# Will Explained Actions Of An Autonomous Vehicle Increase The Trust Level Of Its Human Driver?

Einav Shenwald and Ella Shalom

**Abstract**

*The key for smooth integration of autonomous cars in transportation systems, is human drivers' trust in the abilities of the autonomous vehicles. This project examines whether an autonomous vehicle which explains its actions improves the trust level of its human driver. To test this hypothesis, we conducted a between-subjects study where subjects answered questions about their trust level, after experiencing different situations which may occur while driving. The results show that subjects who experienced an autonomous vehicle that explains its actions expressed a higher level of trust in the autonomous vehicle.*

## 1 Introduction

Automated vehicles (AVs) have the potential to reshape transportation as it exists today. It could benefit the environment by providing more fuel-efficient driving, and our society by reducing driving-related accidents. Moreover, it may have a great effect on our everyday lives by providing us with the time that was previously spent driving. Despite this, there are serious concerns about whether individuals will choose to use AVs. One of the key factors to the success or failure of AV acceptance by human users is trust. Trust is defined as "the reliance by one agent that actions prejudicial to the well-being of that agent will not be undertaken by influential others" [6]. Many studies have researched ways to increase human trust in automation. Specifically, some studies have focused on how the explain-ability of automation increase the trust level of users[3]. Explanations ease people's worry about the unknown, and as a result, they are more inclined to trust the automated systems [7]. However, AV is a unique automated system. By using it, human drivers are required to give up control when there are major consequences, ultimately their lives. Therefore, understanding what will affect humans' trust in AVs is crucial [1]. We designed our study to evaluate whether the explain-ability of the AV's actions before it is taken, helps increase the human drivers' trust in the AV.

# 2 Research Hypothesis

Our research Hypothesis is: will explained actions of an autonomous vehicle increase the trust level of its human driver. The construct we are examining is the trust level of human drivers in the AV's actions. To evaluate the construct we used five short videos depicting various driving situations, observed from the driver's point of view. The controlled variable is the AV's explanations of its actions. Some subjects will watch videos without the AV's explanations of its actions. While others will watch videos where the AV provides a concise explanation that consists of two types of descriptions. "Why" - description of the reasoning for its actions (e.g., "Pedestrian detected") and "how" - description of how the AV is reacting to it (e.g., "The car is slowing down"). The AV reaction could be a certain action or the lack of it if it's not required. For example, when a pedestrian suddenly appears, it may seem to the subject that it is crucial to react to the situation. Therefore, even if an action is not taken, the explanation will be sound in order to show to the human driver that the AV is aware of the situation. The AV way of reacting in this case is continuing as it is. We chose this type of verbal explanation because explaining both "how and why" is important for gaining a high level of trust.[2]

# 3 Experimental Design

## 3.1 Location and setup

The number of people involved in running one trial is one; one subject will answer the questionnaire alone. The subjects will be asked to answer the questionnaire in a quiet place.

## 3.2 Participants

The participants were people with a driving license who has driving experience of over a year, in age 20 to 60. We decided to apply those limitations to neutralize factors related to age or lack of driving experience that could affect the trust level of the subjects. We recruited people from the university, our work, family, and friends.

## 3.3 videos

As mentioned above, the different conditions are whether the subjects are watching a video with or without explanation. While driving, people experience situations that differ on the road type, weather, vehicles and pedestrian they encounter. Therefore, we wanted to use videos that contain diverse situations.

We handpicked YouTube videos that seemed realistic and less commercial. We edited those videos to make them even more realistic: we muted commercial and relaxing audios and cut off commercial parts. After that, we added an

| Description | Explanation | Video w/o explanation | Video with explanation |
|---|---|---|---|
| The AV drives in Interurban road, bypassing a truck | The right lane is clear, moving to the right lane | link | link |
| The AV drives in urban road, crosses a T junction | The junction is clear, turning right. Bicycle rider detected | link | link |
| The AV drives in urban road, across a bicycle and there is a light rain | Bicycle rider detected, slowing down the driving pace | link | link |
| The AV crosses a crossroad | Pedestrian detected. The crossroad is clear, continue driving | link | link |
| The AV crosses a border when a green light appears | Green light detected, continue driving | link | link |

Table 1: Details on the 5 videos used in our research, including the link to the video and the explanation

explanation to them using an artificial voice. Each video has an accompanying explanation, both visible in Table 1:

## 3.4 Questions

We chose to evaluate our construct by asking questions, in which the subject was given a sentence that signifies his trust in the AV's actions. The subjects were asked to rate their level of agreement with the sentences (in the context of the videos of the AV). The agreement level rate options differ from 1 to 5, when 1, and 5 represent the lowest and highest agreement accordingly. We chose this type of method to assess the subjects' trust because we wanted the answers to be comparable, not dependent on the specific way a subject express himself. On the other hand, we wanted to allow the subjects to express their feelings gradually (in contrast to agree or disagree) and to allow them to be neutral, by choosing 3. In order to build our questionnaire, we relied on a validated scale of trust perception[4].

We presented 5 sentences after each video:

- The AV is reliable.

- The AV can make the right choice based on the situation of the road

- I felt trust in the AV's actions

- I felt that the AV acted in a responsible way

- I trusted the decisions the AV was making on the road

Those sentences, like the rest of the questionnaire, were written in Hebrew, as it is the native language of all the subjects.

We added a control question about the content of each video, that was meant to check the subjects' attention to them. If the subject answered it incorrectly, we removed his responses from our results.

## 3.5  Experiment procedure

Subject enters the google forms questionnaire and reads a short explanation about it. The subject watch five short videos, each followed by the questions described above.

## 3.6  Possible risk

There is no chance of physical harm following the experiment as it is done through videos. However, certain people can feel unease or even anxiety after watching the videos.

# 4  Results

We recruited 30 participants. However, we had to disqualify 3 participants who watched videos with AV's explanations, and 5 participants who watched videos without AV's explanations. Overall 7 participants were disqualified, we remained with 10 subjects in one group and 12 in the other one. Each subject who got disqualified answered only one control question wrong, out of the five questions which were presented to him overall. The control questions were about the main content of the video. Therefore, we couldn't rely on the trust level of a participant who answered it wrong.

The age and gender distributions of the remained participants can be seen in Figure 1 and Figure 2 accordingly.

During the design of our experiment, we faced the challenge of finding the best way to test autonomous driving. We wanted the subjects' experience to feel as real as possible but needed to settle for videos, which are practical and yet still effective. After deciding to use videos, we had to find videos that depict various situations while being natural and noncommercial. It was challenging to find these kinds of videos. And the ones we did manage to find needed editing. Also, adding the explanations at the exact moment that would seem as if the
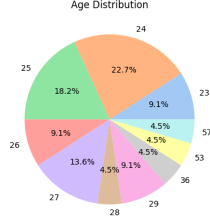
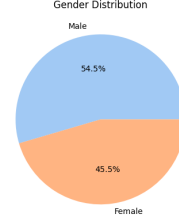Figure 1: Age distribution of
the research participants



Figure 2: Gender distribution
of the research participants

AV explain its actions before taking them wasn't trivial. While we executed our experiment it was difficult to find subjects that were willing to answer our questionnaire. People often asked if we offer them something in return for their participation. The previous challenge was emphasized by the fact that some subjects were disqualified. So we had to recruit additional participants.

As mentioned above we measured the trust level of human drivers in the AV's actions, the possible value of trust level per video is between 5-25, which is a total of 25-125 for all the videos.

We found out that the trend of trust levels for all videos was similar thorough both groups (Figure 3a). Even though, each video was trusted more when heard with explanations than without them.
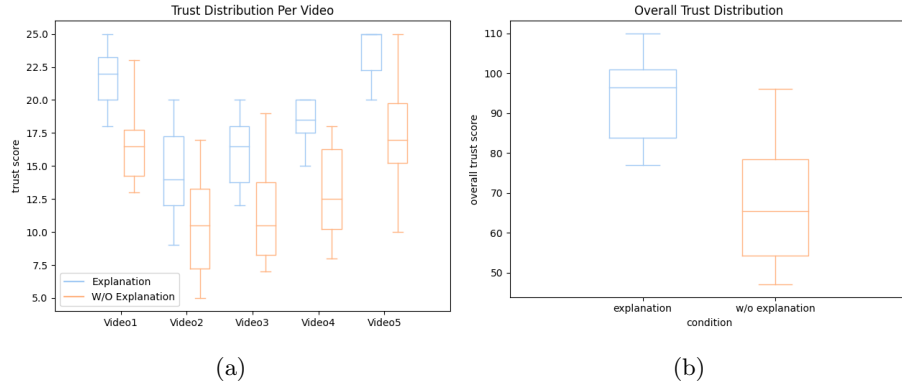


Figure 3: Distribution of the trust level of participants in each group
(a) Distribution of the trust scores per video
(b) Distribution of the accumulated trust scores over all the videos

As shown in Figure 3b, the subjects who watched the videos with the AV's explanations gave higher accumulative trust for all the videos than the participants in the control group who watched the videos without explanations. To test the significance of our results, we used one-tailed T-test [5]) with alpha=0.01.

The 12 participants in the first condition, with explanations (mean=94.17, stdv=10.12) compared to the 10 participants in the control group (mean=68.9, stdv=16.8) ranked the AV with significantly higher trust scores, $t(20) = 4.15$, $p < .001$. Despite the diverse responses from the control group, our hypothesis, will explained actions of an autonomous vehicle increase the trust level of its human driver, was significantly correct.

## 5 Conclusion and Future Work

To conclude, we tested whether the explain-ability of an AV's actions will help to increase the trust level of its human driver. The results show that people tend to trust the AV more when it explains its actions before acting.

The fact that our results were significant indicates that explain-ability is an important factor in gaining human trust. It could mean that it may be useful to incorporate some form of explain-ability of AV's actions into the AV's design. Even though our results were significant, we think it is crucial to test our hypothesis by driving the AV and not by videos that only simulate this experience. There is more to be done to find ways to increase humans' trust in AVs, yet we are definitely encouraged by the results.

There are many directions for expanding our research. As mentioned before, our hypothesis should be tested in a more realistic environment. It could also be important to conduct extensive research, with a higher number of subjects. In our project, We used a specific type of explanation. We believe different ones should also be explored. For example, both "how" and "why" components of the explanations are not necessarily required. And other kinds of explanations may be effective as well. Naturally, some situations are perceived riskier or less trustworthy than others. Further research on this could allow us to understand which situations require an explanation and which do not. It would also be interesting to search for the source of trust or lack of it in different situations and adapt our explanations according to that. It would be beneficial to inspect other factors that may increase trust and compare them with the effect of AV's explanations. For example, time and experience in driving an autonomous vehicle may increase human trust in them, will the same explanations still be needed? An adjustment of the explanations through time and personal customization of it could also be interesting to investigate in the future.

# References

[1] J. Meyerson L.P. Robert Jr D. Tilbury X.J. Yang A.K. Pradhan J. Haspiel, N. Du. Explanations and expectations: Trust building in automated vehicles. *International Conference on Human-Robot Interaction, ACM*, pages 119–120, 2018.

[2] Kwac J. Ju W. Steinert M. Leifer L. Nass C. Koo, J. a 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)*, page 269–275, 2014.

[3] Pearl Pu and Li Chen. Trust building with explanation interfaces. *Proceedings of the 11th International Conference on Intelligent User Interfaces*, page 93–100, 2006.

[4] Kristin E. Schaefer. *Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI"*, pages 191–218. Springer US, Boston, MA, 2016.

[5] Social Science Statistics. How to report a t-test result in apa style. *https://www.socscistatistics.com/tutorials/ttest/default.aspx*, 2022.

[6] D. R. Billings V. Kocsis J. Y. Chen K. E. Oleson and P. A. Hancock. antecedents of trust in human-robot collaborations. *International MultiDisciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA). IEEE*, page 175–178, 2011.

[7] A.M. Bisantz Y. Seong. The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, pages 608–625, 2008.