(Over)Trust in Automated Driving: The Sleeping Pill of Tomorrow?

Thomas Kundinger
AUDI AG, Ingolstadt, Germany
Technische Hochschule Ingolstadt (THI), Germany
Johannes Kepler University (JKU), Linz, Austria
thomas1.kundinger@audi.de

Philipp Wintersberger
CARISSMA
Technische Hochschule Ingolstadt (THI), Germany
Johannes Kepler University (JKU), Linz, Austria
philipp.wintersberger@carissma.eu

Andreas Riener
Technische Hochschule Ingolstadt (THI), Germany
Johannes Kepler University (JKU), Linz, Austria
andreas.riener@thi.de

ABSTRACT

Both overtrust in technology and drowsy driving are safety-critical issues. Monitoring a system is a tedious task and overtrust in technology might also influence drivers' vigilance, what in turn could multiply the negative impact of both issues. The aim of this study was to investigate if trust in automation affects drowsiness. 30 participants in two age groups conducted a 45-minute ride in a level-2 vehicle on a real test track. Trust was assessed before and after the ride with a subjective trust scale. Drowsiness was captured during the experiment using the Karolinska Sleepiness Scale. Results

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312869

depict, that even a short initial system exposure significantly increases trust in automated driving. Drivers who trust the automated vehicles more show larger signs of drowsiness what may negatively impact the monitoring behavior. Drowsiness detection is important for automated vehicles, and the behavior of drowsy drivers might help to infer trust in an unobtrusively way.

KEYWORDS

Automated driving; driver drowsiness; trust; driver state; user study; subjective measures.

INTRODUCTION

Overtrust is a critical challenge for a safe use of automated vehicle (AV) technology [19]. Recent incidents with AVs, such as the fatal accident with Tesla Autopilot [8] or the Uber self-driving Taxi [2] are (at least partly) connected to overtrust, as drivers failed to monitor and intervene properly. Monitoring complex systems for longer periods of time is a challenging task *"even for highly motivated human beings"* (c.f. *"irony of automation"*) [3], and vigilance in such an environment is known to significantly degrade within half an hour [3] - in other words: monitoring is tiring. This is especially relevant as already in manual driving, drowsiness is suspected to be main cause of about 11% of fatal accidents [14]. Although automated driving (AD) promises to increase road safety, current systems mainly operate on level 2 [6] and have the inherent need to be monitored and corrected by human operators. Since drivers of future AVs are typical consumers rather then domain experts (such as permanently trained airplane pilots) [20], we must anticipate different expectations, knowledge and thus also behavior. We hypothesize that a connection exists between trust and the engagement in non-driving related tasks (NDRTs), which in case of drowsiness could become safety critical: Drivers that trust the automation more, may show greater willingness to fall asleep, and vice versa. To evaluate this claim we conducted a field experiment where participants had to monitor a 45-minute lasting trip in a real automated (robot-controlled) vehicle (see Figure 3) [18]. We assessed participant's subjective trust in the system before and after, and as well subjective drowsiness ratings frequent times during the trip. Statistical analysis of the collected data (based on two age groups, young and elderly drivers) allow to evaluate how trust levels change after a single session of real system exposure, as well as if the hypothesized correlation between subjective trust and drowsiness can be accepted.

Level 0 (no automation): Driver continually performs lateral and longitudinal driving task.
Level 1 (driver assistance): Driver continually performs lateral or longitudinal driving task.
Level 2 (partial automation): Driver must monitor the driving environment and driving task at all times.
Level 3 (conditional automation): Driver does not need to monitor driving and driving environment; must always be in position to resume control.
Level 4 (high automation): Driver is not required during defined use case.
Level 5 (full automation): No driver required.

Sidebar 1: Driver's responsibility in the Society of Automotive Engineers (SAE) International's levels of automation for on-road vehicles [6]

RELATED WORK

Lee and See define trust as *"attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"* [12], but also state that trust is an attitude underlain by beliefs, that leads to intentions and thus resulting reliance behavior [12]. Further, to accomplish proper levels, trust should match the "true capabilities of an agent" [12]. Overtrust is thus a situation, where subjective trust exceeds a system's capabilities, which can ultimately lead to misuse of technology [15].

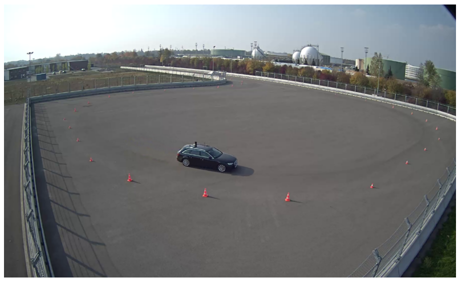

Figure 1: Outdoor test area with test vehicle (Audi A4 Avant)

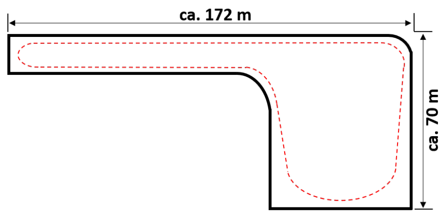

Figure 2: Schematic representation of the test area with dimensions in meters (m) and study track (dashed line)

Various situations can lead to overtrust, including pure performance based (poor calibration, resolution, or specificity [12]), but also pre-existing (dispositional, situational or learned trust [5]) explanations. Trust research often emphasizes the performance based component – for example, recent publications in the domain of AD suggest to make system capabilities and performance transparent to the user [9, 11, 13, 21]. However, we claim that future consumer-oriented users of AVs add an important aspect that has not been addressed yet. In contrast to professional operators, drivers may self-negotiate their trust levels to justify engagement in NDRTs, which is an often mentioned advantage of AVs [16]. We believe drowsiness as ideal candidate for evaluating such behavior since giving way to drowsiness (what can be re-formulated as "the willingness to fall asleep") exposes a high risk. Investigating if and how such NDRT engagement is affected by users' trust levels might not only reveal additional safety risks, it could also allow to behaviorally measure user trust in an unobtrusive way.

FIELD STUDY

To evaluate a) how trust is affected by initial system exposure and b) if a connection between subjective trust and drowsiness exists, we conducted a real driving study on a test track. Subjects had to take a 45-minute lasting trip in a simulated level 2 vehicle. The general study setting, except the transition from simulation to real world, is based on the study described in Kundinger et al. [10].

Vehicle and Test Track

To enable AD, the test vehicle (Audi A4 Avant) was equipped with a driving robot [18], which was attached to both the pedals and the steering wheel to take over vehicle control (see Figure 3), without restricting the normal sitting position or the ability to drive the vehicle manually. The maximum speed of the vehicle on the test track was 25 km/h. The temperature in the vehicle was set to 23 degrees C (A/C), and no radio/music was played during the drive. Three safety precautions were provided to make the ride as safe as possible: an additional set of driving pedals installed on the co-driver seat for the experimenter who was in the vehicle during the study, an emergency stop in the vehicle's center console, and a remote control operated by a person outside. In order to accelerate or at least not to interrupt the test person's drowsiness development during the test drive, and to reduce possible alerting effects, the route (closed loop) was designed as monotonous as possible (see Figure 2).

Method and Procedure

To assess subjective trust, participants had to complete the trust scale (TS) by Jian et al. [7], which provides sub-scales for both trust and distrust, before and after the drive. To assess drowsiness, we utilized the Karolinska Sleepiness Scale (KSS) [1] and issued the participants to rate their experienced drowsiness level in 5-minute intervals during the trip. The nine items of the scale (ranging from level 1 "extremely alert" to level 9 "very sleepy, sleep fighting") were presented vertically at a tablet

Figure 3: Test vehicle setup: tablet (1) for self-ratings with KSS; driving robot on pedals (2) and steering wheel (3)

computer in the vehicle's center console (see Figure 3). To be as little obtrusive as possible when reminding participants to update their drowsiness rating, we increased the tablet's brightness gradually (decreasing again after the rating). Upon arrival, participants completed a short demographic survey as well as the TS to assess their initial trust. Then, each participant completed a short 10-minute test drive where we explained the test vehicle and the procedure. Afterwards, the 45-minute lasting ride in the AV started. To simulate level 2 driving [6], drivers were issued to monitor the environment continuously, and be prepared to take over control in case system failures occur (however, no manual intervention was necessary). Participants were further importuned to not utilize their cell phone, not to eat or drink and avoid chewing gum, and to not close their eyes for a more extended time (or even fall asleep). Additionally, they were advised to abstain from conversing with the experimenter (except in case of health problems or other inconsistencies), and not to perform any secondary task despite rating their subjective drowsiness when prompted. After the drive, participants had to complete the TS again to assess the effect of initial system exposure on their trust levels. In addition, we captured a number of physiological signals and video recordings of the drivers' faces, and participants performed also a manual drive in equal conditions to compare the effects on drowsiness between manual and automated driving (additional results will be published elsewhere).

Participants

30 adults were recruited for this study with a newspaper advertisement. They were separated into two groups as indicated by their suggested sleep need over the life span, based on a report of the Sleep Health Foundation [17]. Half of the persons were selected in the age range 20-25 (7 female; 8 male; M = 24 years; SD = 1.486 years), the other half in the range 65-70 (5 female; 10 male; M = 67 years; SD = 1.831 years). For their efforts they were compensated with €40 in cash. The prerequisites for participating incorporated the ownership of a valid driving license, a good state of health, no sleep disorder, and no confinement in their driving ability. Besides, they were told not to consume any caffeinated beverages within 5 hours before participating.

Table 1: Descriptive statistics: values for mean (M), standard deviation (SD) and Cronbach's alpha (C. α) for trust/distrust items (before/after ride) and KSS-slope

	M	**SD**	**C. α**
Distrust (before)	1.85	1.20	.887
Distrust (after)	1.56	1.11	.846
Trust (before)	3.70	.94	.864
Trust (after)	4.18	1.11	.951
KSS-slope	.49	.05	-

RESULTS

In the following, we report the results of our evaluation. Statistical analysis was conducted using IBM SPSS V.24, effects are reported as significant at $p < .05$. Considering the trust scale, we calculated scale values for both concepts trust and distrust, since all scales showed acceptable reliability (Cronbach's alpha $> .846$ for all scales, see Table 1). Since not all of our data was normally distributed, we conducted Wilcoxon signed ranked tests to evaluate within-subjects effects. Participants rated the sub-scale *distrust* lower after than before the trip with the AV, however the difference is not statistically significant ($p = .130$). Ratings for *trust* on the other hand increased significantly ($p = .002$) after the ride. Considering between subjects-effects with respect to gender or the different age groups, no significant differences could be found. Male drivers rated trust in the vehicle after the ride ($M = 4.58, SD = .86$) higher than female drivers ($M = 3.56, SD = .145$), but a statistical significance was not given ($p = .068$). To quantify the effect of the 45-minute monitoring task on drowsiness, we performed linear regression on the nine subsequent KSS ratings of each participant, and calculated the slope of the regression line. This allowed to express an increase of drowsiness in a single number, while omitting interpersonal differences emerging from the ordinal nature of the scale (c.f. an increase from level 1 to 4 shows an equal slope than an increase from 4 to 7). Statistical evaluation using Mann-Whitney U tests revealed no significant differences, neither between the two age groups, nor regarding gender. Considering a potential correlation between trust and drowsiness, we found a significant positive correlation between the drowsiness increase (KSS-slope) and trust ratings after the drive ($r = .408, n = 30, p = .013$).

DISCUSSION AND CONCLUSION

First, results of our experiment with a "real" (robot-controlled) AV confirm what has been hypothesized in driving simulator studies [4] – even short initial system exposure significantly increases trust in AD technology. Further, considering our correlation analysis of subjective trust and drowsiness, we could validate our hypothesis that drivers who trust the AV more also show larger (self-reported) signs of drowsiness. This is important as the attested safety risk of drowsy driving could become even more critical with AVs, that (at level 2) demand being permanently monitored by the driver. On the other hand, this could allow to include (given this assumption holds for physiological measurements) drowsiness measures as unobtrusive behavioral measure for automation trust, too. Increased signs of drowsiness could thus be interpreted in a way, that drivers of AVs “accept to fall asleep due to high trust in automation”. This could be combined with other physiological/behavioral measures to be able to infer trust levels, since trust in automation is an abstract concept difficult to assess [20]. However, future experiments will be necessary to evaluate if these results preserve over other types of NDRTs, what could ultimately lead to a sophisticated quantification of driver trust from engagement in side activities. Moreover, to validate the results, additional measures other than surveys should be evaluated. Research on both trust in automation and drowsiness will be necessary to prevent misuse, and allow a successful implementation of AD technology.

REFERENCES

[1] T. Akerstedt and M. Gillberg. 1990. Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience* 52 (1990), 29–37. https://doi.org/10.3109/00207459008994241

[2] Andrew J. Hawkins. 2018. Uber’s self-driving car showed no signs of slowing before fatal crash, police say. https://www.theverge.com/2018/3/19/17140936/uber-self-driving-crash-death-homeless-arizona (retrieved on February 15, 2019).

[3] Lisanne Bainbridge. 1983. Ironies of automation. In *Analysis, Design and Evaluation of Man–Machine Systems 1982*. Elsevier, 129–135.

[4] Christian Gold, Moritz Körber, Christoph Hohenberger, David Lechner, and Klaus Bengler. 2015. Trust in automation–Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing* 3 (2015), 3025–3032.

[5] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434.

[6] International SAE and On-Road Automated Driving (ORAD) Committee. 2014. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems (J3016 Ground Vehicle Standard). *SAE International* 2014-01 (2014). https://saemobilus.sae.org/content/J3016{_}201401

[7] Jiun-Yin Jian, Ann M Bisantz, and Colin G Drury. 2000. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics* 4, 1 (2000), 53–71.

[8] Jordan Golson. 2016. Tesla driver killed in crash with Autopilot active, NHTSA investigating. https://www.theverge.com/2016/6/30/12072408/tesla-autopilot-car-crash-death-autonomous-model-s (retrieved on February 15, 2019).

[9] Jeamin Koo, Jungsuk Kwac, Wendy Ju, Martin Steinert, Larry Leifer, and Clifford Nass. 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 (IJIDeM)* 9, 4 (2015), 269–275.

[10] Thomas Kundinger, Andreas Riener, Nikoletta Sofra, and Klemens Weigl. 2018. Drowsiness Detection and Warning in Manual and Automated Driving: Results from Subjective Evaluation *(AutomotiveUI '18)*. 229–236. https://doi.org/10.1145/3239060.3239073

[11] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 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*. ACM, 183–189.

[12] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (2004), 50–80.

[13] Brittany E Noah and Bruce N Walker. 2017. Trust Calibration through Reliability Displays in Automated Vehicles. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 361–362.

[14] J. M. Owens, T. A. Dingus, F. Guo, Y. Fang, M. Perez, J. McClafferty, and B Tefft. 2018. Prevalence of Drowsy Driving Crashes: Estimates from a Large-Scale Naturalistic Driving Study. (Research Brief.). *Washington, D.C.: AAA Foundation for Traffic Safety* (February 2018).

[15] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 2 (1997), 230–253.

[16] Bastian Pfleging, Maurice Rang, and Nora Broy. 2016. Investigating User Needs for Non-driving-related Activities During Automated Driving. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 91–99. https://doi.org/10.1145/3012709.3012735

[17] Sleep Health Foundation. 2015. Sleep Needs Across The Lifespan. http://www.sleephealthfoundation.org.au/files/pdfs/Sleep-Needs-Across-Lifespan.pdf (retrieved on January 6, 2019).

[18] Stähle GmbH. 2019. Automated Driving System SFPHYBRID for cars. https://www.staehle-robots.com/english-1/products/sfphybrid-eng/ (retrieved on January 7, 2019).

[19] Alan R Wagner, J. Borenstein, and A. Howard. 2018. Overtrust in the robotic age. *Commun. ACM* 61, 9 (2018), 22–24.

[20] Philipp Wintersberger and Andreas Riener. 2016. Trust in Technology as a Safety Aspect in Highly Automated Driving. *i-com* 15, 3 (2016), 297–310.

[21] Philipp Wintersberger, Tamara von Sawitzky, Anna-Katharina Frison, and Andreas Riener. 2017. Traffic Augmentation As a Means to Increase Trust in Automated Driving Systems. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter (CHItaly '17)*. ACM, New York, NY, USA, Article 17, 7 pages. https://doi.org/10.1145/3125571.3125600