XAI Online study

The study aimed to investigate the relationship between perceived trust and conformity regarding the action of an autonomous vehicle given a verbal explanation to its action.

Study design

To test whether explanations affect user's trust and perceived conformity, a within-between design was chosen. Participants were exposed to various simulated driving scenarios in which an autonomous vehicle performed an action either with or without an auditive explanation commenting the action of the autonomous vehicle. Both explanation and non-explanation conditions were tested between subjects.
Additionally, two types of driving situations existed in the simulation scenarios: Situations were either ambiguous situations, where it was not visibly obvious to the participant why the autonomous vehicle acted the way it did. To avoid inappropriate actions of the car, ambiguous situations were dissolved at the end of each video, to make sure that the autonomous vehicle could be trusted. The second type were clear situations in which it was visibly obvious to the participant why the autonomous vehicle acted the way it did. The situation conditions ambiguous and clear were tested within subjects.
Participants were randomly assigned to the conditions.

Apparatus

Simulation Videos and Auditive Explanation

For a most realistic driving simulation, driving scenarios were simulated with the software UnityEngine. The driving simulation videos were then presented within a fixed order in form of video content on a survey-website. The order of the videos did also not differ between conditions.
The scenarios showed a one-way German highway with two lanes and a naturalistic landscape and a background motor sound. In the simulation, the situation was viewed from a driver's perspective inside an autonomous vehicle. At the beginning of each scenario the autonomous vehicle was driving at a constant speed either on the left or the right lane of the highway. Situations differed between lane change and braking actions of the autonomous vehicle. No intervention was possible for participants during the view of the driving scenarios.
Auditive verbal explanations were simulated with a female Google Wavenet voice. The explanation was played verbal before the beginning of the action of the autonomous vehicle in each video of the explanation condition. Videos of the non-explanation condition did not differ to the explanation condition in the scenario, but had no explanation and only a motor sound in the background.

Questionnaires

To measure the amount of conformity that participants perceived regarding the action of the autonomous vehicle, a self-constructed one-item-scale was used. Aiming to assess the subject's perceived appropriateness of the performed action, the item addressed the surprise of participants about the action after each simulation scenario. The wording of the item depended on the kind of action showed, which was either braking or lane-change. Example: How did you perceive the lane change of the autonomous vehicle at the time it occurred?
Items were answered on a five-point-Likert-scale ranging from 1 (Not at all surprising) to 5 (Very surprising).
For measuring the construct trust, the *Human-Computer-Trust-Scale* (HCTS) of Madsen und Gregor (2000) was used. For this study, the scale was translated from English to German. Items were answered on a five-point-Likert-scale ranging from 1 (Strongly disagree) to 5 (Strongly agree).
To control for the participant's presence and involvement state during the experiment, Presence was assessed with the German version of the *Igroup Presence Questionnaire* (IPQ) (http://www.igroup.org/pq/ipq/index.php). Items were answered on a five-point-Likert-scale ranging from 1 to 7 with 1 the lowest and 7 the highest possible rating of presence during the participation in the study.
Statistical Analysis was performed with the software IBM SPSS 26. Explanation and situation conditions served as independent variables, whereas trust and surprise served as dependent variables.
Moreover, driving experience (duration of license ownership and driving frequency per week), presence and display size served as control variables.

Tasks?

Procedure

The study was conducted via the Online-Platform SoSciSurvey.de and was promoted as a study in the context of autonomous driving.
Psychology students of the Otto-Friedrich-University Bamberg could earn credit points for participating in the online-study.
After giving consent, monitor size and device information was asked. To guarantee for higher monitor sizes, participants were only allowed to participate in the study with desktop-computers, Notebooks and Laptops, but not smartphones or tablets. Then instructions for the test part about how to adjust the monitor, seat position and instructions to give attention to all aspects of the driving scenario like being in a real car were given.
After one trial-video of a basic driving situation with or without explanation given the condition, participants watched either 20 videos in the explanation or non-explanation condition. Of these 20 videos, 10 were ambiguous scenarios and 10 were clear scenarios. After each video participants rated their perceived conformity regarding the action of the autonomous vehicle on the one-item surprise scale.
After completion of the videos and surprise ratings, participants were given several questionnaires, including IPQ, HCTS, as well as control questions concerning the video, sound quality and demographic data.
No personal or sensitive data was collected, all data was collected anonymous and without the possibility to link data to a specific person.

Hypotheses

Explaining the action of an autonomous vehicle should make the action more understandable and predictable for participants. Therefore, a main effect of explanation on surprise about actions of the autonomous vehicle was expected.

H1: An explanation of the action of the autonomous vehicle leads to less surprise of participants regarding the action of the autonomous vehicle than no explanation of the action of the autonomous vehicle.

Ambiguous situations were expected to induce more surprise in participants than clear situations, due to missing visible information about reasons for actions of the autonomous vehicle. Hence, a main effect of situation on surprise was anticipated.

H2: Ambiguous situations lead to more surprise of participants regarding the action of the autonomous vehicle than clear situations.

Especially in ambiguous situations, where users do not obviously see why the autonomous vehicle is acting, lower surprise was expected to be reported with a given explanation than without explanation. Therefore, an interaction effect between situation and explanation on surprise was expected.

H3: An explanation of the action of the autonomous vehicle in ambiguous situations leads to less surprise of participants regarding the action of the autonomous vehicle than no explanation of the action of the autonomous vehicle.

Surprise was expected to be related to the amount of conformity that participants perceived between their own subjective estimation of appropriateness and the action estimated appropriate and performed by the autonomous vehicle. An association between surprise and trust in the autonomous vehicle was expected.

H4: *Surprise regarding the action of the autonomous vehicle predicts trust in the autonomous vehicle.*

Explanations were expected to reduce insecurities in appropriateness about the actions of the autonomous vehicle, providing valid information about why actions were performed in specific situations. Therefore, a main effect of explanation on trust was expected to be revealed.

H5: An explanation of the action of the autonomous vehicle leads to more trust scores of participants in the autonomous vehicle than no explanation of the action of the autonomous vehicle.

Participants

Subjects had a mean age of 25.09 (SD=8.48) and 44.6% of all subjects were female and 55.4% male. 82.1% were students, 10.7% employees, 5.4% trainee and 1.8% self-employed.
All 56 subjects reported to own a car or truck driver's license.

Of 56 subjects, 7.1% reported to own their license for a time period over 20 years, 3.6% for a time period of 11-20 years, 35.7% for a time period of 6-10 years, 50% for a time period of 2-5 years and 3.6% for a time period less than 2 years. 32.1% reported to drive less than 1 day a week on average, 17.9% reported to drive 1 day a week on average, 16.1% reported to drive 2 days a week on average, 12.5% reported to drive 3 days a week on average, 5 8.9% reported to drive 4 days a week on average, 5.4% reported to drive 5 and 6 days a week on average and 1.8% reported to drive 7 days a week on average.
The mean reported monitor size was 18.868 inches (SD=5.357 min=12 max=32).
Of all 56 subjects, there were 29 subjects (51.8%) in the explanation condition and 27 subjects (48.2%) in the non-explanation condition.

Results

94 participants participated in the study, of which 62 participants completed the study in full length.
6 subjects needed to be removed due to reported insufficient sound quality and extremely fast completion.
56 subjects remained for further analyses.

Correlations

Explorative Analyses on descriptive variables as display size, duration of driver's license ownership, driving frequency and age were performed with correlations on themselves and dependent variables. Significant correlations between duration of ownership and total sum score of surprise (r=-.277, p= .039), age and sum score surprise in ambiguous situations (r=-.307, p= .021) and age and total sum score of surprise (r(56) =-.305, p= .022) were found. Duration of ownership also significantly correlated with age of participants (r =-.859, p< .01). No significant correlations with IPQ and HCTS were found.
Interestingly, overall SCAS scores correlated significantly with driving frequency (r =.418, p= .001), as well as with overall HCTS scores (r =.357, p=.007).

Normal distribution was tested for all dependent variables using Saphiro Wilk tests. Surprise and trust scores were all normally distributed with p > .05 in all four conditions, as well as outside conditions.

For testing H1, H2 and H3 a repeated measures ANCOVA (mixed ANCOVA) was performed to investigate main and interaction effects of situation and explanation on surprise. For the 2x2 design the explanation and non-explanation conditions served as between-subject factors and ambiguous and clear situation conditions as within-subject factors. Due to the correlations of age with several variables, age served as covariate .
There was homogeneity of covariances, as assessed by Box's test (p = .596). There was homogeneity of the error variances of both surprise in ambiguous and appropriate situations, as assessed by Levene's test (p > .05).
Surprise scores were normally distributed for both groups, as assessed by the Shapiro-Wilk test (p > .05).
The mixed ANCOVA yielded no significant main effect for explanation $F(1, 56) = .228$, p =.635, partial $\eta^2 = .004$, meaning that explanation and non-explanation condition did not differ significantly. However, a significant main effect for situation was found $F(1, 56) = 22.008$, p <.05, partial $\eta^2 = .297$, meaning that situation groups did differ significantly. Moreover, no significant interaction effect between situation and explanation $F(1, 56) = .487$, p = .488, partial $\eta^2 = .00$ was shown.
A significant main effect was also found for the covariate age $F(1, 56) = 5.161$, p <.05, partial $\eta^2 = .090$.

For testing H4, that lower surprise ratings lead to higher trust ratings, a univariate linear regression model was used with sum HCTS scores as dependent and total sum surprise scores as independent variables. The linear regression showed no significant relationship between surprise and trust, $F(1, 54) = 1.101$, p =.299, with an R^2 = .020.
Surprise was a nonsignificant predictor for trust measured with overall HCTS scores. Estimated decrease of surprise was $\beta = -.181$; t (28) =-1.049; p =.299.

For testing H5, that explanation leads to higher trust, a one-way ANOVA was performed.
There was homogeneity of the error variances, as assessed by Levene's test (p > .05).
The mixed ANOVA yielded no significant effect for explanation on trust $F(1, 56) = .713$, p = .402, partial $\eta^2 = .013$, meaning that explanation and non-explanation condition did not differ in trust scores.

Discussion