Trust me...Not: Communicating the reliability of automated vehicles influences skin conductance responses to unexpected events

Sarah Faltaous[1], Martin Baumann[2], Lewis Chuang[3]

[1] University of Duisburg-Essen

[2] Ulm University

[3]Ludwig Maximilian University of Munich

Introduction: Many aspects of driving a car are increasingly automated, with the assistance of onboard sensors, telematics, and algorithms. Commercially available automotives (i.e., NHTSA L2/L3; (SAE,2014)) are capable of executing 'control' (e.g., lane-keeping) and even 'maneuvering' (e.g., overtaking) actions (Michon 1985). However, the driver continues to be responsible for 'strategic' (e.g., driving style) decisions. This means that they are still expected to intervene when automation fails. Nonetheless, it is well-established that automation can diminish vigilance and situation awareness (Greenlee, 2018; Greenlee, 2019). Thus, regardless of regulations, an overtrust in automation could induce users to perform non-driving tasks, which could compromise their ability to resume control and, hence, result in fatal accidents (NTSB, 2020).

Accordingly, different designs for takeover requests have been proposed to facilitate a smooth transition for the driver back into the driving loop (Borojeni, 2016; Melcher, 2015; Roche, 2019; Brandenburg, 2019). Notably, some researchers have proposed that automated vehicles could communicate, not only the need to resume control but also, the likelihood of a critical event during this alarm [Faltaous,2018; Hoff, 2015; Sorkin, 1988]. Here, we propose providing continuous feedback on the vehicle's reliability in detecting collision targets. We report skin-conductance responses to the onset of potential collisions to evaluate the sensitivity of users to this ongoing communication of an automated vehicle's reliability. This metric, based on previous work (Storm,2002), correlates to the driver's sudden change in arousal, which we interpret as surprise. This work is motivated by the need to design feedback systems that are able to calibrate the trust between the driver and the autonomous vehicle by providing more transparency regarding the system's certainty (Yang, 2017).

Method: Based on previous work (Faltaous, 2017; Faltaous, 2018), participants (13 males, 6 females) experienced a simulation of operating an SAE L3 vehicle on a two-lane bidirectional highway. The ego-vehicle could encounter slow-moving vehicles ahead of itself. When this occurs, it will attempt to overtake on the opposing lane as soon as possible, unless an oncoming vehicle is detected in the opposing lane. Participants were instructed prior to the experiment that failures could happen and that they were required to step on a foot pedal when this happened. For instance, the automated vehicle might falsely detect an oncoming vehicle when there was none and wait indefinitely. Alternatively, it might fail to detect an oncoming vehicle when there was one, which would result in a collision. Our analysis focused on skin conductance response (SCR) during a system failure that could result in a collision. The study was a 2x2 experiment design for the factors of *Human Reliability* or *Vehicle Reliability*. These were respectively manipulated by the visible range of no fog (995m) vs fog (200m) and by an LED strip attached to the driving wheel that communicated either high (green) or low (red) vehicle reliability. It should be noted that this communication did not reflect the actual likelihood of failure. The four conditions were randomly presented to participants, with an average duration of 15 mins each and 10 encounters. All participants were required to perform an operation span task (i.e., 3 letters sequence) as their primary task, which was presented via a tablet on the right of the driving wheel.

Result: Points of interest were extracted to analyze the SCR by dividing data from each condition into 10 intervals, corresponding to each encounter or trial (Figure 1). The SCR was defined as the difference between the baseline (i.e., mean of the recorded values before an encounter) and the maxima after the occurrence of take-over maneuver. The SCR value for system failure that could result in a collision was submitted to a 2x2 repeated measures ANOVA for Human Reliability (no fog, fog) and Vehicle Reliability (high, low). This revealed a statistically significant interaction ($F(1,19)=5.47$, $p=0.03$, $\omega^2=0.46$, see Figure 2). To explore this interaction, we performed a Bayesian t-test within the levels of low and high vehicle reliability. When vehicle reliability is high, there is substantial evidence for the null hypothesis (BF01=3.56). When vehicle reliability is low, there is inconclusive evidence for either the null or test hypothesis (BF01=0.89).

Discussion: The current results suggest that communicating vehicle reliability has an influence on user expectations. When vehicle reliability is low, our participants were vigilant to the extent of their abilities. Hence, they were not aroused by system failure when there was high visibility and aroused when there was low visibility. When vehicle reliability was high, participants' arousal levels were undifferentiated regardless of the visibility conditions. In other words, they were consistently aroused by system failure. These results provide insights into the impact of communicating the reliability of vehicle automation. Communicating that a vehicle is highly reliable, independent of reality, is likely to result in a lack of user engagement and result in high levels of arousal and surprise during system failure, independent of the user's actual ability to detect this.

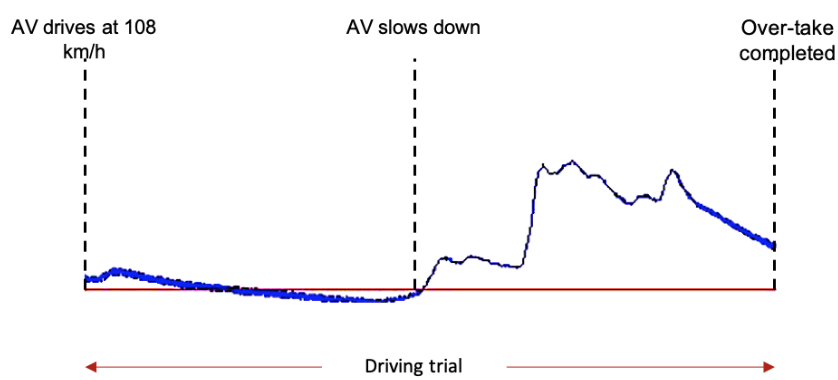

Figure 1. A participant's SCR change across a driving trial. We divided the driving trial into two intervals. The first interval is where no hazards are displayed and the participant has the chance to focus on the non-driving task. This interval is used to compute the baseline for the SCR. The second interval starts when the car starts to slow down, which indicates a potential hazard. The difference between the local maxima in this interval and the baseline is used for the results.

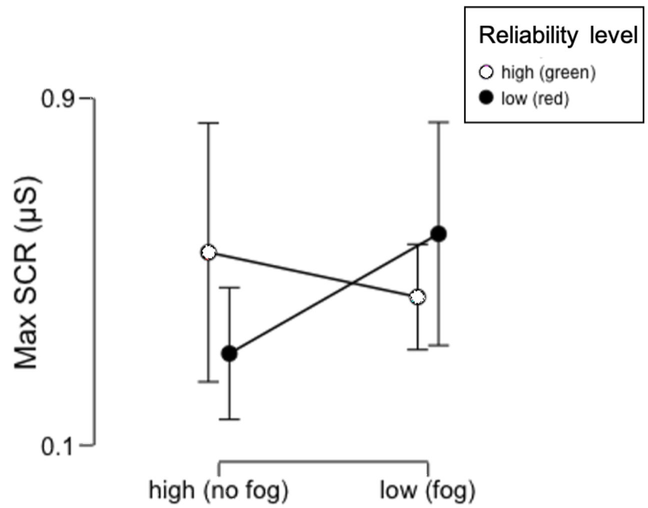

Figure 2. The maximum SCR for the factor of Human Reliability (no fog vs fog) and Vehicle Reliability (high vs low). Error bars indicate 95% confidence intervals.

References:

[Borojeni, 2016] Borojeni, S. S., Chuang, L. , Heuten, W., & Boll, S. (2016). Assisting Drivers with Ambient Take-Over Requests in Highly Automated Driving. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotive'UI 16). Association for Computing Machinery, New York, NY, USA, 237–244. DOI:https://doi.org/10.1145/3003715.3005409

[Brandenburg, 2019] Brandenburg, S., & Chuang, L. (2019). Take-over requests during highly automated driving: How should they be presented and under what conditions?. *Transportation research part F: traffic psychology and behaviour*, *66*, 214-225.

[Faltaous, 2017] Faltaous, S., Machulla, T., Baumann, M., and Chuang, L. (2017). Developing a Highly Automated Driving Scenario to Investigate User Intervention: When Things Go Wrong. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications Adjunct (AutomotiveUI '17). Association for Computing Machinery, New York, NY, USA, 67–71. DOI:https://doi.org/10.1145/3131726.3131766

[Faltaous,2018] Faltaous, S., Baumann, M., Schneegass, S., & Chuang, L. L. (2018, September). Design guidelines for reliability communication in autonomous vehicles. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications* (pp. 258-267).

[Greenlee, 2018] Greenlee ET, DeLucia PR, Newton DC. Driver Vigilance in Automated Vehicles: Hazard Detection Failures Are a Matter of Time. Human Factors. 2018;60(4):465-476. doi:10.1177/0018720818761711

[Greenlee, 2019] Greenlee ET, DeLucia PR, Newton DC. Driver Vigilance in Automated Vehicles: Effects of Demands on Hazard Detection Performance. Human Factors. 2019;61(3):474-487. doi:10.1177/0018720818802095

[Hoff, 2015] Hoff, K.A., Bashir, M.: Trust in automation integrating empirical evidence on factors that influence trust. Human Factors: The Journal of the Human Factors and Ergonomics Society 57 (2015) 407–434.

[Melcher, 2015] Melcher, V., Rauh, S., Diederichs, F., Widlroither, H., & Bauer, W. (2015). Take-over requests for automated driving. *Procedia Manufacturing*, *3*, 2867-2873.

[Michon, 1985] Michon, J. A. (1985). A critical view of driver behavior models: what do we know, what should we do?. In *Human behavior and traffic safety* (pp. 485-524). Springer, Boston, MA.

[NTSB, 2020] National Transportation Safety Board. 2020. Collision Between a Sport Utility Vehicle Operating With Partial Driving Automation and a Crash Attenuator, Mountain View, California, March23, 2018. Highway Accident Report NTSB/HAR-20/01. Washington, DC.

[Roche, 2019] Roche F, Somieski A, Brandenburg S. Behavioral Changes to Repeated Takeovers in Highly Automated Driving: Effects of the Takeover-Request Design and the Nondriving-Related Task Modality. Human Factors. 2019;61(5):839-849. doi:10.1177/0018720818814963

[SAE,2014] Society of Automotive Engineers: Automated driving levels of driving automation are defined in new sae international standard j3016 (2014) [last retrieved: June 2021].

[Sorkin, 1988] Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors*, *30*(4), 445-459.

[Storm,2002] Storm, H., Myre, K., Rostrup, M., Stokland, O., Lien, M.D. and Ræder, J.C. (2002), Skin conductance correlates with perioperative stress. Acta Anaesthesiologica Scandinavica, 46: 887-895. https://doi.org/10.1034/j.1399-6576.2002.460721.x

[Yang, 2017] Yang X. J., Unhelkar V. V., Li K. and Shah J. A. , "Evaluating Effects of User Experience and System Transparency on Trust in Automation," 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI, 2017, pp. 408-416.