Employing Virtual Agents for Building Trust in Driving Automation: A Qualitative Pilot Study

Bernhard Hilpert, Tanja Schneeberger, and Patrick Gebhard
firstname.lastname@dfki.de
German Research Center for Artificial Intelligence (DFKI)
Saarland Informatics Campus, Saarbrücken, Germany

ABSTRACT

The acceptance of driving automation is influenced by user trust. Integrating virtual agents into the interface between automation and user has been found to increase trust, also in driving automation. In this context however, it remains unclear what virtual agents' characteristics influence trust. This paper presents a pilot study about how inexperienced users picture a virtual agent's competencies, communication style, and embodiment to increase trust in driving automation level 4. In semi-structured interviews with 10 participants, we analyzed both the general situation of automated driving and two specific situations. Results show that situational understanding, situational judgment, and emotional relation are the most relevant competencies regarding building trust. Nearly all users preferred a human-like embodiment. This pilot study's results lay the foundations for a larger scale quantitative study aiming to examine various factors that influence user trust.

KEYWORDS

trust, autonomous driving, self-driving vehicles, virtual agents

Reference:
Bernhard Hilpert, Tanja Schneeberger, and Patrick Gebhard. 2021. Employing Virtual Agents for Building Trust in Driving Automation: A Qualitative Pilot Study. In *Workshop on Robo-Identity: Artificial identity and multi-embodiment at the 16th International Conference on Human-Robot Interaction (HRI '21), March 08, 2021, Boulder, CO.*

1 INTRODUCTION

There is a high chance that self driving vehicles (SDVs), as robotic agents [18], will become one of the major fields of interaction between humans and advanced technology [15]. In general, driving automation is expected to have several benefits [20]. However, these benefits will only be achieved if drivers accept the novel technology. As trust seems to be a critical influence factor for acceptance in this context [3, 11, 17], it might be beneficial if the agent is realized so that it enhances the user's trust in driving automation. While a range of different approaches to implement trust-inducing agents in the context of SDVs has already been proposed [10, 16, 21], empirical research on the factors that determine trust in such agents is scarce at this point in time. It remains unclear what role the agent needs to fulfil in order to increase trust in the user and how its embodiment influences the users' perception. In a primary interview study, we assessed the users' fundamental expectations regarding the role of such agents in SDVs to induce trust including its competencies, communication style and and lastly its form of embodiment. The results of this qualitative study will further provide the basis to conduct a series of quantitative studies and collect empirical data on the determinants of trust in a virtual agent in SDVs.

2 THEORETICAL BACKGROUND

2.1 Driving Automation and Trust

Driving automation systems perform part or all of the dynamic driving task on a sustained basis on a range between no driving automation (level 0) to full driving automation (level 5). Level 4, the high driving automation, describes cars that do not require human intervention in most circumstances and can operate in self-driving mode in geographically prescribed areas [13]. Level 4 cars can be seen as a form of intelligent interactive robots [18] and have high potential as so-called robotaxis [2, 19]. According to Hoff and Bashir, trust in automation can be distinguished in dispositional, situational, and learned trust, whereby specific factors influence each. Dispositional trust represents the user's overall tendency to trust automation, independent of context or system (e.g., age). Situational trust is influenced by the external environment (e.g., system complexity) and the internal, context-dependent characteristics of the user (e.g., attentional capacity). Learned trust is dependent on the evaluation of a system drawn from past experience or the current interaction. Dynamically adjusting, the learned trust level is based on the experiences about the system's functionality and performance. Lee and See further define trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 51) [17]. Trust in an automated system therefore might depend on the expectation that the agent possesses the competencies necessary to meet the challenges of the situation. Since there is no existing literature on what constitutes these expectations, a first research question can be formulated:

1) Which competencies should an agent have to increase trust?

According to [17], the level of trust is dynamically calibrated based on the correspondence of the users' expectation and their perception of the agents' capabilities. Without support, it seems challenging for drivers to understand the capabilities and limitations of driving automation [5]. Especially with increasing driving experience, users may find it discomforting to give up control to driving automation systems they do not fully understand [1]. Transparency has proven to play a critical role in accommodating complexity [14]. To reduce system complexity and increase understanding and trust, a transparent and user-friendly communication of the agents' capabilities might help overcoming uncertainty. While recommendations towards the design of automated systems' communication style have been formulated [11], the users' explicit expectations towards a agents' communication style remain unclear. Therefore we formulate the second research question:

2) Which communication style should an agent display to increase trust?

2.2 Embodiment and Trust

Lastly, the embodiment of the agent plays a critical role in inducing trust in the user. Who the agent comes across is consequential for the acceptance and the building of trust in SDVs. The use of embodied virtual agents to increase trust in driving automation has been considered [10]. The authors compared different visualizations of a driving automation's interpretation of the current situation and its corresponding actions aiming to increase trust. Compared to a baseline showing car indicators, a human-like chauffeur avatar did not significantly increase trust. It seems that unspecific human-like appearance might not be enough to increase trust in driving automation per se. Research from other automation contexts, however, has shown that the integration of human-like virtual agents that explain complex facts can lead to an increase of trust in the autonomous intelligent system [22]. An anthropomorphic perception increases users' understanding by allowing them to attribute pre-existing knowledge and structures. This positively influences comprehension and helps to establish an emotional relation [6]. In driving automation, adding human-like voice output as an anthropomorphic feature was found to increase trust [21]. It seems to be unclear what form of embodiment the users expect and how it influences their perception of the agent and the SDV itself. Based on the findings that anthropomorphic agents increase trust [4, 6, 21, 22], we formulate the following *hypothesis*: *An agent aiming to increase trust in driving automation, is expected to have an anthropomorphic appearance.*

2.3 Interview Outline

As a first step to answer these research questions and find first evidence that supports the hypothesis, we conducted qualitative interviews with drivers usually driving cars within the automation levels between 0 or 1. Trust is influenced by general determinants, but also highly situation- and context-dependent and therefore dynamic [11, 12], Moreover, trust is a dynamic variable calibrated along provided information prior to and during the initial drive with an automated vehicle [11, 14] or in high stakes situations[9]. We explore the two research questions and the hypothesis for both the general situation of automated driving and in two trust critical driving scenarios. The first situation describes the very first experience with the automated vehicle and the first interaction with the virtual agent. The second situation describes a high-stakes situation, namely, entering a motorway.

3 METHODS

3.1 Participants and Interviewers

We recruited 10 participants via social network applications, who were carefully selected to be representative concerning certain key variables such as gender, age, and driving habits. The sample was equally distributed in gender (5 female) and age (M = 51.4, , SD = 22.2). Every age group (18-30, 31-45, 46-60, 61-75, above 75) was covered with a female and a male participant each. Participants drove on average 213.7 km per week (SD = 214.8 km, Min = 2 km, Max = 600 km). All participants were German native speakers and acquired their driver's licence and driving experience in Germany. Participation was voluntary.

The four interviewers (2 female) were aged between 22 and 31.

3.2 Procedure

The interviews were either conducted virtually or in-person (especially with the elderly participants). For the video conference interviews, the camera of both the interviewee and the interviewer was enabled. After welcoming the participant, the interviewer explained the interview's goal and outline. To ensure the same level of knowledge regarding driving automation level 4 and virtual agents, a standardized description of both was read aloud. Afterward, participants were asked to imagine themselves in an automated car, supported by Video 1. Then, participants were asked to answer general questions on their expectations towards the agents' appearance, competencies, and communication style that would increase their trust. This was followed by describing two specific, trust critical situations: First contact with an automated vehicle and entering a motorway. For each of these situations, a video was shown (Video 2 + 3), and participants were subsequently asked about their expectations towards the agents' situation-specific competencies and communication style. The interview was concluded by three general questions and the inquiry of the demographic data. In the end, participants were thanked for their participation and released. The interviews were audio-recorded and took on average 38.5 minutes.

3.3 Material

The interviews were conducted in a semi-structured setting, where the interview structure, the set of questions, and their order remained the same for all participants.

Competencies were assessed using five general (e.g., "What do you expect from a virtual agent who creates trust in an automated car?") and four questions for each trust-critical situation (e.g., "What would you expect from the agent to say or do in this situation for you to be able to trust the driving automation?") on the content the agent should provide to increase trust.

Communication style was examined using 4 questions on how the agent should interact with the interviewee (e.g., "How would you like the agent to act in this situation for you to be able to trust it?").

Embodiment included a general picture of the agent, anthropomorphism level, preferred gender, visualization, and presence. In addition to that, the participants were asked specifically if they perceived the agent as a representation of the SDV or a separate entity.

The *demographic data* included among others age, gender, and weekly driven kilometers.

We showed three first person persepctive *videos* to help the participants picture themselves being inside an automated vehicle. Video 1 (90 s) showed a drive with a Tesla driving through different situations without the driver's interference. Video 2 (30 s) showed somebody entering a Tesla and putting on the seat belt. Video 3 (70 s) showed a dashcam video of a car entering a German motorway. All videos were played without sound but explained by the interviewer.

4 RESULTS

To extract meaningful insights concerning competencies, communication style, and appearance, we analyzed the data in a three-step

Table 1: Concepts, subconcepts and respective interviewee quote.

Concept of Competence	Subconcepts	Example
Situational understanding	(1) Context knowledge	Weather/Route/Traffic
	(2) System status	Tire pressure
	(3) Situational awareness	Location of other cars
Situational judgement	(1) Indication of critical situations	"Warning: Pedestrian ahead"
	(2) Explanation of driving behavior	"Adjusting speed to traffic"
Emotional relation	(1) Perspective taking	Understanding a situation is challenging
	(2) Empathy	Support when emotionally challenged
	(3) Adaptivity	Adapt to users' preferences
	(4) Small talk	Greeting/jokes/Asking for feedback

process using the Grounded Theory method [7, 8]. As a first step to analyze interviewees' answers, the recorded audios were transcribed and coded by an independent annotator. In a second step, we categorized the coded answers by content and relevance to the research questions. Thirdly, we classified them to extract concepts and subconcepts of explanatory value regarding the three research questions.

Regarding *Research question 1*, the analysis revealed that participants expected three critical competencies of the agent (see Tab. 1).

Research question 2 focused on the communication style and assessed how the virtual agent should communicate the competencies to increase trust. In total, six attributes could be identified. The mentioned communication styles were "explanatory" (10x), "informative/factual" (9x), closely followed by "reassuring" (8x), "calm" (6x), "friendly" (6x) and "humorous" (3x).

All participants expected the virtual agent to communicate proactively in safety-critical situations. Outside of these situations, two participants explicitly stated that the agent should only talk if directly addressed. The other eight participants expected the virtual agent to be proactive also in non-safety critical situations.

Our *hypothesis* stated that the users would expect an anthropomorphic appearance of the trust increasing agent. The interview answers confirmed this hypothesis. 9 participants requested a human-like visual appearance. The remaining person wanted an agent with a natural language interface and an abstract visualization. 6 participants favored a male agent, 3 a female, and 1 was without preference. Of those participants asking for a human-like appearance, 2 wanted to see a head, 2 wanted a representation of a bust, and 5 wanted to see an upper body including hands. 6 participants wanted the agent to be present all the time, while the rest only when necessary or requested. 8 participants stated that they perceived the agent as a entity separate from the SDV, while the remaining 2 saw no separation. One person specifically expressed the wish to use the assistant of his smart-home system in his SDV.

5 DISCUSSION, CONCLUSION AND FUTURE WORK

This study aimed to examine users' expectations towards a trust increasing virtual agent in driving automation. We focused in 10 interviews on competencies, communication style, and embodiment. Results indicate that the primary expectation towards the virtual agent is to assure the user of the drives' safety. This includes the explanation of situational understanding and judgment. The results regarding the communication style are in line with this, giving the virtual agent a role as explanatory middleman between the driving automation and the user. Furthermore, most of the content that was expected to be proactively communicated referred to safety-critical situations, while most users also emphasized an informative communication style. The agent seems to have a very interactive function between the user and the driving automation. This gives a first hint that users might need an embodiment to understand the driving automation decisions and acquire a feeling of safety and trust. One other important competence of the virtual agent is its ability to build an emotional relation. It seems that users want to be understood, supported, and taken care of to mitigate uncertainty.

The result regarding the embodiment of a trust inducing agent supports other findings that anthropomorphism increases trust in automation [4, 6, 21, 22]. Additionally, our findings suggest that the agent does not seem to represent the car per se, but rather to fulfil the role of an assistant. This poses the question if the migration or integration of already existing smart assistants might be an option and what its effects on trust would be. Overall, the results of this qualitative study seem to show that it might be difficult for users to trust faceless driving automation that comes without an embodiment. An interactive virtual agent with a human-like appearance and transparent as well as empathetic demonstration of its competencies could support the users in building trust into driving automation. It has to be pointed out that the results of this study should be seen as preliminary and a first step into understanding how to increase trust with a virtual agent in driving automation. Moreover, these results only apply to drivers in the German traffic area and might not be generalized to other countries. However, the results lay the foundation for our planned quantitative study with several agent prototypes to further explore the factors influencing user's trust. Moreover, we want to include the analysis of factors that influence dispositional trust as individual's overall tendency to trust automation, independent of context or a specific system [11].

ACKNOWLEDGMENTS

This work is partly funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) within the EmmI project (funding code 19A20008E).

REFERENCES

[1] Hillary Abraham, Chaiwoo Lee, Samantha Brady, Craig Fitzgerald, Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. 2017. Autonomous vehicles and alternatives to driving: trust, preferences, and effects of age. In *Proceedings of the Transportation Research Board 96th Annual Meeting (TRB'17)*.

[2] Douglas A. Bolduc. 2018. *Volvo, Baidu team up for Level 4 autonomous EVs in China.* https://www.autonews.com/article/20181101/MOBILITY/311019997/volvo-baidu-team-up-for-level-4-autonomous-evs-in-china Accessed: 2021-02-09.

[3] Jong Kyu Choi and Yong Gu Ji. 2015. Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction* 31, 10 (2015), 692–702.

[4] Ewart J De Visser, Samuel S Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.

[5] Aaron Edelmann, Stefan Stümper, and Tibor Petzoldt. 2019. Specific Feedback Matters-The Role of Specific Feedback in the Development of Trust in Automated Driving Systems. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2223–2228.

[6] Nicholas Epley, Adam Waytz, and John T Cacioppo. 2007. On seeing human: a three-factor theory of anthropomorphism. *Psychological Review* 114, 4 (2007), 864.

[7] Barney G Glaser and Anselm L Strauss. 1967. Grounded theory: Strategies for qualitative research. *Chicago, IL: Aldine Publishing Company* (1967).

[8] Barney G Glaser and Anselm L Strauss. 2010. *Grounded theory: Strategien qualitativer Forschung*. Huber.

[9] 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.

[10] Renate Häuslschmid, Max von Buelow, Bastian Pfleging, and Andreas Butz. 2017. Supporting Trust in autonomous driving. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 319–329.

[11] 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.

[12] Brittany E Holthausen, Philipp Wintersberger, Bruce N Walker, and Andreas Riener. 2020. Situational Trust Scale for Automated Driving (STS-AD): Development and Initial Validation. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 40–47.

[13] SAE International. 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://www.sae.org/standards/content/j3016_201806/ Accessed: 2021-02-09.

[14] Johannes Kraus, David Scholz, Dina Stiegemeier, and Martin Baumann. 2020. The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction, and system transparency. *Human Factors* 62, 5 (2020), 718–736.

[15] Miltos Kyriakidis, Riender Happee, and Joost CF de Winter. 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour* 32 (2015), 127–140.

[16] David R Large, Kyle Harrington, Gary Burnett, Jacob Luton, Peter Thomas, and Pete Bennett. 2019. To please in a pod: employing an anthropomorphic agent-interlocutor to enhance trust and user experience in an autonomous, self-driving vehicle. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*. 49–59.

[17] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.

[18] Sven Nyholm and Jilles Smids. 2018. Automated cars meet human drivers: responsible human-robot coordination and the ethics of mixed traffic. *Ethics and Information Technology* (2018), 1–10.

[19] Alexandria Sage. 2018. *Waymo unveils self-driving taxi service in Arizona for paying customers.* https://www.reuters.com/article/us-waymo-selfdriving-focus/waymo-unveils-self-driving-taxi-service-in-arizona-for-paying-customers-idUSKBN1O41M2 Accessed: 2021-02-09.

[20] Chris Urmson and William Whittaker. 2008. Self-driving cars and the urban challenge. *IEEE Intelligent Systems* 23, 2 (2008), 66–68.

[21] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52 (2014), 113–117.

[22] Katharina Weitz, Dominik Schiller, Ruben Schlagowski, Tobias Huber, and Elisabeth André. 2019. "Do you trust me?" Increasing user-trust by integrating virtual agents in explainable AI interaction design. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*. 7–9.

Bernhard Hilpert is an assistant researcher at the Affective Computing Group at the German Research Centre for Artificial Intelligence. He received a Master's Degree in Psychology in 2020 at the Saarland University in Saarbrücken, Germany. He is conducting research on cognitive and affective processes as well as mental models in Human-Computer-Interaction.

Tanja Schneeberger is researcher at the Affective Computing Group at the German Research Centre for Artificial Intelligence. She received a Master's Degree in Psychology in 2014 at the department of work and organizational psychology at the Saarland University in Saarbrücken, Germany. She is conducting research on dyadic interactions between humans and social agents.

Patrick Gebhard is head of the Affective Computing Group at the German Research Centre for Artificial Intelligence (DFKI). He has long-term experience in the evaluation, representation, simulation, and display of emotions. His research started over a decade ago simulating emotions for the creation of believable agent behavior in interactive human-agent systems. He uses this expertise to research human interaction with the goal to enable a human-like interaction with computer systems.