I Drive - You Trust: Explaining Driving Behavior Of Autonomous Cars

Gesa Wiegand
fortiss GmbH
LMU Munich
Munich, Germany
wiegand@fortiss.org

Thomas Weber
fortiss GmbH
LMU Munich
Munich, Germany
weber@fortiss.org

Heinrich Hussmann
LMU Munich
Munich, Germany
hussmann@ifi.lmu.de

Matthias Schmidmaier
fortiss GmbH
LMU Munich
Munich, Germany
schmidmaier@fortiss.org

Yuanting Liu
fortiss GmbH
Munich, Germany
liu@fortiss.org

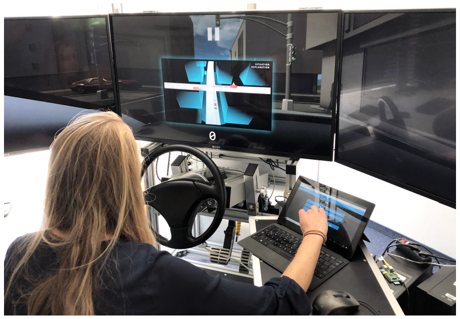

Figure 1: Participant selects elements during the study.

ABSTRACT

Driving in autonomous cars requires trust, especially in case of unexpected driving behavior of the vehicle. This work evaluates mental models that experts and non-expert users have of autonomous

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

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312817

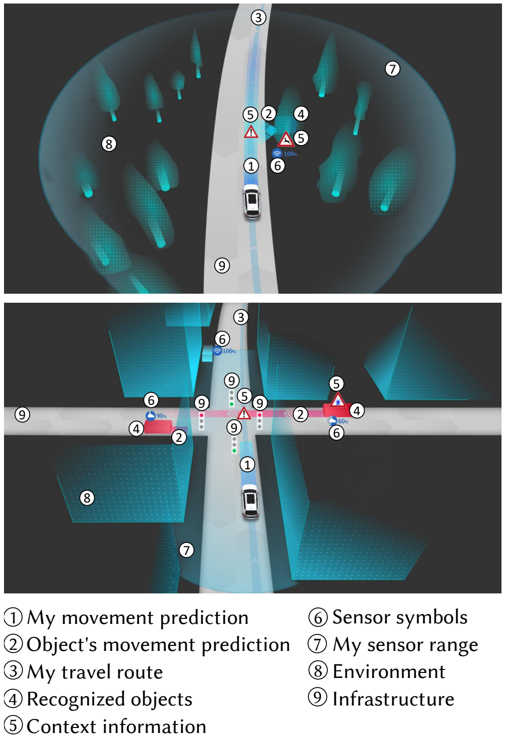

Figure 2: Configurable explanation screen consisting of nine visual components.

driving to provide an explanation of the vehicle's past driving behavior.
We identified a target mental model that enhances the user's mental model by adding key components from the mental model experts have. To construct this target mental model and to evaluate a prototype of an explanation visualization we conducted interviews (N=8) and a user study (N=16). The explanation consists of abstract visualizations of different elements, representing the autonomous system's components. We explore the relevance of the explanation's individual elements and their influence on the user's situation awareness. The results show that displaying the detected objects and their predicted motion was most important to understand a situation. After seeing the explanation, the user's level of situation awareness increased significantly.

KEYWORDS

Autonomous Driving, Explainability, Mental Model, Situation Awareness

INTRODUCTION

To design an explanation component for the driving decisions of an autonomous vehicle, we first need an understanding of the expert and non-expert mental model of autonomous systems. Subsequently, a target mental model can be identified by adding evaluated key components of the user mental model to key components of the expert mental model [3]. This work explores the practical implementation of an explanation interface to answer what needs to be visualized in the car in order for the user to understand autonomous driving behavior (SAE Level 4 or 5 [7]). Situation awareness in current research assesses the awareness of the driver of the surrounding environment while driving autonomously [5]. It calibrates trust in the automation algorithm and can enhance the take over behavior [9]. In our work we measure the improvement of situation assessment with a visual situation explanation. The explanation is intended to improve trust of the driver in driving decisions of the autonomous vehicle that have no obvious cause. In the considered situation, the driver might require an explanation after a non-transparent driving decision of the car. Explanations can increase trust [10] but transparency of systems does not automatically achieve that goal [2]. Therefore the explanation has to be designed in direct agreement to its expected purpose.

RELATED WORK

In order to clarify algorithm decisions, various visualization approaches have been researched. Bojarski et al. analyze a convolutional neural network (CNN) which is trained for autonomous driving by unveiling the learned image features [1]. A work by Koo et al. [8] focuses on the end user of the system. Explaining to the driver how a car behaved led to poor driving performance. In contrast, providing a reason for the driving behavior was preferred and led to better driving performance. Providing both information led to safest driving behavior but increased negative feelings. The explanation was

Situation A: The autonomous vehicle drives through a forest, behind a tree on the right hand side is a deer.
Autonomous vehicle behavior: Slowing down.
Behavior explanation: The deer might run on the street and get hit by the vehicle.
Situation B: The autonomous vehicle drives in a town and stops at a green traffic light at an intersection.
Autonomous vehicle behavior: Stopping at green traffic light.
Behavior explanation: An emergency vehicle is detected by cloud information and might cross the intersection even though the traffic light is red.

Sidebar 1: Situations presented to the participants.

provided verbally. In our work we include the mental models of experts and non-experts to explore an explanation of the driving behavior, using a visual explanation. To measure the effectiveness of an explanation, Gunning [6] suggests to explore mental models and user satisfaction. We set up the target mental model of autonomous systems, based on the methodology of Eiband et al. [3]. Using a methodological analysis of expert and non-expert mental models, a target mental model can be derived.

METHODOLOGY

To have a basic understanding of explanation visualization preferences of users, a questionnaire (N=20) was conducted. We compared statistical, text and speech based explanations, environment model, graphical sensor data and real world image explanations. The preferred visualization was an environmental model, which we chose as visualization output. According to Eiband et al. [3] a deep understanding of what to explain is an advantageous prerequisite to set up a transparent system. In order to achieve that, an expert mental model of the system is first elicited. In the next step a user mental model is evaluated. Those results enable us to set up the target mental model of the system, which answers the question what information should be visualized to explain autonomous driving behavior.

Expert Mental Model. The interviewed experts defined three categories which describe an autonomous driving system: perception, deliberation and action. *Perception* includes all components concerning sensors, object detection, localization, tracking, maps, data fusion and fusion. *Deliberation* includes route planning, environment model, environment prediction and trajectory planning. Conducting the calculated driving *actions* requires control of the car's actuation and a Human-Computer-Interface (HCI). Concerning the comprehension of autonomous systems, the experts stated: âĂIJTo understand the driving decision of the autonomous vehicle an environment model would be sufficient for most end users.âĂIJ and âĂIJTrajectory planning is too complex for normal users.âĂİ.

They additionally stated that the vehicle should apologize to the user in case it made a mistake and it should be triggered and adapted according to type and significance of a situation. However, the experts did not believe that normal users would want to get deep insight into the system's algorithm, as it would be hard to understand.

Target Mental Model. In order to acquire the target mental model, key components of the user mental model need to be enhanced by key components of the expert mental model. Within this work we identify the user mental model in the scope of a study. By defining the differences between the expert mental model and non-expert mental model and by priority allocation, the key components for the target mental model are extracted.

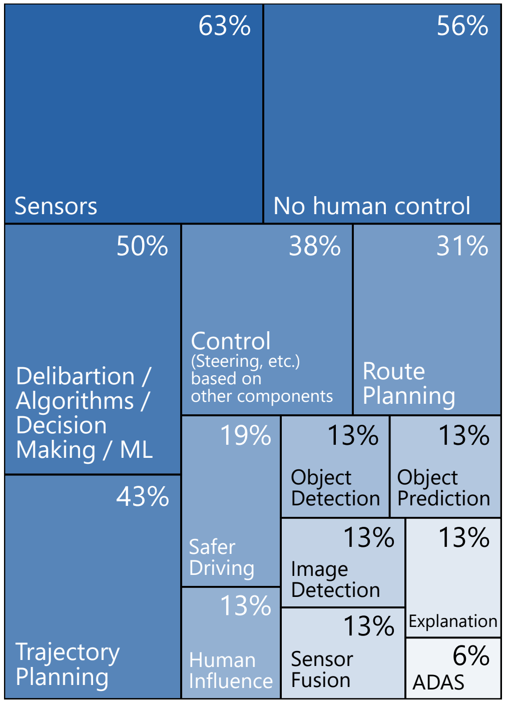

Figure 3: Elements mentioned during User Mental Model evaluation. The numbers indicate the percentage of participants that mentioned each element.

STUDY

The study was conducted in a driving simulator with 16 participants (12 male, 4 female) with the average being between 20-30 years old. The driving simulation was controlled by the supervisor while participants were manipulating the explanation screen via a separate interface on a tablet and answered a questionnaire on an additional laptop (see setup in Figure 1). Before driving in the simulator, the participants were informed that they are driving in an autonomous car, therefore no intervention of them is needed during the study. As secondary task the participants watched a video and counted people therein in order to simulate a realistic autonomous driving experience. As main task, participants were confronted with two different driving situations (see description in Sidebar 1) . After each situation they were asked what happened in the environment and what the vehicle did. Subsequent to the vehicle's reaction, a visual explanation of what happened in the environment was presented to the participant. The explanation consisted of abstract visualizations of different layers (1-9), representing the autonomous system's components (see Figure 2). The element "My movement prediction" (1) was an area in front of the vehicle which indicates, where the vehicle might move next. "Object's movement prediction" (2) was an area in front of the object visualizing the estimated next movement of the object. "My travel route" (3) was a line on the road visualizing the planned route. The detected objects were visualized by the element "Recognized objects" (4). "Context information" (5) included the background information of the situation, abstracting information e.g. from the cloud. Abstract "Sensor symbols" (6) showed from which sensor (e.g. ego-vehicle sensors or car2car information) the information was retrieved. The "Sensor range" (7) information of the vehicle was visualized by a transparent region around the vehicle. "Environment information" (8) included houses or trees in the scenario. "Infrastructure" (9) displayed the road and the traffic lights. The participants chose their preferred explanation by removing the elements that were not necessary for their situation understanding. Afterwards the participants were asked again what happened in the situation. The study was designed as within-subjects study so every participant drove through every situation. The order of the situations was changed so that 8 participants started with Situation A and the other half started with Situation B.

RESULTS

In order to evaluate the results of the user study, we evaluated the situation awareness level [4] of the participants. Each situation is assessed and the situation awareness elements are divided into every level. The participants reached a certain level according to their perception (L1), comprehension (L2) and prediction (L3) of the situation. The statements of the participants were judged according to following levels:

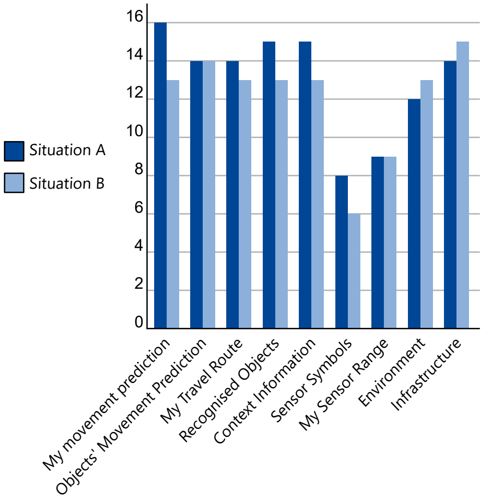

Figure 4: Frequency at which participants chose the explanation elements.

Level 1: Perception: Elements
Situation A: Deer in the forest. Car is slowing down
Situation B: Green traffic light. Car stopped.
Level 2: Comprehension: Dependencies and significance of objects recognized
Situation A: Car is slowing down because there is a deer/object in the forest.
Situation B: Car stops even though traffic light is green.
Level 3: Prediction: Future actions are predicted
Situation A: Car could collide with deer.
Situation B: Car stops because a rescue vehicle could turn around the corner and collide with the ego vehicle. The difference between the level of situation awareness before and after the explanation was evaluated with the Wilcoxon signed rank test. The results (p = 0.0001) showed significant increase in situation awareness. The chosen elements are shown in Figure 4. The least chosen element is sensor symbols which indicates that the interpretation of sensor symbols might not be intuitive.

User Mental Model. The mental model of users composes of elements that are similar to the components of the mental model of experts (see Figure 3). The statement *"An autonomous car is a vehicle, equipped with different types of technologies that work together in a way that allows the vehicle to move and navigate in its environment freely and independently and according to given rules - without human interference."* includes the description of sensors, algorithms, fusion, navigation, trajectory planning and actuation in the words of a non-expert. But the description of an autonomous system in the words of an non-expert also differs significantly from expert knowledge. The ability of an autonomous car is overrated by non-experts ("It is able to foresee dangers") or the user has other expectations of the system ("Explains what it is doing").

Target Mental Model. After the evaluation of the non-expert and expert user mental model a target mental model is identified (see Figure 5). The common notion of the functionality of an autonomous system is displayed in the overlapping area. The key components that are rated by the users as not as important and the wrongly identified elements of the users are excluded from the target mental model. Therefore, even though "Data and Sensor Fusion", "Map" and "Environmental Model" was mentioned by non-experts and experts it is excluded from the target mental model. The element "Sensors" is rated low as chosen element (see Figure 4). Nevertheless, as it is a high priority component during the User Mental Model evaluation (see Figure 3) we added it in the target mental model.

DISCUSSION AND CONCLUSION

Within our work we measure the quality of explainability using the driver's levels of situation awareness. Situation awareness is important if the driver needs to take over control during a driving scenario. Within the scope of our work, the driver is not required to take over the steering wheel and therefore

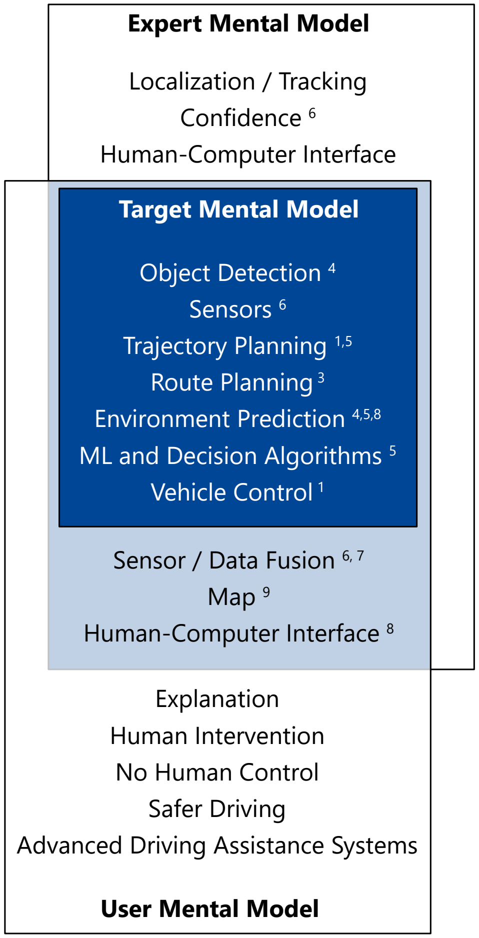

Figure 5: Target mental model consisting of components found in both mental models (shaded) with a selection of the key components (blue). Superscript numbers (1-9) indicate which explanation components are used for visualization (Figure 2).

has no time limit for explanation interpretation. Differences between the assessment of situation awareness and explanation screens should be researched in more depth. As we focused on *what* to visualize, a next step is to evaluate in a structured manner *how* to visualize the explanation. In this context, adaptive explanations can be addressed to provide individual optimization for various user types. Considering the expert level of the user can result in different visualization methods. The technical interested user might need an explanation of the deep learning algorithm to fully understand the vehicle's behavior. The optimal time to show the explanation can lead to the question if the driver wants to influence the autonomous driving behavior. Possible advantages could be the constant control of the driving behavior and adaptions through the driver if the current autonomous driving behavior differs from desired actions. In this work we evaluated what information the driver needs to have explained if an unexpected driving behavior of the autonomous vehicle occurs. That evaluation resulted in a target mental model that includes the combined key elements of expert mental model and user mental model. The explanation significantly improved the situation awareness of the users.

REFERENCES

[1] Mariusz Bojarski, Philip Yeres, Anna Choromanska, Krzysztof Choromanski, Bernhard Firner, Lawrence Jackel, and Urs Muller. 2017. Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car. (April 2017). arXiv:cs.CV/1704.07911

[2] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455.

[3] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM, New York, NY, USA, 211–223. https://doi.org/10.1145/3172944.3172961

[4] Mica R Endsley. 2017. Toward a theory of situation awareness in dynamic systems. In *Situational Awareness*. Routledge, 9–42.

[5] Mica R Endsley. 2019. Situation Awareness in Future Autonomous Vehicles: Beware of the Unexpected. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*. Springer International Publishing, 303–309.

[6] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* (2017).

[7] SAE International. 2016. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles.

[8] 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.

[9] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and Why Not Explanations Improve the Intelligibility of Context-aware Intelligent Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2119–2128.

[10] Nava Tintarev and Judith Masthoff. 2011. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*. Springer, 479–510.