Towards a Personalized Trust Model for Highly Automated Driving

Philipp Wintersberger[1], Anna-Katharina Frison[1], Andreas Riener[1], Linda Ng Boyle[2]

University of Applied Sciences Ingolstadt, Germany; Email: firstname.lastname@thi.de[1]
University of Washington, Seattle, US; Email: linda@uw.edu[2]

Abstract

User acceptance of automated vehicles (and dependent dimensions such as road safety, frequency of use or level of recommendation) is said to be highly dependent on the operator's individual trust in this technology. As a consequence, the development of driving functions and future driver-vehicle interfaces should allow for appropriate trust calibration. To better understand trust and the effect of mis-calibration on the way to a personalized trust model, we propose a set of trust-related research questions derived from related work and our own user studies. Based on preliminary investigation, we recommend examining 1) differences in users and subgroups of users, 2) different levels of trust based on situation or context, 3) methods for quantifying trust in naturalistic driving studies, and 4) definitions for an established/approved trust model and the individual calibration of the model with regard to driving behavior and automotive user interfaces. The final outcome should be a multidimensional trust model that fits the individual passenger/driver by dynamically adapting driving mode and UI representation/feedback.

1 Introduction

In the upcoming "age of automation", the role of the human–machine interface will drastically change. While many high-complexity automated systems today are operated by expert, the operators of automated systems in the future (with self-driving cars being a prominent representative) will be the everyday consumer. They will be able to decide among a range of products, and select options that fit their needs and expectations. In contrast to devices used today, future autonomous systems will perform more complex and perhaps more safety-critical tasks. Looking at the latest advances in vehicular technology and efforts from Google, Apple,

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
B. Weyers, A. Dittmar (Hrsg.):
Mensch und Computer 2016 – Workshopbeiträge, 4. - 7. September 2016, Aachen.

http://dx.doi.org/10.18420/muc2016-ws08-0008

Tesla and other automotive manufacturers, self-driving vehicles could be among the first automated[1] systems available for the broader public.

A main concern for designers and researchers is how people will trust this technology. Trust will play a key role in the success of automated vehicles (and other automated systems), as users actually place their life in the "hands" of a complex computer system. Various studies have shown that users' trust in highly automated driving (HAD) differs a great deal. Also being unaware of the system boundaries can lead to bad trust calibration. For instance, Dickie and Boyle (2009) found, that a reported large group of users had incorrect knowledge about the boundaries of their adaptive cruise control (AAC) system. The consequence is that's drivers' relied on the system even in situations when it could not work (e.g., tight curvature, stop-and-go traffic).

Wrongly calibrated (understood) trust is a safety issue that will be impacted by the level of automation, the operator experience, and system penetration. As a result, it will be necessary for future vehicle HMIs to adapt to the user to create "appropriate trust", allowing him/her to recognize system failures and potentially dangerous situations - a matter of life and death in the driving context. To support drivers in the steering and monitoring process while performing different kinds of non-driving tasks (i.e., secondary/tertiary tasks), technology must be able to measure and influence their individual trust levels as a reaction to various personal (situation awareness, fatigue, stress, driving performance) and environmental (environment, system uncertainty due to traffic volume, weather conditions, sensor outages) factors. With this work, we highlight the need for further trust-related research, propose expected differences in trust between people, discuss recent advances and finally come up with a set of research questions to support trust calibration for automated vehicles.

2 Trust in Automated Vehicles

Trust is an abstract concept that has many definitions. Therefore, we fall back upon the trust definition of Mayer and Schoorman (1995) and the analogy of Lee and See (2004), as done by Ekman et al. (2016), to allow a direct connection between interpersonal trust and trust in automation. Trust is built on the possibility to observe the system's behavior (i.e., performance), understand the intended use of the system (its purpose), as well as understand how it makes its decisions. Already Paul Fitts had suggested in a principle known as "HABA-MABA", that humans are poor at monitoring automated systems (1951) and, according to Wagner (2015), humans may learn to "*unjustifiably trust the car and won't monitor its behavior for errors*". Bainbridge (1983) argued, that humans are not very effective when under time pressure, and Norman (1990) added, that many automation catastrophes result from missing or inappropriate feedback. People's experience with a system can, according to Parasuraman and Riley (1997), be shaped of use, misuse or disuse. Muir (1994) called the "*process of adjusting trust to correspond to an objective measure of trustworthiness*" as

[1] corresponding to the German Verband der Automobilindustrie (VDA), level 3 or higher. At that level, the system works independently without the need for permanent monitoring by a human operator.

"*calibration of trust*", where this calibration should eliminate a potential over- or distrust of the operator by education and retraining. Lee and See (2004) provided design considerations for developers to design for appropriate trust. Ekman and colleagues (2016) state that a holistic approach is necessary as trust development "*starts long before a user's first contact with the system, and continues long thereafter*". Hergeth et al. (2015) could show, by analyzing users' gaze behavior that a connection exists between trust and monitoring frequency. Payre and colleagues (2015) showed that overtrust in automated driving leads to higher reaction times and Helldin et al. (2013) presented system uncertainty to users and demonstrated that this can increase the take-over performance. In contrast to automation in airplanes, where only well trained experts (pilots) are allowed to operate with a system that hardly faces critical situations, piloting an automated vehicle will be performed by many characters as different as possible, without the opportunity of trial and error or complex guided training sessions while always residing in a safety critical environment. Further, trust in automated vehicle is not restricted to a user's view on a single vehicle, but also the whole idea of automated traffic in general. In a recent article, Goodall (2016) discusses the moral issues emerging from automated traffic and states, that the public needs decisions of the systems in order to be rational and comprehensible.

In summary, future vehicle HMI must account for individual differences in terms of the driving and monitoring processes to 1) provide for the best possible performance of the driver–vehicle interaction and 2) demonstrate the overall benefits of automation with varying levels of market penetration.

3 Research Questions and Proposed Methods

In a recent Dagstuhl Seminar[2], more than 30 experts in the field discussed about the future of automated driving. During one week and a lot of controversial discussions, no agreement was found on the definition or agenda of trust research. Nevertheless, the topic per se was found to be very important. To shape a roadmap to further understand trust, we propose a model for personalized trust calibration for vehicles, considering that there are different dimensions of trust, and calibration of personalized trust is highly dependent on the level of automation, the amount of exposure to the automation, and the environment the automated systems are in use. Although trust in automated vehicles is also relevant for other traffic participants like pedestrians or cyclists, this paper only deals with trust calibration for vehicle operators and/or occupants. By defining different dimensions of trust that can be quantified, we could classify users and provide them individual feedback in various situations, thus calibrate their trust. Therefore, we want to find answers to the following research questions:

- **RQ1: What trust-related differences exist between individuals?**
  In a recent user study (N=47, 26 female, age $\overline{xf}$ = 22.5 ± 7.3, 21 male, age $\overline{xm}$ = 24.04 ± 12.04, all Computer Science students) comparing mental and emotional states

[2] Dagstuhl Seminar 16262 "Automotive User Interfaces in the Age of Automation", http://www.dagstuhl.de/16262.

of passengers in a high-fidelity driving simulator (2016), we could show that people accept HAD similarly as if they were passengers of a human driver (either female or male gender). Nevertheless, qualitative interviews conducted with the individual subjects revealed a difference in the personal trust they attribute to HAD – while some stated an automated vehicle to be much more trustworthy than human drivers, others had a completely different opinion and claimed to trust humans fundamentally more than "computers". Our position is that we did not expect peoples' trust in HAD to be equally distributed, rather we expected differences to be observed between age groups (Fig. 1) and other relevant driver characteristics such as education level, gender, cognitive impairment, and even risk taking behavior. The validity of our assumptions will be evaluated with additional studies in a driving simulator environment, supported by subjective evaluation. The goal is to examine the effects of different levels of automation on trust of different user groups.

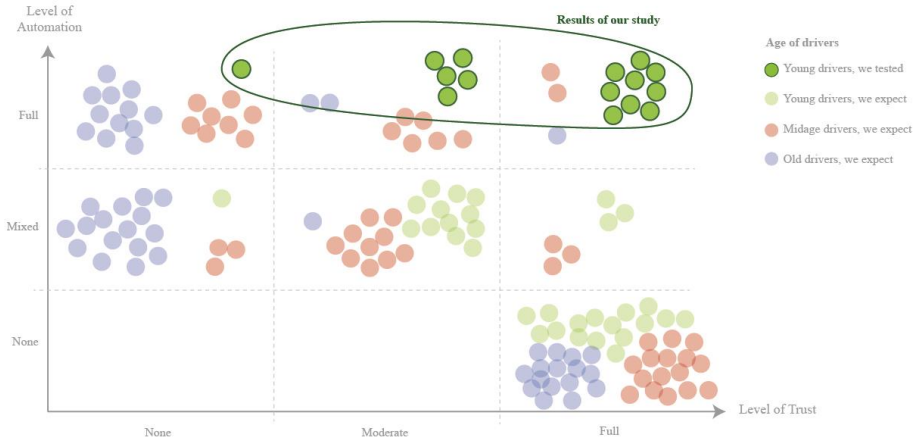

Fig. 1: Results of qualitative interviews showed that many young people trust automated vehicles, but some do not. Mid age or older people are not as familiar with technology as younger generations. Differences should be evaluated also by comparing other aspects like highest education, gender, disabilities, etc.

- **RQ2: What dimensions of trust are important for HAD?**
  There is no single dimension of "trust in automated vehicles." Hence, we divided trust into more concrete statements such as "trust in navigation", "trust in correct driving maneuvers", "trust in proper take-over timing and strategy.". A question thus would be which aspects are especially interesting for the overall trust model, and which can be combined or removed? To model the space, existing theories of trust and user acceptance will be investigated. The results will be then both discussed with experts and evaluated using different user groups.

- **RQ3: How can we unobtrusively quantify trust in real time?**
  Trust calibration is a continuous process, and trust levels of a single person might change even within a short trip. To create appropriate trust, somehow the current state

of trust must be quantified. This mechanism has to work in real time while being unobtrusive – consumers most likely will not be pleased by regular trust surveys issued by their vehicles. As Hergeth et al. (2015) stated, there might exist a connection between trust and a user's monitoring frequency, what could be a good starting point for further investigations. But not only gaze behavior, also approaches for classifying situation awareness or mental workload will be needed to deduce wrong calibrated trust. The individual measurements then can build a feature vector that determines the user's position in the model.

- **RQ4: How can user interfaces be designed to ensure an appropriate level of trust?**
  When we are able to quantify a user's trust levels, we can think about the actual process of trust calibration – which methods do exist to increase or decrease trust? Trust calibration will be a form of communicating with the driver, what will require intensive HMI research. We think that this cannot be achieved with a general method, different users or user groups might be best targeted differently based on their trust related features – the question is, which strategies and modes of communication will lead to the best results. Also, trust is much more easily to destroy than to build up, thus such methods have to be taken carefully. Nevertheless, both will be necessary to calibrate trust levels of a user.

4 Conclusion

Calibrating trust to the passengers of automated vehicles will be important to prevent misuse, disuse or wrong expectations such as not knowing the system's boundaries. Based on the individual situation and user, different forms of feedback will be appropriate to re-calibrate trust to the proper levels. By continuing research in this domain, we might be able to build a multidimensional trust model, where different measures can be quantified to instantly assign appropriate system functionality and user feedback. This could lead to better system usage, a more pleasurable user experience, and finally to less critical driving situations.

Acknowledgements

This paper is based, in part, on discussions with participants of the Dagstuhl Seminar 16262 "Automotive User Interfaces in the Age of Automation", http://www.dagstuhl.de/16262.

References

Bainbridge, L. (6 1983). Ironies of Automation. *Automatica Vol. 19*, S. 775-779.

Billings, D. R., Schaefer, K. E., Chen, Y. Y., & Hancock, P. A. (2012). Human-Robot Interaction: Developing Trust in Robots. *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, S. 109-110.

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. *Advances in human performance and cognitive engineering research*.

Coeckelbergh, M. (2012). Can we trust robots? *Ethics and information technology 14.1*, S. 53-60.

Dickie, D. A., & Boyle, L. D. (10 2009). Drivers' Understanding of Adaptive Cruise Control Limitations. *Proceeding of the Human Factors and Ergonomics Society Annual Meeting*.

Ekman, F., Johansson, M., & Sochor, J. L. (2016). Creating Appropriate Trust for Autonomous Vehicle Systems: A Framework for HMI Design. *Proceedings of the 95th Annual Meeting of the Transportation Research Board*.

Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system.* National Research Council, Div. of.

Goodall, N. J. (2016). Can you program ethics into a self-driving car? *IEEE Spectrum 53.6*.

Guy, W. H., Stanton, N. A., & Young, M. S. (2001). Where is computing driving cars? *International Journal of Human-Computer Interaction 13.2*, S. 203-229.

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 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM*, S. 210-217.

Hergeth, S., Lorenz, L., Krems, J. F., & Toenert, L. (2015). Effects of Take-Over Requests and Cultural Background on Automation Trust in Highly Automated Driving. *8th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society 46.1*, S. 50-80.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of management review*, S. 709-734.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the studiy of trust and human intervention in automated systems. *Ergonomics 37.11*, S. 1905-1922.

Norman, D. A. (1990). The Problem with automation: inappropriate feedback and interaction, not over-automation. *Philosophical Transactions of the Royal Society of London*, S. 585-593.

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*, S. 230-253.

Payre, W., Cestac, J., & Delhomme, P. (2015). Fully Automated Driving Impact of Trust and Practice on Manual Control Recovery. *Human Factors: The Journal of the Human Factors and Ergonomics Society*.

Wagner, M. (2015). *A Philosophy for Developing Trust in Self-driving Cars, in Road Vehicle Automation 2.* Springer International.