# Explanations and Expectations: Trust Building in Automated Vehicles

Jacob Haspiel
University of Michigan
haspiel@umich.edu

Na Du
University of Michigan
nadu@umich.edu

Jill Meyerson
University of Michigan
jillmeye@umich.edu

Lionel P. Robert Jr.
University of Michigan
lprobert@umich.edu

Dawn Tilbury
University of Michigan
tilbury@umich.edu

X. Jessie Yang
University of Michigan
xijyang@umich.edu

Anuj K. Pradhan
University of Michigan
anujkp@umich.edu

## ABSTRACT

Trust is a vital determinant of acceptance of automated vehicles (AVs) and expectations and explanations are often at the heart of any trusting relationship. Once expectations have been violated, explanations are needed to mitigate the damage. This study introduces the importance of timing of explanations in promoting trust in AVs. We present the preliminary results of a within-subjects experimental study involving eight participants exposed to four AV driving conditions (i.e. 32 data points). Preliminary results show a pattern that suggests that explanations provided before the AV takes actions promote more trust than explanations provided afterward.

## CCS CONCEPTS

• **Human-centered computing → Human-computer interaction (HCI)**; **HCI theory, concepts and models**;

## KEYWORDS

Trust in AVs; Human-Machine Interface; Transparency

## 1 INTRODUCTION

Trust in automated vehicles (AVs), or one's openness to being subjected to the actions of an automated vehicle, has emerged as a key determinant of drivers' acceptance of AVs and other robotic technologies [1, 7, 10]. Despite major investments in AVs, drivers continue to express concerns with handing over the driving task to the automation [7]. Simply put, many drivers do not trust AVs enough to feel comfortable handing over the driving task [7].

Expectations and explanations are often at the heart of trust. Humans have learned to trust agents that behave in an expected or predictable manner [9]. Unfortunately, driving conditions can often necessitate the need for an AV to engage in unanticipated or unpredictable actions [7]. However, research has shown that an AV can potentially mitigate this negative effect by providing an explanation about why it has taken specific actions [3, 4].

This study introduces the importance of timing with regard to the effectiveness of explanations in promoting trust in the AV. We assert that providing the explanation before rather than after the action is likely to lead to greater trust in the AV. We also assert that providing drivers with an option to decide whether the AV will take the action should lead to more trust beyond just providing an explanation. To test these assertions, we are conducting a study. Next, we present details of the study and our preliminary results.

## 2 METHOD

### 2.1 Participants

We have examined eight participants (3 females), with an average age of 25.7 years (standard deviation [SD] = 2.36 years). Participants were screened for various inclusion criteria including driver's license status and susceptibility to simulator sickness. Participants were paid $20 for participating in the 60- to 75-minute study.

### 2.2 Experimental Apparatus

The study was conducted in a high-fidelity vehicle simulator at the University of Michigan Transportation Research Institute. The simulator consists of a full Nissan Versa sedan located in a dedicated lab space. Road scenes were projected on three screens in front of the vehicle and one screen in the back of the vehicle. An integrated four-camera eye-tracking system provides head-pose, eye-blink, and gaze data. Heart rate, heart rate variability and galvanic skin response were collected using a Shimmer wearable device. The driving simulator was programmed as an SAE Level 3 [8], where the driving was conducted by the AV, and the driver was responsible for monitoring the environment.

### 2.3 Experimental Design

The study employed a within-subjects design with four driving conditions (DC): DC1 - the AV provides no explanation about its actions; DC2 - explanations are presented 7 seconds prior to the AV actions; DC3 - explanations are presented within 1 second after actions have been taken by the AV; and DC4 - explanations are


*HRI '18 Companion, March 5- 8, 2018, Chicago, IL USA*

ACM ISBN 978-1-4503-5615-2/18/03.
https://doi.org/10.1145/3173386.3177057

presented 7 seconds before the AV takes action and the driver is asked to approve or disapprove the AV's proposed action. If the driver disapproves, the AV does not engage in the action. The four conditions are counterbalanced using a Latin square design.

In each driving condition, participants engaged in a 6- to 8-minute drive. Each drive contained three unexpected events, one from each of three categories, which forced the AV to take unexpected actions. The three categories were: events by other drivers, events by police vehicles, and unexpected re-routes. Events occurred at prescribed times, were unique to each condition, and were balanced by type across conditions.

The dependent variables presented in this study are the participants' subjective attitudes. The attitudinal measures include trust [5], cognitive workload [2], anxiety [4, 6], and driver preference [4, 6]. Participants also ranked each driving condition on trust from 1 (most trust) to 4 (least trust). We also collected objective physiological measures, which are not presented in this paper.

## 2.4 Procedure

Upon arrival, participants signed an informed consent and completed a demographic survey. After completing a training session, participants experienced each of the four driving conditions. Participants engaged the automated driving mode at the beginning of each driving condition. After each driving condition, participants completed a questionnaire measuring their trust, workload, and preferences. At the end of the experiment, participants ranked the trustworthiness of the AV in the four driving conditions from highest (1) to lowest (4).

## 2.5 Analysis

We tested both assertions using data from the eight participants. We used repeated measures analysis of variance (ANOVA) to analyze the relationship between the independent variable (driving conditions) and dependent variables (subjective attitude in automated vehicle and rank order).

# 3 RESULTS AND DISCUSSION

We did not expect to find significant results with only eight participants and we did not. However, as seen in Table 1, a pattern can be seen. Our preliminary results show that driving condition 2 (DC2), where the explanation is given before the action, yielded the highest measure of trust both by attitude and by rank order (note: 1 is highest and 4 is lowest in rank order of trust). DC2 also produced the highest preference rating and lowest cognitive load. Driving condition 4 (DC4), before explanation and with control over the AV's actions, had the lowest anxiety and the highest cognitive load.

As we stated, none of the results were significant. Therefore, we refrain from drawing conclusions. However, results do suggest that providing the driver with an explanation before the action produces the highest trust scores. This seems especially true with regard to the rank order. Our results suggest that providing the driver with an explanation and control over the AV's actions increases the workload but does not seem to provide any more trust in the AV.

**Table 1: Preliminary Results**

| | Trust | R.O. Trust | Anxiety | Prefer. | Cog.Load |
|---|---|---|---|---|---|
| DC1 | 5.6 | 3.3 | 3.1 | 5.2 | 20.3 |
| DC2 | 5.9 | 1.7 | 2.9 | 5.5 | 16.6 |
| DC3 | 5.6 | 2.7 | 2.8 | 5.4 | 19.3 |
| DC4 | 5.6 | 2.3 | 2.6 | 5.2 | 25.6 |

DC = Driving Conditions; R.O. = Rank Order
Prefer. = Driver Preference; Cog. Load = Cognitive Load

# 4 CONCLUSION

In summary, we present preliminary results of a study examining the impacts of timing on the effectiveness of explanations to promote trust in AVs. Preliminary results show a pattern suggesting that explanations provided before actions promote more trust than explanations provided afterward. However, providing the driver with additional control might not provide additional benefits. We hope the final results can be used to inform the design of AVs and other human-machine interfaces.

# 5 ACKNOWLEDGMENT

This work was supported by University of Michigan Mcity.

# REFERENCES

[1] Chandler Creech, Suresh Kumaar Jayaraman, Lionel Robert, Dawn Tilbury, Xi Jessie Yang, Anuj Pradhan, and Kate Tsui. 2017. Trust and Control in Autonomous Vehicle Interactions. *Presented at the Morality and Social Trust in Autonomous Robots workshop at the 2017 Robotics: Science and Systems (RSS 2017)* (2017). http://hdl.handle.net/2027.42/137661
[2] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.
[3] 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 (01 Nov 2015), 269–275.
[4] Jeamin Koo, Dongjun Shin, Martin Steinert, and Larry Leifer. 2016. Understanding driver responses to voice alerts of autonomous car operations. *International Journal of Vehicle Design* 70, 4 (2016), 377–392.
[5] B.M. Muir. 1987. Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies* 27(5-6) (1987), 527–539.
[6] Clifford Nass, Ing-Marie Jonsson, Helen Harris, Ben Reaves, Jack Endo, Scott Brave, and Leila Takayama. 2005. Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1973–1976.
[7] Luke Petersen, Dawn Tilbury, Xi Jessie Yang, and Lionel Robert. 2017. Effects of Augmented Situational Awareness on Driver Trust in Semi-Autonomous Vehicle Operation. *Autonomous Ground Systems (AGS) Technical Session of the 2017 Ground Vehicle Systems Engineering and Technology Symposium* (2017).
[8] On road Automated Vehicle Standards Committee et al. 2016. SAE J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. *SAE International* (2016).
[9] Lionel P Robert, Alan R Denis, and Yu-Ting Caisy Hung. 2009. Individual swift trust and knowledge-based trust in face-to-face and virtual team members. *Journal of Management Information Systems* 26, 2 (2009), 241–279.
[10] Sangseok You and Lionel Robert. 2018. Human-Robot Similarity and Willingness to Work with a Robotic Co-Worker. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM.