

Automation Transparency: Implications of Uncertainty Communication for Human-Automation

Interaction and Interfaces

Alexander Kunze

Loughborough University

Stephen J. Summerskill

Loughborough University

Russell Marshall

Loughborough University

Ashleigh J. Filtness

Loughborough University

a.kunze@lboro.ac.uk, LDS.1.23, Loughborough Design School, Loughborough University, LE11

3TG Loughborough, United Kingdom

Abstract (150 words)

Operators of highly automated driving systems may exhibit behaviour characteristic for overtrust issues due to an insufficient awareness of automation fallibility. Consequently, situation awareness in critical situations is reduced and safe driving performance following emergency takeovers is impeded. A driving simulator study was used to assess the impact of dynamically communicating system uncertainties on monitoring, trust, workload, takeovers, and physiological responses. The uncertainty information was conveyed visually using a stylised heart beat combined with a numerical display and users were engaged in a visual search task. Multilevel analysis results suggest that uncertainty communication helps operators calibrate their trust and gain situation awareness prior to critical situations, resulting in safer takeovers. Additionally, eye tracking data indicate that operators can adjust their gaze behaviour in correspondence with the level of uncertainty. However, conveying uncertainties using a visual display significantly increases operator workload and impedes users in the execution of non-driving related tasks.

Keywords: vehicle ergonomics; attitudes and behaviour; attention and vigilance; human-computer-interaction; information displays

Practitioner summary (50 words): This paper illustrates how the communication of system uncertainty information helps operators calibrate their trust in automation and, consequently, gain situation awareness. Multilevel analysis results of a driving simulator study affirm the benefits for trust calibration and highlight that operators adjust their behaviour according to multiple uncertainty levels.

Automation Transparency: Implications of Uncertainty Communication for Human-Automation
Interaction and Interfaces

Interaction with imperfect, inanimate agents

Knowingly or subliminally, humans routinely react to and account for uncertainties. Our decisions and actions are, to some degree, affected by the uncertainties associated with the information they are based on (Preuschhoff, Mohr, & Hsu, 2013).

The impact of uncertainties on decision-making and action selection is thereby not limited to humans but extends to inanimate, automated agents in equal measure (Gal, 2016). Incorrect or incomplete inputs and imperfect or simplified algorithms induce uncertainties into automated systems, promoting unexpected or erratic behaviour. Considering that automated technologies are integrated into safety-critical systems such as aircraft, vehicles, and power plants whose failures can lead to devastating consequences, the operator's knowledge of the system's inherent uncertainties is essential for safe performance.

Yet, cases of inappropriate trust and over-reliance indicate that humans are not sufficiently aware and critical of the fallibility and limitations of automated agents when interacting with them (Manzey, Reichenbach, & Onnasch, 2012; Parasuraman & Riley, 1997). The human operator, however, cannot be identified as the primary culprit. Rather, the fault lies with the designed human-automation interaction, in particular its opacity (Endsley, 2017; Norman, 1989). Informing operators about the system's inherent uncertainties could help operators identify when they may be required to take over the system, thereby improving safety.

This paper investigates the influence of communicating automated system uncertainties on operator behaviour within a driving context. An anthropomorphic interface for dynamically communicating system uncertainties and a driving simulator protocol were used to

comprehensively evaluate the impact of uncertainty communication under consideration of workload, eye movement and physiological measures in addition to trust, takeover performance and situation awareness (SA).

Uncertainty Communication

The term *uncertainty* signifies *doubt* and implies a lack of exact knowledge (BIPM, 2008), whereby its specific definition depends on the usage context (Uggirala, Gramopadhye, Melloy, & Toler, 2004; Zimmermann, 2000). Within this publication, the term includes all uncertainties induced to automated systems during data acquisition, transformation, and output generation (see Figure 1) (Endsley & Jones, 2012; Kiureghian & Ditlevsen, 2007; Pang, Wittenbrink, & Lodha, 1997; Uggirala et al., 2004).

Implications for Appropriate Trust, Attention Allocation and SA

Automating systems under the expectation that a fallback-ready user can resume the previously automated task in case of system failures entails several human factors challenges, foremostly a lack in SA at the time of a system failure (Endsley, 1995; Endsley & Kiris, 1995).

Several models for SA exist (Stanton, Salmon, Walker, Salas, & Hancock, 2017), with the Endsley (1995) model of SA being the most cited and widely used (Golightly, Wilson, Lowe, & Sharples, 2010). In this model, SA is interpreted as three ascending, but not necessarily linear levels (Endsley, 2015): the first level refers to the spatial and time-specific perception of elements and events in the environment; level 2 is the comprehension of these elements and events; level 3 concerns the projection of their future state.

In the context of driving automation, SA can be gained by glancing towards the field relevant for driving (FRD) and by surveying the operational state of the automated system, supported by information provided by the automation interface (Baumann & Krems, 2007). To assist users

with the acquisition and maintenance of SA, their attention management can be influenced through trust calibration (Hergeth, Lorenz, Vilimek, & Krems, 2016; Lee & See, 2004; Parasuraman & Manzey, 2010).

Attention can be classed into two disparate functions, namely top-down, endogenous attention and bottom-up, exogenous attention (Katsuki & Constantinidis, 2014). The former is an internally induced process based on present goals, conscious plans, and prior knowledge while the latter is an externally induced attention guidance process driven by the relative salience of external stimuli (Connor, Egeth, & Yantis, 2004; Corbetta & Shulman, 2002; Itti, Koch, Way, & Angeles, 2001). Thus, operators can actively allocate their attention to the automation interface and the FRD (endogenous) or the interface can invoke the attentional shift itself, for instance through an increase in the salience of system parameters (exogenous).

As trust is often associated with a willingness to act, it affects endogenous attention (Johns, 1996; Mayer, Davis, & Schoorman, 1995). Lee and See (2004) identified three essential aspects for achieving appropriate trust: calibration, resolution, and specificity. Calibration refers to the agreement between the user's trust in automation and the capabilities of the automation. Resolution is the degree to which the user's trust judgement differentiates different capability levels. Specificity refers to the degree of differentiation between different components or aspects of the trustee. A combination of calibrated trust as well as a high resolution and specificity facilitates appropriate trust and thereby helps to alleviate both disuse and misuse of automation. Systems can support the user with the buildup of appropriate trust by providing information along the dimensions abstraction and detail (Lee & See, 2004). Abstraction refers to information regarding the performance, process, and purpose of the automation (Lee & Moray, 1992), with the former being the most crucial (Hoff & Bashir, 2015). Detail describes the entity that is to be

trusted, ranging from the overall system to its single functions and modes (Lee & See, 2004).

The provision of information along the two presented dimensions does, however, not exclusively determine the trust of the operator. As a result of a meta-analysis of studies, Hoff & Bashir (2015) define three distinct layers of trust, namely dispositional, situational, and learned trust.

By providing information along the dimension abstraction, for instance about a system's inherent uncertainties, to a certain detail, for instance regarding the overall system, the dynamic learned component of trust can be affected to achieve a more appropriate trust calibration (Hoff & Bashir, 2015). Dynamic learned trust refers to the trust component that is affected by the users' current interaction with the system as mediated by design features and prior knowledge about the system's capabilities. The remaining trust components are not considered to be varying during interactions but are rather determined in advance of the interaction (Hoff & Bashir, 2015).

A closely related concept is that of complacency, which 'represents an active reallocation of attention away from the automation to other manual tasks in cases of high workload (Parasuraman & Manzey, 2010)'. Complacent behaviour is thereby thought to be influenced by trust (Parasuraman & Manzey, 2010) and was shown to be particularly significant in highly automated systems that fail only on a single occasion (Bailey & Scerbo, 2007; de Waard, van der Hulst, Hoedemaeker, & Brookhuis, 1999; Molloy & Parasuraman, 1996). An attention allocation strategy that places more attention on competing tasks rather than the automated system may be the result of an initially high trust in system capabilities which is further reinforced if the system performs faultlessly, thereby promoting complacent behaviour (Parasuraman & Manzey, 2010).

Complacency is a predictor of visual attention (Bailey & Scerbo, 2007; Metzger & Parasuraman, 2005; Parasuraman & Manzey, 2010) and, as a result, affects failure detection. Similar to trust, complacency is directly related to automation reliability (Bagheri & Jamieson,

2004; Wickens, Sebok, Li, Sarter, & Gacy, 2015) and, additionally, to the fluctuation of that reliability (May, Molloy, & Parasuraman, 1993). Even if operators exhibit appropriate trust levels without complacent behaviour and therefore maintain an optimal attention allocation strategy, failures may be missed (Moray & Inagaki, 2000). Particularly highly reliable automated systems that justify increased trust levels afford users to allocate more attention to concurrent tasks and therefore decrease the likelihood of an overlap between system failures and monitoring. Thus, systems should ideally support the operator with exogenous stimuli that shift the users' attention towards the automation in critical situations, whereby detrimental secondary effects of alarms, such as the cry wolf effect (Yang, Unhelkar, Li, & Shah, 2017), are to be avoided.

An additional challenge relevant to automated driving is the presence of concurrent tasks, which further promote the attention allocation away from the automated task (Wickens & Dixon, 2007). Thereby, the attention allocation to concurrent tasks may be expanded with an increasing degree of automation (Carsten, Lai, Barnard, Jamson, & Merat, 2012). Moreover, workload affects the attention allocation strategy of operators (Young & Stanton, 2002), as does the type of concurrent tasks (Diekfuss, Ward, & Raisbeck, 2017). This emphasises that not only the implications of uncertainty communication on trust are relevant, but also those on operator workload under consideration of concurrent tasks.

Figure 2 summarises the outlined relationships between the factors influencing SA with a focus on trust-related processes. For the purpose of comprehensiveness and in accordance with the model proposed by Endsley (2017), the influence of an operator's mental model on SA is taken into account. Further, specific aspects of the human-machine interface that were shown to influence trust in a meta-analysis of studies are depicted (Hoff & Bashir, 2015).

The flow of the continuous, green line depicts the five principal steps how the communication of system uncertainties affects the dynamic learned component of trust (1), appropriate trust (2), attention allocation (3), and SA (4,5) under consideration of principal concepts affecting this functional chain, such as concurrent tasks or workload.

Uncertainty Communication Methods

Multiple studies have indicated the specific benefits of conveying uncertainty or reliability information of automated systems to human operators (Beller, Heesen, & Vollrath, 2013; Chen et al., 2018; Helldin, Falkman, Riveiro, & Davidsson, 2013; Louw et al., 2017; Louw & Merat, 2017; Mercado et al., 2016). In a driving context, Beller, Heesen & Vollrath (2013) investigated the impact of displaying a schematised uncertain face in the instrument cluster in unclear situations. The results indicate that the communication of system uncertainties increases driving safety, specifically the time to collision (TTC).

Expanding upon these findings, Helldin, Falkman, Riveiro, & Davidsson (2013) explored the communication of uncertainty using seven bars in the instrument cluster, each bar representing one level. The results show that users who were presented with the uncertainty information could afford to allocate their attention away from the FRD and perform concurrent tasks for a longer time than the control group. Nonetheless, the participants of the experimental group were able to take over the driving task faster than those of the control group.

In line with these findings, studies in aviation and the military confirmed the benefits of presenting uncertainty information on task performance (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Finger & Bisantz, 2002; McGuirl & Sarter, 2003; L. Wang, Jamieson, & Hollands, 2009).

Implications for uncertainty communication on factors other than takeover performance and trust have yet to be investigated. Specifically, the monitoring of an additional display is expected to increase the workload of operators and may therefore affect their attention allocation (Young & Stanton, 2002). Further, the communication of uncertainties may lead to feelings of unease in human operators due to a confrontation with the fallibility of the system and the associated risks, prompting the necessity to take physiological measurements. As suggested by Lee & See (2004), a higher resolution of trust may positively impact its appropriateness, however no studies known to the authors have explored this in a driving context.

Anthropomorphic Uncertainty Display

One way to communicate uncertainty information is by anthropomorphic features, which are a suitable means to promote greater trust in automation (de Visser et al., 2012; Hoff & Bashir, 2015; Pak, Fink, Price, Bass, & Sturre, 2012; Waytz, Heafner, & Epley, 2014). This begs the question which anthropomorphic features are suitable to communicate several levels of uncertainty. While facial expressions, as employed by Beller et al. (2013), may invoke the strongest responses when the intensity difference between expressions is large, subtle differences for communicating multiple uncertainty levels require more attentional resources and are more difficult to identify (Gollan, McCloskey, Hoxha, & Coccaro, 2010). Instead, other natural human responses must be considered. When humans experience (mentally) stressful situations such as making critical decisions based on uncertain information, a series of events is set off amongst which is an increase in heart rate (Hubbard & Workman, 1998; Steptoe & Voegle, 1991). Further, increases in focus and alertness were shown to raise human heart rate (X. Wang, Pinol, Byrne, & Mendelowitz, 2014). While not obvious in other people, the simulated heart rate of a

system can be visualised, resulting in intuitive, natural levels of uncertainty that map human responses to stressful, uncertain situations to an inanimate agent.

Objectives and Hypotheses

The current study differs from previous work in that it (a) evaluates the impact of uncertainty communication more comprehensively, i.e. under consideration of workload, eye movement and physiological measures in addition to trust, takeover performance and situation awareness (SA), (b) analyses the behavioural implications of presenting uncertainty content with higher resolution, and (c) uses a heartbeat animation to dynamically convey uncertainties with multiple levels in a natural, anthropomorphic manner. The following hypotheses were investigated:

1. The experimental group adapts their behaviour better to the visibility range than the control group as indicated by the solving rate of concurrent tasks, ratio of correctly answered SAGAT questions, trust values, and attention allocation.
2. The experimental group performs a safer takeover than the control group as indicated by MTTC, acceleration, and TTC values.
3. The experimental group experiences higher subjective workload than the control group due to having to monitor an additional display.
4. There is an observable relationship between trust and visual attention.
5. Physiological effects differ between groups as indicated by heart rate.

Method

A driving simulator study was conducted to investigate the impact of presenting uncertainty information using a stylised heartbeat on operator trust, attention allocation, SA,

takeover performance, workload, and physiological parameters. Ethical approval was granted by the Ethics Approvals Sub-Committee of Loughborough University.

Participants

A total of 34 participants (14 female) with an average age of 30.82 years (SD = 9.52) volunteered to take part in the experiment. Participants held a driving license for a mean time of 11.18 years (SD = 9.70, range: 1-36) and averaged 6248.53 miles per year (SD = 5122.15).

Experiment Design

The study employed a 2 (uncertainty display, between) \times 4 (fog conditions, within) mixed design. The first factor was the uncertainty display (UD) and had two levels (UD unavailable; UD available). The second factor was the visibility range as a consequence of fog density that was varied in four equidistant levels between 91.44m (thick fog)¹ and 1005.84m (no/low fog) (Met Office UK, 2018). Fog was chosen as an indicator for uncertainty as its small water droplets scatter transmit pulses of lidar radar systems and affect the sensors' reliability (Rasshofer, Spies, & Spies, 2011). Further, fog was considered to be relatable for novice users of an automated vehicle as a reason for system uncertainties. The group to which the UD was available, hereafter referred to as *experimental group*, was made aware of the increase in uncertainty through a change in system heart rate. The conveyed system heart rate changed linearly with the visibility range, with a lower limit of 50bpm indicating a high visibility range of more than 1005.84m and an upper limit of 140bpm signifying a visibility range of less than 91.44m (see Figure 3).

In contrast to closely related work, the current study does not involve multiple takeovers as part of the scenario as this cannot be considered representative of the failure rate of real-world systems. Rather, it follows other publications, e.g. (Molloy & Parasuraman, 1996), that

investigated automated systems that fail on only one occasion. This failure rate is still to be considered higher than realistic, however, it allows for empirically testing participants in a controlled laboratory environment. Further, prior research implies that the successful detection of automation failures decreases with a reduction in failure rates (Davies & Parasuraman, 1982; Parasuraman & Manzey, 2010).

The system failure was communicated to the operator with a takeover request (TOR). The TOR was issued with 7 s TTC remaining (Gold, Damböck, Lorenz, & Bengler, 2013) using combined visual and abstract auditory cues (Bazilinskyy, Petermeijer, Petrovych, Dodou, & de Winter, 2018; Politis, Brewster, & Pollick, 2015).

To account for the influence of concurrent tasks, participants were instructed to engage in a visual search task while the automation was activated. In the visual search task, participants had to identify if a certain target is present within a field of distractors (Treisman & Gelade, 1980), mimicking real-world situations in which operators are engaged in concurrent tasks.

Apparatus

The study was conducted in a static driving simulator consisting of a mock-up Land Rover cockpit and three 160cm × 90cm projection screens, placed 160cm away from the steering wheel with the outer screens inclined towards the driver and generating a 135° horizontal field of view. The simulator employed the STISIM Drive system and three projectors with a combined resolution of 5760 × 1080 pixels projected the rendered images onto the planes. Further, a digital instrument cluster with a resolution of 1280 × 480 pixels was used to display the animated heart beat. It consists of a stylised, animated heartbeat and a number indicating the heart rate in beats per minute (see Figure 4). To incorporate the need for exogenous stimuli, the screen turned red for 500 ms to indicate an increase in heart rate.

The visual search task was performed on an Apple iPad Air 2 firmly attached to the centre console. Eye-tracking data was recorded using Tobii Pro 2 glasses with a gaze sampling frequency of 100 Hz. The data was subsequently analysed with Tobii Pro Lab (Tobii, 2018). The heart rate of participants was measured with a frequency of 1 Hz using a Polar A370 with optical heart rate monitoring capabilities.

Scenario and Procedure

Participants were briefed about the experiment using a pre-recorded video and were subsequently asked to fill in a consent form and a demographic survey. Throughout the experiment, all instructions were presented in a consistent manner to prevent effects of differing introductions on dependent variables such as trust (Körber, Baseler, & Bengler, 2018). As part of the introduction, participants were made aware of the fallibility of automated driving systems and the specific consequences of system failures, particularly TORs. The experimental group was further introduced to the UD.

Following the introduction, participants adjusted seat and steering wheel position and completed several practice scenarios with a fixed distance and a total duration of 30 to 40 minutes that involved multiple emergency manoeuvres and ensured a basic familiarisation with the simulation vehicle.

Immediately after the training sessions, participants were rebriefed about the experiment and subsequently commenced with a 20-min scenario in a vehicle equipped with a conditionally automated driving system (SAE International, 2016). In the scenario, the system drove autonomously on a two-lane UK motorway with moderate traffic, performing a series of lane switches and overtaking manoeuvres while keeping the speed constant at 70 mph. Figure 3

depicts how the visibility range changed due to fog and shows the resulting heart rate on the UD. Further, the graphic indicates when measures of trust and SA were taken.

Takeover

Figure 3 illustrates the takeover situation which was modelled after recent automation failures in which a stopped vehicle caused crashes (Shepardson, 2018; Stewart, 2018). For a successful takeover, operators had to perceive both the stopped vehicle and the vehicle passing on the right-hand side, understand the consequences, and project the future state of the own and surrounding vehicles – thus possess SA. During a test of situation awareness (see dependent variables) midway through the scenario, the situation was replicated, with exception of the stopped vehicle. This way, participants' SA could be directly measured in a situation similar to that of the takeover while avoiding to interfere with the takeover itself.

Immediately following the scenario, participants were asked to fill in a workload questionnaire and were briefly interviewed about how they experienced the experiment.

Dependent Variables

In accordance with related work (Beller et al., 2013; Hergeth et al., 2016), subjective trust was measured using single-item trust ratings on a percentage scale. Thereby, the experimenter prompted participants to answer the following question during dynamic simulation: 'To what percentage do you currently trust the automated system to safely perform the driving task?' A single-item trust scale was preferred over alternative, more comprehensive measures (Jian, Bisantz, & Drury, 2000) as it allows the repeated, time-efficient collection of trust data and thus interferes less with participant behaviour.

Further, the visual attention allocation of operators was recorded employing eye tracking (direct measurement) and a visual search task (indirect). Eye tracking data was analysed using monitoring frequency and ratio (Hergeth et al., 2016):

$$\text{Monitoring frequency} = n_{\text{monitoring glances}} / t_{\text{section}}$$

$$\text{Monitoring ratio} = \sum_{i=1}^n t_{\text{fixation } i} / t_{\text{section}}$$

Monitoring glances were thereby defined as any fixations on driving-related areas of interest (AOIs), including FRD, instruments, and UD. Consecutive fixations to the same AOI were combined if they were separated by blinks of less than 120 ms or if the fixations were shorter than 120 ms (Hergeth et al., 2016; Jacob & Karn, 2003). The monitoring data was analysed for a period of 80 s prior to the TOR. In the first 40 s, i.e. between 80 s and 40 s before the TOR (TOR-80-40), the visibility was high (≥ 3300 ft) and the corresponding heart rate on the UD at its minimum. For the following 40 s up until the TOR (TOR-40-0), the visibility was at its minimum and the heart rate at its maximum. One participant was excluded from the analysis of eye tracking data due to a failure of the processing unit, resulting in a sample size of 33.

Regarding the visual search task, the achieved solving rate was recorded. Additionally, the user heart rate was measured as an indicator for trust (Waytz et al., 2014) as well as stress and participants were asked to fill in a subjective workload assessment questionnaire using 7-point Likert scales in accordance with the NASA-TLX (Hart, 2006; Hart & Staveland, 1988).

To achieve a direct and objective measure of SA, the situation awareness global assessment technique (SAGAT) was used (Endsley, 2000; Nguyen, Lim, Duy Nguyen, Gordon-Brown, & Nahavandi, 2018). For this, the simulation was frozen without prior notification and all displays were blanked before the participant was queried about the current state of the driving environment. To avoid the introduction of noise to other dependent variables such as operator

heart rate, subjective trust or gaze behaviour as a result of freezing the simulation, the SAGAT test was followed by a recovery section where no dependent measures were taken (see Figure 3).

Following the TOR, minimum TTC (MTTC) as well as lateral and longitudinal acceleration values were recorded to determine driving safety and takeover quality. Further, time-to-takeover (TTT) was measured as the time between TOR and first driver input, i.e. change in steering angle, brake or accelerator input.

Data Analysis

To increase validity, multiple measures of dependent variables on individual participants were taken. While this reduces the risk that results are due to chance, it leads to non-independent data points. Instead of aggregating the data, which would not take advantage of all collected data points, multilevel analysis (Snijders & Bosker, 2012) using R and lme4 (Bates et al., 2018) was applied. Unless stated otherwise, random effects were implemented using intercepts for *subjects*. The visibility range was treated as a continuous variable to account for its ordinal characteristics while the presence of the UD was implemented as a categorical variable. The assumption of homoscedasticity was validated through visual inspection of residual plots. Degrees of freedom and p values for t values were approximated based on Satterthwaite's method using lmerTest (Kuznetsova, Brockhoff, & Christensen, 2018).

Results

Hypothesis 1: Behavioural implications

Solving Rate of Concurrent Tasks

The impact of UD and visibility range on the solving rate of concurrent tasks was assessed using a linear mixed model (LMM) under consideration of the interaction between the two independent variables (see Table 3). The interaction between UD and visibility range had a

significant effect ($t(304.00) = 4.914, p < .001$), indicating that the visibility range affected the solving rate depending on the presence of the UD. Figure 5 depicts the effects of the LMM, showing that the solving rate was significantly impacted by the visibility range if the UD was present. Irrespective of the visibility range, the mean solving rate was lower for the experimental group, indicating significant differences between groups in addition to the interaction.

SAGAT responses

Table 1 shows the ratio of correctly answered SAGAT questions for each group. For all individual questions, the experimental group achieved a higher correct response ratio. Across questions there was a significant difference regarding the frequency of correct replies between groups, as indicated by the results of a chi-square test with Yates' continuity correction, $\chi^2(1) = 4.349, p = .037$.

Subjective Trust

The impact of UD and visibility range on trust was evaluated using a LMM under consideration of interactions. The interaction between visibility range and UD was significant ($t(236.00) = 6.795, p < .001$), whereby subjective trust ratings only varied with the visibility range when the UD was present (see Figure 6). Further, there was a main effect of UD on subjective trust ratings. However, the effect cannot be considered meaningful as the trust ratings are not lower for the experimental group regardless of the visibility range (see Figure 6).

Visual Attention

Results of a LMM with monitoring frequency as a dependent variable and UD as well as Section (TOR-80-40, TOR-40-0) as predictors showed a significant main effect of UD ($t(44.16) = 6.877, p < .001$) as well as a significant interaction, $t(32.00) = -4.752, p < .001$. Figure 8 illustrates the effects of the interaction, showing that the monitoring frequency

increased only for the experimental group between the two sections. Further, it confirms the main effect of UD on monitoring frequency.

Figure 9 depicts the monitoring ratio as another indicator for visual attention. For the control group, the distribution of attention did not significantly change between the two analysed sections. The experimental group, on the other hand, allocated more attention to the FRD (MD=0.219), Instruments (MD=0.009) and UD (MD=0.073) while focusing less on the concurrent task (MD=-0.221) in the high uncertainty section (TOR-40-0).

Hypothesis 2: Takeover performance

Minimum Time-to-Collision

Participants of the experimental group (M=2.765s) achieved significantly higher MTTC values than the control group (M=1.641s), $t(29.329) = -2.251, p = .032, d = 0.772$ (medium).

Maximum Lateral and Longitudinal Acceleration

Lateral and longitudinal acceleration can be considered as indicators for takeover quality (Bubb, Bengler, Gruenen, & Vollrath, 2015), whereby a value of 0.3m/s^2 delineates the limit of comfortable lateral acceleration. Maximum lateral acceleration values were significantly higher for the control group (M=2.450 m/s^2) compared to the experimental group (M=1.635 m/s^2), $t(21.91) = 2.809, p = .010, d = 0.963$ (large). Longitudinal acceleration did not significantly differ between groups.

Time-to-Takeover

Participants of the experimental group (M=2.090s) averaged a faster TTT than those of the control group (M=2.356s), however without statistical significance.

Hypothesis 3: Operator workload

Unpaired (two sample) t-tests were used to analyse the results of each workload survey question individually (Sullivan & Artino, 2013). Table 2 summarises the results, whereby only the perceived effort differed significantly between the groups, with the experimental group reporting higher values, $t(24.123) = -4.352, p < .001$.

Hypothesis 4: Relationship between trust and visual attention**Solving Rate Relative to Subjective Trust**

Figure 7 depicts the relationship between solving rate and subjective trust. Implementing UD and trust as fixed effects, subjective trust significantly affected the solving rate ($t(268.07) = 2.770, p = .006$), raising it by 0.064 1/min (SE=0.023) for each percent of trust.

Subjective Trust Relative to Visual Attention

Entering group and the frequency of monitoring glances as fixed effects into the model, monitoring frequency significantly affected subjective trust ($t(12.31) = -2.157, p = .0389$), lowering it by 29.184 (SE = 13.529) for each glance per second.

Solving Rate Relative to Visual Attention

With group and monitoring frequency as fixed effects, monitoring frequency significantly affected the solving rate of concurrent tasks ($t(62.95) = -3.861, p < .001$), lowering it by 22.385 1/min (SE=5.798) for each glance per second to driving-related AOIs.

Hypothesis 5: Physiological effects

The absolute, offset, and relative heart rate of participants was not significantly affected by either subjective trust ratings or the visibility range, neither did it significantly differ between groups. Similar to prior publications (Waytz et al., 2014), the heart range change as a consequence of an event (here: TOR) was analysed with the two fixed effects group and TOR,

whereby TOR had two levels (20s interval before TOR; 20s interval after TOR). The TOR had a significant effect on heart rate, raising it by 1.010bpm (SE=0.129), $t(1359) = -7.840, p < .001$. The factor UD did not have a significant effect.

Discussion

The current study was conducted to examine the effects of dynamically communicating uncertainties on behavioural, physiological as well as performance-related outcome variables. Consistent with prior publications from the automotive (Beller et al., 2013; Helldin et al., 2013) and other domains (McGuirl & Sarter, 2003), the results of the current study suggest a number of significant implications of communicating uncertainty information on human-automation interaction.

Hypothesis 1: Behavioural implications

The results of the study confirmed Hypothesis 1 as there was a significant impact of uncertainty communication on the solving rate of concurrent tasks, SAGAT responses, subjective trust, and visual attention.

As predicted, the experimental group solved less concurrent tasks with a lower visibility range – an effect which could not be observed in the control group. This indicates that the presentation of uncertainty information affects the (endogenous) attention allocation of human operators (Katsuki & Constantinidis, 2014). This is reaffirmed by the analysed eye tracking data, whereby both the monitoring frequency and monitoring ratio indicate an attentional shift from the concurrent task to driving-related AOIs in critical situations. SAGAT measures confirmed that this had positive implications for SA. Further, the results indicate that presenting uncertainty information in a higher resolution, i.e. not binary (Beller et al., 2013) but in several distinct degrees (in this case 50, 80, 110, and 140bpm), helps users to differentiate between different

system capability levels and may lead to more appropriate behaviour. This is emphasised by the progressive increase in both trust scores and solving rate with higher visibility ranges in the experimental group (see Figure 5 and Figure 6).

However, the communication of uncertainties did entail side-effects. As such, the experimental group solved significantly less concurrent tasks than the control group, even if only considering periods for which the UD indicated the lowest uncertainty. As indicated by the monitoring frequency, the experimental group glanced more frequently to the UD in high visibility situations than the control group. The findings highlight a drawback of presenting uncertainty information using visual displays as this requires operators to constantly monitor its state to perceive possible changes. Specifically, the experimental group monitored the UD with a ratio of 12.06% prior to the TOR. This share contributes to the glance time off road and thereby increases the risk of missed changes and crashes (Seppelt et al., 2017). The ratio could be reduced through improved display design, for instance peripheral awareness displays (Kunze, Marshall, Summerskill, & Filtner, 2017; Kunze, Summerskill, Marshall, & Filtner, 2018a, 2018b). Additionally, prior research has indicated that the prolonged monitoring of visual uncertainty displays leads to missed uncertainty changes (Large, Burnett, Morris, Muthumani, & Matthias, 2017), thus reaffirming the adverse consequences of using displays that require focal attention for uncertainty communication.

Further, the experimental group did not trust the system significantly more than the control group in situations of low uncertainty, as could have been predicted due to the use of an anthropomorphic interface and based on the findings of previous studies (Beller et al., 2013). This finding supports prior research which indicated that effects as a result of anthropomorphism are only activated if automated agents possess easily recognisable physical human-like attributes

such as faces (de Visser et al., 2017). The subjective trust ratings in conjunction with the solving rate of the concurrent task and the recorded monitoring behaviour do, however, indicate that participants of the experimental group were better able to calibrate their trust: In situations with a reduced visibility range, they focused more towards the FRD and solved less concurrent tasks while reporting lower trust scores, indicating more appropriate reliance and showing less signs of automation misuse.

Hypothesis 2: Takeover performance

The current study confirms the prediction that the communication of uncertainty information improves both takeover performance and quality. Consistent with prior findings, MTTC values were significantly higher for the experimental group, indicating that the provision of uncertainty information prepared drivers for potential automation failures. Similarly, significantly lower maximum lateral acceleration values indicate a less abrupt takeover for the experimental group.

In contrast to previous research (Helldin et al., 2013), TTT values were not significantly affected. Potentially, the engagement in concurrent tasks prevented participants of both groups from showing signs of drowsiness (Miller et al., 2015) and they were equally quick to react. The recorded TTT values correspond to data recorded by other research projects (Zhang, Varotto, & Happee, 2018).

Hypothesis 3: Operator workload

For both groups, mental and temporal demand were higher than optimal values of around half the range of the scale (Stanton, Dunoyer, & Leatherland, 2011). Brief interviews following the experiment indicated that participants perceived a high mental workload due to the execution of a visually demanding task while having to regularly monitor the driving scene. Temporal

demand was perceived as high due to the emergency TOR. Further, the experimental group indicated a significantly higher perceived effort. Thereby, participants particularly disliked that they constantly had to monitor the UD to notice changes, which was also confirmed by the eye tracking data. For extended usage, operators may not be willing to put that degree of effort in monitoring the display, particularly if the system is performing safely (Large et al., 2017). In contrast to the phrased concerns, the UD did not make participants feel more uneasy, as suggested by the scores for *frustration* (see Table 2).

Hypothesis 4: Relationship between trust and visual attention

The current study confirms the predicted relationship between trust and attention, as indicated by significant effects of subjective trust indications on both the solving rate of concurrent tasks as well as the monitoring frequency. As such, the solving rate increases with subjective trust while the monitoring frequency is lowered with trust. This affirms the results of prior publications (Hergeth et al., 2016; Muir & Moray, 1996).

Hypothesis 5: Physiological effects

The heart rate of participants of the experimental group was not significantly affected by the heart rate on the UD and did not differ significantly from that of the control group. While there was a significant change in heart rate due to the TOR, the experimental group did not show a different response from that of the control group.

Limitations

Analogue to prior publications (Beller et al., 2013; Helldin et al., 2013; Hergeth et al., 2016), all participants drove through the same scenario. Thus, the order of the visibility range levels may have had an impact on the results. However, this was deliberately accepted as the current design assured that all participants experienced a scenario that would likely invoke the

strongest complacency effects as the most critical conditions were at or near the end of the simulation (Lee & Moray, 1992, 1994).

Further, the study only varied the factor visibility range due to fog as an indicator for uncertainty to achieve internal validity. Other factors such as unclear lane markings, traffic density, or other weather conditions may also affect dependent variables such as trust of participants.

Additionally, the current study may have benefited from the tracking of other physiological measures, such as heart rate variability (HRV). Although this did not significantly predict trust in prior publications (Wintersberger, von Sawitzky, Frison, & Riener, 2017), it will be explored in future experiments to specifically investigate HRV in the context of uncertainty communication.

Conclusion and Outlook

The results of the study confirm the outlined implications (see Figure 2) of uncertainty communication on attention allocation, situation awareness, and takeover performance. As such, dynamically presenting uncertainty information leads operators to employ more appropriate attention allocation strategies. This is indicated by more monitoring glances to driving-related AOIs as well as a reduced solving rate of concurrent tasks with lower visibility range and therefore higher uncertainty. Thereby, participant behaviour differed between low, intermediate, and high uncertainty situations, indicating benefits of communicating uncertainty information with several levels.

The visual uncertainty display required users to regularly check the system state, leading to a higher perceived effort and more time spent off concurrent tasks and the field-relevant for driving. Future work should concentrate on how to best present uncertainty information, for

instance by unobtrusive means that do not increase operator workload and require fewer additional glances (Kunze et al., 2017). Acknowledging the uncertainty of automated agents may aid the shift from viewing automation as a tool to collaborative automation (de Visser, Pak, & Shaw, 2018).

References

- Bagheri, N., & Jamieson, G. A. (2004). The Impact of Context-related Reliability on Automation Failure Detection and Scanning Behaviour. In *2004 IEEE International Conference on Systems, Man and Cybernetics* (pp. 212–217). IEEE.
- Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8(4), 321–348.
<https://doi.org/10.1080/14639220500535301>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... Green, P. (2018). *Linear Mixed-Effect Models using “Eigen” and S4*. Version 1.1-17. Retrieved from <https://cran.r-project.org/web/packages/lme4/lme4.pdf>
- Baumann, M., & Krems, J. F. (2007). Situation Awareness and Driving: A Cognitive Model. In P. C. Cacciabue (Ed.), *Modelling Driver Behaviour in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems* (pp. 253–265). London: Springer London. https://doi.org/10.1007/978-1-84628-618-6_14
- Bazilinskyy, P., Petermeijer, S., Petrovych, V., Dodou, D., & de Winter, J. C. F. (2018). Take-over requests in highly automated driving: A crowdsourcing survey on auditory, vibrotactile, and visual displays. *Transportation Research Part F: Traffic Psychology and Behaviour*, 56, 82–98. <https://doi.org/10.7717/peerj-cs.13>
- Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the Driver Automation Interaction: An Approach Using Automation Uncertainty. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 55(6), 1130–1141.
<https://doi.org/10.1177/0018720813482327>

BIPM. Evaluation of measurement data - Guide to the expression of uncertainty in measurement (2008).

Bubb, H., Bengler, K., Gruenen, R. E., & Vollrath, M. (2015). *Automobilergonomie*. Wiesbaden: Springer Vieweg. <https://doi.org/10.1007/978-3-642-13505-7>

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control Task Substitution in Semiautomated Driving: Does It Matter What Aspects Are Automated? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5), 747–761. <https://doi.org/10.1177/0018720812460246>

Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*, 19(3), 259–282. <https://doi.org/10.1080/1463922X.2017.1315750>

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology*, 14(19), 850–852. <https://doi.org/10.1016/j.cub.2004.09.041>

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201–215. <https://doi.org/10.1038/nrn755>

Davies, D. R., & Parasuraman, R. (1982). *The Psychology of Vigilance*. Academic Press.

de Visser, E. J., Krueger, F., McKnight, P., Scheid, S., Smith, M., Chalk, S., & Parasuraman, R. (2012). The world is not enough: Trust in cognitive agents. *Proceedings of the Human Factors and Ergonomics Society*, 263–267. <https://doi.org/10.1177/1071181312561062>

de Visser, E. J., Monfort, S. S., Goodyear, K., O'Hara, M., Lee, M. L., Parasuraman, R., & Krueger, F. (2017). A Little Anthropomorphism Goes a Long Way: Effects of Oxytocin on Trust, Compliance, and Team Performance With Automated Agents. *Human Factors: The*

Journal of the Human Factors and Ergonomics Society, 59(1), 116–133.

<https://doi.org/10.1177/0018720816687205>

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From ‘automation’ to ‘autonomy’: the importance of trust repair in human–machine interaction. *Ergonomics*, 0139, 1–19.

<https://doi.org/10.1080/00140139.2018.1457725>

de Waard, D., van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver Behavior in an Emergency Situation in the Automated Highway System. *Transportation Human Factors*, 1(1), 67–82. https://doi.org/10.1207/sthf0101_9

Diekfuss, J. A., Ward, P., & Raisbeck, L. D. (2017). Attention, workload, and performance: A dual-task simulated shooting study. *International Journal of Sport and Exercise Psychology*, 15(4), 423–437. <https://doi.org/10.1080/1612197X.2015.1121508>

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human Computer Studies*, 58(6), 697–718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32–64. <https://doi.org/10.1518/001872095779049543>

Endsley, M. R. (2000). Direct measurement of situation awareness: validity and use of SAGAT Development of Queries. In *Situation Awareness Analysis and Measurement* (pp. 1–21). Mahwah, N.J.: Lawrence Erlbaum Associates.

Endsley, M. R. (2015). Situation Awareness Misconceptions and Misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1), 4–32. <https://doi.org/10.1177/1555343415572631>

- Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned From Human-Automation Research. *Human Factors The Journal of the Human Factors and Ergonomics Society*, 59(1), 5–27. <https://doi.org/10.1177/0018720816681350>
- Endsley, M. R., & Jones, D. G. (2012). *Designing for Situation Awareness*. CRC Press.
- Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors and Ergonomics Society*, 37(2), 381–394. <https://doi.org/10.1518/001872095779064555>
- Finger, R., & Bisantz, A. M. (2002). Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, 2(1), 1–25. <https://doi.org/10.1080/14639220110110324>
- Gal, Y. (2016). *Uncertainty in Deep Learning*. PhD Thesis. University of Cambridge.
- Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). “Take over!” How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1938–1942. <https://doi.org/10.1177/1541931213571433>
- Golightly, D., Wilson, J. R., Lowe, E., & Sharples, S. (2010). The role of situation awareness for understanding signalling and control in rail operations. *Theoretical Issues in Ergonomics Science*, 11(1–2), 84–98. <https://doi.org/10.1080/14639220903009961>
- Gollan, J. K., McCloskey, M., Hoxha, D., & Coccaro, E. F. (2010). How do depressed and healthy adults interpret nuanced facial expressions? *Journal of Abnormal Psychology*, 119(4), 804–810. <https://doi.org/10.1037/a0020234>
- Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meeting*, 904–908. <https://doi.org/10.1037/e577632012-009>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results

of Empirical and Theoretical Research. *Advances in Psychology*, 52(C), 139–183.

[https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. *Proceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '13*, 5, 210–217. <https://doi.org/10.1145/2516540.2516554>

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(3), 509–519. <https://doi.org/10.1177/0018720815625744>

Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>

Hubbard, J. R., & Workman, E. A. (1998). *Handbook of Stress Medicine: An Organ System Approach*.

Itti, L., Koch, C., Way, W., & Angeles, L. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, 2(3), 194–203. <https://doi.org/10.1038/35058500>

Jacob, R. J. K., & Karn, K. S. (2003). Commentary on Section 4 – Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises. In *The Mind's Eye* (pp. 573–605). Elsevier. <https://doi.org/https://doi.org/10.1016/B978-044451020-4/50031-1>

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated System. *International Journal of Cognitive Ergonomics*, 4(1),

53. https://doi.org/10.1207/S15327566IJCE0401_04

Johns, J. L. (1996). A Concept Analysis of Trust. *Journal of Advanced Nursing*, 24(1), 76–83.

<https://doi.org/https://doi.org/10.1046/j.1365-2648.1996.16310.x>

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *Neuroscientist*, 20(5), 509–521.

<https://doi.org/10.1177/1073858413514136>

Kiureghian, A. Der, & Ditlevsen, O. (2007). Aleatoric or Epistemic? Does it matter? In *Special Workshop on Risk Acceptance and Risk Communication*. Stanford University.

<https://doi.org/10.1016/j.strusafe.2008.06.020>

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, 66, 18–31.

<https://doi.org/10.1016/j.apergo.2017.07.006>

Kunze, A. (2018). Heartbeat Animation. Retrieved May 21, 2018, from

<https://dribbble.com/shots/4136090-Heartbeat-Animation>

Kunze, A., Marshall, R., Summerskill, S. J., & Filtness, A. J. (2017). Enhancing driving safety and user experience through unobtrusive and function-specific feedback. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct* (pp. 183–189). Oldenburg, Germany: ACM.

<https://doi.org/10.1145/3131726.3131762>

Kunze, A., Summerskill, S. J., Marshall, R., & Filtness, A. J. (2018a). Evaluation of Variables for the Communication of Uncertainties Using Peripheral Awareness Displays. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. Toronto, ON, Canada: ACM.

<https://doi.org/10.1145/3239092.3265958>

Kunze, A., Summerskill, S. J., Marshall, R., & Filtner, A. J. (2018b). Preliminary Evaluation of Variables for Communicating Uncertainties Using a Haptic Seat. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct*. Toronto, ON, Canada: ACM.

<https://doi.org/10.1145/3239092.3265959>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2018). *Tests in Linear Mixed Effects Models*. Version 3.0-1. Retrieved from <https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf>

Large, D. R., Burnett, G., Morris, A., Muthumani, A., & Matthias, R. (2017). Design Implications of Drivers' Engagement with Secondary Activities During Highly - Automated Driving – A Longitudinal Simulator Study. *Road Safety & Simulation International Conference 2017*, (October), 1–10.

Lee, J. D., & Moray, N. (1992). Trust, Control Strategies and Allocation of Function in Human-Machine Systems. *Ergonomics*, *35*(March), 1243–1270.

<https://doi.org/10.1080/00140139208967392>

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human Computer Studies*, *40*, 153–184.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, *46*(1), 50–80.

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis and Prevention*, *108*(September), 9–18.

<https://doi.org/10.1016/j.aap.2017.08.011>

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C, 90*(November), 1–18. <https://doi.org/10.1016/j.trc.2017.01.001>

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making, 6*(1), 57–87. <https://doi.org/10.1177/1555343411433844>

May, P., Molloy, R., & Parasuraman, R. (1993). Effects of automation reliability and failure rate on monitoring performance in a multi-task environment. *Paper Presented at the Annual Meeting of the Human Factors Society, Santa Monica, CA.*

Mayer, R. C., Davis, J. H., & Schoorman, D. F. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review, 20*(3), 709–734.

McGuirl, J. M., & Sarter, N. B. (2003). How are we doing?: Presenting System Confidence Information to Support Trust Calibration and Adaptive Function Allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 47*(3), 538–542. <https://doi.org/10.1177/154193120304700360>

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management. *Human Factors, 58*(3), 401–415. <https://doi.org/10.1177/0018720815621206>

Met Office UK. (2018). What is fog? Retrieved May 25, 2018, from <https://www.metoffice.gov.uk/learning/clouds/fog>

Metzger, U., & Parasuraman, R. (2005). Automation in Future Air Traffic Management : Effects

- of Decision Aid Reliability on Controller Performance and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(1), 1–15.
- Miller, D., Sun, A., Johns, M., Ive, H., Sirkin, D., Aich, S., & Ju, W. (2015). Distraction Becomes Engagement in Automated Driving. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting*, 2, 2–6. <https://doi.org/10.1177/1541931215591362>
- Molloy, R., & Parasuraman, R. (1996). Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(2), 311–322. <https://doi.org/10.1177/001872089606380211>
- Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science*, 1(4), 354–365. <https://doi.org/10.1080/14639220052399159>
- Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429–460. <https://doi.org/10.1080/00140139608964474>
- Nguyen, T., Lim, C. P., Duy Nguyen, N., Gordon-Brown, L., & Nahavandi, S. (2018). A Review of Situation Awareness Assessment Approaches in Aviation Environments. *Eprint ArXiv:1803.08067*, 1–16. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1803/1803.08067.pdf>
- Norman, D. A. (1989). *The “Problem” of Automation: Inappropriate Feedback and Interaction, Not “Overautomation.”* San Diego.
- Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059–1072. <https://doi.org/10.1080/00140139.2012.691554>
- Pang, A. T., Wittenbrink, C. M., & Lodha, S. K. (1997). Approaches to uncertainty visualization.

The Visual Computer, 13(8), 370–390. <https://doi.org/10.1007/s003710050111>

Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: an attentional integration. *Human Factors*, 52, 381–410.

<https://doi.org/10.1177/0018720810376055>

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse.

Human Factors The Journal of the Human Factors and Ergonomics Society, 39(2), 230–253.

Politis, I., Brewster, S., & Pollick, F. (2015). To Beep or Not to Beep?: Comparing Abstract versus Language-Based Multimodal Driver Displays. In *Proceedings of the ACM CHI'15 Conference on Human Factors in Computing Systems* (pp. 3971–3980). Seoul, Republic of Korea. <https://doi.org/10.1145/2702123.2702167>

Preuschhoff, K., Mohr, P. N. C., & Hsu, M. (2013). Decision making under uncertainty. *Frontiers in Neuroscience*, 7(November), 2012–2013. <https://doi.org/10.3389/fnins.2013.00218>

Rasshofer, R. H., Spies, M., & Spies, H. (2011). Influences of weather phenomena on automotive laser radar systems. *Advances in Radio Science*, 9, 49–60. <https://doi.org/10.5194/ars-9-49-2011>

SAE International. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (J3016_201609)*.

Seppelt, B. D., Seaman, S., Lee, J., Angell, L. S., Mehler, B., & Reimer, B. (2017). Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-Car data.

Accident Analysis and Prevention, 107(December 2016), 48–62.

<https://doi.org/10.1016/j.aap.2017.07.021>

Shepardson, D. (2018). Tesla hits parked California police vehicle; driver blames “Autopilot.”

Retrieved May 30, 2018, from <https://www.reuters.com/article/us-tesla-autopilot/tesla-hits-parked-california-police-vehicle-driver-blames-autopilot-idUSKCN1IU2SZ>

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis*. SAGE Publications Ltd.

Stanton, N. A., Dunoyer, A., & Leatherland, A. (2011). Detection of new in-path targets by drivers using Stop & Go Adaptive Cruise Control. *Applied Ergonomics*, *42*(4), 592–601. <https://doi.org/10.1016/j.apergo.2010.08.016>

Stanton, N. A., Salmon, P. M., Walker, G. H., Salas, E., & Hancock, P. A. (2017). State-of-Science: Situation Awareness in individuals, teams and systems. *Ergonomics*, *0139*(February), 1–33. <https://doi.org/10.1080/00140139.2017.1278796>

Stephens, A., & Voegelé, C. (1991). Methodology of mental stress testing in cardiovascular research. *Circulation*, *83*(4 Suppl), 14–24.

Stewart, J. (2018). Why Tesla's Autopilot Can't See a Stopped Firetruck. Retrieved May 23, 2018, from <https://www.wired.com/story/tesla-autopilot-why-crash-radar>

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. <https://doi.org/10.4300/JGME-5-4-18>

Tobii. (2018). Tobii Lab Pro (Version 1.86). Tobii AB.

Treisman, A. M., & Gelade, G. (1980). A Feature-Integration of Attention. *Cognitive Psychology*, *12*(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)

Uggirala, A., Gramopadhye, A. K., Melloy, B. J., & Toler, J. E. (2004). Measurement of trust in complex and dynamic systems using a quantitative approach. *International Journal of Industrial Ergonomics*, *34*, 175–186. <https://doi.org/10.1016/j.ergon.2004.03.005>

Wang, L., Jamieson, G. A., & Hollands, J. G. (2009). Trust and reliance on an automated combat

identification system. *Human Factors*, 51(3), 281–291.

<https://doi.org/10.1177/0018720809338842>

Wang, X., Pinol, R. A., Byrne, P., & Mendelowitz, D. (2014). Optogenetic Stimulation of Locus Ceruleus Neurons Augments Inhibitory Transmission to Parasympathetic Cardiac Vagal Neurons via Activation of Brainstem α_1 and α_2 Receptors. *Journal of Neuroscience*, 34(18), 6182–6189. <https://doi.org/10.1523/JNEUROSCI.5093-13.2014>

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

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues in Ergonomics Science*, 8(3), 201–212. <https://doi.org/10.1080/14639220500370105>

Wickens, C. D., Sebok, A., Li, H., Sarter, N., & Gacy, A. M. (2015). Using Modeling and Simulation to Predict Operator Performance and Automation-Induced Complacency with Robotic Automation. *Human Factors*, 57(6), 959–975. <https://doi.org/10.1177/0018720814566454>

Wintersberger, P., von Sawitzky, T., Frison, A.-K., & Riener, A. (2017). Traffic Augmentation as a Means to Increase Trust in Automated Driving Systems. *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter - CHIItaly '17*, (September), 1–7. <https://doi.org/10.1145/3125571.3125600>

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017). Evaluating Effects of User Experience and System Transparency on Trust in Automation. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, 408–416.

<https://doi.org/10.1145/2909824.3020230>

Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, 3(2), 178–194.

<https://doi.org/10.1080/14639220210123789>

Zhang, B., Varotto, S. F., & Happee, R. (2018). Determinants of take-over time from automated driving: A meta-analysis of 93 studies, (May). <https://doi.org/10.13140/RG.2.2.33648.56326>

Zimmermann, H. J. (2000). Application-oriented view of modeling uncertainty. *European Journal of Operational Research*, 122(2), 190–198. [https://doi.org/10.1016/S0377-2217\(99\)00228-3](https://doi.org/10.1016/S0377-2217(99)00228-3)

Footnotes

¹ The simulation was run using imperial units. Figures show the visibility range in feet.

Tables

Table 1

Ratio of correctly answered SAGAT questions

| Group | SAGAT1 | SAGAT2 | SAGAT3 | SAGAT4 |
|--------------|---------------|---------------|---------------|---------------|
| Control | 0.47 | 0.59 | 0.41 | 0.47 |
| Experimental | 0.76 | 0.88 | 0.53 | 0.53 |

Table 2

Mean and standard deviation of NASA TLX scores

| Group | Mental demand | Physical demand | Temporal demand | Performance | Effort | Frustration |
|--------------|----------------------|------------------------|------------------------|--------------------|-------------------|--------------------|
| Control | 4.65 (1.66) | 2.24 (1.48) | 4.59 (1.54) | 5.29 (0.99) | 3.71*** (1.53) | 3.24 (2.22) |
| Experimental | 4.88 (1.41) | 2.47 (1.42) | 5.12 (1.41) | 5.00 (1.62) | 5.53*** (0.80) | 3.29 (1.83) |

*** $p < .001$. ** $p < .01$. * $p < .05$

Table 3

Analysis of variance table with approximated degrees of freedom and p values based on Satterthwaite's method

| Variables | Estimate | Standard Error | df | t | p |
|----------------------------------|-----------------|-----------------------|-----------|-----------------------|-----------------------|
| Solving rate of concurrent tasks | | | | | |
| Visibility range | 0.368 | 0.460 | 304.00 | 0.800 | .424 |
| UD | -15.007 | 2.396 | 91.35 | -6.264 | < .001*** |
| Visibility range \times UD | 3.196 | 0.650 | 304.00 | 4.914 | < .001*** |
| Solving rate of concurrent tasks | | | | | |
| Subjective trust | 0.06387 | 0.02306 | 268.07 | 2.770 | .006* |

| Variables | <i>Estimate</i> | <i>Standard Error</i> | <i>df</i> | <i>t</i> | <i>p</i> |
|----------------------------------|-----------------|-----------------------|-----------|----------|-----------|
| UD | -6.8737 | 1.7043 | 32.36 | -4.033 | < .001*** |
| Subjective trust | | | | | |
| Visibility range | 0.605 | 1.591 | 236.00 | 0.380 | .704 |
| UD | -48.052 | 7.484 | 147.500 | -6.420 | < .001*** |
| Visibility range × UD | 15.286 | 2.250 | 236.00 | 6.795 | < .001*** |
| Monitoring frequency | | | | | |
| UD | 0.385 | 0.0560 | 44.16 | 6.877 | < .001*** |
| Section | 0.00441 | 0.0322 | 32.00 | 0.137 | 0.892 |
| UD × Section | -0.216 | 0.0455 | 32.00 | -4.752 | < .001*** |
| Subjective trust | | | | | |
| Monitoring frequency | -29.184 | 13.529 | 12.31 | -2.157 | .0389* |
| UD | 3.288 | 4.803 | 31 | 0.685 | .499 |
| Solving rate of concurrent tasks | | | | | |
| Monitoring frequency | -22.385 | 5.798 | 62.950 | -3.861 | < .001*** |
| UD | -1.648 | 3.079 | 38.680 | -0.535 | .596 |

***p<.001. **p<.01. *p<.05

Figures

Sources of uncertainties

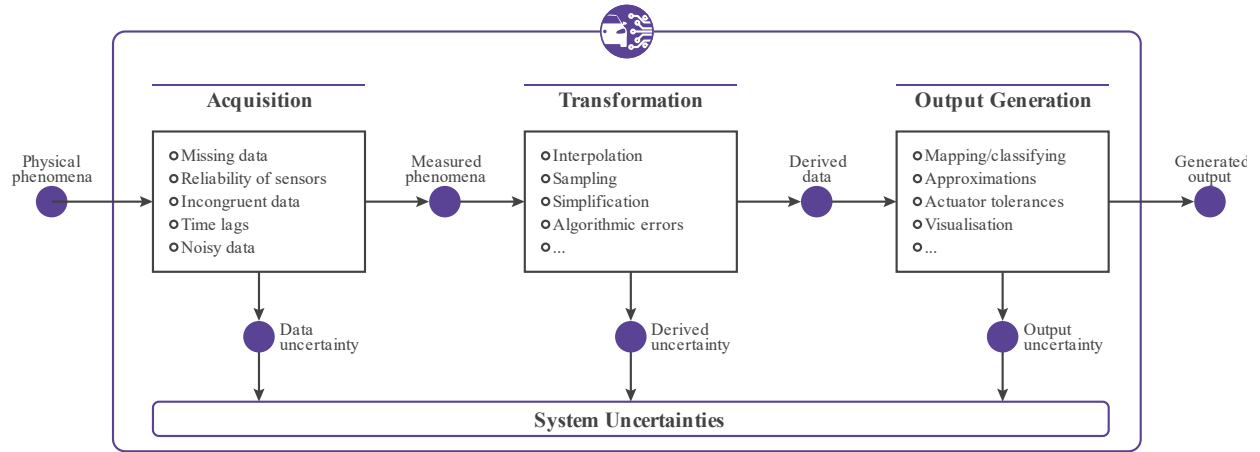


Figure 1: Sources of system uncertainties (Endsley & Jones, 2012; Kiureghian & Ditlevsen, 2007; Pang et al., 1997; Uggirala et al., 2004)

Overview Model

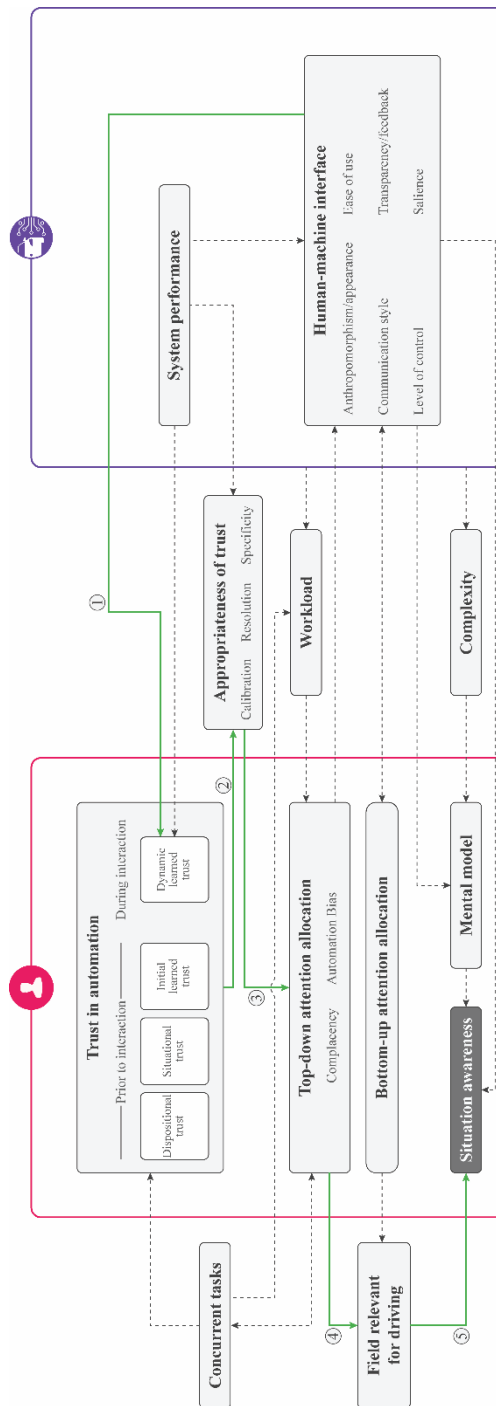


Figure 2: Overview model depicting the functional net of trust, attention, concurrent tasks and automation attributes. The continuous green lines indicate how the communication of uncertainties can affect situation awareness (Diekfuss et al., 2017; Endsley, 2017; Hoff & Bashir, 2015; Lee & See, 2004; Parasuraman & Manzey, 2010; Wickens & Dixon, 2007; Young & Stanton, 2002)

Experimental scenario and Takeover Situation

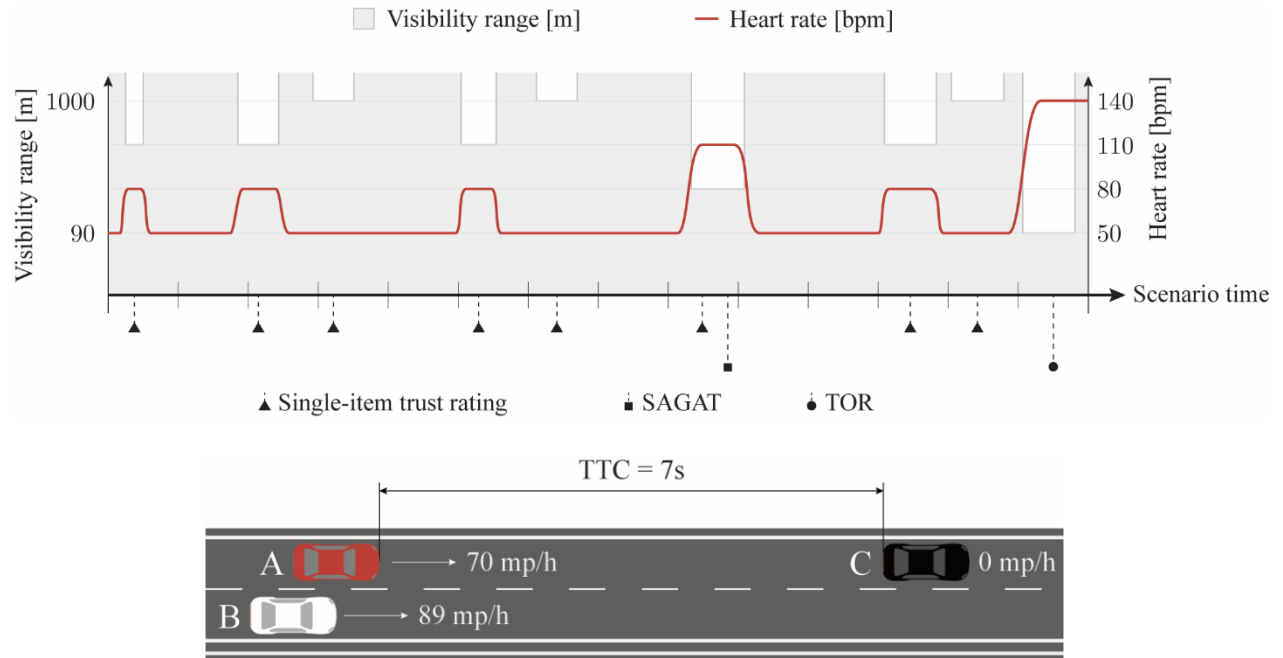


Figure 3: Visibility range and system heart rate across the experimental scenario (top); location of own vehicle (A), passing vehicle (B), and stopped vehicle (C) at the time of the TOR (bottom)

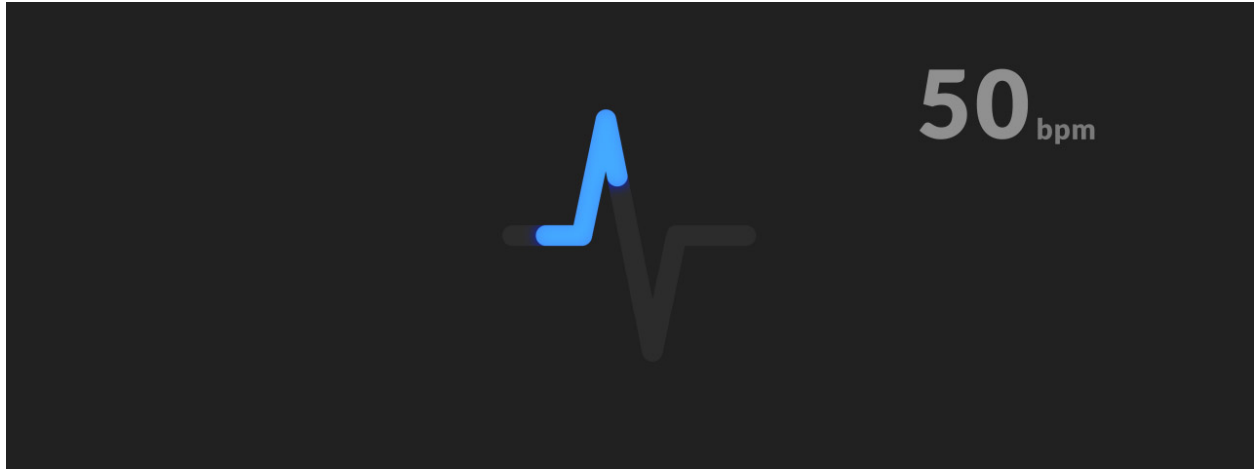


Figure 4: Uncertainty display consisting of a stylised heartbeat and a numeric indication for beats per minute (Kunze, 2018)

Effects of linear mixed model

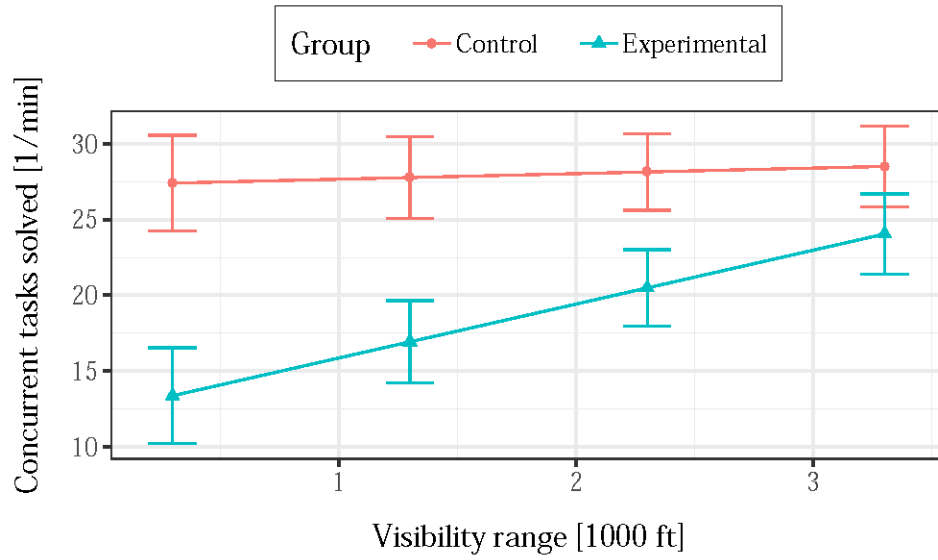


Figure 5: Effects of the linear mixed model predicting the impact of visibility range and UD on the solving rate of concurrent tasks

Effects of linear mixed model

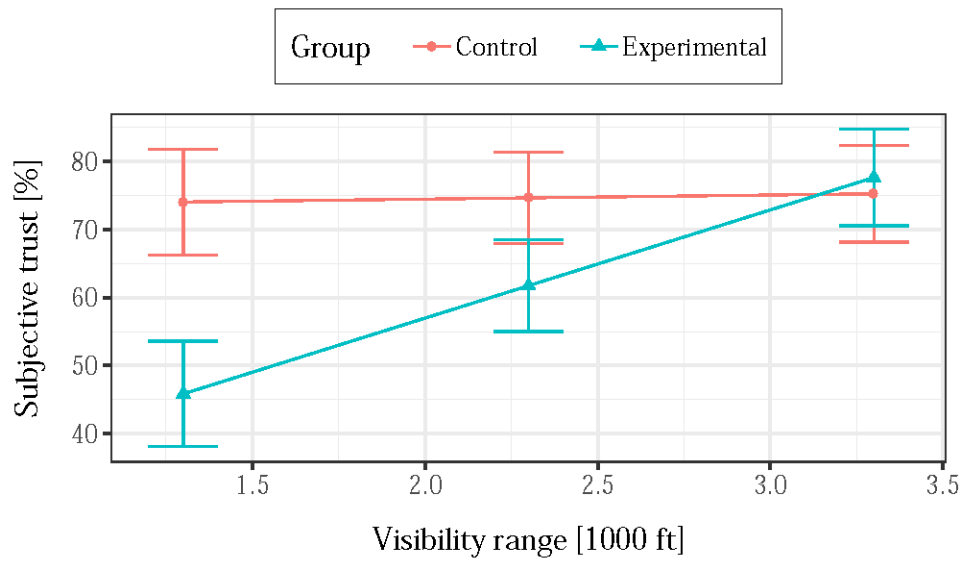


Figure 6: Effects of the linear mixed model predicting the impact of visibility range and UD on subjective trust ratings

Concurrent tasks solved relative to subjective trust

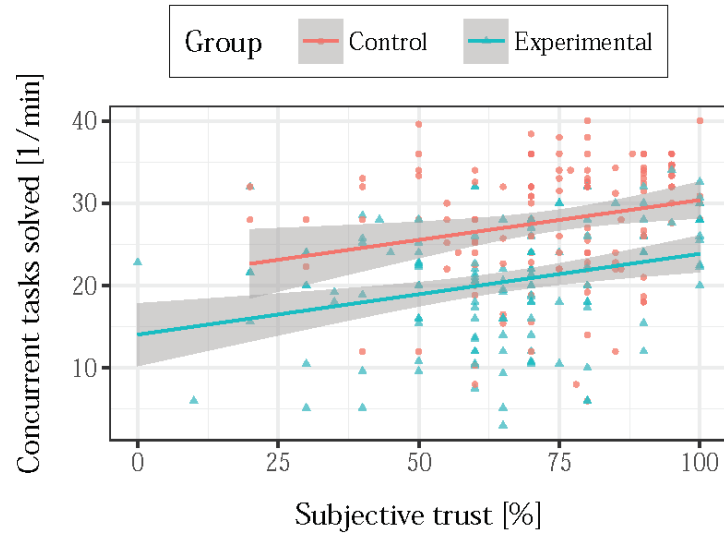


Figure 7: Concurrent tasks solved per minute relative to subjective trust scores

Monitoring frequency relative to section prior to TOR

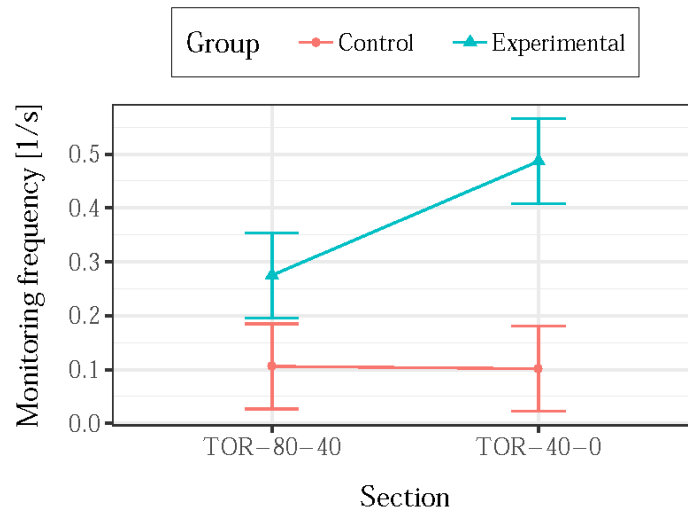


Figure 8: Monitoring frequency towards driving-related AOIs in a low uncertainty section (TOR-80-40) and high uncertainty section (TOR-40-0)

Relative glance duration for each area of interest between groups

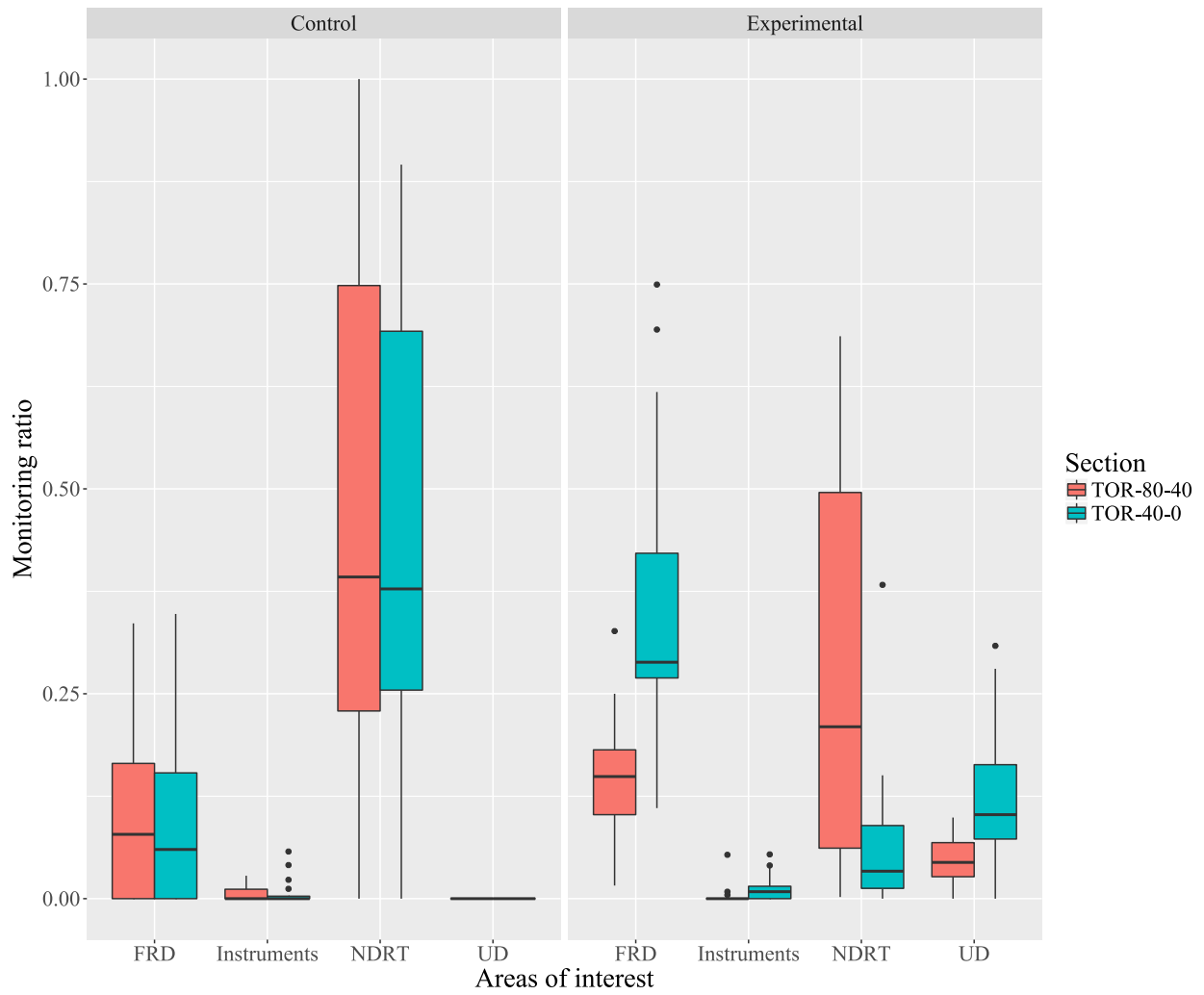


Figure 9: Monitoring ratio for each AOI and two sections between groups, FRD = field relevant for driving, NDRT = non-driving related / concurrent task, UD = uncertainty display