis of the EDA, the signal can be subdivided into a tonic and a phasic component. The skin conductance level (SCL) is a measure for slow tonic shifts in the EDA signal and therefore in arousal. Skin conductance responses (SCRs) are the phasic changes in the EDA signal and indicate a reaction to a stimulus (Kyriakou et al., 2019). A high number of SCRs indicates high perspiration during a time period. A high amplitude and a long rise time of an SCR act as indicators of a high intense eliciting stimulus (Kyriakou et al., 2019). These tonic and phasic components can be analysed to detect the amount of perspiration, which is an indication of the activity of the sympathetic autonomous nervous system and hence an indication of emotional arousal and psychological stress or even fear.

Emotions and arousal are related to information that is presented. Koo et al. (2015), for example, showed with the help of subjective evaluations in questionnaires that why-information, given during a driving simulation with events, decreased anxiety. Physiological data can additionally strengthen the subjectively reported information given by the participants and provide insight into individual emotional reactions and wished for information. Only both objective measurable physiological data as well as subjective evaluations of the participants together can give a holistic picture of people's emotions and their need for information.

1.4 Current study

In this current study, the focus lies on investigating a person's need for information when travelling in an autonomously driving vehicle with the aim to improve future mobility concepts. To examine a person's need for information, physiological data (ECG and EDA data) as well as subjective evaluations were analysed. The study is embedded in the "ViVre" project of the "Deutschen Zentrum für Luft- und Raumfahrt" (DLR). The aim of the project is to develop and investigate new functions of autonomously driving vehicles within a future publicly available mobility concept. The current study was conducted as a real-world Wizard-of-Oz study in which an on-demand drive in an automated vehicle of automation level four (SAE) was simulated. During the drive, two events – an abrupt braking event and an unexpected detour - occurred. Shortly after the start of the events, half of the participants received an error notice and the other half received an error notice followed by content information (what-and-why information). Using real-world studies instead of simulations to investigate the impact of automated vehicle explanations could reduce the potential for biased research outcomes, since people might behave differently in driving simulators than in real-world driving conditions (Bella, 2008; Zhang et al., 2021c).

Since the study was conducted prior to my thesis, my contribution is the analysis of the physiological and psychological data. Cardiological data, electrodermal data and subjective ratings of pleasantness were collected during the drive in the vehicle and retrospective evaluations of the drive were gathered in a post-questionnaire. The analyses of the participants' physiological changes and changes in subjective ratings during the events, especially when information was displayed, aimed at identifying whether the participants preferred an information type (error notice vs error notice followed by content information).

Four hypotheses were examined:

- H1: An unexpected event whilst travelling in an autonomously driving vehicle has an impact on the passenger's physiology and subjective evaluations of the situation.
- H2: Participants differ in their physiological reaction and subjective evaluation when an error message together with content information is displayed to when only an error message without content information is displayed during the events.
- H3: The physiological reaction and subjective evaluation to the information display during an event is different for different event types.
- H4: Retrospectively, the participants preferred the display of content information over merely an error message to help them understand the actions of the vehicle, to feel less insecure during an event and to reduce the wish for more information.

The first three hypotheses refer to the data collected during the drive, whereas the fourth hypothesis refers to the post-questionnaire. The first hypothesis serves as a control that the events as such had an impact on the participants. The second and third hypotheses examine the effect of the information type and the event type on the participants. The fourth hypothesis examines the participants' retrospective desire for specific information during the events and the whole drive in the vehicle.

2 Methods

2.1 Study design

The real-world study was designed as a so-called Wizard-of-Oz study in which the participants were made believe that they are travelling in an autonomously driving vehicle. Whilst travelling in the vehicle, two events occurred: an abrupt braking event and an unexpected detour. During the events, the participants received different information, depending on what group they were assigned to:

***withInfo group**:* During the drive, the participants received an error message followed by content information whenever an event occurred.

***noInfo group**:* During the drive, the participants received an error notice without content information whenever an event occurred.

A between-group comparison of the participants' reactions to different information displays was enabled by this design.

2.2 Participants

42 subjects signed up for the study. Four participants had to be excluded from the analyses due to malfunctioning technical equipment. One further participant was excluded because he was under time pressure for private reasons during the whole experiment and the physiological data was therefore not comparable to that of the other participants. 37 participants remained for the analyses (11 female, 26 male). The mean (M) age over all participants was 33 years with a standard deviation (SD) of 11.0 and an age range of 36 years (min=21 years, max=57 years). The participant groups consisted of 19 participants for the withInfo group (ten female, *M* = 33 years, *SD* = 11.5, *range* = 35 years) and 18 participants for the noInfo group (one female, *M* = 34 years, *SD* = 10.7, *range* = 33 years). All but one participant had a driver's license. Most of the participants were civil servants or employees (43.2%) followed by students (40.5%). The rest of the participants were either self-employed, pensioners, looking for work or they did not say. Participants were recruited via a participant database of the DLR as well as via Facebook. For each started half hour, the participants received 5€. All participants were informed about the process of the study and gave their informed consent to participate.

2.3 Technical set-up

The vehicle that was used for the study was the VIEWCar II, a VW Passat GTE Plug-in Hybrid, from the DLR (Appendix A1, Figure A1.1). The VIEWCar II is a research vehicle that is technically able to drive fully autonomously in selected urban areas. A GPS sensor (NAVILOCK GPS NL-402U USB u-blox 5 Chipsa) was used to record the location coordinates and the vehicle speed during the drive. In the vehicle, a tablet (Microsoft Surface Pro 7, 12.3') served as a display of information.

Physiological data was collected with the ECG sensor EcgMove 3 and two EDA sensors EdaMove 3 from movisens GmbH (Karlsruhe, Germany). Subjective ratings during the drive were gathered via an application on a Google Pixel 4a smartphone. The application could additionally be used to book drives with the vehicle, to fill in questionnaires and to navigate.

2.4 Experimental set-up

2.4.1 Wizard-of-Oz study

In the real-world Wizard-of-Oz study, the participants were made believe that they were travelling in an automated vehicle. The vehicle as such was technically able to drive the route of the experiment fully autonomously but shortly before the start of the study, a large construction site opened up on the route, such that the trip had to be driven manually. To be able to examine the driving experience of the participants in an autonomously driving vehicle, the participants were still instructed that they would travel in an autonomously driving vehicle and that the person behind the steering wheel was merely present for safety reasons to interfere if necessary. To prevent the participants from realizing that the person was actually driving, a partition was set up between the front and back seats. The sight through the front window was therefore blocked (Appendix A1, Figure A1.2). The participants were told that in the future, autonomously driving vehicles would drive without a safety driver and that that feeling was a central aspect that was going to be investigated in this study which should be reinforced by the partition. Additionally, a video was shown to the participants, in which the vehicle drove a section of the route completely autonomously. At the end of the study, the participants were informed that the vehicle did not drive autonomously.

2.4.2 Scenario

The experiment took place in a scenario which embedded the trip in an autonomously driving vehicle in a possible everyday situation. Each participant received the instruction to imagine that they could not work for the upcoming three hours due to technical difficulties at their workplace. Therefore, they decide to return an old book to the nearest sidewalk library and to borrow a new one. To get from the workplace to the sidewalk library and back, they book an autonomous shuttle-car with the help of the application on the smartphone and is picked-up from a virtual stop.

2.4.3 Route

Figure 1: Driving route

The driving route of the vehicle is depicted in this figure. The light blue and orange lines indicate the way from the DLR campus to the sidewalk library. The orange sections mark the sections of the route that were used as baseline routes for the analyses of the physiological data. The dark blue line indicates the way back from the sidewalk library to the DLR campus. On the way back from the sidewalk library, two events occurred which are marked in red.

The starting point of the drive in the vehicle was close to the DLR campus in Brunswick. The starting point was chosen in such a way that the virtual stop, at which the automated vehicle awaited the participant, was not apparent and that navigating to the stop with the application was necessary. The participant got into the right back seat of the vehicle and the vehicle started to drive and stopped near the sidewalk library (Figure 1). With the help of the application, the participant navigated to the sidewalk library. Then the participant booked a vehicle via the application for the way back to the DLR campus. He/she navigated via the application to the next virtual stop and the vehicle drove back in the direction of the DLR campus (Figure 1). On the way back, two events occurred (Figure 1): The first was that the vehicle halted abruptly in a parking lot at the side of the road and remained parked for about one minute. Then the driving continued. The second instance occurred when the vehicle deviated from the route that was indicated via the application on the smartphone. The detour was about one kilometre long and took about two minutes. Finally, the vehicle arrived at the DLR campus, the participants left the vehicle and navigated via the application to the starting point of the experiment on the DLR campus. The whole trip in the vehicle took about 30 min and the drive was about 15 km long.

2.4.4 Information display

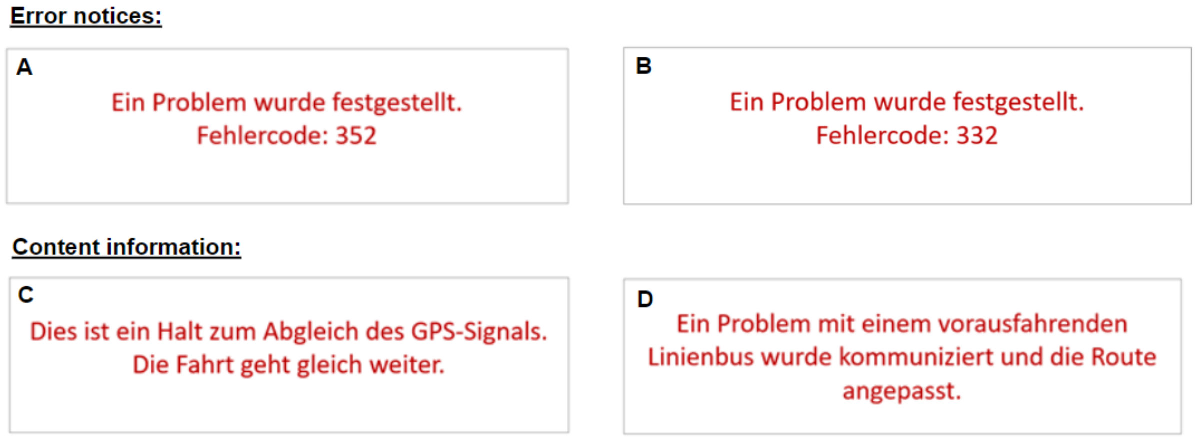


Figure 2: Information display
The error notices that were displayed on the screen in front of the participants in the vehicle during the instances are shown. (A) + (B) are error notices that every participant received at the beginning of an event: "A problem was detected. Error code: 352" for the abrupt braking event and "A problem was detected. Error code: 332" for the unexpected detour. (C) + (D) are content information notices which were displayed after an error notice for participants in the withInfo group. (C) is the notice for the first event: "This is a stop to synchronize the GPS data. The drive will continue soon." (D) is the notice for the second event: "A problem with a public-transit bus driving in front was communicated and the route was adapted."

During the whole drive, a display in the participant's visual field showed the starting and end address of the trip. Information that the vehicle is driving autonomously was also displayed (Appendix A2). Depending on which participant group the passenger in the car was assigned to, different information was displayed on the screen in front of the participant in the car when the events occurred: Participants assigned to the withInfo group received an error notice followed by content information on what happened and why the vehicle reacted to the incident in the specific way (Figure 2). Participants assigned to the noInfo group only got an error notice without content information on the incident (Figure 2, A+B). The information display was started and stopped manually by the driver of the vehicle.

2.5 Measures

During the drive in the automated vehicle, physiological data from ECG and EDA-sensors recordings and subjective ratings of the participants were gathered. After the drive, a post-questionnaire was filled in by the participants.

2.5.1 Physiological recordings

To measure cardiological activity during the drive, the participants wore an ECG sensor on their skin as a chest belt.
Electrodermal activity was measured with EDA-sensors during the drive. Two EDA sensors were worn on each wrist and the two electrodes were applied to the skin on the palm of each hand. Since emotional sweating is most profound in the palms of the hands (Christopoulos et al., 2019), recordings from these sites were taken (Appendix A1, Figure A1.3).

2.5.2 Subjective rating

Figure 3: Subjective rating
A screenshot of the application on the participant's smartphone is depicted. At the bottom of the screen, the participant's well-being can be rated ranging from unpleasant to pleasant.

During the whole experiment, the participant was told to rate his/her well-being on a 10-point Likert scale from unpleasant (one) to pleasant (ten) with a slider at the bottom edge of the smartphone screen (Figure 3).

2.5.3 Post-questionnaire

The post-questionnaire contained Likert-scaled questions and open-ended questions concerning the participants' information needs. The questionnaire was given to the participants in German. The following questions are therefore translations from German to English. Please find the original German questions in the Appendix (Appendix A3).

Quantitative questions were posed to the participants regarding the event of the abrupt braking, the event of the unexpected detour and the whole drive in the vehicle. For each of these three categories, three questions regarding the information that was given to the participants during the drive had to be rated on a 5-point Likert scale. The rating with one meant total disagreement and five meant total agreement. The three questions were the following:

1) The display in the autonomous shuttle helped me to understand the present actions of the shuttle.
2) In that situation / during the drive I felt insecure due to missing information.
3) In that situation / during the drive I wished for more information.

"The following information would have helped me to understand the situation better" was the open-ended question that was posed to the participants for the two events. To evaluate the whole drive in the vehicle, the participants were asked to describe up to three situations in which they would have wished for more information during the drive in the shuttle.

2.6 Study Procedure

For the study, the participants came to the DLR campus. First, documents regarding data privacy, information about the study and a consent form were handed over to the participant and were signed. Then the participant was equipped with a mobile ECG, two EDA-sensors – one for each wrist - and a smartphone. A demographic questionnaire, a questionnaire regarding the technical affinity and a questionnaire to evaluate the participant's acceptance of autonomous driving were handed to the participant and filled in. Then the scenario of having to return a book to the sidewalk library by using an autonomously driving vehicle to get there was introduced. Due to the complexity of the study, a summary with the most important information was given to the participants to take with them on the trip and the procedure was once again explained by the experimenter. The participant then navigated to the virtual stop with the help of the application, got into the vehicle and was driven to a location near the sidewalk library. The participant left the vehicle and navigated with the help of the application to sidewalk library to return a book and to select a new one. Here, the participant filled in the acceptance questionnaire regarding autonomous driving a second time on the smartphone. Then the participant booked a vehicle via the application for the way back to the DLR campus. He/she navigated to the next virtual stop and the vehicle drove back in the direction of the DLR campus. On the way back, two events – an abrupt braking event and an unexpected detour – occurred. Depending on which group the participant was assigned to, he/she either only received an error notice or an error notice followed by content information during the events. Finally, the vehicle arrived at the DLR campus, the participants left the vehicle and navigated back to the starting point of the experiment on the DLR campus. There, the participant filled in the acceptance form for the third time and filled in a post-questionnaire. Then the participants were debriefed and informed that the vehicle did not drive autonomously and questions regarding the payment for participating in the study were answered. The whole experiment took about two hours. The conductance of this study is in accordance with the Helsinki Declaration and was accepted by the ethical committee of the DLR. The participants gave their informed consent to partaking in the study and a hygiene concept according to the present Covid-19 rules was designed and implemented.

3 Data processing

3.1 Data preparation

3.1.1 Cardiological data

The HR and the HRV are the cardiological parameters of the ECG recordings which are considered in the following. The ECG data of only 23 out of 37 participants could be examined (Table 1) because in 13 cases the data was not recorded due to the device being unfunctional and had to be sent in for repair and in one case because of temporal gaps in the signal recording.

HR

The HR of the participants was recorded and the HR at beats per minute was calculated by the algorithm of movisens. In the case of the physiological parameter HR of the ECG data, a strong individual component has to be considered. Age, gender (Umetani et al., 1998), physical fitness (Grant et al., 2013) etc. influence the baseline heart rate of every individual making the absolute values hard to compare between participants. A first exploration of the physiological data indicated a strong inter-individual difference in the baseline HR (Appendix A5). Therefore, a common scale is needed. An often used transformation of the values is the z-transformation of the HR values. Two baseline route segments (Figure 1) were selected to perform the z-transformation with per participant. One route was a section of the trip where there were no traffic lights and where there was little interaction with other road users ("Forststraße"), while the other route consisted of a normal urban situation with traffic lights, crossings etc. ("Bienroder Weg"). Together, these two routes represent a normal everyday driving scenario.

$$MD = \frac{MD_{Forststraße}(HR) + MD_{Bienroder\ Weg}(HR)}{2} \tag{1}$$

The median (MD) HR of each participant was calculated per baseline route and then the mean of the two MDs was taken to account for the different lengths of the baseline routes (1).

$$MAD = \frac{MAD_{Forststraße}(HR) + MAD_{Bienoder\ Weg}(HR)}{2} \tag{2}$$

The same procedure was performed for the calculation of the median absolute deviation (MAD): the median absolute deviations of the heart rates of the baseline routes were calculated and the mean of the two MAD values was taken (2).

$$zHR = \frac{HR - MD}{MAD} \tag{3}$$

The z-transformation of every HR value was then performed by subtracting the MD and dividing by the MAD of the baseline route heart rates (3).

The MD and MAD were chosen instead of the typically used M and SD because they are more robust measures. Following analyses will be performed with the z-transformed HR values (zHR).

HRV

The heart rate variability of each participant was analysed with the time-domain metric Root Mean Square Successive Difference (RMSSD). This measure was chosen to analyse short-term changes in the HR. Time-domain measures of the HRV quantify for the amount of variance in the inter-beat intervals (McCraty & Shaffer, 2015). Since sudden changes in the HR are primarily mediated parasympathetically via the vagus nerves and the RMSSD measure estimates the vagally mediated changes (Shaffer et al., 2014; Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, 1996), the RMSSD measure is a suitable statistically robust measure (Salahuddin et al., 2007) to analyse the HRV. Baek et al. (2015) showed in their study that the RMSSD metric is reliable for 30s time windows. Frequency-domain indices of the HRV require at least 60s windows to be considered robust metrices (Baek et al., 2015). Since the duration of the abrupt braking event is only about 60s long, frequency-domain measures will not be considered and a focus on the more robust time-domain measure RMSSD of the HRV is laid. The normalized RMSSD (nRMSSD) values were calculated manually for each participant.

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2} \tag{4}$$

The RMSSD values were calculated by taking the temporal difference of successive RR-intervals. An RR-interval is the distance between two R-peaks in the ECG signal. By squaring the difference of successive RR-intervals, one receives only positive values. Then the mean of these quadrated differences is calculated and the square root is taken (4).

$$nRMSSD = \frac{RMSSD}{RR} \tag{5}$$

To take account of individual differences in HRV, the RMSSD values were normalized by dividing the RMSSD values by the mean RR-intervals (Grant et al., 2013) of the baseline routes (5). Only these nRMSSD values will be used for the following analyses. For further details on how the nRMSSD calculations were performed for later analyses please view Appendix A4.

3.1.2 Electrodermal data

The tonic and phasic component of the EDA signal are examined in the following. The slow tonic shifts are analysed with the SCL and the phasic changes are examined with the SCRs. Six participants had to be excluded after visual inspection of the EDA signal due to an extremely noisy signal. In two further cases, participants had to be excluded for the analysis of the first event but not for the second event for the same reason. Therefore, 29 out of 37 participants' data could be used for the abrupt braking event and 31 out of 37 participants could be considered for the unexpected detour (Table 1).

SCL – tonic component

The SCL was pre-processed by the algorithm provided by movisens and then further normalized manually. The algorithm of movisens calculated the SCL (in µS) on the chosen output interval of one second and the signal was cleaned with a low-pass filter (cut-off: 0.1 Hz).
Braithwaite et al. (2013) point out that a general climbing drift may appear in the SCL data which may be a problem for long experiments. However, these trends can be artefacts as well as real tonic shifts. Visual inspection of the participants' SCL signals indicated no to a minimal drift in the data. Since only short successive time intervals during the two events are compared to each other, the effect of a trend in the signal is neglectable. However, the following analysis refrains from comparing the SCL of the two events with each other because of a temporal difference between the events, that might be more strongly influenced by a drift in the SCL. If a trend was present, the second event would have a higher SCL than the first event.
To take account of individual differences in the SCL signal, a range correction was performed as proposed by Lykken & Venables (1971). The range of the tonic response of an individual was calculated during the baseline routes and the SCL was then divided by this range. This normalization enabled group comparisons. Since the SCL of both hands was measured, the mean normalized SCL was calculated from both hands. These values were used for the following analyses.

SCRs – phasic component

The algorithm of movisens detected and counted the number of SCRs and measured the amplitudes and rise times of the SCRs for the data recorded by the sensors of each hand. Additionally, the movement acceleration of each hand was measured.
Further processing of the data was performed manually. Thammasan et al. (2019) found in their study that touching pinkie and thumb induced movement artefacts, although that was one of the conditions in which the accelerator of the movisens EDA-sensor which was placed on the wrist was moved least. Similar movements were performed by the participants in this study when they manipulated the smartphone. To diminish the possibility of hand movements distorting the signal, SCRs that coincided with high movement acceleration were discarded. To further reduce the impact of artefacts due to movement, the SCRs of both hands were analysed together. Similarities in fast EDA fluctuations in both hands indicate a reliable signal while dissimilarities indicate that one of the sensors suffers from an artefact, according to Thammasan et al. (2019). The detailed procedure of data processing is described below.

For each participant:

1) For each hand, the mean movement acceleration was calculated for the time interval (see section 3.1.6 Time intervals for analysis) from the start time of the before-event time interval of the first event to the end of the after-event time interval of the second event. The mean was used as a cut-off value. SCRs that coincided temporally with movement acceleration values equal to or above the mean were discarded for the respective hand. If the mean movement acceleration was below 0.05g, 0.05g was used as a cut-off since the participant's movement was little. The value of 0.05g was determined by visual inspection of the participants' signals.
2) If the remaining SCRs of both hands occurred at the same time, the mean amplitude and the mean rise time of the SCRs of both hands was calculated. If the SCRs occurred at the same time but one of the SCRs was discarded due to surpassing the movement acceleration cut-off value, the amplitude and rise time of the other hand was taken. The hand that did not move therefore

represents the undistorted electrodermal reaction. If the remaining SCRs occurred with one second deviation between the hands, the mean amplitude and rise times of the two hands were calculated. A deviation of the SCRs of one second between the two hands is considered to be a reaction to the same event but the time lag is due to technical inaccuracy of the EDA sensors. The time lag of one second was decided on after visual inspection of the data.

The number, mean amplitude, maximum amplitude, mean rise time and maximum rise time of the SCRs are parameters which were considered. For the comparison of the event types in the following analysis, the parameter number of SCRs was normalized with the duration of the event for every participant to take account of time differences in the duration of the two event types.

3.1.3 Subjective rating

The 10-point Likert-scaled questions did not need any preparation before the quantitative analyses. Seven participants had to be completely excluded from the analyses of the subjective ratings. In six cases, the recording of the subjective ratings stopped at the time points of the events and in one case the rating remained at one level during the whole recording. Three further participants were only considered during the first event and excluded for the second event due to missing recordings. Therefore, 30 participants were examined for the first event and 27 participants were examined for the second experiment (Table 1).

3.1.4 Post-questionnaire

All 37 participants completed the post-questionnaire (Table 1). The 5-point Likert scaled questions in the post-questionnaire did not need any further preparation before the quantitative analyses.

The statements to the open-ended questions were categorized into a two-level category system. The statements were categorized on a first level into the categories “Display of route”, “Display of time” and “Display of further information” and on a second level in to the categories “on the tablet”, “Duration of event”, “Duration of the rest of the trip until arrival”, “More context” and “Need for action or not”. Residual categories contain statements that could not be fully assigned to the level-1 or level-2 categories.

3.1.5 Overview of participant constellations for analyses

Table 1: Overview of participant constellations

	Participants	**Total amount**	**withInfo group**	**noInfo group**
ECG	P03, P04, P05, P07, P08, P09, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20, P21, P22, P23, P24, P39, P40	23	11	12
EDA	P03, P04, P06[1], P07, P08, P09, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P21, P22, P23, P24, P26, P27, P29, P30, P31[1], P32, P33, P34, P38, P39, P40	31 (2[1])	16 (1[1])	15 (1[1])
Subjective ratings	P04, P05, P06[2], P09, P10, P13, P14, P15, P16, P17, P18, P19, P21, P22, P23, P24, P25, P26, P27, P28[2], P30, P31, P32, P33, P34, P35, P36, P38, P39[2], P40	30 (3[2])	16 (3[2])	14 (0[2])
Post-questionnaire	P03, P04, P05, P06, P07, P08, P09, P10, P11, P12, P13, P14, P15, P16, P17, P18, P19, P20, P21, P22, P23, P24, P25, P26, P27, P28, P29, P30, P31, P32, P33, P34, P35, P36, P38 P39, P40	37	19	18

[1] participants were excluded for the first event

[2] participants were excluded for the second event

The table (Table 1) displays which participants were considered for the data preparation and the later analysis of the ECG data, EDA data, subjective ratings and post-questionnaire. The total amount of participants as well as the number of participants assigned to the withInfo group (participants that received an error notice followed by content information) and noInfo group (participants that only received an error notice without content information) for the respective analyses is presented in the right three columns. Because of the different participants considered per analysis, the number of participants that were assigned to the withInfo group and to the noInfo group also differ. However, the group sizes do not differ greatly (Table 1). By allowing different participant constellations per analysis, the maximal amount of data could be preserved to perform robust analyses.

3.1.6 Time intervals for analysis

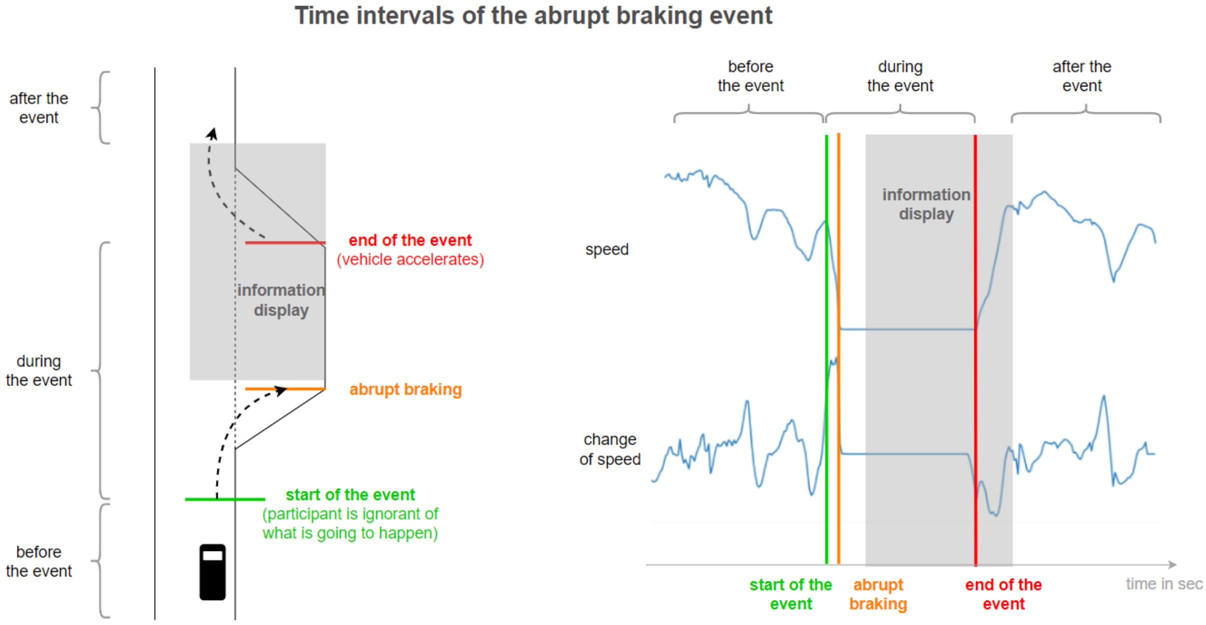


Figure 4: Time intervals of the abrupt braking event

The two figures depict the time intervals of the first event. The figure on the left depicts the scenario of the abrupt braking with the time intervals while the figure on the right displays the time interals according to the recorded data of speed and the calculated change of speed.

For each participant and each event, time intervals are defined to perform analyses of the physiological data and subjective ratings during the events and during information display. The mean values of the physiological parameters and the mean values and the range of subjective ratings of every time interval were then calculated to perform comparative analyses in the next step.

To identify the impact of the events on the participants, the time intervals before the event, during the event and after the event were examined for each of the two events on trip. The GPS coordinates shortly before the vehicle turned into the parking lot for the abrupt braking event were used to define the start of the first event. The start time of the first event was therefore the time point at which the participants reached the GPS coordinates. The end of the first event was the time point at which the vehicle started to accelerate again. Next, the duration of the event for each participant was calculated. To receive the start time of the before-event time interval, the duration of the event was subtracted from the start time of the event interval. The end point of the before-event interval was therefore the start of the event interval. The time interval after the event was equally long as that of during the event. The start time of the after-event interval was defined as the time point at which the information about the event on the participant's display was switched off. For most participants, this occurred a few seconds after the vehicle accelerated again (Figure 4, left). If the information display ended before the vehicle accelerated, the end of the event was taken as the start time for the after-event interval. The end time point of the after-event time interval was calculated by adding the duration of the event to the start point of the after-event time interval. To ensure that the abrupt braking event was contained in the during-event time interval, the speed and the change of speed of the vehicle were examined. The abrupt braking was indicated by a decrease in speed and by the maximal change in speed (Figure 4, right).

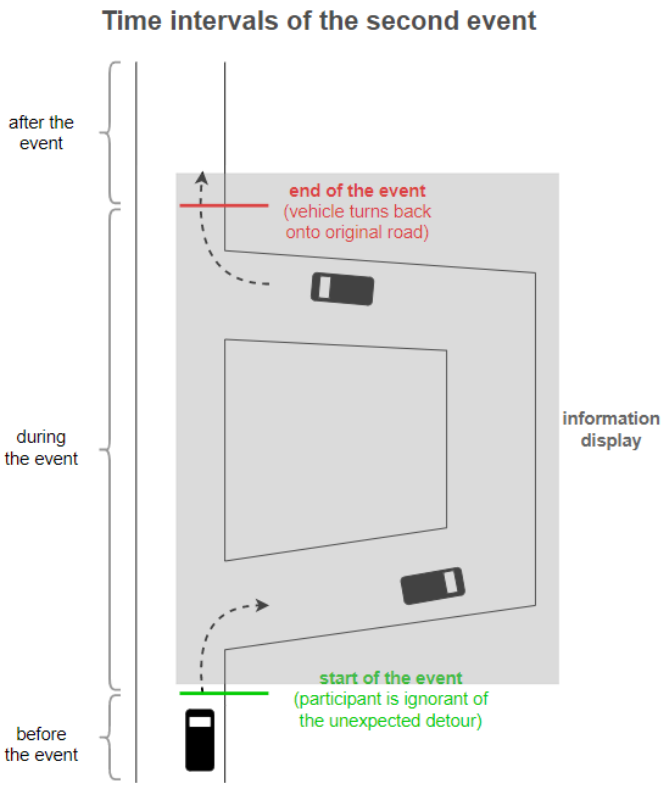


Figure 5: Time intervals of the unexpected detour
The figure depicts the scenario of the unexpected detour with the time intervals.

The time intervals for the second event – the unexpected detour – was calculated similarly to the first event: The GPS coordinates were used to define the start and the end of the detour. If the information on the event was displayed earlier on the screen than the calculated time point according to the GPS data, then the time of the information display was used as the start time of the event. After that, the duration of the event was calculated and was then subtracted from the start time point of the event to receive the start time of the before-event time interval. In contrast to the first event, not the end of the information display but the end of the during-event interval was chosen as a start point of the after-event time interval (Figure 5). The reason was that the participants might have realised that they were on the original route again, even though the information was still displayed. The end of the after-event interval was defined by adding the duration of the second event to the end time point of the during-event interval.

The time intervals during the event while at the same time information was displayed were considered to perform group comparisons on the type of information that was displayed during the events. The start of the information display was at the same time or slightly after the beginning of an event. The end of the time interval was set equal to the end of the event. The information display in some cases exceeded the end of the events but only the time intervals in which both the event and the information display occurred at the same time are relevant for the following analyses (Figure 4, Figure 5, grey background).

3.2 Data analysis and statistics

3.2.1 Linear mixed effects analysis

Linear mixed effects analyses were performed to systematically analyse the physiological reactions and subjective ratings of the participants during the two events. R and the package "nlme" (Pinheiro et al., 2019) with the function lme were used for the analysis. The results are going to be reported separately for the event of abrupt braking and the event of the unexpected detour.

$$value \sim interval + group + \ interval|P \tag{6}$$

The time interval (before event|during event|after event) and the group (withInfo: passenger received error notice and content information| noInfo: passenger received error notice but no content information) were entered as fixed effects. The time interval was seen as a categorical fixed effect to compare within-subject data (Magezi, 2015). The participants (P) were examined as random effects with random slopes for the effect time interval. The reason to take account of the time interval as random slopes is the assumption that the responses differ between participants for the time intervals. Some might be more effected by the events than others. Since an effect of the subdivision of the participants into groups is only expected in the time interval during the event and not in the other time intervals, the assumption that responses of the participants differ between groups is neglected at this point. For this reason and because of too little data, the effect group will therefore not be modelled as random slopes. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity. Since linear models are said to be robust to the violation of the normality assumption when it comes to hypothesis testing even for relatively small sample sizes (Gelman & Hill, 2007; Knief & Forstmeier, 2021), the normality assumption will not be further considered in the following analyses. P-values were obtained by likelihood ratio tests with an analysis of variance (ANOVA) of the full model with the effect in question (6) against the model without the effect in question.

3.2.2 Mann-Whitney U-test

The non-parametric Mann-Whitney U-test was used for group comparisons to analyse the impact of different information displays on the participants. The Mann-Whitney U-test was used for the physiological data, the subjective ratings and the quantitative questions in the post-questionnaire. The test was calculated in Python. For each participant, the mean value of the time interval from the start of the information display to the end of the information display or, if the information display exceeded the time of the event, the end of the event was considered.

3.2.3 Wilcoxon sign rank test

The Wilcoxon sign rank test was used as a non-parametric measure for paired sample tests and was executed in Python. The measure was used to compare the physiological reaction and subjective evaluation of the participants between the two events. For each participant, the mean value of the time interval from the start of the information display to the end of the information display or the end of the event was considered. The two events were compared for the withInfo group and the noInfo group separately. Some participants had to be excluded for one of the events during the analysis of the EDA data and the subjective ratings (Table 1). Since group sizes have to be equal to be able to calculate the Wilcoxon sign rank test, the participants that were excluded for one event were excluded from both events for this analysis. This way a comparison of the event types was still possible.

3.2.4 Spearman's correlation

Spearman's rank correlation calculated in Python and was used as a non-parametric measure to investigate the correlation between the participants' heart rates and the driving speed of the vehicle, which was found in a study by Yanagida et al. (2016). A Spearman's rank correlation coefficient was calculated for each participant during the baseline routes, during the first event and during the second event. According to Cohen (1988), a correlation coefficient between 0.1 and 0.3 is thought to represent a weak association between 0.3 and 0.5 a medium correlation and equal or above 0.5 a high correlation. A positive correlation coefficient represents a positive linear correlation while a negative coefficient indicates a linear inverse correlation between two variables.

3.2.5 Frequency analysis

A frequency analysis was performed on the categorized statements of the post-questionnaire to examine what information was retrospectively most desired by the participants during the events and the whole drive. The qualitative data was analysed in accordance with the qualitative content analysis procedure by Mayring & Fenzl (2019). Instead of counting the number of specific statements, the number of participants that gave specific statements was analysed. This way, a frequency bias due to some participants answering the questions more extensively than others is prevented.

3.2.6 Overview of hypotheses with respective statistical analyses

Table 2: Overview of hypotheses with respective statistical analyses

Hypotheses	**Statistics**	**Parameters**
H1: An unexpected event whilst travelling in an autonomously driving vehicle has an impact on the passenger's physiology and subjective evaluation of the situation	Linear mixed effects analyses - in case of significance: post-hoc Tukey test	X^2-value, p-value p-value
H1.1: Correlation between driving speed and zHR	Spearman's rank correlation	Spearman's rho (ρ)
H2: Participants differ in their physiological reaction and subjective evaluation when an error message together with content information is displayed to when only an error message without content information is displayed during the events	Two-sided Mann-Whitney U-test	U-value, p-value
H3: The physiological reaction and subjective evaluation to the information display during an event is the same for different event types	Two-sided Wilcoxon sing rank test	W-value, p-value
H4: Retrospectively, the participants preferred the display of content information over merely an error message to help them understand the actions of the vehicle, to feel less insecure during an event and to reduce the wish for more information	Quantitative data: one-sided Mann-Whitney U-test Qualitative data: frequency analysis	U-value, p-value Number of participants that gave a specific statement

To analyse the impact of the events on the participants' physiology and ratings, linear mixed effects analyses were performed (Table 2, **H1**). To further examine whether physiological changes are due to the occurred events or due to a correlation between a person's physiology and the driving speed, the correlation between the zHR and the driving speed during the events and during the baseline routes was calculated with Spearman's rank correlation (Table 2, **H1.1**). A two-sided Mann-Whitney U-test was used to analyse differences between different types of information displays (Table 2, **H2**). To compare event types, the two-sided Wilcoxon sign rank test was employed (Table 2, **H3**). The subjective evaluations of the information display in the post-questionnaire were analysed with a one-sided Mann-Whitney U-test in the case of the quantitative data and a frequency analysis was performed on the categorized qualitative statements (Table 2, **H4**).

4 Results

In the following, the results of the analyses are reported for each of the four hypotheses in separate subsections. For each hypothesis, an overview of the results is given via descriptive analyses before the results of the inferential statistics are presented. The results of the inferential statistics will be reported separately for the cardiological data from the ECG recordings, the electrodermal activity from the EDA sensors and the subjective ratings. Additionally, the results will be reported separately for the first and second event (abrupt braking and unexpected detour). Finally, an overview of the results according to the verification of the hypotheses is given.

4.1 Hypothesis 1 – the effect of an event

Hypothesis 1: An unexpected event whilst travelling in an autonomously driving vehicle has an impact on the passenger's physiology and subjective evaluation of the situation.

Descriptive statistics

Table 3: Descriptive statistics during the three time intervals of each event

	First event						**Second event**					
	before		**during**		**after**		**before**		**during**		**after**	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
ECG												
zHR	0.05	1.06	-1.03	1.04	-0.56	1.11	-0.53	1.06	-0.42	1.54	-0.07	1.36
nRMSSD	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02	0.04	0.02
EDA												
SCL	9.38	12.52	9.80	13.79	9.28	13.81	8.75	13.69	9.01	13.81	9.16	14.42
No. SCRs *	3.24	3.41	3.14	3.26	2.17	2.04	6.81	6.39	6.65	5.57	6.74	5.81
Mean Amp	0.35	0.29	0.41	0.29	0.42	0.33	0.47	0.29	0.46	0.27	0.43	0.26
Max Amp	0.59	0.49	0.60	0.45	0.59	0.47	0.83	0.49	0.78	0.45	0.76	0.48
Mean Rise time	0.95	0.67	1.17	0.66	1.17	0.73	1.32	0.61	1.39	0.57	1.29	0.65
Max Rise time	1.18	0.84	1.39	0.78	1.33	0.82	1.72	0.75	1.74	0.73	1.66	0.82
Subj. ratings												
Mean	9.04	1.15	7.95	1.81	8.64	1.27	9.01	0.92	8.64	1.59	9.09	0.90
Range	0.23	0.50	2.13	2.65	1.23	2.01	0.52	1.23	1.00	1.94	0.63	1.42

*** The number of SCRs are not normalized with the duration of the respective event per participant in this table because values would be smaller < 0.00. One therefore has to keep in mind that since the second event was about twice as long as the first event, the number of SCRs also has to be about double if no stronger reaction to one of the two events was present.

On examining the cardiological data, for the first event – the abrupt braking event – one can observe a decrease in the mean zHR during the event compared to before the event. After the event, the mean zHR is higher than during the event but lower than before the event (Table 3). These results indicate a decrease in HR when the abrupt braking event occured. For the second event – the unexpected detour – an increase in mean zHRs can be seen from before to after the event. The mean zHR values increased but remained comparable between before and during the second event (Table 3). According to the mean zHR values, the HR increased when the second event occurred. The mean and SD nRMSSD values during all time intervals of the first and second event remained stable (Table 3), indicating that the events did not influence the HRV.

The means of the parameters that indicate electrodermal activity give a mixed picture of the data. For the first event, the mean SCL was higher during the event than before and after the event, the mean number of SCRs was highest before the event and the mean values of the mean and maximum amplitudes and rise times were slightly higher during and after the event than before the event (Table 3). The tonic component of the EDA signal indicates increased perspiration during the first event, while the phasic component does not show major differences between the mean values of the time intervals. For the second event, the mean SCL increased slightly from before to during and from during to after the event, the mean number of SCRs were lowest during the event but comparable to the other two time intervals and the mean values for the amplitudes and rise times were similar during all three time intervals (Table 3). According to the mean values during the time intervals of the second event, EDA remained comparable.

Table 4: Descriptive statistics of the subjective ratings during the abrupt braking event

	Subj. ratings – first event											
	withInfo group						**noInfo group**					
	before		**during**		**after**		**before**		**during**		**after**	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Mean	8.90	1.31	8.17	1.49	8.67	1.38	9.21	0.94	7.70	2.09	8.58	1.12
Range	0.19	0.39	1.81	2.53	0.88	1.41	0.29	0.59	2.50	2.75	1.64	2.47

Table 5: Descriptive statistics of the subjective ratings during the unexpected detour

	Subj. ratings – second event											
	withInfo group						**noInfo group**					
	before		**during**		**after**		**before**		**during**		**after**	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Mean	9.23	0.80	9.08	0.91	9.20	0.78	8.81	0.99	8.23	1.95	8.99	0.99
Range	0.08	0.27	0.69	1.43	0.15	0.53	0.93	1.58	1.29	2.28	1.07	1.79

In the case of the first and the second event, the mean subjective ratings were lowest and the mean range of the ratings was highest during the events. These findings are more pronounced in the case of the first event compared to the second event (Table 3). These findings are also present when one examines the subjective ratings for the two participant groups separately (Table 4, Table 5). For the noInfo group, the decrease in mean ratings and the increase in the mean range of the ratings is more pronounced than for the withInfo group (Table 4, Table 5). In the case of the second event, the mean ratings are in general lower and the range of ratings are higher in the noInfo group than in the withInfo group.

Inferential statistics

ECG

Event 1 – abrupt braking

zHR

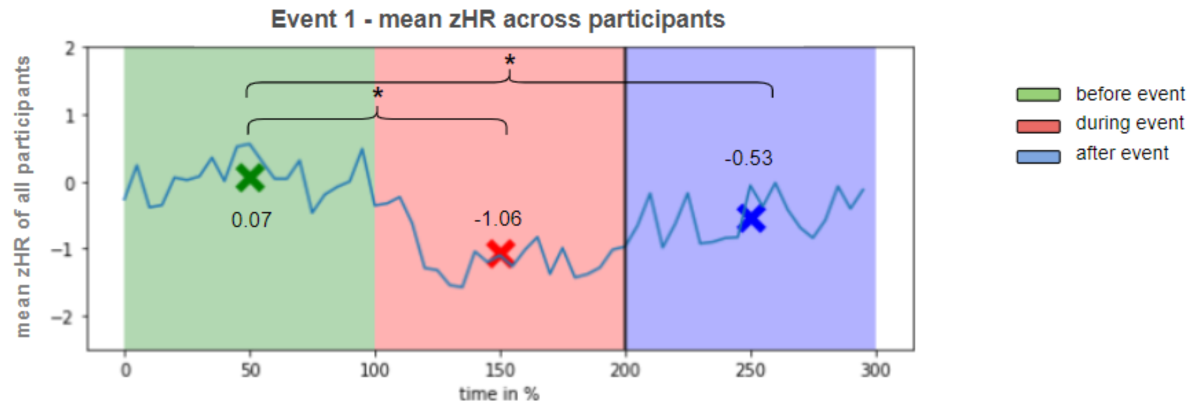


Figure 6: Mean zHR across participants for the abrupt braking event
The mean zHR across participants (n=23) for the first event is depicted (blue line). The duration of the event differed slightly between participants, so the average zHRs across participants were calculated for time in percent (5% time sections). The time intervals before the event (green), during the event (red) and after the event (blue) together span the time in percent from 0% to 300%. The mean z-scores over all participants in the respective time interval is marked with a cross in the matching colour of the time interval. Note that the mean values in this figure differ slightly from the mean values which were reported in the descriptive section above (Table 3) because zHR values had to be resampled for this figure. The black vertical line between the time intervals during and after the event indicates a discontinuity in the data between the two intervals. The asterisk denotes a significant difference between the time intervals ($p<0.05$).

In the linear mixed effect analysis, the fixed effect interval was examined to evaluate whether the heart rates of the passengers in the vehicle differed between the time intervals. The zHR differed significantly between the time intervals ($X^2(7)=16.3117$, $p=0.0224$). A Tukey post-hoc test was performed to analyse which time intervals differed significantly. The test revealed that the zHR was significantly lower ($p<0.001$) during the event than before the event. 20 out of 23 participants had lower heart rates during compared to before the event. No significant difference in zHR could be detected between the two groups ($X^2(1)=0.3854$, $p=0.5347$). While the groups of participants did not differ, the time intervals during the first event did with respect to the zHR.

The results of the linear mixed model analysis are supported by the descriptive analyses (Table 3) and visualised by Figure 6: The zHR is highest before the abrupt braking, decreases abruptly at the beginning of the event interval, i.e. during or shortly after the abrupt braking, remains low during the event time interval and then slightly increases again after the event (but not significantly). In sum, both the descriptive and inferential statistics showed that the participants' HRs was lower during the abrupt braking event compared to before and after the event.

nRMSSD

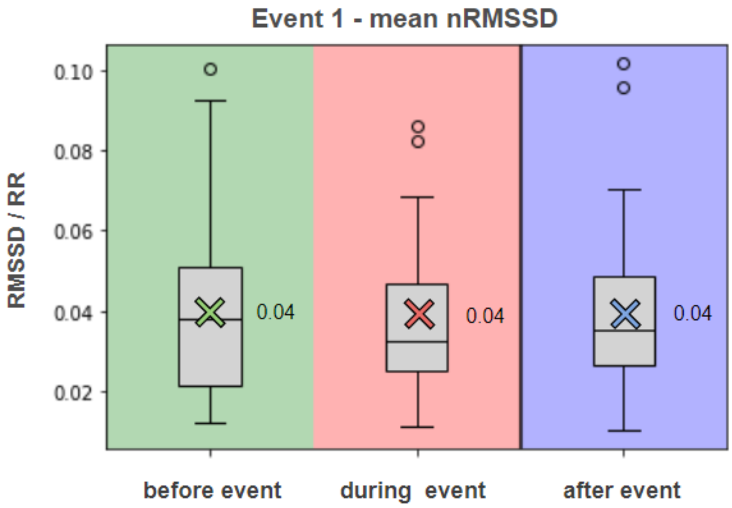


Figure 7: Mean nRMSSDs for the abrupt braking event
The mean nRMSSD values of all 30 second windows for each participant (n=23) per time interval are contained in the boxplots for the first event of abrupt braking. The time intervals are before the event (green), during the event of abrupt braking (red) and after the event (blue). The black line indicates the temporal discontinuity between the time intervals. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

The linear mixed effects analysis of the effect interval revealed no significant difference in nRMSSD between the time intervals before the event, during the event and after the event. This result is supported by the descriptive statistics, since the mean nRMSSD values did not change between the time intervals (Table 3). The boxplots in Figure 7 indicate a slightly more compact boxplot during the event time interval than before and after the event. Although no significant difference between the time intervals could be found, the slight decrease in HRV (Figure 7) during the event points towards a slight change in physiological arousal. The analysis of the fixed effect group did not reveal a significant difference in nRMSSD values between the two participant groups.

Event 2 – unexpected detour

Heart rate

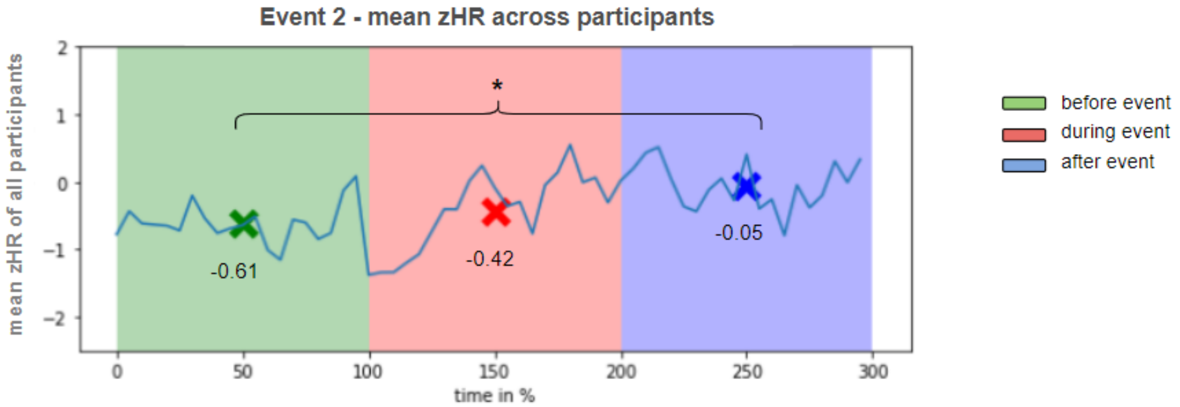


Figure 8: Mean zHR across participants for the unexpected detour event

The mean zHRs across participants (n=23) for the second event is depicted (blue line). The duration of the event differed slightly between participants, so the average zHRs across participants were calculated for time in percent (5% time sections). The time intervals before the event (green), during the event (red) and after the event (blue) together span the time in percent from 0% to 300%. The mean z-scores over all participants in the respective time interval is marked with a cross in the matching colour of the time interval. Note that the mean values in this figure differ slightly from the mean values which were reported in the descriptive section above (Table 3) because zHR values had to be resampled for this figure. The asterisk denotes a significant difference between the time intervals ($p<0.05$).

The fixed effect time interval was examined to compare the heart rates of the participants before, during and after the detour. The heart rates differed significantly between the time intervals ($X^2(7)=15.848$, $p=0.0265$). The Tukey post-hoc test revealed no significant difference in the zHR between before and during ($p=0.807$) nor between during and after the second event ($p=0.138$). The zHR was significantly higher after the event compared to before the event ($p=0.017$). In the course of the examined time intervals, the zHR increased (Figure 8). The fixed effect group was analysed and did not indicate a difference in heart rates between the groups ($X^2(1)=0.1802$, $p=0.6712$).

The results of the linear mixed effects analysis of the participants' zHR correspond to the findings of the descriptive statistics (Table 3) and the visualisation in Figure 8: The zHR increases during the event such that the zHR is significantly higher after the event than before the event. The significant difference between the time intervals before and after the event indicate a change in the participants' HRs during the unexpected detour.

Heart rate variability – nRMSSD

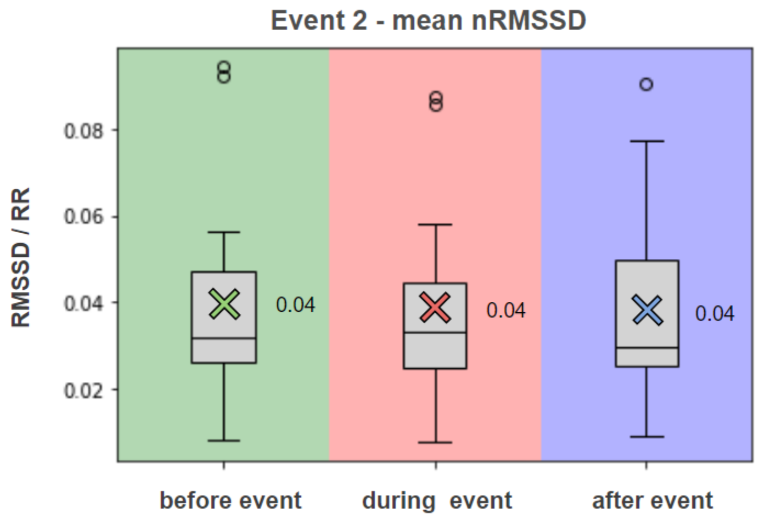


Figure 9: Mean nRMSSDs for the unexpected detour event
The mean nRMSSD values of all 30 second windows for each participant (n=23) per time interval are contained in the boxplots for the second event of the unexpected detour. The time intervals are before the event (green), during the event of abrupt braking (red) and after the event (blues). Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

The nRMSSD values were also examined for the second event. No significant difference between the time intervals, nor between the groups could be detected when performing a linear mixed effects analysis. The mean nRMSSD values of all participants were comparable during the three time intervals of the second event (Table 3, Figure 9). The participants’ HRV did not change during the second event.

Correlation between heart rate and speed

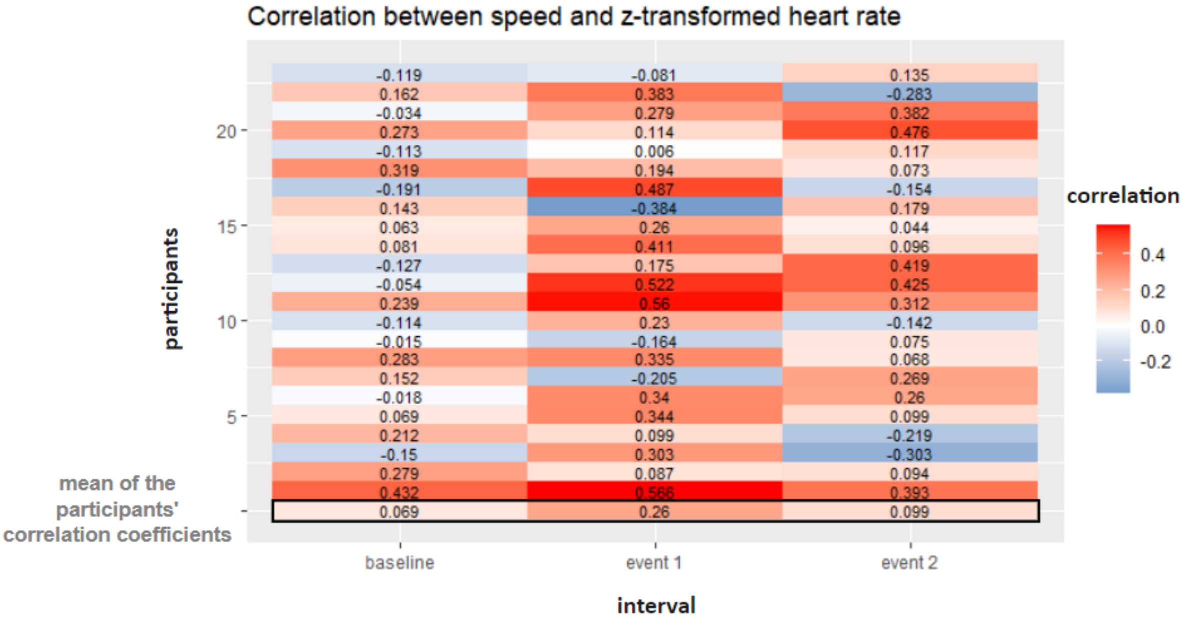


Figure 10: Correlation between driving speed and zHR
This correlation matrix depicts the Spearman's rank correlation between the driving speed and the z-scores of the heart rate. The correlation coefficients for the baseline routes, the first and the second event were calculated for each participant (n=23). The correlations in the framed row represent the mean correlations of all participants per interval.

The results show a neglectable correlation between the driving speed and the heart rate during a normal driving situation as in in the baseline interval (ρ=0.069) (Figure 10). A weak positive correlation between speed and heart rate can be observed during the first event (ρ=0.26). Low speed is therefore weakly correlated with low HRs while high speed is weakly correlated with high heart rates. This weak correlation is a product of the occurred abrupt braking event since a correlation was not present during the baseline section. The unexpected detour showed a minimal correlation (ρ=0.099), which does not even represent a weak correlation according to Cohen (1988) and can therefore be neglected. The results exclude the possibility that the above reported results are a mere product of a correlation between driving speed and HR.

EDA

Event 1 – abrupt braking

SCL

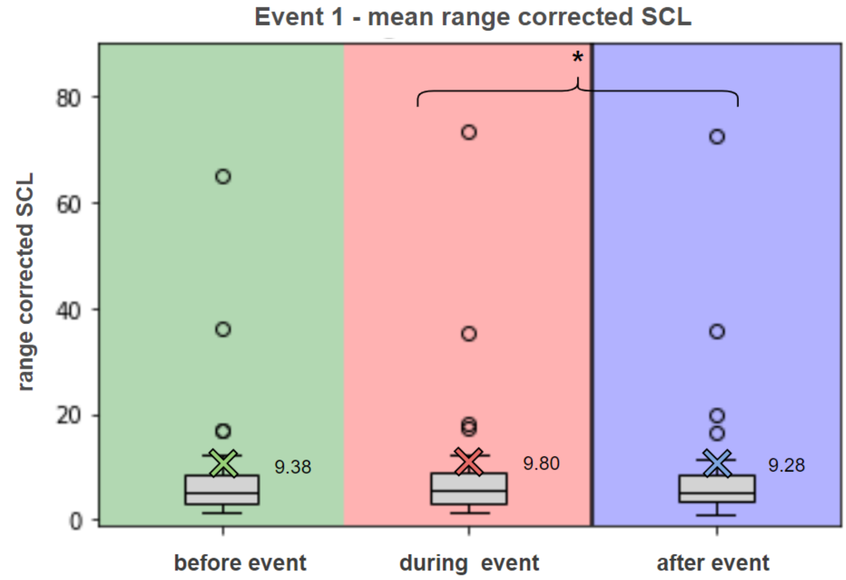


Figure 11: Mean range corrected SCL for the abrupt braking event

The intra-individual range corrected mean SCLs of the participants (n=31) are depicted for the time intervals before (green), during (red) and after (blue) the first event. The vertical black line indicates a temporal discontinuity between time intervals. The asterisk indicates a significant difference between the time intervals ($p<0.05$). Boxplot parameters: box = middle 50%, line = median, cross = mean length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

The tonic activity between the time intervals was compared with the help of the SCL measure. Results from the linear mixed effects analysis revealed a significant difference between the time intervals ($X^2(7)=45.8121$, $p<0.0001$). While the time intervals before the event and during the event and before the event and after the event did not differ significantly according to Tukey post-hoc test ($p_{before\ vs.\ during}=0.3381$, $p_{before\ vs\ after}=0.9570$), the time intervals during and after the event differed significantly ($p=0.0103$). During the first event, the SCL was significantly higher than after the event (Figure 11). These findings are supported by the descriptive statistics with mean values $M_{before}=9.38$, $M_{during}=9.80$ and $M_{after}=9.28$ (Table 3). No difference between the groups of participants could be detected ($X^2(1)=0.0156$, $p=0.9006$). Arousal seems to have been higher during than after the first event.

Number of SCRs

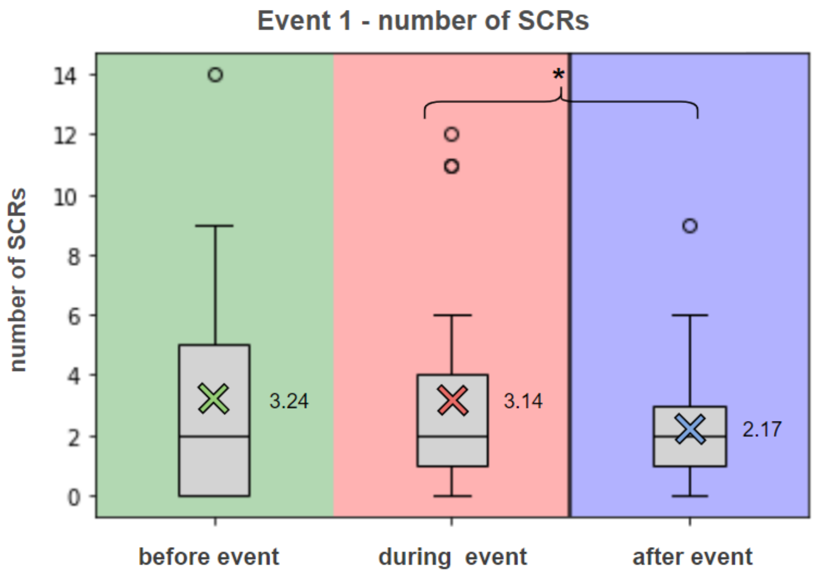


Figure 12: Number of SCRs for the abrupt braking event

The Number of SCRs of each participant (n=31) is depicted for the time intervals before (green), during (red) and after (blue) the first event. The vertical black line indicates a temporal discontinuity between time intervals. The asterisk indicates a significant difference between the time intervals ($p<0.05$). Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

A linear mixed effects analysis was performed to examine whether there is a difference in the number of SCRs between before, during and after an event. A significant difference between the time intervals could be observed for the first event ($X^2(7)=31.041$, $p=0.0001$). The Tukey post- hoc test revealed no significant difference between before and during the event ($p=0.6124$), but a significant difference between during and after the event ($p=0.0251$) (Figure 12). No significant difference could be perceived when comparing the number of SCRs before and after the event ($p=0.1304$) (Figure 12). The groups of participants did not differ in the number of SCRs ($X^2(1)=0.9789$, $p=0.3225$). Electrodermal reaction seems to have been higher during the first event than after the event.

SCR Amplitude

For the time intervals before, during and after the event, liner mixed effects analyses with the mean amplitude as well as the maximum amplitude were calculated. No significant difference between the time intervals nor between the groups could be found in either case. The abrupt braking seems to have had no impact on the amplitude of the SCRs.

SCR Rise Time

The linear mixed effects analysis was performed with the mean rise time and the maximal rise time of the SCRs during the three time intervals. No significant difference between the time intervals nor between the groups could be observed.

Event 2 – unexpected detour

SCL

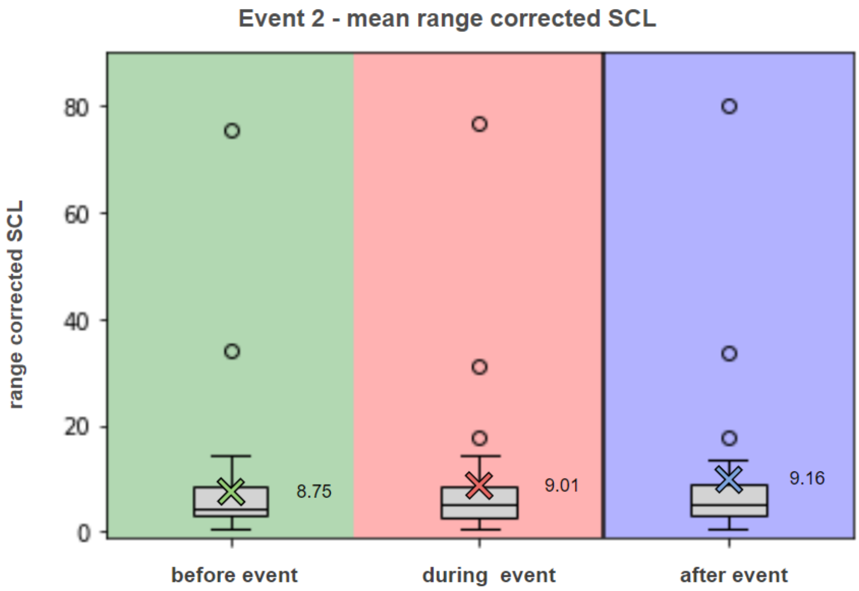


Figure 13: Mean range corrected SCL for the unexpected detour event

The intra-participant range corrected mean SCLs of the participants (n=31) are contained in the boxplots for the time intervals before, during and after the second event. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

While the linear mixed effects analysis indicated a significant difference between the time intervals ($X^2(7)=14.2806$, $p=0.0464$), the Tukey post-hoc test did not reveal significant differences between two specific time intervals. The discrepancy between the linear mixed effects analysis and the Tukey post-hoc test lies in the more conservative nature of the latter. The SCL of the groups did not differ ($X^2(1)=0.5979$, $p=0.4394$). There seem to be general differences between the SCL of the time intervals but no specific differences between two time intervals (Figure 13).

Number of SCRs

The linear mixed effects analyses of the number of SCRs of the time intervals before, during and after the second event showed no significant differences between the time intervals. Neither did the groups differ. The unexpected detour did not have a significant effect on the participants' number of SCRs.

SCR Amplitude

The linear mixed effects analyses of the mean rise amplitude and the maximum amplitude did not indicate significant differences between the time intervals, nor between the groups.

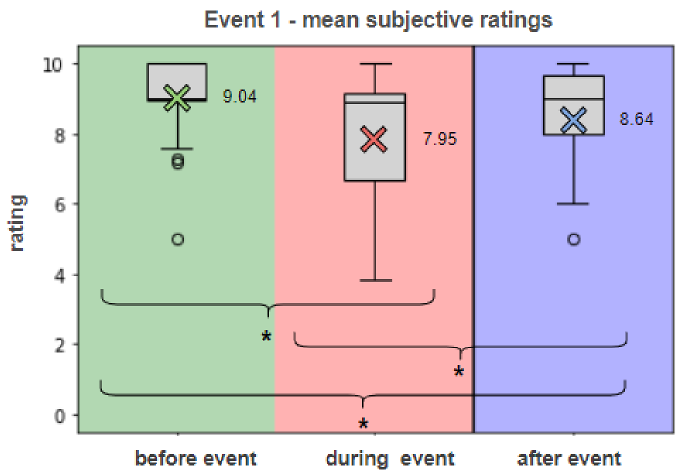
SCR Rise Time

Analysing the SCR rise times with linear mixed effects analyses, a significant difference between the time intervals could be observed on comparing the mean rise times ($X^2(7)=14.2075$, $p=0.0476$) and an almost significant difference when comparing the maximum rise times ($X^2(7)=13.898$, $p=0.053$). However, the post-hoc Tukey tests did not reveal a significant difference between two specific time intervals, which can be attributed to the more conservative nature of the Tukey test. No significant difference between the groups was present both in regards to the mean rise times and the maximum rise times. The unexpected detour seems to have led to differences between the time intervals in general with regards to the SCR rise times, but not between two specific intervals.

<u>Subjective rating</u>

Event 1 – abrupt braking

Mean rating



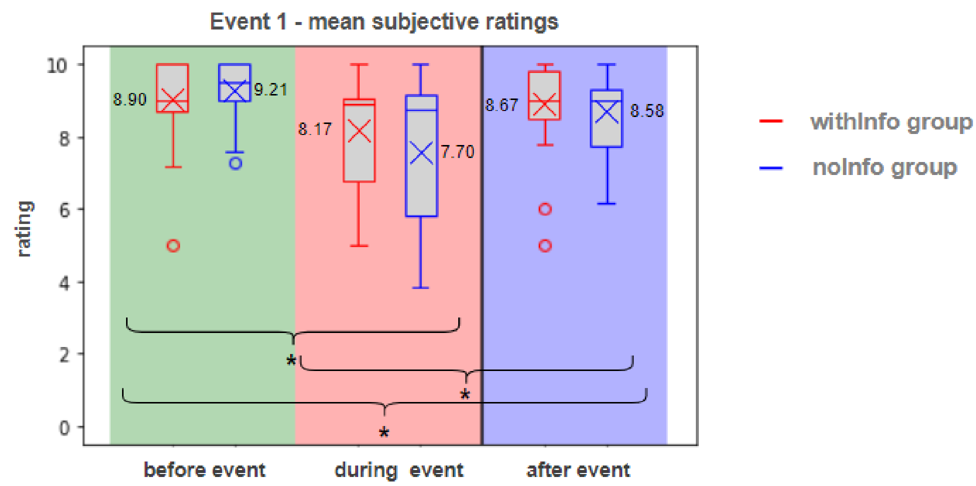
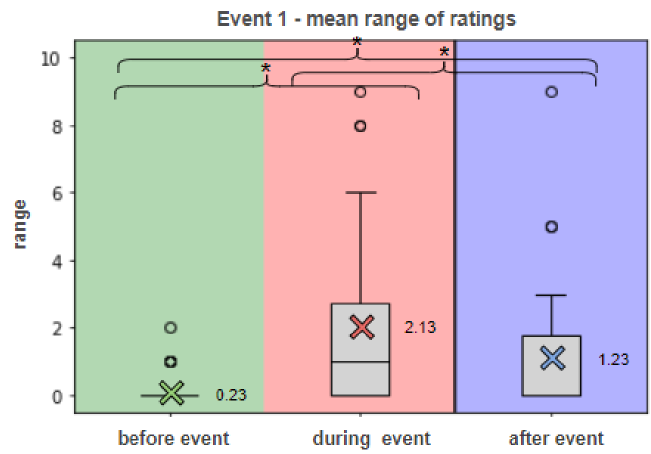


Figure 14: Mean subjective ratings for the abrupt braking event
Both figures show the mean ratings of the participants during the time intervals before the event (green), during the event (red) and after the event (blue). In the upper figure, the boxplots contain the participants of both groups (n=30) and in the second figure the participants are subdivided into their groups ($n_{withInfo}=16$, $n_{noInfo}=14$). The red boxplots contain the mean ratings of the withInfo group – the participants that received content information – and the blue boxplots contain the mean ratings of the

noInfo group – the participants that did not receive content information during the event. The black line indicates the temporal discontinuity between the time intervals. The asterisks indicate significant differences between the time intervals. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/- 1.5 with the last data point defining the end of the whisker, circle = outlier.

A linear mixed effects analysis was performed with the ratings, which the participants gave during the drive. The mean rating during the time intervals before, during and after the event were compared. The time intervals differed significantly ($X^2(7)=54.9866$, $p<0.0001$). The post-hoc Tukey test revealed significantly lower ratings of the participants during the first event compared to before ($p<0.0001$) and after ($p=0.0033$) the event. After the event, the ratings were significantly lower compared to before the event ($p=0.0066$). The mean ratings of all participants per time intervals underline the results of the linear mixed effect analysis ($M_{before}=9.04$, $M_{during}=7.95$, $M_{after}=8.64$) (Table 3, Figure 14, left). No significant difference could be observed between the two groups ($p=0.6288$). After the event, the lower ratings directly during the event did not fully return to the high ratings of before the event.

Range of ratings



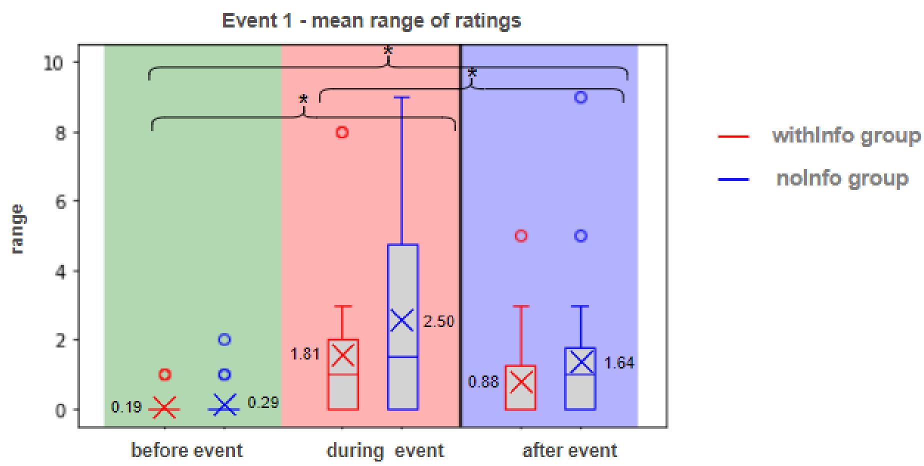


Figure 15: Mean range of subjective ratings for the abrupt braking event
The figure shows the range of the participants' subjective ratings during the time intervals before the event (green), during the event (red) and after the event (blue). In the upper figure the boxplots

contain the participants of both groups (n=30) and in the second figure the participants are subdivided into their groups ($n_{withInfo}$=16, n_{noInfo}=14). The red boxplots contain the range of ratings of the withInfo group – the participants that received content information – and the blue boxplots contain the range of ratings of the noInfo group – the participants that did not receive content information during the event. The black line indicates the temporal discontinuity between the time intervals. The asterisks indicate significant differences between the time intervals. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

Next to the mean rating, the intra-participant range of the ratings during the time intervals was analysed with a linear mixed effects analysis. The time intervals differed significantly ($X^2(7)$=106.181, $p<0.0001$). The Tukey test showed a significantly higher range of the ratings during the event than before ($p<0.0001$) or after the event ($p=0.0019$). After the event, the range was significantly larger than before the event ($p=0.0244$). The mean ratings of all participants per time intervals underline the results of the linear mixed effect analysis (M_{before}=0.23, M_{during}=2.13, M_{after}=1.23) (Table 3**,** Figure 15). The groups did not differ significantly ($p=0.3119$). The larger changes in the participants ratings during and after the event seem to be induced by the abrupt braking event, since almost no change in rating was present before the event (Figure 15).

Event 2 – unexpected detour

Mean rating

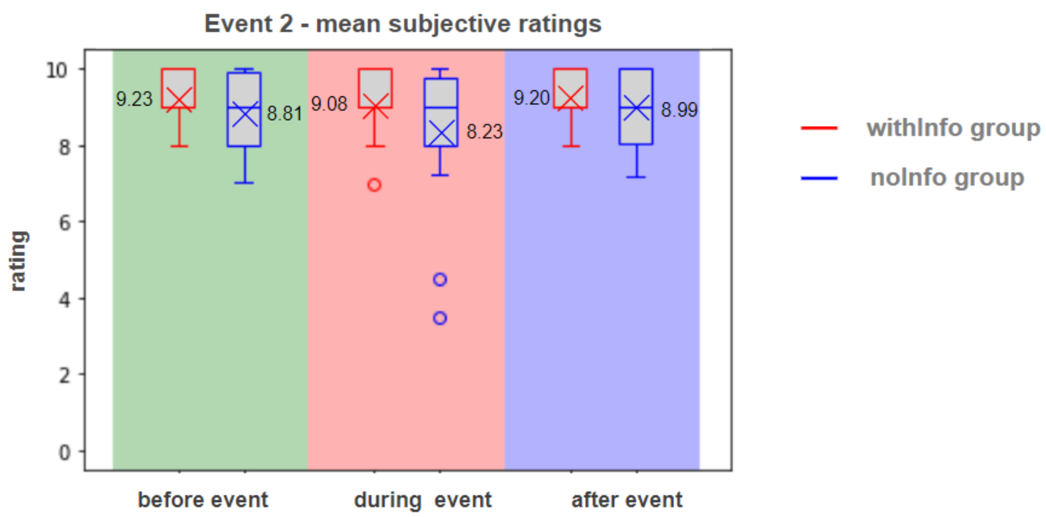


Figure 16: Mean subjective ratings for the unexpected detour event

The figure shows the mean ratings of the participants during the time intervals before the event (green), during the event (red) and after the event (blue). The red boxplots contain the mean ratings of the withInfo group ($n_{withInfo}$=13) – the participants that received content information – and the blue boxplots contain the mean ratings of the noInfo group (n_{noInfo}=14) – the participants that did not receive content information during the event. The black line indicates the temporal discontinuity between the time intervals. The asterisks indicate significant differences between the time intervals. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

The time intervals of the second event were compared by utilizing a mixed linear effects analysis. While the analysis indicated a significant difference between the time intervals ($X^2(7)=50.7342$, $p<0.0001$), the more conservative post-hoc Tukey test did not show significant differences between specific time intervals. A significant difference for the fixed effect group was found during the mixed linear effects analysis ($X^2(1)=0.5432$, $p=0.0461$) but no significant difference was revealed by the Tukey test ($p=0.431$). The Figure 16 and Table 5 show that the withInfo group – the group that received content information – has overall higher ratings than the noInfo group. General differences between the time intervals and between the groups seem to exist during the second event.

Range of ratings

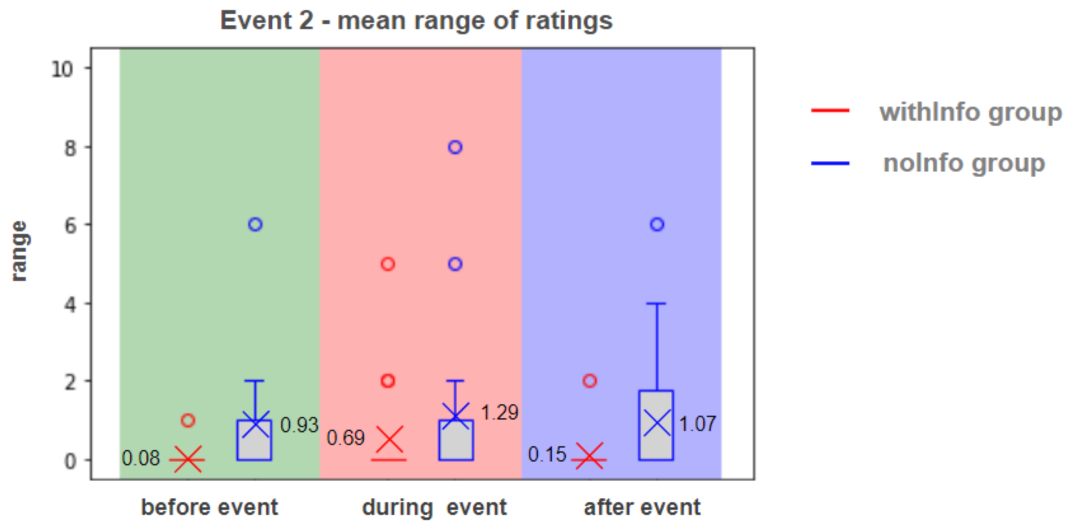


Figure 17: Mean range of subjective ratings for the unexpected detour event
The figure shows the range of the participants' subjective ratings during the time intervals before the second event (green), during the event (red) and after the event (blue). The red boxplots contain the range of ratings of the withInfo group ($n_{withInfo}=13$) – the participants that received content information – and the blue boxplots contain the range of ratings of noInfo group ($n_{noInfo}=14$) – the participants that did not receive content information during the event. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

As was the case when comparing the mean ratings between the time intervals, the linear mixed effects analysis of the range of ratings indicated a significant difference between the time intervals ($X^2(7)=34.1252$, $p<0.0001$). However, the more conservative Tukey test did not show specific significant differences between two time intervals. The groups differed significantly according to the linear analysis ($X^2(1)=9.9454$, $p=0.0016$) and also according to the post hoc Tukey test ($p=0.0005$). The noInfo group had a significantly higher range compared to the withInfo group (Table 5**,** Figure 17). The time intervals seem to differ in the range of ratings in general, whilst the withInfo group, which received content information, shows significantly lower ranges in rating than the group that did not receive content information.

The hypothesis that an unexpected event whilst travelling in an autonomously driving vehicle has an impact on the passenger's physiology could only be partly confirmed. The HR varied when events occurred but the HRV did not. On examining the EDA data, the first event led to an increased SCL and to a higher number of SCRs and therefore to more perspiration while the second event did not. The events had an impact on the subjective evaluations of the participants during the drive. Lower ratings could be observed during the abrupt braking event compared to before and after the event. During the unexpected detour, the participants that received content information showed overall higher subjective ratings than participants that did not receive this information.

4.2 Hypothesis 2 – type of information display

Hypothesis 2: Participants differ in their physiological reaction and subjective evaluation when an error message together with content information is displayed to when only an error message without content information is displayed during the events.

Descriptive Statistics

Table 6: Descriptive statistics for group and event comparisons

	First event				**Second event**			
	withInfo group		**noInfo group**		**withInfo group**		**noInfo group**	
	M	SD	M	SD	M	SD	M	SD
ECG								
zHR	-1.26	0.90	-1.00	1.16	-0.39	1.46	-0.36	1.63
nRMSSD	0.03	0.01	0.04	0.03	0.03	0.01	0.03	0.00
EDA								
SCL	11.68	18.42	7.66	5.05	11.90	18.92	6.84	4.55
No. SCRs *	3.07	3.49	2.79	2.96	6.20	5.22	4.29	4.08
Mean Amp	0.31	0.28	0.37	0.30	0.50	0.24	0.29	0.23
Max Amp	0.49	0.46	0.50	0.41	0.88	0.44	0.45	0.40
Mean Rise time	0.89	0.70	1.07	0.77	1.41	0.60	1.30	0.59
Max Rise time	1.12	0.90	1.27	0.90	1.65	0.68	1.57	0.78
Subj. ratings								
Mean	8.12	1.57	7.72	2.16	9.04	0.94	8.12	2.15
Range	1.85	2.74	2.36	2.74	0.69	1.43	0.86	1.36

***** The number of SCRs are not normalized with the duration of the respective event per participant in this table because values would be smaller < 0.00. One therefore has to keep in mind that since the second event was about twice as long as the first event, the number of SCRs also has to be about double if no stronger reaction to one of the two events was present.

A comparison of the mean HR and HRV values between the two groups of participants, which received different types of information, indicates comparable cardiological reactions in the case of both events.

With regards to the electrodermal data, slight differences in the mean values between the two groups for the abrupt braking event can be observed. The mean SCL of the withInfo group is higher than that of noInfo group. However, the withInfo group has a more than three times larger SD than the noInfo group, which makes the mean values hard to compare between the two groups. The mean number of SCRs is also higher for the withInfo group than for the noInfo group. Contrary to the mean SCL and mean number of SCRs values, the mean values of the parameters mean amplitude, maximum amplitude, mean rise time and maximum rise time are slightly higher for the noInfo group than for the withInfo group. The mean values of the electrodermal parameters for the first event do not clearly indicate whether one group perspired more during the abrupt braking event with different information display than the other. When information was displayed during the second event, the withInfo group had higher electrodermal activity than the noInfo group according to the mean values of the parameters. A comparison of the mean SCL values between the groups is influenced by a more than four times higher SD in the case of the withInfo group than the noInfo group, which makes a comparison difficult. The mean values of the EDA parameters indicate higher perspiration during content information display than during no content information display when the unexpected detour occurred.

According to the mean values of the subjective rating parameters, the withInfo group gave higher mean ratings and had lower range in the ratings than the noInfo group in both events.

Inferential statistics

To ensure that the groups do not differ significantly because of the participant constellation, the physiology (ECG and EDA) as well as the subjective ratings of the groups were compared during the baseline routes. An almost significant difference ($U=90.0$, $p=0.0517$) between the zHR of the groups could be found with the withInfo group having lower heart rates than the noInfo group. No difference in HRV could be detected. On examining the electrodermal activity of the participants, one could detect no difference between the groups when analysing the SCL and the five parameters of the SCRs. The mean subjective ratings and the ranges of the ratings also did not differ significantly between the groups. Since the groups did not differ during the baseline routes, significant results when analysing the times periods of the events are not due to general physiological difference between the groups due to the participant assignment to the groups. The following cardiological or electrodermal effects as well as the effects found in the subjective ratings can be attributed to the type of information display.

<u>ECG</u>

Event 1 – abrupt braking

Passengers that received content information did not differ in their zHR compared to passengers that did not receive content information. The nRMSSD values also did not differ between the groups.

Event 2 – unexpected detour

As in the case of the first event, passengers did not differ in their heart rate when one group received content information and the other group did not. The nRMSSD values were also comparable.

EDA

Event 1 – abrupt braking

The electrodermal activity of the two groups during the first event was compared. No differences between the groups could be found with respect to the tonic and phasic activity.

Event 2 – unexpected detour

During the second event, differences in the participants' electrodermal activity between the group that received content information and the group that did not were found.

Number of SCRs

The withInfo group and the noInfo group did not differ in terms of the number of SCR during the second event.

SCR Amplitude

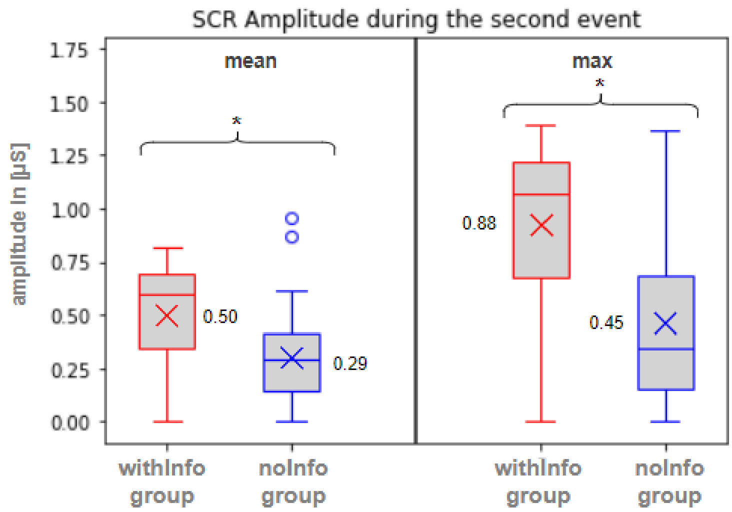


Figure 18: SCR amplitude during the second event

The participants' SCR mean (left) and maximum (right) amplitudes during the second event is depicted. Participants which received content information during the drive (withInfo group, $n_{withInfo}$=16) are grouped into the red boxplots while those participants that did not receive content information (noInfo group, n_{noInfo}=15) are grouped into the blue boxplots. The asterisk indicates a significant difference between the two groups of participants. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

Comparing the two groups concerning mean and the maximum amplitudes during the second event, one can find significant or almost significant differences between the two (U_{mean}=154.0, p_{mean}=0.0342; U_{max}=159.0, p_{max}=0.0195). Figure 18 and Table 6 indicate that the noInfo group had lower amplitudes than the withInfo group during the second event. Participants that received content information therefore had higher SCR amplitudes during the event.

SCR Rise Time

Analysing the mean and maximum SCR rise time did not reveal differences between the groups.

<u>Subjective rating:</u>

Event 1 – abrupt braking

When comparing the two groups during the first event according to the mean ratings and the range of ratings, no significant difference could be observed.

Event 2 – unexpected detour

As in the case of the first event, no significant difference between the two groups could be detected when performing a U-test for the mean ratings and the range of ratings respectively.

The hypothesis that participants differ in their physiological reaction when an error message together with content information was displayed to when only an error message without content information was displayed during the events was mostly falsified. Content information did not affect the participants' physiology during the first event. During the second event, no cardiological but an electrodermal reaction to the content information display was found. The content information seems to have led to a more intensive experience of the second event than when only an error message was displayed, according to the results of the mean and maximum amplitude analysis of the SCRs and the mean values of the EDA parameters. The subjective ratings did not differ in the case of both events despite of whether content information was presented or not.

4.3 Hypothesis 3 – type of event

Hypothesis 3: The physiological reaction and subjective evaluation to the information display during an event is different for different event types.

Descriptive statistics

A comparison of the mean values of the two events for the withInfo group and the noInfo group separately indicate higher zHR during the second event with information display than during the first event. The mean nRMSSD values are comparable between the two events for both groups (Table 6). Higher mean HRs during the second event in both groups point towards general higher cardiological activity during the unexpected detour event compared to the abrupt braking event independent of the type of information display.

For the withInfo group, all mean values of the tonic and phasic EDA parameters are higher during the second event than the first event. In the case of the noInfo group, the mean values of the EDA parameters indicate slightly higher EDA during the first compared to the second activity, except for the mean rise time and maximum rise time which are higher during the second event than during the first event (Table 6). The mean values are comparable between the two events for the noInfo group. According to the mean values of the EDA parameters, the unexpected detour induced more perspiration than the abrupt braking event when content information was present. When only an error without content information was displayed, the perspiration was similar.

For both groups, the mean values of the subjective ratings are higher and the range of ratings is lower during the second event compared to the first event when information is displayed (Table 6). The mean values therefore indicate that the unexpected detour was experienced as more pleasant than the abrupt braking event.

Inferential statistics

<u>ECG</u>

In the withInfo group as well as in the noInfo group no significant cardiological difference between the two events could be found when analysing zHr and nRMSSD. The events induced comparable cardiological reactions independent of the type of information display.

<u>EDA</u>

The electrodermal reaction when comparing the two events differed between the groups.

Number of SCRs

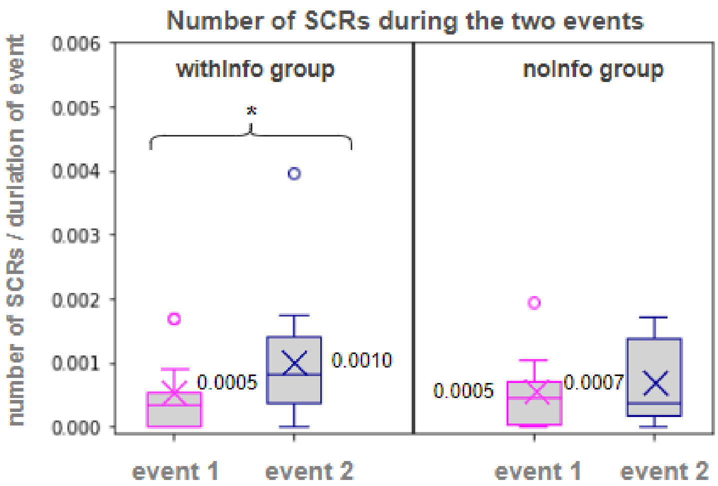


Figure 19: Number of SCRs during the two events with information display
The number of SCRs is compared between the two events for the group of participants that received content information (withInfo group, left, $n_{withInfo}$=16) and those participants that did not receive content information (noInfo group, right, n_{noInfo}=15). The first event of abrupt braking is depicted in magenta and the second event of the unexpected detour is depicted in dark blue. The asterisk indicates a significant difference between the two groups of participants. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

The participants that received content information had significantly differing numbers of SCRs during the two events (W=11.0, p=0.0159). The number of SCRs was lower during the first event than during the second event (Figure 19). The number of SCRs of the participants that did not receive content information did not differ between the two events (W=33.0, p=0.3824). The second event seems to have elicited more perspiration than the first event when content information was shown to the participants. These findings are supported by the descriptive statistics in Table 6.

SCR Amplitude

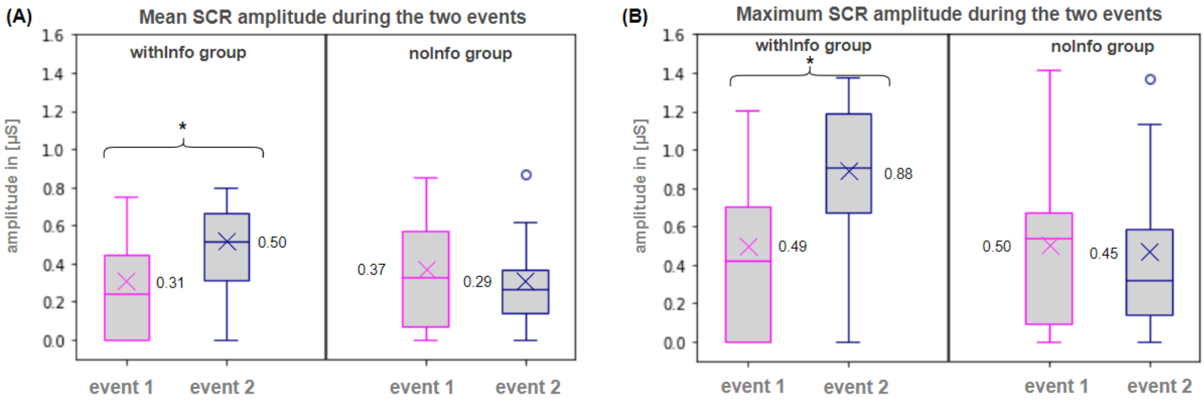


Figure 20: SCR mean and maximum amplitudes during the two events with information display

In (A), the mean SCR amplitudes of the two events are compared, while in (B) the maximum SCR amplitudes are compared. In each figure A and B, the boxplots for the withInfo group (left) contain the data of the participants that received content information ($n_{withInfo}$=16) and the boxplots for the noInfo group (right) contain the data of the participants that did not receive content information (n_{noInfo}=15). The first event of abrupt braking is depicted in magenta and the second event of the unexpected detour is depicted in dark blue. The asterisk indicates a significant difference between the two events. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

A significant difference between the two events with respect to the mean and maximum amplitudes can be detected for the withInfo group (W_{mean}=19.0, p_{mean}=0.0355; W_{max}=15.0, p_{max}=0.0186). During the second event, the amplitudes are higher for the group that received content information than during the first event (Figure 20**,** A+B, left). No significant difference in the size of the amplitudes could be detected for the noInfo group (Figure 20, A+B, right). These findings for the withInfo and noInfo groups are supported by the descriptive statistics in Table 6. The results indicate that when content information was present, the unexpected detour was experienced more intensely than the abrupt braking.

SCR Rise Time

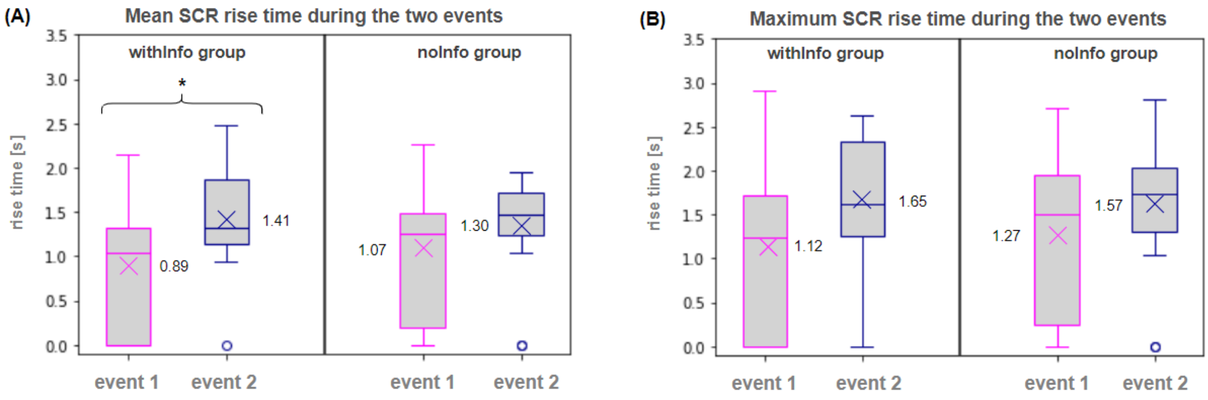


Figure 21: SCR mean and maximum rise times during the two events with information display
In (A), the mean SCR rise time of the two events are compared, while in (B) the maximum SCR rise time are compared. In each figure A and B, the boxplots for the withInfo group (left) contain the data of the participants that received content information ($n_{withInfo}$=16) and the boxplots for the noInfo group (right) contain the data of the participants that did not receive content information (n_{noInfo}=15). The first event of abrupt braking is depicted in magenta and the second event of the unexpected detour is depicted in dark blue. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

The two events differed significantly when examining the mean SCR rise time of the withInfo group (W=15.0, p=0.0186). Rise times were longer for the second event than for the first event (Figure 21, A, left, Table 6). The maximum rise time did not differ significantly between the two events (W=27.5, p=0.1160) but the rise times were also higher for the second compared to the first event (Figure 21, A, left, Table 6). No difference in rise times between the two events could be observed for the noInfo group that did not receive content information.

<u>Subjective rating:</u>

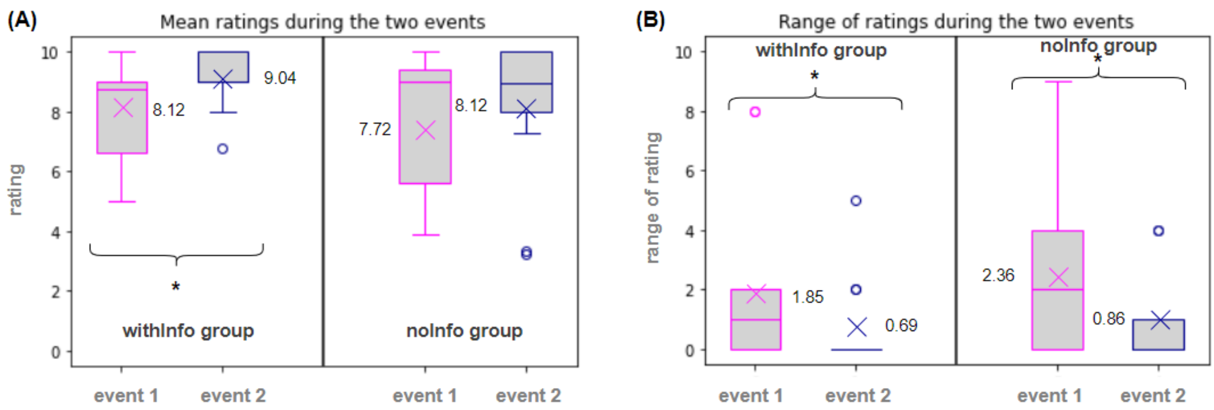


Figure 22: Mean and range of subjective ratings during the two events with information display
In (A), the mean subjective ratings of the two events are compared, while in (B) the range of subjective ratings are compared. In each figure A and B, the boxplots for the withInfo group (left) contain the data of the participants that received content information ($n_{withInfo}$=13) and the boxplots for the noInfo

group (right) contain the data of the participants that did not receive content information (n_{noInfo}=14). The first event of abrupt braking is depicted in magenta and the second event of the unexpected detour is depicted in dark blue. Boxplot parameters: box = middle 50%, line = median, cross = mean, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

A significant difference in the participants ratings between the events could be observed for the withInfo group: The mean ratings differed significantly between the two events (W=0.0, p=0.0117). Ratings were more positive during the second event than during the first event (Figure 22, A, left, Table 6). No significant difference between the participants' mean ratings could be seen for the noInfo group (W=21.0, p=0.5076) (Figure 22, A, right, Table 6). Both groups had significantly different ranges in their ratings when comparing the two events (W_A=0.0, p_A=0.0422, W_B=2.0, p_B=0.0240). The ranges in the ratings were lower during the second event compared to the first event (Figure 22, B).

The hypothesis that the physiological reaction to and subjective evaluation of the information display during an event is different for different event types could partly be verified. Only for the group that received an error message without content information, the hypothesis can be fully falsified with respect to the physiological reaction. The range of the subjective ratings was smaller during the second event than during the first event. In the case of the withInfo group, no cardiological but an electrodermal difference between the events was found. Electrodermal activity was higher and the intensity of the perspiration was higher during the second event compared to the first event for the participants that received content information. A significant difference in mean subjective ratings between the two events could be found in the withInfo group with the mean ratings being higher and the range being lower for the second event. During the unexpected detour the participants which received content information perspired more and rated the detour higher in terms of pleasantness than the abrupt braking event.

4.4 Hypothesis 4 – retrospective evaluation

Hypothesis 4: Retrospectively, the participants preferred the display of content information over merely an error message to help them understand the actions of the vehicle, to feel less insecure during an event and to reduce the wish for more information.

Quantitative analysis

Descriptive statistics

Table 7: Descriptive statistics of the quantitative ratings in the post-questionnaire

	First event				**Second event**				**Whole drive**			
	withInfo group		**noInfo group**		**withInfo group**		**noInfo group**		**withInfo group**		**noInfo group**	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Q1	4.00	1.20	2.28	1.18	3.95	1.13	1.89	1.13	3.89	1.05	2.50	1.10
Q2	2.26	1.19	3.17	1.25	2.11	1.29	3.33	1.03	1.95	0.91	2.78	1.06
Q3	3.16	1.30	3.89	1.41	3.00	1.56	4.11	1.37	2.79	1.44	3.72	1.36

Q1: "The display in the autonomous shuttle helped me to understand the present actions of the shuttle."
Q2: "In that situation / during the drive I felt insecure due to missing information."
Q3: "In that situation / during the drive I wished for more information."

The mean ratings of the participants to the first Likert-scaled question in the post-questionnaire regarding the helpfulness of the display to understand the actions of the vehicle (Table 7, **Q1**) was higher for the withInfo group than for the noInfo group. Participants that received content information therefore found the display during the first event, the second event and the whole drive more understandable than those participants that did not receive content information according to the mean values.

Agreement to the second statement of feeling insecure due to missing information was higher in the noInfo group than in the withInfo group according to higher mean values for the first event, the second event and the whole drive (Table 7, **Q2**). Receiving no content information seems to have induced a stronger feeling of insecurity due to missing information.

According to the mean ratings, more information was wished for especially by the noInfo group since the mean ratings were higher than for the withInfo group. Especially for the second event, the participants that only received an error notice wished for more information (Table 7, **Q3**).

Inferential statistics

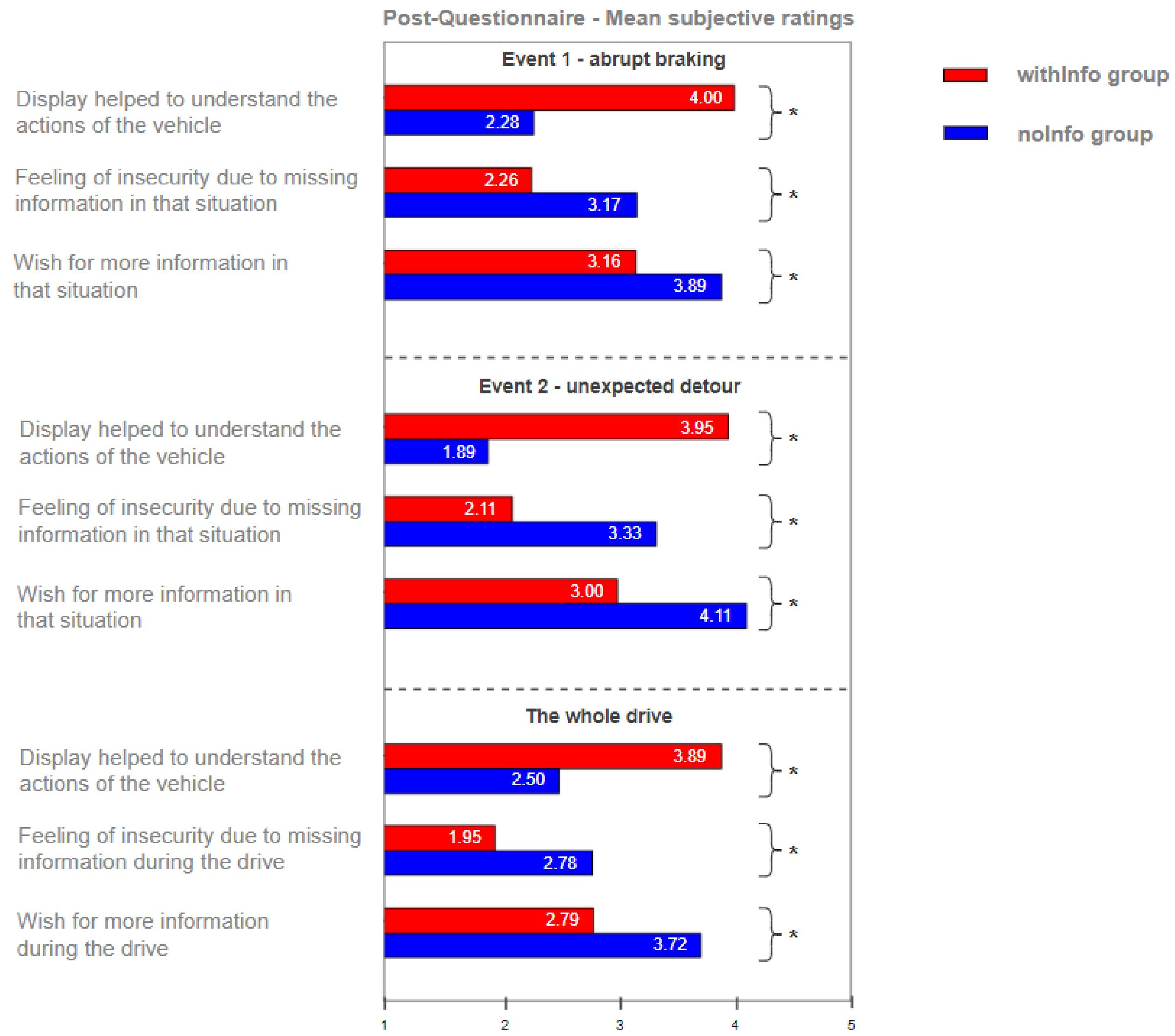


Figure 23: Subjective ratings to the quantitative Likert-scaled questions in the post-questionnaire

The mean participant ratings of the qualitative 5-point Likert-scaled questions of the post-questionnaire are depicted in the bar plots. The figure is subdivided into three parts: the three questions regarding the first event, the three questions regarding the second event and the three questions regarding the whole drive in the vehicle. The red bars represent the mean values of the withInfo group ($n_{withInfo}$=19) – the participants that received content information - and the blue bars represent the mean values of the noInfo group – the participants that did not receive content information (n_{noInfo}=18). The asterisks indicate significant differences between the groups as a result of a one-sided calculated Mann-Whitney U-test.

The question whether the display helped the participant to understand the actions of the vehicle was affirmed significantly more strongly by the withInfo group, which received content information, than by the noInfo group in all three categories ($U_{event\ 1}$=279.0, $p_{event\ 1}$=0.0003; $U_{event\ 2}$=304.5, $p_{event\ 2}$<0.0001; $U_{whole\ drive}$=277.5, $p_{whole\ drive}$=0.0005) (Figure 23). Results further showed that the withInfo group felt significantly less insecure than the noInfo group during the first (U=103.0, p=0.0167), the second (U=80.0, p=0.0022) and the whole drive (U= 97.0, p=0.0101) (Figure 23). The withInfo group also significantly less affirmed the question whether more information was wished for during the first

(U=113.0, p=0.0351), the second event (U=92.0, p=0.0063) and the whole drive (U=107.0, p=0.0230) than the noInfo group (Figure 23). When comparing the answers of the two events, the participant groups do not differ significantly in their ratings. Content information was therefore in general preferred over merely an error message. This is the case for both event types equally.

Qualitative analysis

Table 8: Number of participants that wished for further information

Participants	**withInfo group**			**noInfo group**		
	First event	*Second event*	*Whole drive*	*First event*	*Second event*	*Whole drive*
No of participants that answered the question	8 / 19	7 / 19	12 / 19	16 / 18	12 / 18	13 / 18

First, the number of participants that answered the open-ended questions and thereby indicated their wish for further information was analysed according to which group they belonged to (Table 8). More participants in the noInfo group answered the open-ended questions concerning the wish for more information during the first event, the second event and the whole drive than the withInfo group (Table 8). People that did not receive content information therefore had the wish for more information during the events and during the whole drive.

Table 9: Frequency analysis of the categorized qualitative statements in the post-questionnaire

Categories		**withInfo group**			**noInfo group**		
Level 1	*Level 2*	*First event*	*Second event*	*Whole drive*	*First event*	*Second event*	*Whole drive*
Display of route	*on the tablet*	2	1	8	1	3	3
	Residuals				1		
Display of time	*Duration of event*	2	1	4	1	1	1
	Duration of the rest of the drive until arrival	1	1	2	1	1	3
	Residuals			1	2	1	3
Display of further Information	*More context*	3	4	5	10	10	14
	Need for action or not	3	1	1	4	1	
	Residuals	1	1	1	1		3
Residuals		1			1		2

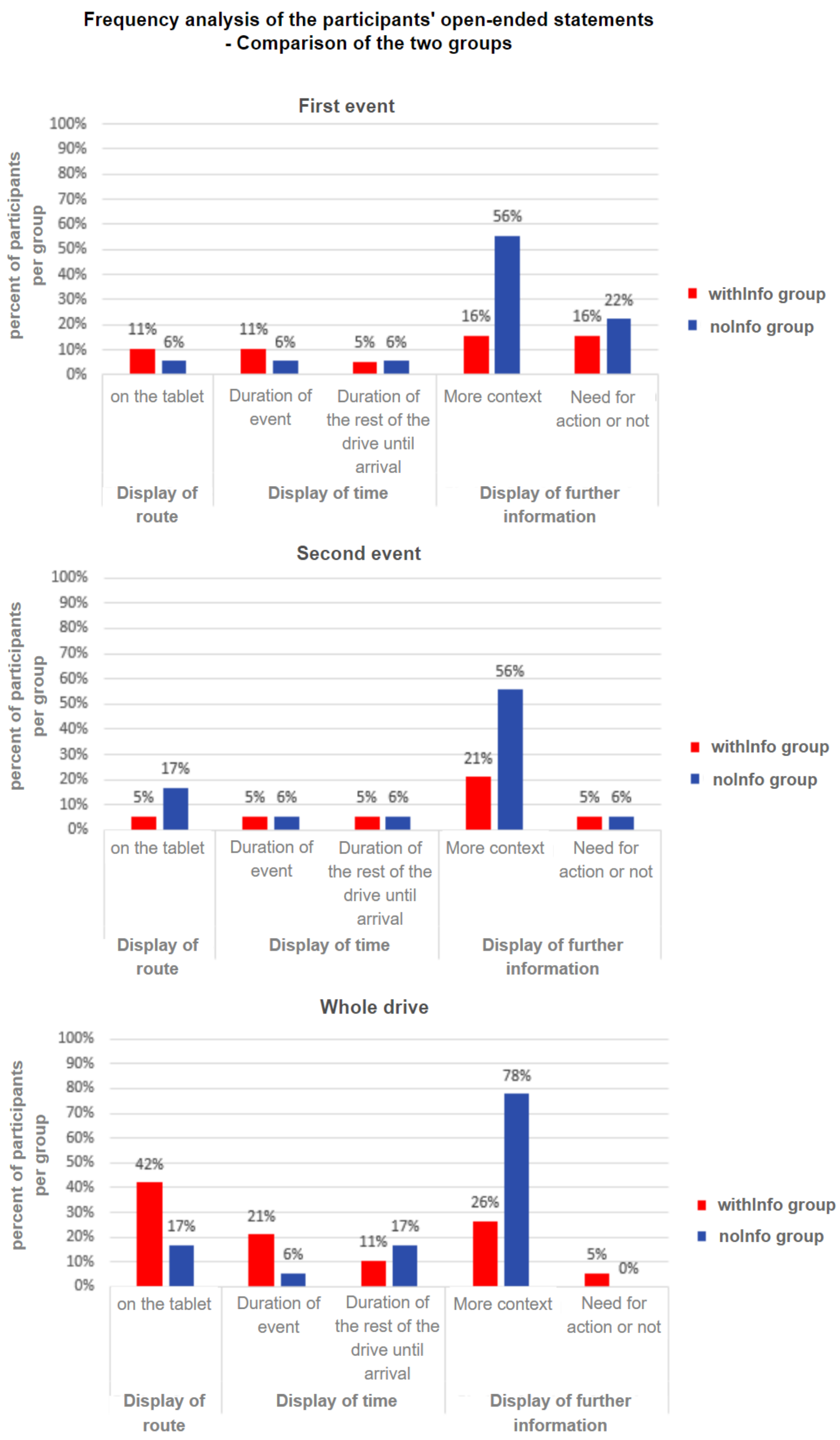


Figure 24: Percentage of participants that wished for information of a certain category

The results of the frequency analysis are depicted separately for the first event, the second event and the whole drive. The percentages of participants per group that commented on the level-1 and level-2 categories are depicted in the bar plots. The bars of the withInfo group (content information) are displayed in red and the bars of the noInfo group (no content information) are displayed in blue.

During the first event, the second event and the whole drive, participants in the noInfo group wished for more context information than participants in the withInfo group (Table 9**,** Figure 24). During the first event, more than three times, during the second event more than double and during the whole drive almost three times as many participants of the noInfo group wished for more content information than the withInfo group (Table 9). More than half the participants during the events and more than 75% of the participants during the whole drive that were assigned to the noInfo group wished for more context information (Figure 24). More information on the reason why the vehicle acted in a specific way when an event occurred was wished for by the participants that only received an error message and no content information. Although participants in the withInfo group received content information, participants declared that more context information would have helped them understand the event situations better and/or wished for more context information during the drive (Table 9**,** Figure 24). More information on the exact reason for e.g. the stop during the first event that exceeded the content information already presented was asked for. A participant proposed to add a button *"with 'further information' for those people, who would like to have more background information"* (P6). During the first event, both groups would have liked to receive information on the need of them taking action or not (Table 9**,** Figure 24). P27 (noInfo group) would have liked to receive a notice whether safety is at risk. The results indicate that the error messages did not suffice for the participants to understand the event situations and that more explanatory information was whished for by participants mainly of the noInfo group and by some participants of the withInfo group.

While the wish for more context information was the strongest in the noInfo group considering the whole drive, the strongest wish with 42% of the participants in the withInfo group was a display of the route on the display in front of them in the vehicle (Figure 24). Participants in both groups, but slightly more strongly so in the withInfo group, wished for more information on the duration of the event. Participants of both groups wished for a time display that shows the duration of the rest of the drive until arrival (Table 9**,** Figure 24). One participant who received content information and two participants that did not receive content information wished for a sign on the tablet or an acoustical sign that the drive in the vehicle started and ended. Those participants which received content information during the event tend to wish more for information concerning the whole drive, such as the current route and time displays than participants that did not receive content information. The focus of interest of the noInfo group was to gain more information on the cause and actions of the vehicle during the events.

The forth hypothesis could be verified with both the quantitative as well as the qualitative analysis of the post-questionnaire. Participants preferred the display of content information over merely an error message to help them understand the actions of the vehicle, to feel less insecure during an event and to reduce the wish for more information. However, participants that received content information also partly still wished for more context information as well as further information concerning the procedure of the whole drive, according to the open-ended statements.

4.5 Overview of the results according to the verification of the hypotheses

The results of the analyses are summarized according to the verification of the hypotheses in the table below (Table 10). The events had an impact on the participants' physiology and subjective ratings, so the first hypothesis can in general be verified. No difference in the reaction to the information type during the drive can be observed between the participant groups, except for the mean and maximum amplitudes during the unexpected detour. The second hypothesis can therefore not be verified. No difference in cardiological but a difference in electrodermal reaction to the event types can be observed for the withInfo group, which received content information. Also, a difference in the subjective ratings between the two event types were found. The third hypothesis can be verified for the withInfo group. The forth hypothesis can be fully verified: the participants preferred content information over only receiving an error notice as the ratings and statements in the post-questionnaire showed.

Table 10: Summary of the results according to the verification of the hypotheses

Hypotheses	**Verified by ECG**		**Verified by EDA**						**Verified by subjective ratings**		**Verified by post-questionnaire**	
	zHR	*nRMSSD*	*SCL*	*No. of SCR*	*Mean Amp*	*Max Amp*	*Mean Rise*	*Max Rise*	*mean*	*range*	*Quanti-tative*	*Quali-tative*
H1: An unexpected event whilst travelling in an autonomously driving vehicle has an impact on the passenger's physiology and subjective evaluation of the situation	✓	✘	✓ (✓) (2nd event)	✓ (1st event)	✘	✘	(✓) (2nd event)	✘	✓ (✓) (2nd event)	✓ (✓) (2nd event)		
H2: Participants differ in their physiological reaction and subjective evaluation when an error message together with content information is displayed to when only an error message without content information is displayed during the events	✘	✘	✘	✘	✓ (2nd event)	✓ (2nd event)	✘	✘	✘	✘		
H3: The physiological reaction and subjective evaluation to the information display during an event is different for different event types	✘	✘		✓ (withInfo group)	✓ (withInfo group)	✓ (withInfo group)	✓ (withInfo group)	✓	✓ (withInfo group)	✓		
H4: Retrospectively, the participants preferred the display of content information over merely an error message to help them understand the actions of the vehicle, to feel less insecure during an event and to reduce the wish for more information											✓	✓

✓: verification of the hypothesis; (✓): significant results in the linear mixed effects analysis but not in the post-hoc Tukey test;

✘: no verification of the hypothesis; text in brackets indicates for which event or which group the hypothesis holds

5 Discussion

5.1 Summary of the results

In the analyses of this study, a focus was laid on examining the effect of information display during unexpected events when passengers were driving in an, as they believed, autonomously driving vehicle. To disentangle the effect of the events from the effect of the information display on the participants physiology and subjective ratings, first the impact of the events by comparing time intervals before during and after the events was analysed followed by a group comparison to investigate the effect of different types of information. In general, the participants reacted physiologically and psychologically to the occurred events but less so to the type of information display during the drive. Retrospective subjective evaluations of the events and the drive indicated a preference for content information.

The results of the analyses revealed that the event situations had an impact on the participants' physiology and subjective ratings during the events (Table 10, **H1**). Furthermore, the mean ratings in the three time intervals of the unexpected detour event were in general lower for those participants that did not receive content information compared to those that did (Table 5). Presenting additional content information to an error message during the events, however, did not lead to significantly differing physiological responses or subjective ratings to when participants only received an error notice. An exception was the significant difference between the two groups in the participants' maximal amplitudes of the SCRs in the EDA data during the second event (Table 10, **H2**). Additionally, the mean values of the EDA parameters indicated higher perspiration when content information was displayed than when merely an error notice without content information was displayed during the unexpected detour (Table 6). The event type does not induce a significantly different cardiological reaction but a significantly different electrodermal reaction for participants that received content information (Table 10, **H3**). There are indications that the withInfo group experienced the second event as more intense than the first event. At the same time the withInfo group rated the second event as being more pleasant than the first event.

A discrepancy between the physiological results and subjective ratings during the drive in the vehicle and the retrospective subjective evaluation of the event situations and the whole drive in the post-questionnaire seems to exist: While the data collected during the drive does not reveal a difference in information type preference (Table 10, **H2**), the retrospective evaluations indicate a wish for content information during incidents and during the whole drive in an automated vehicle to understand event situations and feel less insecure (Table 10, **H4**).

In sum, physiological and psychological reactions to the events were observable. Indications of higher perspiration during the unexpected detour event when content information was displayed to when no content information was displayed were reflected in the mean values and high SCR amplitudes. The event type elicits differing electrodermal reactions when content information is present with stronger reactions for the unexpected detour. From the retrospective evaluations one can conclude that explanatory information during events and during the whole drive is highly desired by the participants. In general, physiological reactions may stand in contrast to subjective evaluations.

5.2 Influence of the events on passengers in automated vehicles

The events had an impact on the passenger' physiology and subjective ratings whilst driving in an automated vehicle. The HR changed during both events and decreased subjective ratings could be observed especially during the abrupt braking event. The HRV, however, was not influenced by the events. The tonic component of the EDA signal indicated increased perspiration during the events, while the phasic components together do not give a coherent picture of a reaction to the events.

During the abrupt braking event, the time intervals before, during and after the event differed significantly which shows that the event had an impact on the passengers' cardiology. Contrary to expectations, the zHRs of the participants decreased significantly from before to during the event. A mere correlation effect between driving speed and heart rate, which would explain the reduced HRs when the vehicle stopped for the abrupt braking, could be dismissed as an explanation: No correlation between driving speed and HR during an eventless everyday urban driving situation in the automated vehicle was observed. Therefore, the abrupt braking event as such had an influence on the participants' HRs. In their study, Beggiato et al. (2018) also witnessed a decrease in heart rates when participants where in discomfort periods during a simulation of automated driving. The discomfort situation in their study consisted of an automated vehicle approaching a lorry rapidly and braking very late to avoid a collision. By analysing the time periods before, during and after the events and finding increased heart rates after the events, Beggiato et al. (2018) explained their results with the "preparation for action" hypothesis. The hypothesis refers to a decrease in HR before a planned action (Andreassi, 2010; Cooke et al., 2014). The current results are similar to those of Beggiato et al. (2018) in the sense that the zHR deceased significantly from before to during the event but no significant - only a slight - increase in zHR between the time intervals during and after the event can be observed. An "action"-reaction after the event was therefore not or only very mildly present. The "preparation for action" hypothesis might only partly hold here. An emotional fear reaction to the events, which led to this physiological response may be a better explanation and is discussed below (see 5.4.1 Physiological reactions as signs of emotions). The analyses showed that HRs were comparably high before the abrupt braking event. An explanation for this could be a carry-over effect of heightened HRs, since the participants were physically active when they walked to the sidewalk library about two to three minutes before the abrupt braking event took place. The HR is affected to a varying degree between participants to recent physical activity (Brouwer et al., 2018), depending on the recovery rate due to the degree of physical fitness (Grant et al., 2013), age and gender (Umetani et al., 1998) etc. of the participants. It is therefore likely that the HR was still enhanced for some participants in the time interval before the first event and that therefore the significant difference in HR between before and during the event appears more pronounced in the current results. However, since the HR decreased quite abruptly when the abrupt braking event occurred, one can assume that the decrease in HR was a reaction to the event. The HRV was not affected by the abrupt braking event. Only a slightly lower mean nRMSSD values over all participants during the event compared to before and after the event points towards a slight change in physiological arousal. The analyses therefore did not reveal inversely correlated HR and HRV as would be physiologically expected (Beggiato et al., 2018; Brouwer et al., 2018; McCraty & Shaffer, 2015) but HR and HRV seem to be slightly correlated with a decrease in HR and HRV during the event. In sum, cardiological reactions to the abrupt braking event became evident in the analyses despite the influence of physical activity shortly before the event occurred.

During the unexpected detour, zHRs between the time intervals before the event and after the event differed significantly, with higher HRs after the event. The difference in HRs can be explained with the unexpected detour that occurred between the two time intervals. At the beginning of the detour, an abrupt decrease in zHRs can be observed when averaging over all participants. This pattern in the HR is similar to that at the beginning of the abrupt braking event. One can hypothesise that the physiological impact of the event is present mainly at the beginning of the second event which is observable by a decrease in HR and that the HR activity is then enhanced during the remainder of the detour and after the detour following the “preparation for action” hypothesis. Since the whole ~2 min time period of the detour and not only the beginning of the event was examined and compared to the time intervals before and after the event, the hypothesis concerning the direct reaction of the participants to the event can neither be verified nor falsified. The unexpected detour did not have a significant impact on the HRV of the participants. Only a slight decrease in the mean nRMSSD values over all participants from before to during and from during to after the event points towards a slight increase in physiological arousal. Both events had a cardiological impact on the participants by changing their HR. However, HRV did not serve as an indicator for cardiological reactions to the events.

A reaction to the events could also be partly observed in the results of the electrodermal data. In the event of abrupt braking, a significant difference between during and after the event in terms of the SCL and the number of SCRs was present. Both the tonic and the phasic component of the EDA signal indicated higher electrodermal activity during compared to after the event. These results indicate that perspiration and therefore arousal declined after the event. The comparable electrodermal activity between before and during the first event may also be partly attributed to a carry-over effect of prior physical activity, as mentioned above for the ECG data. No significant differences in the SCR amplitudes and the SCR rise times between the time intervals could be found. The abrupt braking event therefore did not lead to an increased SCR as associated with an intense experience. Together the results indicate that an electrodermal reaction to the abrupt braking event was present but the intensity of the perspiration was comparable to the time intervals before and after the event.

A general difference in the participants’ SCLs between the time intervals the mean rise time of the SCRs was found for the unexpected detour event. No significant difference between two specific time intervals but a general difference between the time intervals of the second events indicates a slight impact of the event on the passengers’ perspiration. Electrodermal reactions to the unexpected detour were minimal. In general, both events induced an electrodermal reaction with higher perspiration induced by the abrupt braking event. The intensity of the events was, however, not (strongly) reflected in the amplitudes and rise times.

The subjective ratings of the participants during the drive indicate a negative reaction to the abrupt braking event. The mean ratings during the event were significantly lower than before and after the event. The passengers found the abrupt braking event less pleasant than the eventless drive in the autonomous vehicle. After the event, the ratings remained significantly lower than before the event. The abrupt braking event seems to have had a prolonged negative effect on the participants’ feelings of pleasantness. These results are supported by the results of the analyses of the range in subjective ratings. The range acts as a measure for a participant’s change in rating and therefore a change in participants perceived pleasure in the drive in a time interval. The range of the ratings was significantly higher during the first event compared to the other time intervals. A high range during the event indicates a discrepancy in the ratings of pleasantness and therefore indicates a reduction of the perceived pleasantness at some point during that time interval. The significant difference in the range between before and after the event together with the significantly lower mean ratings after the event indicate that the abrupt braking event induced a general prolonged reduction in the perceived pleasantness of the drive. As the significant difference in the range of ratings between during and after the event show, a prolonged time was needed to partly regain the feeling of pleasantness after the event occurred. The results of the mean ratings and those of the range in the ratings support each other by indicating that the abrupt braking event had a negative impact on the participants’ experience of pleasantness of the drive. The pleasantness in the drive was not directly restored after the event occurred.

Only the results of the linear mixed effects analysis revealed significant differences between before, during and after the unexpected detour with respect to the mean and range in ratings during the drive. No significant difference could be detected between two specific time intervals. The second event seems to have had an influence on the experienced pleasantness of the drive but only the comparison on the mean values of the time intervals indicate a slight negative impact of the event. While the abrupt braking event had a prolonged negative impact on the participants experienced pleasantness of the drive, the unexpected detour only slightly decreased the pleasantness of the event.

In general, both events induced increased physiological arousal and a decreased experience of pleasantness in the passengers that drove in the automated vehicle. Decreased ratings of pleasantness were more pronounced and prolonged during the abrupt braking event than during the unexpected detour.

5.3 Influence of the type of information on passengers in automated vehicles

The main focus of the analyses was to investigate the impact of the type of information on the passengers' physiology and emotions during a drive in an automated vehicle. The influence of displaying solely an error message compared to an error message followed by content information was analysed. The participants' subjective evaluations during the drive and retrospectively showed a positive impact of content information when events occurred.

The analysis of the first hypothesis – the impact of the events on the participants – revealed a significant difference between the two groups of participants during the three time intervals of the unexpected detour. The group, that received an error notice followed by what-and-why information gave significantly higher ratings and had a significantly lower range in the ratings than the group that only received an error notice. As discussed above, lower mean ratings and higher range indicate a reduced feeling of pleasantness during the drive. The group that received only the error notice therefore found the second event as well as before and after the event less pleasant than the group that received content information. This difference between the two groups was not present during the time intervals of the abrupt braking event. An explanation for the results is that the experience of the first event could have influenced the general pleasure in the drive. While the group that received content information regained their pleasure in the drive, the participants that received merely an error notice did not fully regain their pleasure. The difference between the two groups cannot be attributed to a bias in the participant constellation in the groups since the subjective ratings were comparable between the two groups during the baseline conditions. The change in subjective ratings and therefore the decreased experience of pleasantness during the drive can only be attributed to the type of information which the participants received. Content information during the abrupt braking event seems to have induced a recovery of the experience of pleasantness in the drive before the unexpected detour occurred. Only the display of an error message during the abrupt braking event seems to have decreased the pleasantness of the following drive in the automated vehicle. These two effects result in a significant difference between the groups of participants during the three time intervals of the unexpected detour event.

However, the comparison of the two participant groups during the time intervals of only the events while information was displayed did not revealed strong differences. During the abrupt braking event the type of information display did not lead to differing physiological reactions nor differing subjective ratings. In the case of the unexpected detour, neither a cardiological difference nor a difference in the subjective ratings between the two types of information display could be observed. Only a significant electrodermal difference with respect to the mean and maximum amplitudes of the SCRs could be found between the two participant groups. The group that received content information had higher amplitudes than the group of participants that only received an error notice. The amplitudes indicate SCRs to high intensity stimuli and therefore indicate event bound reactions. The second event with content information display seems to have been experienced more intensely than without the content information display according to these results. Since, however, no other significant differences in the tonic or phasic components of the EDA data could be detected between the types of information during the second event, the result should not be weighted too strongly. An intense stimulus usually elicits a high amplitude together with a high rise time but this effect was not visible during the second event. In sum, examining the specific time intervals during the events when information was displayed, one cannot observe strong indications that the type of information display had a physiological or emotional influence on the passengers in the automated vehicle.

The reason why no effect of the information display on the subjective ratings was observed during the drive may be due to technical errors in the application on the smartphone. For the analysis of the subjective data, seven participants and further two participants during the unexpected detour had to be excluded because the recording of the ratings stopped. Most likely the error in the data recording occurred when the navigation in the application was closed. Since the recordings stopped during the time of the events, it is likely that the participants were looking for further information on the events in the application. Therefore, those participants that had to be excluded for the analyses might have been those that would have reacted strongest with low subjective ratings during the events.

In contrast to the subjective ratings during the events, which indicated no difference in the perceived pleasantness between the type of information that was displayed, the retrospective evaluations of the participants clearly revealed a preference for more information – especially explanatory context information. The quantitative analyses of the post-questionnaire showed that what-and-why information was preferred over only an error message to help understand the actions of the automated vehicle and to feel less insecure during an unexpected event. Both the quantitative and the qualitative analyses revealed that the group that received only an error notice wished for more information in an event situation and during the whole drive in the vehicle. The wish was significantly stronger than in the group that received content information. However, also the participants that already received what-and-why information expressed the wish for even more context information. The type of information that was retrospectively wished for partly differed between the two participants groups, that received differing information during the drive: The main wish of the group that received merely an error message was to received further context information, while the wishes of the participants that received content information were more distributed over further categories, e.g. the wish to have the route displayed on the tablet in the vehicle and/or to receive information on time such as the duration of the event.

During the abrupt braking, next to the wish for more context information, the wish to receive information whether there is a need for the passenger to act was expressed by both participant groups in the post-questionnaire. The need for action was not so much a question during the unexpected detour. These results indicate, that some passengers would have preferred instructions on how to behave during the abrupt braking event. Such a wish points towards the participants not feeling in control of the situation.

In general, the participants' physiology did not (strongly) react to the type of information display. Directly during the events, there was no indication that the type of information had an impact on the experienced pleasantness during the events. However, over the time course of the two events, one could observe that the participants that received an error message without content information found the drive less pleasant after the first event occurred. This finding, together with the retrospective evaluations of the participants indicate a preference and wish for explanatory what-and-why information during a drive in an automated vehicle. The drive becomes more pleasurable, the feeling of insecurity due to missing information is lower and occurring events are perceived as more understandable when more context information is given.

5.4 Influence of the event type on the passengers in automated vehicles

The comparison of the two event types abrupt braking and unexpected detour with respect to the participants' physiological reaction and subjective ratings revealed differences when content information was displayed.

Whilst analysing the effect of the events on the participants' physiology and subjective ratings, a stronger physiological response and reaction in the subjective ratings can be observed for the abrupt braking event compared to the unexpected detour event. Analyses of the physiological data were significant between specific time intervals, as were the subjective rating in the case of the abrupt braking event compared to the unexpected detour. Additionally, more open-ended questions in the post-questionnaire regarding the wish for more information were answered with respect to the first event than to the second event. Especially in the group of participants that did not receive content information, more participants gave statements on the first compared to the second event. Together, these results indicate that the abrupt braking event had a more intense impact on the participants' physiology, experience of pleasantness of the drive and wish for information than the unexpected detour. A more sudden event seems to elicit a higher wish for explanation. This finding is supported by Shen et al.'s (2020) study in which near-crash situations led to the highest wish of explanations.

However, analyses of the event types separately for the different types of information revealed that the reactions to the events differed when content information was displayed but did not differ when only an error message was displayed. Differences in electrodermal activity and subjective ratings between different event types but no cardiological differences were found when content information was displayed. The participants that received content information had a higher number of SCRs which indicates higher electrodermal activity during the second event compared to the first event. Additionally, the mean and maximum amplitudes were significantly higher and the mean rise times were significantly longer. These results indicate that the unexpected detour, whilst receiving content information, was experienced as more intense than the abrupt braking event. Contrarily, the mean subjective ratings during the drive were significantly higher and the range of the ratings were significantly lower during the unexpected detour than during the abrupt braking event when content information was displayed. The analyses therefore indicate that the participants were more highly physiologically aroused while at the same time their rating of pleasantness was high which in contrast points towards an experience of little stress. A possible explanation for these at first sight conflicting results could be the following: The participants that received content information knew that they were currently on a detour. The knowledge of the detour might have been welcome information, such that the participants did not reduce their ratings of the pleasantness of the situation. At the same time the knowledge of going on a detour might have led to further questions. The still remaining uncertainty of the specifics of the situation might be the reason for the electrodermal arousal. The greater arousal during the second event compared to the first event therefore might be explained by the information during the abrupt braking event leaving less unanswered questions than the information that was displayed during the unexpected detour. This explanation is supported by slightly lower agreement of the participants to the statement in the post-questionnaire that the display helped to understand the action of the vehicle during the second event compared to the first event. However, the agreement to the statement did not differ significantly between the two events. Furthermore, the longer duration of the detour (~2 min) event compared to the abrupt braking event (~ 1 min) might have led to more arousal because the time period of missing information was longer.

When only an error message was shown to the participants, no electrodermal difference but a difference in the range of subjective ratings between the event types could be found. The ranges were significantly lower during the second compared to the first event while the mean ratings were comparable between the events. A lower range in ratings during the second event compared to the first event indicates that the second event was experienced as more pleasant than the first event. Together with the discussed results in the above paragraphs one can conclude that the first event therefore had a stronger negative impact on the experienced pleasantness than the second event regardless of what information was displayed.

In sum, the abrupt braking event was experienced as more intense than the second event, according to the physiological reactions and subjective ratings of all participants together. When content information was presented, the unexpected detour led to higher EDA than the abrupt braking event, which might be explained with a less clear explanation of the second event compared to the first event together with the longer duration of the detour with missing information.

5.5 Comparison of the finding of this study with other studies

5.5.1 Physiological reactions as signs of emotions

Both the abrupt braking event and the unexpected detour elicited a physiological response. Since emotions have a physiological consequence (McCraty et al., 2009), the question arises which specific emotions elicited the increased arousal reactions of the participants during the events.

In the case of the abrupt braking event, the HR of the participants decreased, the HRV showed no reaction, the SCL and the number of SCRs increased and the amplitude and rise times of the SCRs remained comparable to before and after the events. Bradley & Lang (2007) exposed their participants to cues that threated shock. They too found an increase in SCL and an initial decrease in HR, with the HR remaining slower compared to the safe conditions. The authors explained their results with a sign for increased attention to a threating stimulus. In this current study, the threatening stimulus would be the sudden braking which might have induced a brief moment of fear. Kreibig (2010) reported in a literature review that fear paradigms which elicit a high degree of self-involvement lead to the feeling of higher imminence of threat (Bradley & Lang, 2000). The participants in the present study retrospectively expressed their wish for information whether they had to act in the case of the abrupt braking event. This wish indicates that the participants felt somewhat involved in that situation. According to Kreibig (2010) this involvement could have led to an imminent fear response which can be characterized by immobilization rather than by an active coping response. A freezing reaction occurs when both the sympathetic and the parasympathetic pathways of the autonomous nervous system are active with a dominance of the parasympathetic pathway (Roelofs, 2017). The sympathetic pathway would therefore be inhibited, leading to a decrease in HR (Kreibig, 2010). According to the Kreibig's (2010) literature research, imminent fear is expressed physiologically by a decrease in HR, no change in HRV, an increase in SCR an in the case of fear an increase in SCL. These findings were each based on at least three studies, according to Kreibig (2010). The physiological responses to the abrupt braking event could therefore be an indication of the participants' experience of fear. Also the feeling of surprise, which is characterized by short duration SCRs of medium response size and rise time and increased SCL (Kreibig, 2010), matches the electrodermal reaction of the participants in the present study. The intensity of the abrupt braking event was not reflected in significantly higher amplitudes or rise times compared to the time intervals before and after the event. However cardiologically, surprise elicits an increase in HR (Kreibig, 2010), which is not the case for the abrupt braking event. One can conclude, that the abrupt braking event took the participants by surprise with also indications for emotions of fear due to a feeling of self-involvement in the situation. The unexpected detour led to minimal reactions in the participants' HR and SCL, such that one can assume that no strong emotions elicited the arousal.

5.5.2 Implications for passenger – autonomous vehicle interaction

Emotions and arousal are related to information that is given. The participants' physiological arousal in the vehicle was, however, mainly a reaction to the event as such than to the type of information that was displayed. The emotions elicited by the type of information display were therefore not so strong as to lead to a physiological reaction.

The emotions of the participants could be deduced from the participants' ratings of pleasantness of the drive. The effect of information display was reflected in the subjective evaluations of the participants. Those participants that did not receive why-and-what information during the abrupt braking event, found the following drive less pleasant than the participants that received the content information. Information on what action the vehicle took and why it took it had a positive impact on the experienced pleasantness of the drive. Koo et al. (2015) also found a positive effect of why-information in the case of semi-automated driving. Why-information decreased anxiety during the drive. However, they found that what-and-why information could also increase anxiety and annoyance (Koo et al., 2015, 2016). Increased anxiety or annoyance would be indicated by decreased ratings of pleasure in the drive when participants received content information compared to when they only received an error notice in this study. This was not the case. What-and-why information therefore did not lead to a negative response but a positive experience of the drive in the current study.

Self-explanatory systems are seen as most satisfactory and trustworthy (Schraagen et al., 2021). Why-and-what information can also make automated driving more accessible to the passengers, because they do not have to monitor the driving environment to understand the intentions of the vehicle (Forster et al., 2017; Zhang et al., 2021c). Only why-information can also induce a feeling of acceptance and trust in the vehicle (Koo et al., 2015) and can increase the perceived understandability (Wiegand et al., 2019). Higher understandability of the actions of the vehicle in event situations as well as in the course of the whole drive was also testified by the participants that received content information compared to those that only received an error notice in the present study. The feeling of insecurity due to missing information was lower in those participants that received content information compared to those that did not. Less insecurity points towards higher trust in the automated vehicle. The present results together with the results of previous studies (Forster et al., 2017; Koo et al., 2015; Wiegand et al., 2019; Zhang et al., 2021c) indicate that explanatory information given by the vehicle increases the understandability and trust in the automated vehicle.

Besides content information, the passengers in the vehicle retrospectively wished for the route, information on how long an event would take and when they would arrive at their destination to be displayed on the tablet in front of them. Meurer et al. (2020) found in their Wizard-of-Oz study that awareness of the route was important for the passengers to relax and feel comfortable. The authors found that participants checked on the performance of the vehicle and the route in their study. Participants in the current study also examined the driving route of the vehicle in the navigation application on their smartphone. Although explanatory information increased the participants' trust in the vehicle, the passengers still seem to wish for information to be able to control for the vehicle's actions. As Zhang et al. (2021c) stated, insufficient information may lead to a felt passive role and a feeling of failing to maintain a sense of control. Indications for Zhang et al.'s (2021c) statement were both present in Meurer et al.'s (2020) and the present study: When incidents occurred, passengers in an automated vehicle wished for information on whether they had to act or not and where therefore not sure whether they were bystanders or somewhat involved (Meurer et al., 2020). Additionally, in both Wizard-of-Oz studies, the passengers wanted information on their safety when an event occurred. A participant in the present study proposed to add an opportunity for passengers on the tablet to access further background information if whished for. Some participants in the current study wished for a clear start and end signal of the drive either on the tablet or acoustically. This wish was also present in participants in Meurer et al.'s (2020) study. In sum, more information on the route, the time and even more explanatory information on the passengers' need to take action or not may increase trust in the vehicle, maintain the passengers' sense of control and may increase the pleasantness of the drive. An interaction with the automated vehicle in the sense of receiving information is therefore highly desired by the passengers. The wish for interfering in the automated drive was not expressed by the participants.

Although the content information as such had a positive impact on the passengers, the timing of the explanation given by the automated vehicle has an impact on the effectiveness Zhang et al., (2021c). To increase the feeling of trust and reduce anxiety, information should be provided before an event occurs (Du et al., 2019; Koo et al., 2015, 2016; Ruijten et al., 2018; Zhang et al., 2021a). Du et al. (2019) and Körber et al. (2018) showed that information display after an event led to the same or even decreased trust in and acceptance of an autonomous vehicle. In the present study, the information display was started manually at the same time or a few seconds after the start of an event. The information was therefore presented during the events. To increase the positive effect of displaying content information to the participants, information should be displayed earlier than was done in this study. Additional information given after an event can increase the passenger's understanding of the event (Körber et al., 2018; Schneider et al., 2021; Shen et al., 2020). Announcing an event before it occurs, giving live explanations not as text displays (Schneider et al., 2021) and giving additional explanatory information after an event could lead to the greatest trust in and understanding of an automated vehicle.

To encourage the use of autonomous vehicles in the future, one needs to further improve the passenger – autonomous vehicle interaction. To change the view that technology dehumanizes (Tussyadiah et al., 2017), trust in the functioning of the autonomous vehicle must be present while at the same time the feeling that the passenger is still somewhat in control of what is happening has to be conveyed. Optimizing the time and content of information display can contribute to enhance the passenger – vehicle interaction.

5.5.3 Implications for (shared) autonomous mobility concepts

People who own a car show willingness to make use of an autonomous vehicle in the future (Maghraoui et al., 2020; Pakusch et al., 2018). The challenge is to motivate a shift from autonomous vehicle "ownership" to shared "usership" (Machado et al., 2018) to meet problems such as increasing traffic congestion, pollution and too few parking spaces as the population living in cities grows (Dobbs et al., 2011; Fagnant & Kockelman, 2015; Martinez et al., 2015; Rigole, 2014). Next to environmental concerns (Machado et al., 2018), travel time, travel cost, comfort, flexibility, availability, reliability and safety are factors that influence a person's decision for choosing one mode of transport over another (Pakusch et al., 2018). A private car enables high flexibility, comfort and availability (Martinez & Viegas, 2017), which makes this transportation mode more attractive than on-demand services in which waiting time and travel time is longer and comfort and convenience is lower (Santos, 2018). To encourage the use of on-demand shared mobility services in the future, the favourable features of owning a private car should be focussed on and optimized for shared mobility services. The feeling of comfort and safety can be conveyed by the type and way information is given to the passengers during the drive. In this study, content information led to an overall higher experience of pleasantness and feeling of less insecurity during the drive than when no explanatory information was present. Comfort for the passenger may also be enhanced by allowing the use of the tablet in the car for work-related or free-time activities. Although the wish was not expressed in the present study, participants in Meurer et al.'s (2020) Wizard-of-Oz study did. Stevens et al. (2019) found in their study that people will have the desire to use their travel time to complete tasks so time can be saved for normally neglected activities. Optimizing the passenger – autonomous vehicle interaction to increase the pleasantness and the feeling of safety during the drive and enabling people to make use of their time which they would normally spend with driving could encourage the use of autonomous vehicles in the future. Environmental advantages and travel cost may additionally motivate the use of shared autonomous mobility.

In the study set-up, on-demand mobility with virtual stops was implemented to model a flexible, available and reliable autonomous driving experience. How well the on-demand service functioned was not the focus of the present analysis but needs to be investigated in the future.

5.6 Relevance for present research and future prospects

The current study was conducted as a real-world Wizard-of-Oz study in which an on-demand drive in an automated vehicle of automation level four was simulated. The experimental set-up enabled to examine the impact of autonomous driving on passengers in a real-world setting. In the past, most studies on automated driving were conducted in driving simulators but people may behave differently in driving simulators than in real-world driving conditions (Bella, 2008; Zhang et al., 2021c). The real-world setting of this study decreases the bias of driving simulation studies in the research area and enables a comparison of the results of driving-simulation and real-world studies.

Psychophysiological data was gathered during the study. Cardiological and electrodermal measures served to objectively measure the physiological and emotional reaction of a person during the drive in an automated vehicle. However, the physiology of the participants in the current study mostly reacted to the unexpected events and not to the type of information display. The participants’ subjective evaluations gave deeper insight into the type of information that was wished for. The objective physiological data and the subjective data together gave a holistic picture of the impact of the drive in an autonomous vehicle on a passenger.

A passenger’s information need during unexpected events was mainly reflected in the quantitative and qualitative data of the post-questionnaire. Future research will need to investigate whether optimizing the content and timing of the information display will lead to a decreased physiological reaction to the event situation. The present results together with the results of prior research indicate a preference for explanatory information during events and during the whole drive. The exact timing of the information display prior to an event and the interaction between the time and the content of the display still needs to be investigated in future research (Zhang et al., 2021c). Using live explanations when events occur should also be considered in the future (Schneider et al., 2021).

5.7 Proposals for improvement

Room for improvement can be seen in the experimental design as well as in the analysis of the present study. Additionally, ideas for possible future studies are proposed.

A limitation in the experiment design was that the abrupt braking event occurred shortly after the participants had been physically active by walking to and from the sidewalk library. The physical activity which increased the physiological activity of the participants can, in some cases, have interfered with the physiological analyses of the abrupt braking event. Especially in the time interval shortly before the abrupt braking event, a carry-over effect can have influenced the results. In the future, a greater time gap than ~2 min between physical activity and an event should be kept.

For the analysis of the physiological data, further information on the participants should have been gathered either prior to or after the experiment. For the later analysis of the cardiological data, one should have asked the participants whether they had any cardiological disorder. A disorder is characterized by arrythmia (Jambukia et al., 2015) which can have an impact on the analysis of the cardiological data. A visual inspection of the participants' ECG signal did not reveal any apparent abnormal structures in the present study. Therefore, one may assume that the results of this study were not influenced by arrythmia. When measuring electrodermal data, it is useful to know the participant's handedness. One can suppose that the dominantly used hand is more prone to movement artefacts than the non-dominant hand. This knowledge may imply to use only one electrodermal sensor on the non-dominant hand. If both hands are recorded, as in the present study, one can consider the non-dominant hand more strongly in the analyses that the dominant hand. One can also gain insight into whether electrodermal activity differs between the dominant and non-dominant hand. This would also affect the analyses of the electrodermal data. Cardiological and electrodermal data analysis would have profited from further information.

During the drive in the vehicle, the question for the subjective rating could be formulated more specifically. Instead of asking for the current well-being of the participant, which could be rated from pleasant to unpleasant, the question could be directed at the participant's feeling of safety or the degree of anxiety. This way the subjective ratings would match the quantitative questions in the post-questionnaire better. One might also consider choosing a smaller Likert scale for the ratings, since a differentiation between 10 degrees of an emotion is very fine-grained.

A limitation in the analyses is that different participant groups had to be considered. A further limitation during the analyses of the EDA data, was that some signals of participants were excluded after visual inspection and finding irregularities in the signal. An objective method should in future be employed. In future studies, a baseline recording of the participants wearing the ECG device and the EDA sensors in a relaxed state should be conducted prior to the study. This would facilitate the normalization and detrending of the physiological data.

For an even more realistic experimental set-up, one could have modelled a payment scheme to investigate the impact of travelling costs the on-demand drive. However, since the aspect of travelling cost would have been an additional variable to consider in this already quite complex experimental set-up, a future study may investigate this aspect in greater detail.

To crystallize the effect of autonomous driving on the passengers, a study with the same set-up but with the participants knowing that a driver is manipulating the vehicle could be conducted and the results be compared to the current Wizard-of-Oz study. Differences in the participants' physiological and psychological reactions might be expected when events occur and when information is displayed. Only then it is possible to make a statement on the connection between autonomous driving and the information need which is associated with it.

6 Conclusion

In my thesis, a person's need for information when travelling in an autonomously driving vehicle was analysed with psychophysiological data. The results showed that passengers travelling in an automated vehicle prefer to receive explanatory content information over merely an error notice. While the information need was not directly reflected in the passenger's physiological arousal during unexpected events, the retrospective subjective evaluations of the participants emphasized the benefit of having received information on the events during the drive. The drive becomes more pleasurable, the feeling of insecurity due to missing information is lower and occurring events are perceived as more understandable when context information is given. Furthermore, the passengers that already received information pronounced a wish to receive even more context information, information on the route and time information. Further optimizing the time and content of information display may contribute to enhance the passenger-vehicle interaction.

The physiological arousal could be mainly attributed to the events. Both events induced a physiological and emotional reaction of a passenger in the automated vehicle. Decreased ratings of pleasantness were more pronounced during the abrupt braking event than during the unexpected detour. The results point towards the abrupt braking event taking the participants by surprise with also indications for emotions of fear due to a feeling of self-involvement in the situation. The type of event therefore also has an impact on the passenger's arousal.

Future research will need to investigate whether displaying more context information and displaying the information before an event decreases the physiological reaction to an event. The analysis of psychophysiological data in a real-world Wizard-of-Oz setting allowed to examine multiple facets of emotions and wishes of passengers travelling in an automated vehicle.

References

Andreassi, J. L. (2010). Psychophysiology: Human behavior and physiological response. In *Psychophysiology: Human Behavior and Physiological Response*. https://doi.org/10.4324/9780203880340

Atasoy, B., Ikeda, T., Song, X., & Ben-Akiva, M. E. (2015). The concept and impact analysis of a flexible mobility on demand system. *Transportation Research Part C: Emerging Technologies*, *56*. https://doi.org/10.1016/j.trc.2015.04.009

Baek, H. J., Cho, C. H., Cho, J., & Woo, J. M. (2015). Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemedicine and E-Health*, *21*(5). https://doi.org/10.1089/tmj.2014.0104

Banks, S. J., Bellerose, J., Douglas, D., & Jones-Gotman, M. (2012). Bilateral skin conductance responses to emotional faces. *Applied Psychophysiology Biofeedback*, *37*(3). https://doi.org/10.1007/s10484-011-9177-7

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

Bella, F. (2008). Driving simulator for speed research on two-lane rural roads. *Accident Analysis and Prevention*, *40*(3). https://doi.org/10.1016/j.aap.2007.10.015

Boucsein, W., & Backs, R. W. (2000). Engineering psychophysiology as a discipline: Historical and theoretical aspects. *Engineering psychophysiology: Issues and applications.*

Bradley, M., & Lang, P. (2007). Motivation and emotion. In J. T. Cacioppo, L. G. Tassinary, & G. G. Bernston (Eds.), *Handbook of Psychophysiology* (2nd ed.), 581–608.

Bradley, M. M., & Lang, P. J. (2000). Measuring Emotion: Behavior, Feeling, and Physiology. In R., D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion*, 242–276. Oxford University Press.

Braithwaite, J. J., Derrick, D., Watson, G., Jones, R., & Rowe, M. (2013). *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments {via the Biopac MP36R & AcqKnowledge software}*.

Brost, M., Gebhardt, L., Karnahl, K., Deißer, O., Steiner, T., Ademeit, A., Brandies, A., Sippel, T., Velimsky, J., Müller, A., & Ulmer, F. (2019). *Reallabor Schorndorf - Entwicklung und Erprobung eines bedarfsgerechten Bussystems*. www.dlr.de

Brouwer, A. M., van Dam, E., van Erp, J. B. F., Spangler, D. P., & Brooks, J. R. (2018). Improving real-life estimates of emotion based on heart rate: A perspective on taking metabolic heart rate into account. *Frontiers in Human Neuroscience*, *12*. https://doi.org/10.3389/fnhum.2018.00284

Chitnis, M., Sorrell, S., Druckman, A., Firth, S. K., & Jackson, T. (2013). Turning lights into flights: Estimating direct and indirect rebound effects for UK households. *Energy Policy*, *55*. https://doi.org/10.1016/j.enpol.2012.12.008

Christopoulos, G. I., Uy, M. A., & Yap, W. J. (2019). The Body and the Brain: Measuring Skin Conductance Responses to Understand the Emotional Experience. *Organizational Research Methods*, *22*(1), 394–420. https://doi.org/10.1177/1094428116681073

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. Hillside. *NJ: Lawrence Earlbaum Associates*.

Cooke, A., Kavussanu, M., Gallicchio, G., Willoughby, A., Mcintyre, D., & Ring, C. (2014). Preparation for action: Psychophysiological activity preceding a motor skill as a function of expertise, performance outcome, and psychological pressure. *Psychophysiology*, *51*(4). https://doi.org/10.1111/psyp.12182

Correia, G. H. D. A., Jorge, D. R., & Antunes, D. M. (2014). The added value of accounting for users flexibility and information on the potential of a station-based one-way car-sharing system: An application in Lisbon, Portugal. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *18*(3). https://doi.org/10.1080/15472450.2013.836928

Detjen, H., Geisler, S., & Schneegass, S. (2020). *Maneuver-based Driving for Intervention in Autonomous Cars*.

Dia, H., & Javanshour, F. (2017). Autonomous Shared Mobility-On-Demand: Melbourne Pilot Simulation Study. *Transportation Research Procedia*, *22*, 285–296. https://doi.org/10.1016/j.trpro.2017.03.035

Dobbs, R., Smit, S., Remes, J., Manyika, J., Roxburgh, C., & Restrepo, A. (2011). *Urban world: Mapping the economic power of cities*.

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

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*. https://doi.org/10.1016/j.tra.2015.04.003

Fagnant, D. J., & Kockelman, K. M. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. *Transportation*, *45*(1). https://doi.org/10.1007/s11116-016-9729-z

Fleury, S., Tom, A., Jamet, E., & Colas-Maheux, E. (2017). What drives corporate carsharing acceptance? A French case study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *45*. https://doi.org/10.1016/j.trf.2016.12.004

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

Gainotti, G. (2019). A historical review of investigations on laterality of emotions in the human brain. *Journal of the History of the Neurosciences*, *28*(1). https://doi.org/10.1080/0964704X.2018.1524683

Gelman, Andrew., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Grant, C. C., Murray, C., Janse van Rensburg, D. C., & Fletcher, L. (2013). A comparison between heart rate and heart rate variability as indicators of cardiac health and fitness. *Frontiers in Physiology*, *4*. https://doi.org/10.3389/fphys.2013.00337

Gupta, S., Buriro, A., & Crispo, B. (2019). DriverAuth: Behavioral biometric-based driver authentication mechanism for on-demand ride and ridesharing infrastructure. *ICT Express*, *5*(1). https://doi.org/10.1016/j.icte.2018.01.010

Hahn, A., Pakusch, C., & Stevens, G. (2020). Die Zukunft der Bushaltestelle vor dem Hintergrund von Mobility-as-a-Service – Eine qualitative Betrachtung des öffentlichen Personennahverkehrs in Deutschland. *HMD Praxis Der Wirtschaftsinformatik*, *57*(2), 348–365. https://doi.org/10.1365/s40702-020-00589-9

Howard, D., & Dai, D. (2013). Public Perceptions of Self-driving Cars: The Case of Berkeley, California. *93rd Annual Meeting of the Transportation Research Board*.

International Transport Forum. (2010). *Transportation Outlook 2010: The Potential for Innovation*.

International Transport Forum. (2014). *Autonomous Driving. Regulatory Issues*.

Jambukia, S. H., Dabhi, V. K., & Prajapati, H. B. (2015). Classification of ECG signals using machine learning techniques: A survey. *Conference Proceeding - 2015 International Conference on Advances in Computer Engineering and Applications, ICACEA 2015*. https://doi.org/10.1109/ICACEA.2015.7164783

Kasos, K., Kekecs, Z., Kasos, E., Szekely, A., & Varga, K. (2018). Bilateral electrodermal activity in the active-alert hypnotic induction. *International Journal of Clinical and Experimental Hypnosis*, *66*(3). https://doi.org/10.1080/00207144.2018.1460551

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing*, *9*(4), 269–275. https://doi.org/10.1007/s12008-014-0227-2

Koo, J., Shin, D., Steinert, M., & Leifer, L. (2016). Understanding driver responses to voice alerts of autonomous car operations. *International Journal of Vehicle Design*, *70*(4). https://doi.org/10.1504/IJVD.2016.076740

Körber, M., Prasch, L., & Bengler, K. (2018). Why Do I Have to Drive Now? Post Hoc Explanations of Takeover Requests. *Human Factors*, *60*(3). https://doi.org/10.1177/0018720817747730

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394-421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Krueger, R., Rashidi, T. H., & Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *69*. https://doi.org/10.1016/j.trc.2016.06.015

Kuno, Y. (1965). The mechanism of human perspiration and sweating in exercise. *Japanese Journal of Physical Fitness and Sports Medicine*, *14*(2). https://doi.org/10.7600/jspfsm1949.14.1

Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., Liedlgruber, M., Wilhelm, F., Osborne, T., & Pykett, J. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors (Switzerland)*, *19*(17). https://doi.org/10.3390/s19173805

Lanatà, A., Valenza, G., Greco, A., Gentili, C., Bartolozzi, R., Bucchi, F., Frendo, F., & Scilingo, E. P. (2014). How the Autonomic nervous system and driving style change with incremental stressing conditions during simulated driving. *IEEE Transactions on Intelligent Transportation Systems*, *16*(3). https://doi.org/10.1109/TITS.2014.2365681

le Vine, S., & Polak, J. (2019). The impact of free-floating carsharing on car ownership: Early-stage findings from London. *Transport Policy*, *75*. https://doi.org/10.1016/j.tranpol.2017.02.004

Li, S., Blythe, P., Guo, W., & Namdeo, A. (2019). Investigating the effects of age and disengagement in driving on driver's takeover control performance in highly automated vehicles. *Transportation Planning and Technology*, *42*(5). https://doi.org/10.1080/03081060.2019.1609221

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology*, *8*(5). https://doi.org/10.1111/j.1469-8986.1971.tb00501.x

Machado, C. A. S., Hue, N. P. M. de S., Berssaneti, F. T., & Quintanilha, J. A. (2018). An overview of shared mobility. *Sustainability, 10*(12). https://doi.org/10.3390/su10124342

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00002

Maghraoui, O. al, Vosooghi, R., Mourad, A., Kamel, J., Puchinger, J., Vallet, F., & Yannou, B. (2020). Shared Autonomous Vehicle Services and User Taste Variations: Survey and Model Applications. *Transportation Research Procedia*, *47*, 3–10. https://doi.org/10.1016/j.trpro.2020.03.066

Martinez, L. M., Correia, G. H. A., & Viegas, J. M. (2015). An agent-based simulation model to assess the impacts of introducing a shared-taxi system: An application to Lisbon (Portugal). *Journal of Advanced Transportation*, *49*(3). https://doi.org/10.1002/atr.1283

Martinez, L. M., & Viegas, J. M. (2017). Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal. *International Journal of Transportation Science and Technology*, *6*(1). https://doi.org/10.1016/j.ijtst.2017.05.005

Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. *Handbuch Methoden der empirischen Sozialforschung*. https://doi.org/10.1007/978-3-658-21308-4_42

McCraty, R., Atkinson, M., Tomasino, D., & Bradley, R. T. (2009). The Coherent Heart Heart-Brain Interactions, Psychophysiological Coherence, and the Emergence of System-Wide Order. *REVIEW December, 5*(2).

McCraty, R., & Shaffer, F. (2015). Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. *Global Advances in Health and Medicine, 4*(1), 46-61. https://doi.org/10.7453/gahmj.2014.073

Meurer, J., Pakusch, C., Stevens, G., Randall, D., & Wulf, V. (2020). A wizard of oz study on passengers' experiences of a robo-taxi service in real-life settings. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 1365–1377. https://doi.org/10.1145/3357236.3395465

Miller, J., & How, J. P. (2016). *Predictive Positioning and Quality Of Service Ridesharing for Campus Mobility On Demand Systems*. http://arxiv.org/abs/1609.08116

Nordhoff, S., de Winter, J., Payre, W., van Arem, B., & Happee, R. (2019). What impressions do users have after a ride in an automated shuttle? An interview study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *63*, 252–269. https://doi.org/10.1016/j.trf.2019.04.009

On-Road Automated Driving (ORAD) Committee. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*.

Pakusch, C., Bossauer, P., Shakoor, M., & Stevens, G. (2016). Using, sharing, and owning smart cars: A future scenario analysis taking general socio-technical trends into account. *ICETE 2016 - Proceedings of the 13th International Joint Conference on e-Business and Telecommunications*, *2*. https://doi.org/10.5220/0005960900190030

Pakusch, C., Stevens, G., Boden, A., & Bossauer, P. (2018). Unintended effects of autonomous driving: A study on mobility preferences in the future. *Sustainability*, *10*(7). https://doi.org/10.3390/su10072404

Pham, T., Lau, Z. J., Chen, S. H. A., & Makowski, D. (2021). Heart rate variability in psychology: A review of HRV indices and an analysis tutorial. *Sensors, 21*(12). https://doi.org/10.3390/s21123998

Picard, R. W., Fedor, S., & Ayzenberg, Y. (2016). Multiple Arousal Theory and Daily-Life Electrodermal Activity Asymmetry. *Emotion Review*, *8*(1), 62–75. https://doi.org/10.1177/1754073914565517

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., van Willigen, B., & Maintainer, R. (2019). Package 'nlme'. Linear and Nonlinear Mixed Effects Models. *Version*.

Rehme, M., Wehner, U., Döring, J., & Rother, S. (2021). Smarte Mobilitätsketten als Zukunftsbild für den ländlichen Raum. *26. Interdisziplinäre Wissenschaftliche Konferenz Mittweida*, 170-174.

Requia, W. J., Higgins, C. D., Adams, M. D., Mohamed, M., & Koutrakis, P. (2018). The health impacts of weekday traffic: A health risk assessment of PM2.5 emissions during congested periods. *Environment International*, *111*. https://doi.org/10.1016/j.envint.2017.11.025

Rigole, P.-J. (2014). *Study of a Shared Autonomous Vehicles Based Mobility Solution in Stockholm*.

Rode, P., Floater, G., Thomopoulos, N., Docherty, J., Schwinger, P., Mahendra, A., & Fang, W. (2017). *Accessibility in Cities: Transport and Urban Form*. https://doi.org/10.1007/978-3-319-51602-8_15

Roelofs, K. (2017). Freeze for action: Neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1718). https://doi.org/10.1098/rstb.2016.0206

Ross, E. D. (2021). Differential hemispheric lateralization of emotions and related display behaviors: Emotion-type hypothesis. *Brain Sciences*, *11*(8). https://doi.org/10.3390/brainsci11081034

Rothenbücher, D., Li, J., Sirkin, D., Mok, B., & Ju, W. (2016). Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*. https://doi.org/10.1109/ROMAN.2016.7745210

Ruijten, P. A. M., Terken, J. M. B., & Chandramouli, S. N. (2018). Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction*, *2*(4). https://doi.org/10.3390/mti2040062

Salahuddin, L., Cho, J., Jeong, M. G., & Kim, D. (2007). Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. *29th Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*. https://doi.org/10.1109/IEMBS.2007.4353378

Santos, G. (2018). Sustainability and shared mobility models. *Sustainability, 10*(9). https://doi.org/10.3390/su10093194

Schneider, T., Hois, J., & Rosenstein, A. (2021, May 6). ExplAIn Yourself! Transparency for Positive UX in Autonomous Driving. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3411764.3446647

Schraagen, J. M., Lopez, S. K., Schneider, C., Schneider, V., Tönjes, S., & Wiechmann, E. (2021). The Role of Transparency and Explainability in Automated Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 27–31. https://doi.org/10.1177/1071181321651063

Selye, H. (1936). A syndrome produced by diverse nocuous agents. *Nature, 138*(3479). https://doi.org/10.1038/138032a0

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01040

Shaheen, S., Cohen, A., & Zohdy, I. (2016). Shared Mobility: Current Practices and Guiding Principles. *Fhwa-Hop-16-022 2.*, *Washington D.C.*

Shen, Y., Jiang, S., Chen, Y., Yang, E., Jin, X., Fan, Y., & Campbell, K. D. (2020). *To Explain or Not to Explain: A Study on the Necessity of Explanations for Autonomous Vehicles*. http://arxiv.org/abs/2006.11684

Stanton, N. A., & Young, M. S. (1998). Vehicle automation and driving performance. *Ergonomics*, *41*(7). https://doi.org/10.1080/001401398186568

Stevens, G., Bossauer, P., Vonholdt, S., & Pakusch, C. (2019). Using time and space efficiently in driverless cars: Findings of a co-design study. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300635

Sztajzel, J. (2004). Heart rate variability: A noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Medical Weekly, 134*(35-36), 514-522.

Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). Heart rate variability artefacts in electrodermal activity. *Circulation*, *93*(5), 1043–1065.

Thammasan, N., Stuldreher, I., Wismeijer, D., Poel, M., van Erp, J., & Brouwer, A. M. (2019). A novel, simple and objective method to detect movement artefacts in electrodermal activity. *2019 8th International Conference on Affective Computing and Intelligent Interaction, ACII 2019*. https://doi.org/10.1109/ACII.2019.8925512

Tussyadiah, I. P., Zach, F. J., & Wang, J. (2017). *Attitudes toward Autonomous on Demand Mobility System: The Case of Self-Driving Taxi*.

Umetani, K., Singer, D. H., Mccraty, R., & Atkinson, M. (1998). Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender Over Nine Decades Downloaded. *Journal of the American College of Cardiology, 31*(3), 593-601.

UN HABITAT III. (2016). *Policy Paper 9: Urban Services and Technology*, 31.

Walker, G. H., Stanton, N. A., & Salmon, P. (2016). Trust in vehicle technology. *International Journal of Vehicle Design*, *70*(2). https://doi.org/10.1504/IJVD.2016.074419

Wiegand, G., Weber, T., Hussmann, H., Schmidmaier, M., & Liu, Y. (2019). I drive - You trust: Explaining driving behavior of autonomous cars. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290607.3312817

Wyczesany, M., Capotosto, P., Zappasodi, F., & Prete, G. (2018). Hemispheric asymmetries and emotions: Evidence from effective connectivity. *Neuropsychologia*, *121*. https://doi.org/10.1016/j.neuropsychologia.2018.10.007

Yanagida, R., Takahashi, K., Miura, M., Nomura, M., Ogawa, Y., Aoki, K., & Iwasaki, K.-i. (2016). Speed ratio but cabin temperature positively correlated with increased heart rates among professional drivers during car races. *Environmental Health and Preventive Medicine*, *21*(6), 439–445. https://doi.org/10.1007/s12199-016-0544-0

Zhang, Q., Jessie Yang, X., & Robert Jr, L. P. (2021). Individual Differences and Expectations of Automated Vehicles. *International Journal of Human-Computer Interaction*, *2*.

Zhang, Q., Jessie Yang, X., & Robert, L. P. (2021). What and When to Explain? A Survey of the Impact of Explanation on Attitudes Toward Adopting Automated Vehicles. *IEEE Access*, *9*, 159533–159540. https://doi.org/10.1109/ACCESS.2021.3130489

Zhang, Q., Yang, X. J., & Robert, L. P. (2021). Drivers' age and automated vehicle explanations. *Sustainability (Switzerland)*, *13*(4), 1–19. https://doi.org/10.3390/su13041948

Zhang, R., Spieser, K., Frazzoli, E., & Pavone, M. (2015). Models, algorithms, and evaluation for autonomous mobility-on-demand systems. *2015 Proceedings of the American Control Conference*, 2573-2587. https://doi.org/10.1109/ACC.2015.7171122

Appendix

A1 Image of the vehicle

Figure A1.1: The VIEWCar II, in which the drive took place, is depicted. Available: https://www.dlr.de/content/de/grossforschungsanlagen/view-car.html [accessed 11.07.2022]

Figure A1.2: The inside of the vehicle where the participant sat is shown in the image. A partition between the front and back seats blocked the sight through the front window and the sight on the driver. The tablet functioned as a screen for information display.

Figure A1.3: The participant sat on the right back seat during the drive. One of the EDA-sensors can be seen on the participant's left wrist with two electrodes in the palm of the hand. In the other hand she is holding the smartphone that was used for navigation and subjective ratings.

A2 Experimental Set-up

Figure A2.1: Information display on the way to the book case.

Figure A2.2: Information display on the way back from the book case.

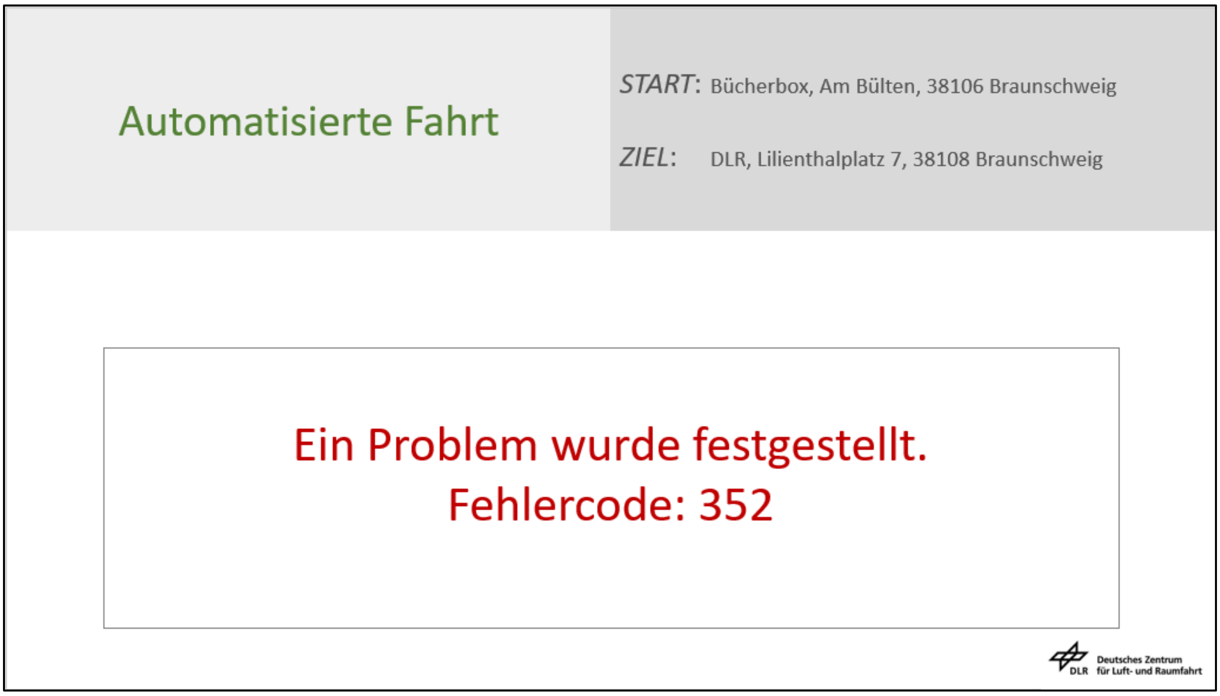


Figure A2.3: Error notice during the abrupt braking event.

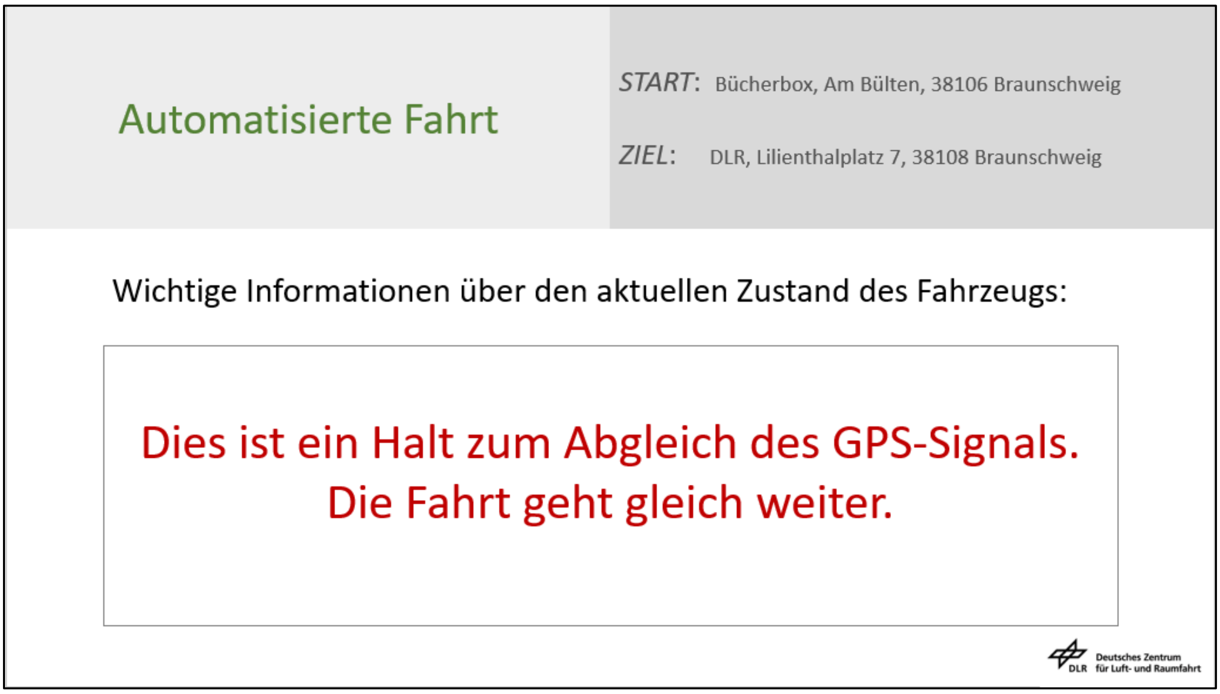


Figure A2.4: Content information display during the abrupt braking event.

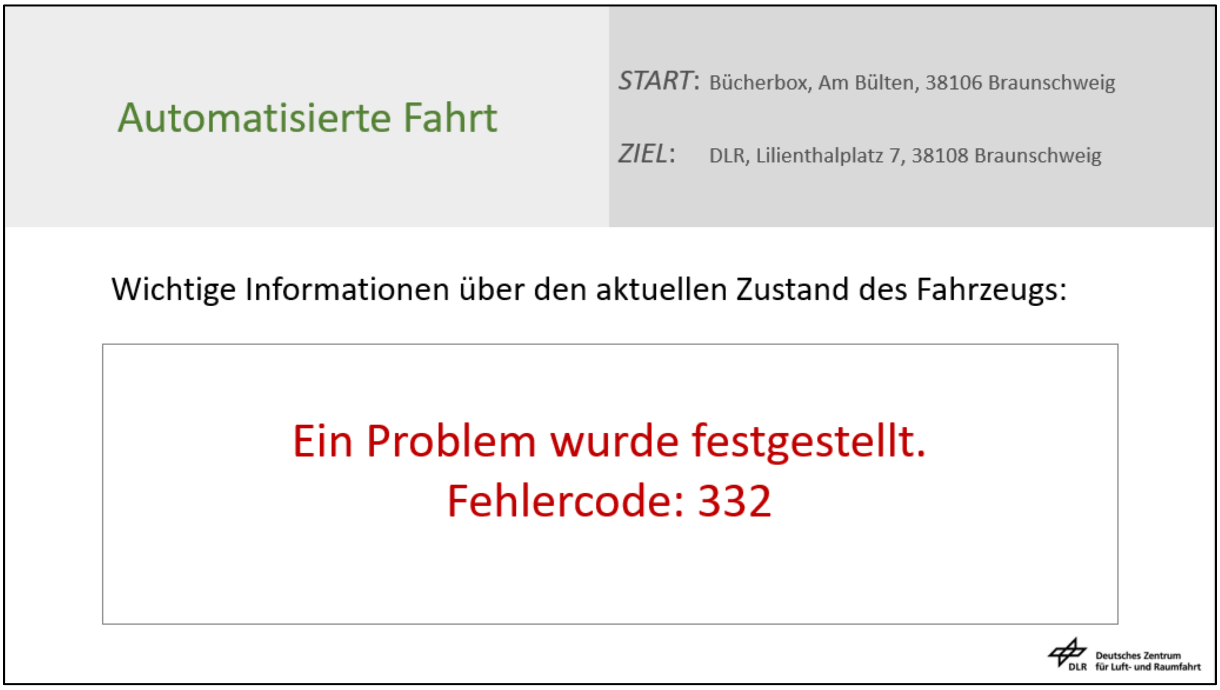


Figure A2.5: Error notice during the unexpected detour.

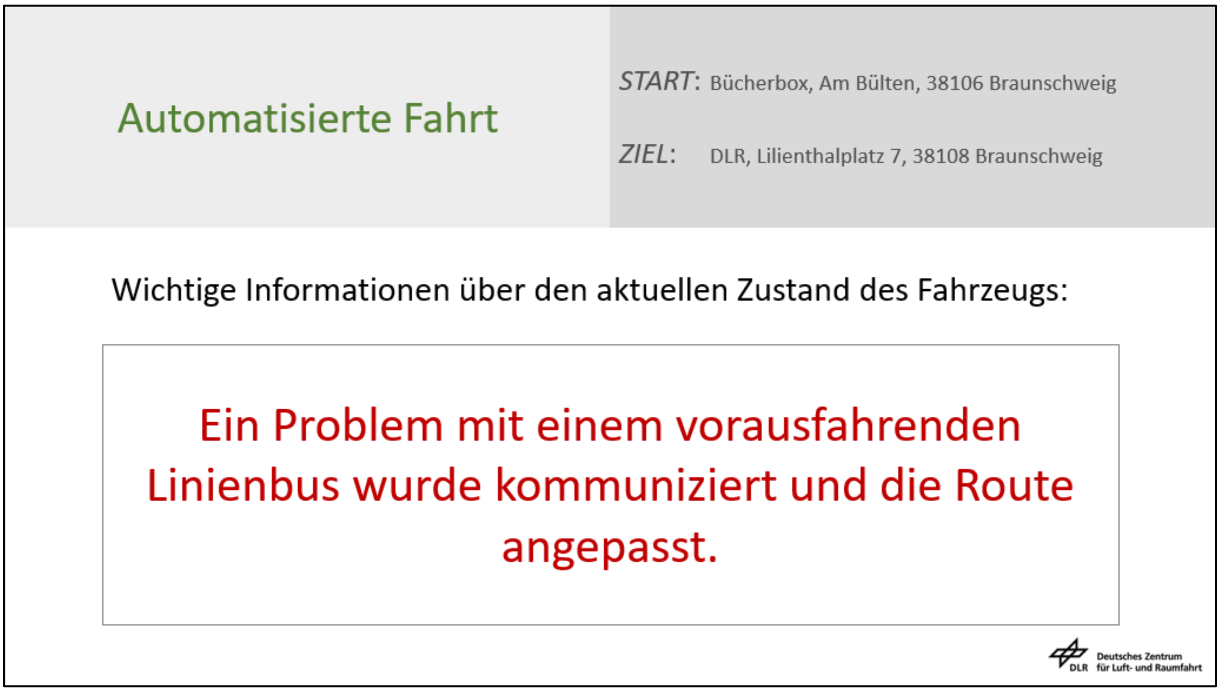


Figure A2.6: Content information display during the unexpected detour.

A3 Post-questionnaire

Bitte bewerten Sie die folgenden Aussagen bezüglich Ihres Empfindens. Es geht um die Situation, als das automatisierte Shuttle auf der Fahrt von der Bücherbox zum DLR rechts rangefahren ist und kurz angehalten hat.

- Das Display im automatisierten Shuttle hat mir geholfen, die aktuellen Handlungen des Shuttles zu verstehen.
- In der Situation fühlte ich aufgrund fehlender Informationen unsicher.
- In der Situation habe ich mir mehr Informationen gewünscht.

Offene Frage: Die folgenden Informationen hätten mir geholfen, die Situation besser zu verstehen

Bitte bewerten Sie die folgenden Aussagen bezüglich Ihres Empfindens. Es geht um die Situation, als das automatisierte Shuttle auf der Fahrt von der Bücherbox zum DLR einen Umweg durch das Wohngebiet fahren musste.

- Das Display im automatisierten Shuttle hat mir geholfen, die aktuellen Handlungen des Shuttles zu verstehen.
- In der Situation fühlte ich aufgrund fehlender Informationen unsicher.
- In der Situation habe ich mir mehr Informationen gewünscht.

Offene Frage: Die folgenden Informationen hätten mir geholfen, die Situation besser zu verstehen

Bitte bewerten Sie die folgenden Aussagen bezüglich Ihres Empfindens. Es geht um die gesamte Fahrt mit dem automatisierten Shuttle.

- Das Display im automatisierten Shuttle hat mir geholfen, die aktuellen Handlungen des Shuttles zu verstehen.
- Während der Fahrt fühlte ich auf Grund fehlender Informationen unsicher.
- Während der Fahrt habe ich mir mehr Informationen gewünscht.

Offene Frage: Beschreiben Sie kurz bis zu 3 Situationen, in denen Sie sich während der Fahrt im Shuttle mehr Informationen gewünscht hätten.

<u>Notice:</u> The statements, apart from the open-ended questions, had to be rated by the participants on a 5-point Likert scale from one, meaning total disagreement, to five, meaning total agreement.

A4 Calculation of nRMSSD for the time intervals of the events

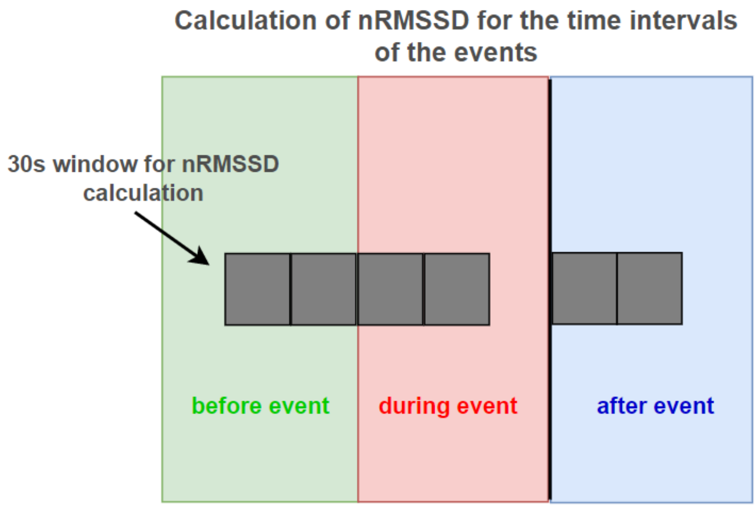


Figure A4: The figure depicts how the nRMSSD values per participant were calculated. Complete 30s time windows were used to calculate the nRMSSD values for the time intervals of the events.

For the linear mixed effects analysis, the participants' physiological reaction during the time intervals before, during and after an event were compared. The nRMSSD values were calculated manually for each participant, as described in the data preparation section, to ensure that as many complete 30s windows could be used when comparing the time intervals. The last complete windows of the before event interval and the first complete windows of the during event time interval are considered. For the time interval after the event, the first complete windows were examined. (Figure A4). For the abrupt braking event which took about 60s, two complete windows and therefore two nRMSSD values per time interval were considered. The duration of the unexpected detour (~2min) varied between the participants such that for some participants three and for others four nRMSSD values per time interval could be considered.

A5 Data exploration – a first impression of the physiological data

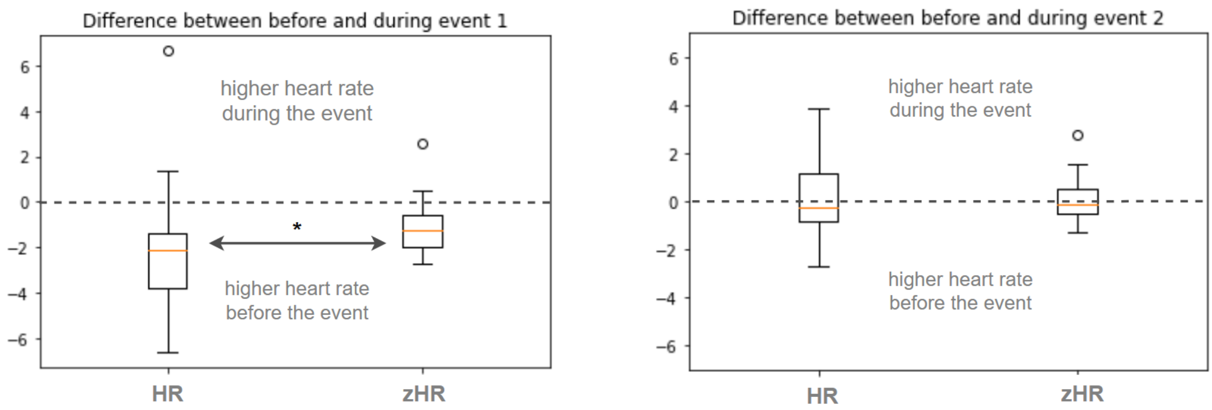


Figure A5.1: The difference between the mean heat rates of the time interval before and during the events of each participant is contained in the boxplots. The left figure depicts the results for the first event and the right figure shows the results for the second event. In each figure, the left boxplot contains the difference values which were calculated with the mean HR. The right boxplot contains the difference values which were calculated with the zHR. Negative differences indicate higher heart rates in the time interval before the event compared to during the event, while positive differences indicate the opposite. Boxplot parameters: box = middle 50%, line = median, length of whiskers = interquartile range (IQR) between first and third quartile multiplied by +/-1.5 with the last data point defining the end of the whisker, circle = outlier.

To get a first impression of the physiological reactions of the participants, the difference in the participants' heart rates between before and during the events were examined. A difference would indicate a change in the participants' cardiological activity as a response to an occurred event. At the same time, the aim was to discover whether considering the zHR over the HR would ameliorate the quality of the analysis for the current data set. Physiological data normally has a strong individual component which can be reduced through z-transformation. The extent of the inter-participant difference, which indicates the importance of z-transforming the HR was explored. For each participant and each event, the mean HR and zHR was calculated for the time intervals before and during the event and the difference was taken.

To investigate the latter, a Wilcoxon sign-rank test for each event separately was performed with the time interval differences. There exists a significant difference between the HR and zHR data in the case of the first event ($p=0.0006$). Examining the boxplots in Figure A5.1 (left), the boxplot zHR is more compact compared to the boxplot containing the HR, which indicates less inter-participant variation in the data. No significant difference between HR and zHR could be found in the case of the second event ($p=0.9881$). However, the boxplot containing the zHR differences is also more compact than the boxplot with the HR, as was the case in the first event (Figure A5.1 (right)). A high inter-participant difference in HR exists such that a normalization of physiological data seems to be advisable to reduce the individual component and receive robust results for the set of participants in the current study.

The change in the participants' cardiological activity from before to during the event can be seen in Figure A5.1. In the case of both events, the median of the differences between the time intervals is negative. The HR or zHR was therefore higher before the event than during the event. The events seem to have led to a decrease in heart rate. This observation was further analysed with a linear mixed effects analysis.