University of Twente
MSc. Psychology
Master Thesis – Human Factors and Engineering
Department of Cognitive Psychology and Ergonomics
Prof. Dr. Willem Verwey
Dr. Simone Borsci
External Supervisor: Dr. Francesco Walker (Leiden University)

Adapting automated vehicle behavior to user trust: Algorithm development and driving simulator study

Melina Kowalski
S1940732

Abstract

This thesis presents the results of two experiments investigating trust in adaptive automation. The first experiment aimed to develop an algorithm for processing continuous trust data collected with a slider for the use of adaptive automation. The experiment involved a fixed-speed driving simulation with seven algorithms using data frequency binning by mode and one FiFo algorithm, i.e., first value in, first value out. The results showed no significant difference between the algorithms. The most changes were detected by the FiFo algorithm, which was used for Experiment 2.

The aim of Experiment 2 was to investigate how trust increases in initially low trusting participants after experiencing driving automation in which the speed adapts to trust changes compared to a fixed-speed simulation. Experiment 1 trust data was used as the control condition and the adaptive speed data as the experimental condition. The results showed a significant difference between pre- and post-test trust scores but not between conditions. Average slider data was not significantly different between conditions.

A strong positive Spearman's rank-order correlation was found between verbal and slider trust data indicating that the slider was able to measure the concept of trust. The strong positive correlation and the high number of detected trust changes suggest that the FiFo algorithm may be useful in the design of adaptive automation, since it is able to identify significant trust changes in real time. However, we did not find that experiencing adaptive automation can increase trust in users with low initial trust levels.

Acknowledgments

First and foremost, I would like to thank my supervisors Professor Dr. Willem Verwey, Dr. Francesco Walker, and Dr. Simone Brosci, for giving guidance, expertise, and valuable feedback. I am grateful for the BMS Lab for providing me with the resources and support to conduct my research. I would like to extend a special thanks to Lucia Rabago Mayer, who went above and beyond to help me with advice on shaping the topic of my thesis after developing the algorithms for the driving simulator. I would like to express my deepest gratitude to Nobaiha Zaman Rayta for co-developing the algorithms with me. I would also like to thank Arda Temel for the building the foundation for the driving simulator used.

I am also thankful to Karel Kroeze for helping me with advice on my data analysis. Lastly, I would like to thank my fellow students, family and friends for their support and feedback on my work.

Table of Contents

Introduction **6**

Autonomous vehicles and levels of driving automation *6*

Trust as a determinant for use of automated vehicles *7*

Trust Calibration *8*

Adaptive automation *9*

Measuring trust and the feasibility of continuous measurement tools *10*

Research Questions *11*

Experiment 1 .... **12**

Methods **16**

Participants *16*

Materials *17*

Algorithm development *21*

Task *25*

Procedure *25*

Data Analysis *26*

Results **26**

Discussion **27**

Experiment 2 .... **28**

Methods **29**

Participants *29*

Materials *29*

Task *29*

Data Analysis *30*

Results **30**

Discussion 32

Post-Test Trust Levels 32

Validity of slider 33

General Discussion and Conclusion 33

Limitations and future research 33

Conclusion 34

References 36

Appendices 39

Appendix A: Pre-test questionnaire English and German 39

Appendix B: Post-test questionnaire English and German 45

Appendix C: Script used to extract speed changes per algorithm 47

Appendix D: Script used for Analysis 1 in SPSS 50

Appendix F: Script used for Analysis 2 in SPSS 55

Appendix G: Script used for Analysis 3 in SPSS 56

Introduction

Autonomous vehicles and levels of driving automation

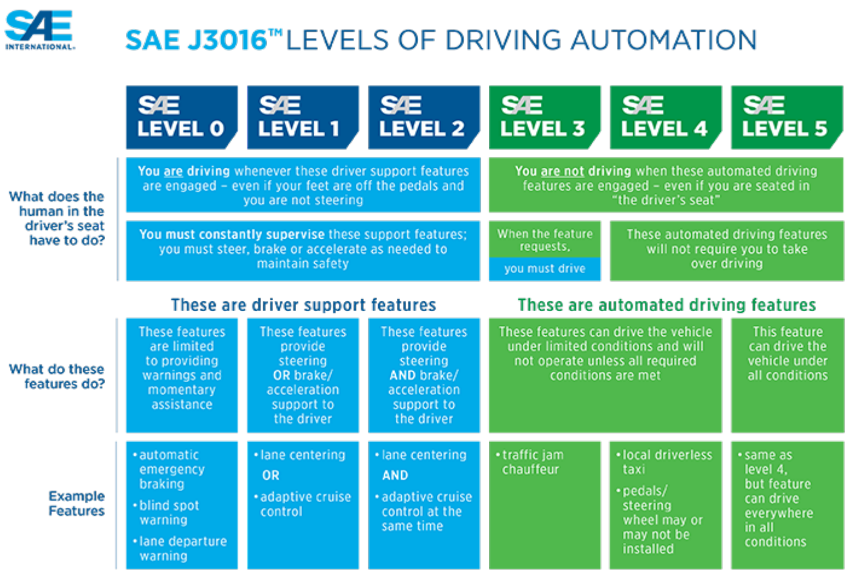
The development of automated vehicles (AVs) increases the need for responsible use and adequate trust. AVs are vehicles that use advanced technologies to assist human drivers and sometimes even handle all aspects of driving (Wang et al., 2020). Commercially available vehicles now already use assistive technologies. For instance, adaptive cruise control helps drivers maintain a safe distance from preceding vehicles (Raats et al., 2020). Relying on automatic features is becoming increasingly common.

AVs promise advantages such as safe traffic, reduced energy consumption, and improved traffic flow (Wang et al., 2020). However, AVs cannot provide complete safety. Yet, it is expected that road safety will increase, and fatal accidents decrease as more AVs are used (Duarte & Ratti, 2018). Most AV accidents are not caused by the AV itself, but by being involved in accidents caused by other drivers (Wang et al., 2020). AVs could reduce road accidents as assisted technologies and machine decision-making become more widespread, but the correct application of these technologies is crucial to preventing fatalities.

Based on the level of automation, a variable degree of human intervention is necessary. There are six levels (Level 0 to Level 5) of driving automation (see Figure 1) (SAE, 2018).

Figure 1

Levels of driving automation



Note. This figure summarizes the differences between levels of automation (SAE, 2018). Level 0 to 2 place the main responsibility for the driving task with the driver itself. Levels 3 to 5 are placing the main responsibility on the AV. Currently, vehicle automation reaches from Level 1 to Level 2. Thus, all current vehicles require constant supervision of the driving task and merely assist the driver.

The correct amount of trust is essential to make sure that this technology is not abused by either trusting the system too much or not enough. This becomes increasingly important as Level 3 and Level 4 automated vehicles are developed since the driver can perform non-driving related tasks while the vehicle controls the driving task.

Trust as a determinant for use of automated vehicles

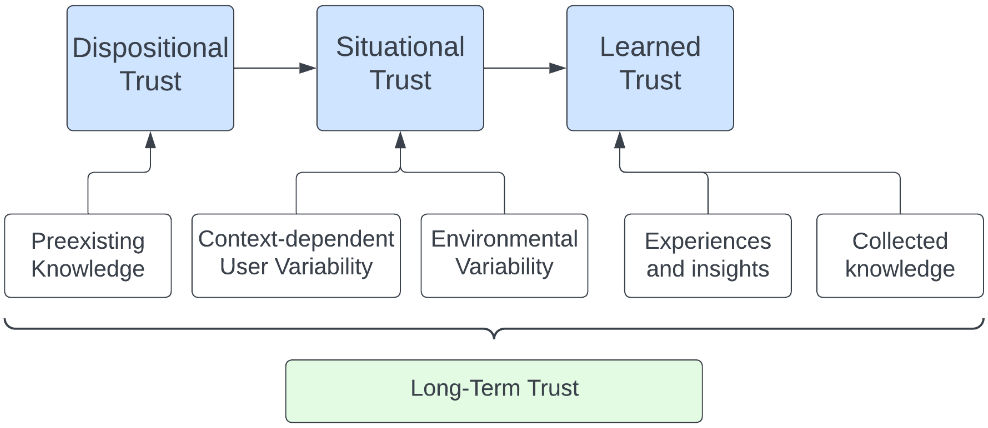
Trust in technology and automated vehicles is a key determinant for a prospective driver's use of AVs (Hoff & Bashir, 2014). However, trust as a concept does not have one agreed-upon definition. Moving forward, the definition by Lee and See (2004) is used. Trust is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p.51). Moreover, trust is not stable and independent, but an attitude that fluctuates.

Trust varies based on *appropriateness* and *context* (Lee & See, 2004). The appropriateness of trust refers to the distance between the true and perceived capabilities that the system has. Appropriateness captures whether the belief in agents' ability is based on real capabilities. The context may change the amount of trust despite inherent attitudes toward the system. The context can be any external or internal factors other than the driver's predisposition (Lee & See, 2004). For instance, a high-risk situation can lead to reduced trust even if the user would otherwise trust the system's capabilities.

There are three types of trust that shape beliefs and attitudes towards AVs (see Figure 2) (Hoff & Bashir, 2013, p. 117).

Figure 2

Factors influencing trust in automation



Note. The figure shows a theoretical model for trust in automated systems. Adapted from "A Theoretical Model for Trust in Automated Systems" by K. Hoff and M. Bashir, 2013, *CHI '13 Extended Abstract on Human Factors in Computing Systems*, p. 117.

The first type, *dispositional trust*, is what a user feels towards AVs in general. It is unrelated to the context or the system at hand (Hoff & Bashir, 2013). This type captures the initial attitude towards the autonomous system. This type of trust cannot change in the short term.

While dispositional trust is context-independent, *situational trust* is dependent on the situation. Within a specific situation, environmental factors can shape trust. Such factors can be the system's complexity, the task to be completed, as well as apparent risks, or advantages of the system. Another influencing factor may be context-dependent user variability (Hoff & Bashir, 2013), which refers to changes in the user that depend on the situation. For instance, confidence, or physical well-being can change the trust levels of participants despite inherent beliefs.

Long-term differences in the form of changed attitudes and behavior can be attributed to *learned trust*. Learned trust is shaped by experiences, insights from current situations, and knowledge collected about autonomous systems (Hoff & Bashir, 2013). Positive experiences, in which the user feels safe can contribute to an increase in long-term trust. A stable and positive attitude toward AVs builds the intention to use the system. The intention to use can ultimately result in the usage of AVs.

Trust Calibration

An appropriate amount of trust can be achieved by calibrating trust. Trust calibration can be defined as the spectrum between misuse and disuse, both forms of abuse (Parasuraman & Riley, 1997). Misuse is defined as over-trust in the system. Users over-rely on features and use technology beyond its intended use (Parasuraman & Riley, 1997). Not monitoring the vehicle while driving within higher levels of automation is an example of misuse.

Disuse refers to users not trusting the system enough, leading them to not use all the offered features (Parasuraman & Riley, 1997). This may result in missed opportunities for increased road safety. For instance, when services such as blind-spot detection are available, but not used. This could cause the driver to miss a car while switching lanes resulting in a crash that would have been detected by the system.

Lee and See (2004) define the optimal level of trust on the trust calibration spectrum as appropriate reliance. Appropriate reliance means that the user fully uses the system including the automated features, but also recognizes boundaries in the system (Wintersberger et al., 2016).

Appropriate reliance in the context of automation can be divided into calibration, resolution, and specificity (Lee & See, 2004). Calibration is the difference between beliefs in a system and its true capabilities. Resolution is the measurement of how precise the capabilities of an AV can be judged. The specificity of user trust describes the degree to which the trust refers to specific aspects of the system. Optimal trust calibration consists of accurate calibration, a high level of resolution, and high specificity. Furthermore, trust calibration is dynamic. With each new system, changes in the environment, and changes inherent to the user, calibration takes place (Hoff & Bashir, 2014).

Adaptive automation

Appropriate reliance can be a challenge to reach. Therefore, flexible approaches that adapt to the dynamic nature of trust are needed. However, the currently implemented autonomous features or stages of full automation are static. Static automation refers to a complete transition between manual and autonomous driving (Kidwell et al., 2012). Vehicles with an automated driving option either drive without driver input or require full driver control.

Static automation has the disadvantage that it can promote distrust due to errors in mode awareness (Calhoun, 2022). Mode awareness refers to users being aware in what state the system is currently in. For instance, if autopilot is activated the user is aware the system is in the autonomous mode. When mode errors occur however, the user cannot accurately depict the mode the system is in (Kidwell et al., 2012). For instance, mode awareness errors can occur when the AV prevents overtaking while the driver intends to switch lanes. In this moment the vehicle takes over the control while the driver may not be aware of the transition. In order to adapt to users' fluctuations in trust, there are two types of flexible automation that can be used, adaptive and adaptable automation. The goal of adaptive and adaptable automation is more effective and efficient human-computer performance.

Adaptable automation gives full authority to the human and the user can adjust the level of automation (Kidwell et al., 2012). The benefit of this automation type is a heightened sense of control over the vehicle (Calhoun, 2022). A sense of control is especially important in uncertain situations. Kidwell et al. (2012) found that adaptable automation can increase confidence in decision making. However, the results also showed an increased workload for participants.

Adaptive automation lies within the authority of the AV. The system allocates tasks between the driver and the vehicle, depending on factors such as user performance or measured workload (Kidwell et al., 2012). Adaptive automation cannot decrease unpredictability, but it can decrease concerns about automation surprise and mode awareness. It also reduces the workload of users, as the system is dividing tasks (Calhoun, 2022).

De Visser and Parasuraman (2011) found that both, adaptable and adaptive automation can increase user trust compared to static automation. The more flexible approach increases Human-machine performance increases with approaches that are more flexible in dividing autonomous tasks. Yet, adaptive automation may be better embedded in the system than adaptable automation. User workload is reduced, and implementation is based on measuring actual performance instead of perceived control over features. Thus, adaptive automation may be more effective than adaptable automation.

Measuring trust and the feasibility of continuous measurement tools

To embed adaptive automation within a system, the AV must use a clear information processing algorithm to change and react to new information. One way to adapt to user trust is to use trust data as feedback to adjust driving performance. It is important, however, that the procedure with which the system processes the driver's trust is consistent and reliable.

There are different ways to measure trust. One is with self-indicated trust questionnaires (Kalayci et al., 2021). Empirically determined trust scale questionnaires (Jian et al., 2000) and their modified versions (Walker, Wang et al., 2019) are typically used to measure trust in AVs. Researchers use questionnaires before, during, or after an experiment. However, researchers rarely use data that captures fluctuations in trust. Using a combined approach could lead to a more holistic assessment of user trust that reflects the development of trust accurately (Kalayci et al., 2021; Walker, 2021).

One way to assess fluctuations in trust is to collect continuous data during an experiment with a physical slider (Walker, Dey et al., 2019). Participants can move the slider from 0 to 100 to show their current attitude. Physical sliders are easy to use and provide real-time data. The collected data can then be processed for feedback in real-time (Walker, Dey et al., 2019). To use feedback as a basis for adaptive automation, the processing algorithm should work reliably with a variety of participants. It should also capture a high number of fluctuations and have clearly defined concepts of what a trust change is.

Research Questions

The aim of the study was to investigate the effect of adaptive automation on user trust. For this purpose, two experiments were conducted. In Experiment 1 the aim was to develop an algorithm to process continuous data collected with a slider (Walker, Dey et al., 2019). Eight algorithms were compared through a within-subject design in a fixed speed simulation. Since the raw slider data solely collected the current trust value, it was necessary to compare trust with each other by using an algorithm to determine if a change in trust was significant or not.

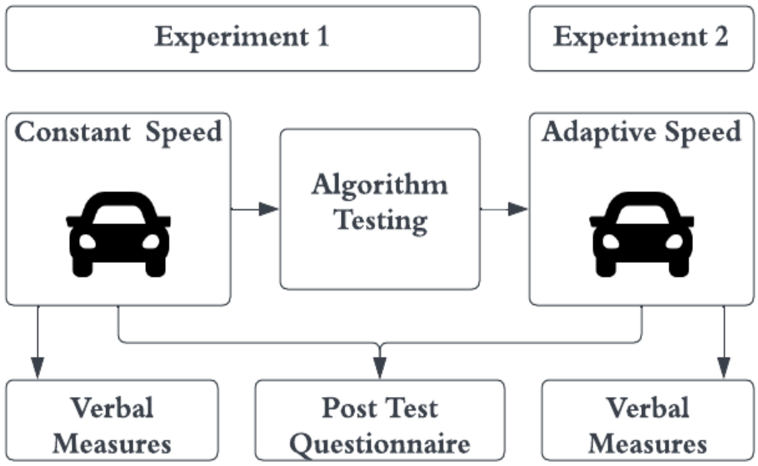
There was no algorithm that could be used as a gold standard. Therefore, we defined a versatile algorithm as an algorithm that led to the most trust changes detected by comparing values with a predefined trust change formula. Due to individual differences in sliding speed, the highest number of changes detected meant a higher versatility in terms of temporal sensitivity. For instance, a low temporal sensitivity and fast slider movements could have led to missed detected changes, while a high sensitivity and slow slider movements could have led to multiple detected changes per movement.

The aim of Experiment 2 was to assess how experiencing adaptive automation may increase trust towards automated vehicles in initially low-trusting users. The adaptive simulation itself consisted of a vehicle driving a predetermined route. In the adaptive condition the speed changed based on the slider values of participants. In the fixed speed simulation, the speed was set to a constant 20 km/h. The most versatile algorithm of Experiment 1 was used as the method to induce speed changes in Experiment 2, which tested the adaptive speed simulation.

The hypothesis was that post-test trust scores would be significantly higher after exposure to the adaptive speed simulation compared to the fixed speed one, since it was expected that adjusting the vehicle behavior to the user's trust would increase trust. Additionally, we aimed to validate the slider values by correlating the scores with verbal indications of trust measured in both experiments. It was hypothesized that verbal trust indications would show high positive correlations with the slider trust data. The fixed speed data of Experiment 1 was used as the control condition for Experiment 2. Participants in both experiments filled out a pre-test questionnaire, and a post-test questionnaire. After both experiments were completed, the post-test trust questionnaire results were compared between-subjects of Experiment 1 and 2 (see Figure 3).

Figure 3

Study Set-Up



Note. The figure shows the relationship between Experiment 1 and Experiment 2. The car represents the simulation part present in both experiments. The same measures are used in both experiments (verbal, post-test questionnaire).

Experiment 1

Eight algorithms were developed for the adaptive driving simulator. Seven algorithms were based on data frequency binning, i.e., sorting the most frequent occurring values over a span of, for instance 0.1s into groups. The group modes, in this case all values collected in 0.1s, would be compared with the next group (the next 0.1s). If the modes were significantly different, this was counted as one trust change. The algorithms covered 0.1s, 0.5s, 1s, 2s, 3s, 4s, and 5s. It was expected that subjects would move the slider at different speeds. By comparing statistics modes to check for significant trust changes, a higher or lower temporal resolution was covered.

The main advantage of using modes for data binning is that this method of processing data is more robust to outliers and noise than other methods. The main disadvantage of this algorithm is the possibility of temporal delays. Especially for algorithms that cover a longer timeframe, the calculation of whether a change is detected can only be completed after all the data is collected. If the mode is calculated every 1s, for instance, a significant change would only be detected after a total of 2s. In the case that change happened within the first 0.5s, it would be registered with a 1.5s delay.

The variety of slider movement speeds could lead to either multiple recognized significant trust changes that should be counted as one or could lead to misses. Both would result in an algorithm that is not able to capture trust changes that are intended as one change by participants. The range of algorithms was developed to aim for a time frame that accounted for the number of detected changes despite subjects' individual differences. The hypothesis was that one algorithm version would result in significantly more detected trust changes compared to the others, due to the individual sliding differences.

The eighth algorithm was developed as an approach to cover both shorter and longer intervals. This algorithm compared every value with all following values within the next three seconds by using a method in which the first value that is recorded is the first value that is removed at the end of the predefined timeframe (FiFo), e.g., three seconds. It did not make a difference if the change occurred within the first 0.5s or after 2s. Therefore, this solution was theorized to show more changes despite individual differences (see Table 1).

Table 1

Example of difference in temporal sensitivity

Time	Trust
0.025	50
0.050	50
0.075	50
0.100	50
0.125	50
0.150	50
0.175	50
0.200	50
0.225	55
0.250	55
0.275	55
0.300	52
0.325	52
0.350	52
0.375	52
0.400	52

Time	Trust	0.1s
0.025	50	50
0.050	50	
0.075	50	
0.100	50	
0.125	50	50
0.150	50	
0.175	50	
0.200	50	
0.225	55	55
0.250	55	
0.275	55	
0.300	52	
0.325	52	52
0.350	52	
0.375	52	
0.400	52	

Note. On the left side algorithm 8 was applied while on the right-side algorithm 0.1s was used on the same dataset. Both algorithms were able to identify one significant trust change. The FiFo algorithm showed a higher temporal sensitivity by detecting the change at second 0.225, while algorithm 0.1s was able to detect the same change later at 0.300s.

Another benefit of the FiFo algorithm is that the changes could be recognized immediately. The disadvantage of this algorithm is that noise and outliers can influence the data. Additional advantages and disadvantages, as well as expectations about the temporal sensitivity of all algorithms tested are listed in Table 2.

Table 2

Algorithm overview

Algorithm	Method of comparing values	Temporal sensitivity	Advantages	Disadvantages
0.1s	Mode	Fast movements	Able to capture rapid changes, minimal delays robust against noise and outliers	Not able to capture longer slider movements
0.5s	Mode	Fast movements	Able to capture rapid changes, minimal delays robust against noise and outliers	Not able to capture longer slider movements
1s	Mode	Fast and slow movements	Able to capture rapid and longer changes, short delays robust against noise and outliers	
2s	Mode	Fast and slow movements	Able to capture rapid and longer changes, short delays robust against noise and outliers	
3s	Mode	Slow movements	Able to capture longer changes robust against noise and outliers	Long delays
4s	Mode	Slow movements	Able to capture longer changes robust against noise and outliers	Long delays
5s	Mode	Slow movements	Able to capture longer changes, robust against noise and outliers	Long delays

Fifo (3s Time Frame)	Fifo	Fast and slow movements	Able to capture rapid and longer changes, no delays	Prone to noise and outliers

The total number of algorithms tested was eight. For results in Experiment 1, it was expected that comparing the trust values of a slider with different algorithms would result in one algorithm leading to more accurate results, i.e., higher versatility despite individual differences. The criterion for measuring accuracy was the highest number of detected changes among all algorithms. Less detected changes would have meant that the time frame was either too short or too long, resulting in misses. The eighth algorithm, which was specifically designed to detect trust changes accurately on participants with different sliding speeds, was expected to show statistically significant differences compared to the other seven algorithms.

Methods

Participants

There were 10 participants consisting of 7 males, and 3 females. Nine participants were students. The minimum age was 18 and the maximum age was 27 ($M = 21.4$, $SD = 2.46$). Participants were from Germany ($n = 4$), the Netherlands, ($n = 4$), Cyprus ($n = 1$), Moldova ($n = 1$). The driving experience ranged from 0.5 to 10 years ($M = 3.8$, $SD = 2.58$). Subjects also indicated their current prevalence of driving activity with the options of every day ($n = 2$), twice per week ($n = 2$), once per week ($n = 2$), once per month ($n = 2$), and never ($n = 2$).

To participate, subjects needed to have a valid driver's license, and sufficient English or German skills to understand the questionnaires. Additionally, participants should have no known motion sickness. Only participants with no prior experience with driving or being a passenger in an automated car were permitted.

The participants were recruited through convenience sampling and through SONA, the test subject pool website used by the University of Twente, in which participants received one credit in exchange for participating in the study. Scores of four or lower on the shortened Empirically Determined Trust Scale for Automated Systems (Verberne et al., 2012) were considered low enough for participation. The ethics committee of the University of Twente approved the study (request no. 201080).

Materials

Pre-test questionnaire

A questionnaire was provided to participants through Google Forms (see Appendix A). The first part of the form consisted of informed consent, in which the aim of the study was described. The second part included demographic questions concerning the age, gender, and nationality. Participants were also asked to state their frequency of driving on European roads and how long they have had their driving license.

The last part consisted of the pre-test trust questionnaire and a 7-item Likert scale, based on the trust rating scale by Jian et al. (2000). The initial scale (Verberne et al., 2012) was shortened to seven items to fit the purpose of this study (Walker, Wang et al., 2019). Answers were ranging from 1 (not at all) to 7 (extremely). Additionally, the word "systems" was adjusted to "self-driving cars" to better fit the context. The two items "I am cautious about self-driving cars", and "Self-driving cars can have harmful consequences" were reverse scored. The mean of the 7-item Likert scale was used to calculate the trust score.

Post-test questionnaire

After the simulation, participants received the post-test questionnaire (see Appendix B). Similar to the pre-test questionnaire, a 7-item Likert scale based on Jian et al. (2000) was used to assess participant trust. Compared to the pre-test questionnaire, items were stated in past tense and were specific to the car used in the simulation (e.g., "I was cautious about the self-driving car"). The same two items were reversed as in the pre-test questionnaire.

Simulator

The driving simulation was built in Unity (Version 2020.38.8) by the BMS Lab at the University of Twente. The asset EasyRoads 3D was used to create the driving scene. The scene consisted of a road that included a steep hill (driving up and down), a bridge with a steep downward road, narrow curves, and flat route parts (see Figure 4).

Figure 4

The route of the autonomous vehicle

Note. The starting position is marked in blue. An exemplary flat part of the route is marked red, an incline before a mountain white, and a drop before a bridge is highlighted in yellow. The vehicle drove clockwise.

Traffic within the scene was created through the Unity asset "Mobile Traffic System" (see Figure 5). Oculus Rift S VR glasses were used to simulate the AV and the environment. Participants were placed in a PlayStation driving simulation seat (see Figure 6). The set-up included a LogiTech steering wheel and pedals, and a Next Level motion and traction platform. The platform, wheel and pedals were not used in this experiment.

Figure 5

The participant view including traffic

Figure 6

Next Level Racing seat including Logitech steering wheel and pedals

Trust Slider

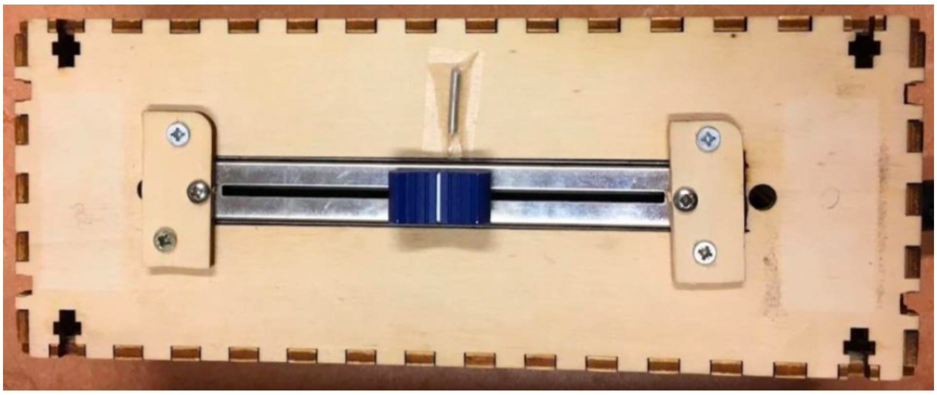
Trust was measured with a physical input device in the form of a slider (see Figure 7). The trust slider was initially developed to measure pedestrian's feelings of safety towards crossing a road (Walker, Dey et al., 2019). In the current study, the feasibility of the slider for measuring trust towards AVs was assessed.

The trust slider was used to capture trust continuously and in real-time with a sampling rate of 66Hz. The scale reached from 0 (no trust at all) to 100 (complete trust). The halfway point was indicated through a metal bump alongside the slider. Since participants took part in VR and were not able to see their position on the slider, this bump served as a haptic orientation point, in addition to the bumps at the two extreme positions.

The slider was built as a motorized slide-potentiometer (see Figure 8), with its extremities mapped to 0 and 100 through an Arduino. The motor was not used in this study. Therefore, participants were able to move the slider freely throughout the experiment. Furthermore, the researchers thoroughly tested the slider output for noise, which was not detected.

Figure 7

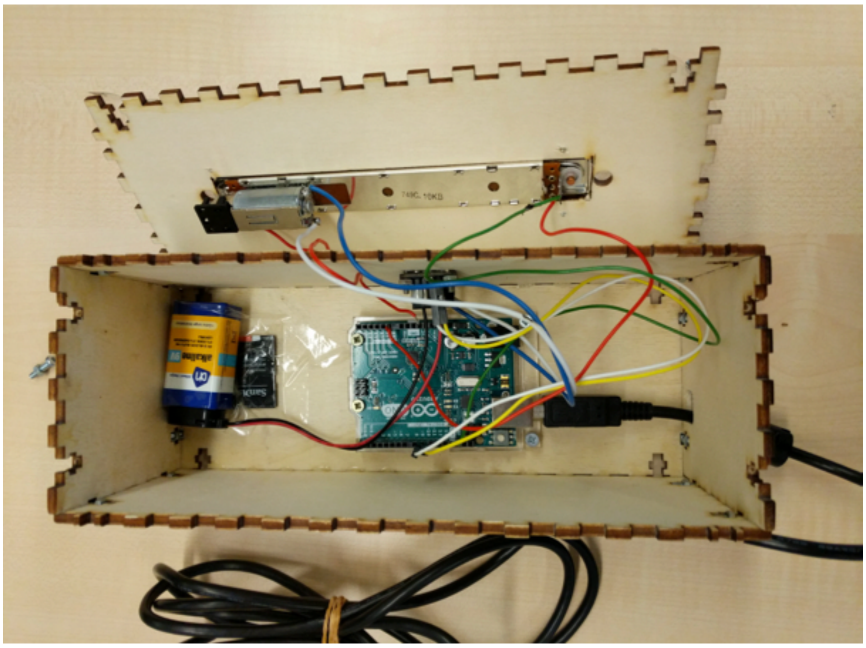
The trust slider



Note. The slider was designed without the halfway point by Walker, Dey et al., 2019.

Figure 8

The inside of the slider



Algorithm development

Trust change calculation

A significant trust change was defined as a change of at least ten percent of any previous values within the time frame, indicated by moving the slider (Equation 1).

$$\text{Trust change} = \text{abs (previous value} - \text{current value)} > 0.1 * \text{previous value} \tag{1}$$

In order to use this equation to compare values with each other and define whether or not a significant change was detected, an algorithm had to be used. A formula would only compare two values at a time. However, the slider collected 66 values per second, therefore a gradual increase from e.g., 50 to 55 in increments of 1, would have led to a mere comparison of 50 to 51, 51 to 52, etc. The algorithm made it possible to compare values over bigger time periods, such as all values collected within one second i.e., frequency data binning.

Another issue was that participants were expected to differ in their sliding speeds. One person may take one second to increase their trust from 50 to 55, while another would take four seconds. Therefore, different time frames had to be compared to analyze which algorithm would be able to detect the highest number of trust changes despite sliding speeds, i.e., an ideal medium capable of detecting a high number of significant changes for the majority of participants. Seven algorithms were developed using a comparison of group means. The eighth one was a FiFo algorithm that used a list method.

The first algorithm compared modes across seven different time periods. A mode is the value that appears most often in a set of data values. The mode was used to ensure a better representation of user trust during those time frames. The seven algorithms covered 0.1, 0.5, 1, 2, 3, 4 and 5 seconds, respectively. Based on the sliding speed of each participant, a higher or lower temporal resolution may result in less missed trust changes. For instance, the 0.1s algorithm measured the mode for all values collected within 0.1 seconds. Following, the mode of the next 0.1 seconds was calculated, and the two modes were compared by using Equation 1. Afterwards, the second mode was compared to the mode of the following 0.1s, etc. Below, there are three examples of how the 0.1s algorithm and the 0.2s algorithm could lead to different results. Table 3 demonstrates how the 0.2s algorithm could detect trust changes that the 0.1s algorithm missed, and how the reverse may occur.

Table 3

Exemplary data highlighting difference in detected trust changes depending on algorithm

<table>
<tr><th>Time</th><th>Trust</th><th>0.1s</th><th>0.2s</th><th>Time</th><th>Trust</th><th>0.1s</th><th>0.2s</th></tr>
<tr><td>0.025</td><td>50</td><td rowspan="4">50</td><td rowspan="8">50</td><td>0.025</td><td>50</td><td rowspan="4">50</td><td rowspan="8">50</td></tr>
<tr><td>0.050</td><td>50</td><td>0.050</td><td>50</td></tr>
<tr><td>0.075</td><td>50</td><td>0.075</td><td>50</td></tr>
<tr><td>0.100</td><td>50</td><td>0.100</td><td>50</td></tr>
<tr><td>0.125</td><td>50</td><td rowspan="4">52</td><td>0.125</td><td>50</td><td rowspan="4">50</td></tr>
<tr><td>0.150</td><td>52</td><td>0.150</td><td>50</td></tr>
<tr><td>0.175</td><td>52</td><td>0.175</td><td>50</td></tr>
<tr><td>0.200</td><td>52</td><td>0.200</td><td>50</td></tr>
<tr><td>0.225</td><td>53</td><td rowspan="4">53</td><td rowspan="8">55</td><td>0.225</td><td>55</td><td rowspan="4">55</td><td rowspan="8">52</td></tr>
<tr><td>0.250</td><td>53</td><td>0.250</td><td>55</td></tr>
<tr><td>0.275</td><td>53</td><td>0.275</td><td>55</td></tr>
<tr><td>0.300</td><td>55</td><td>0.300</td><td>52</td></tr>
<tr><td>0.325</td><td>55</td><td rowspan="4">55</td><td>0.325</td><td>52</td><td rowspan="4">52</td></tr>
<tr><td>0.350</td><td>55</td><td>0.350</td><td>52</td></tr>
<tr><td>0.375</td><td>55</td><td>0.375</td><td>52</td></tr>
<tr><td>0.400</td><td>55</td><td>0.400</td><td>52</td></tr>
</table>

Note. The left dataset shows an example of a faster slider movement. The 0.1s algorithm and the 0.2s algorithm were applied to the same data. The 0.1s algorithm was not able to detect any trust changes (red), while the 0.2s algorithm was able to detect one trust change (green). The right dataset shows an example of a slower slider movement. The 0.1s algorithm was able to detect a change on trust between the mode comparisons of 50 and 55 (green), while the 0.2s algorithm did not detect this change (red).

The second algorithm type was a FiFo algorithm. It involved another approach of calculating trust changes, in which all values were stored in one list. A list is a sequence of values stored under a shared name. After the first twenty values, i.e., 3 seconds, the first item in the list was removed. Then, every time a new value was added to the bottom of the list, the first entry was removed again. This algorithm ensured that the list would cover exactly three seconds. The 3 second time frame was chosen as a medium length that accounts for a variety of sliding speeds. This algorithm was designed to be highly sensitive to trust changes of all sliding speeds, as a significant trust change could be detected as soon as the values were added to the list by the slider, which is also why only one FiFo algorithm was tested. Table 4 demonstrates how the FiFo algorithm detects trust changes with the same exemplary values as in Table 3 and is able to detect the changes correctly and without a delay. The algorithms in Table 3 were able to detect one correct change each while missing the change in the other data.

Table 4

Exemplary data highlighting the detection of trust changes by the FiFo algorithm

Time	Trust	Time	Trust
0.025	50	0.025	50
0.050	50	0.050	50
0.075	50	0.075	50
0.100	50	0.100	50
0.125	50	0.125	50
0.150	50	0.150	52
0.175	50	0.175	52
0.200	50	0.200	52
0.225	55	0.225	53
0.250	55	0.250	53
0.275	55	0.275	53
0.300	52	0.300	55
0.325	52	0.325	55
0.350	52	0.350	55
0.375	52	0.375	55
0.400	52	0.400	55

Note. The same exemplary datasets as in Figure 9 were used. The FiFo algorithm compared every value following the first value within three seconds for comparisons until a significant trust change was found. No misses or multiple trust changes were detected.

In all algorithms, once a change based on the formula in Equation 1 was detected, the list was cleared, and a new list was started. Without clearing the list, the algorithm would have compared all values with each other. For instance, an increase of 5 to 15 would count as significantly different. Without clearing the list, a following value of 17 would also be counted as a change, since 17 is significantly higher than 5. By clearing the list, 17 would only be compared to 15, thus not resulting in a detected trust change.

Design

In Experiment 1, a within-subject design, participants experienced a fixed driving simulation. The algorithms were applied to participant data post-hoc.

Task

No manual take-over behavior was expected by participants, but they were told to actively try to imagine driving in the AV. It was emphasized that the study aimed to improve the driving behavior of the car. The researchers also mentioned that they were interested in trust towards the vehicle throughout the experiment, for which the questionnaires and the slider were used.

The task included participants indicating their current trust level on the slider. Participants continuously moved the slider for ten minutes of simulation. They moved the slider as often and as much as they wanted to indicate trust. Additionally, the researchers played a sound at three predetermined points along the route. Participants then verbally stated their trust on a scale from 0 to 10. After ten minutes of simulation, participants filled out the post-test questionnaire.

Procedure

Participants filled out the pre-test questionnaire before arriving at the simulator. Depending on their trust score, they received an invitation to schedule an appointment to take part in the study. Alternatively, they received an email explaining that their trust scores were not suited for the study.

After arriving at the laboratory facilities and receiving the above-mentioned instructions, participants were placed in the driving simulator. The VR glasses were placed on their heads and the slider was given to participants. They were asked to move the position to the minimum and maximum level to practice moving without being able to see the position. After the set-up, the researchers started the driving simulation. In the first experiment, participants experienced a constant speed of 20km/h. Verbal measures of trust were collected at predetermined parts in the route, which were the beginning of a steep hill, the start of a steep decline at a bridge, and a straight part of the street without steep incline or decline, as indicated in Figure 3. After ten minutes of simulation, participants filled out the post-test questionnaire. Participants were then informed about the goal of the study. They were then thanked for their participation. The whole experiment took 30-40 minutes.

Data Analysis

Number of Detected Trust Changes between Algorithms

Before analyzing the data, the *Number of Detected Trust Changes* per participant and *Algorithm* were determined by dividing the data into equal groups covering 0.1s, 0.5s, 1s, 2s, 3s, 4s, and 5s, respectively. The group sizes referred to the seven algorithms using the mode for comparisons. Next, the mode was calculated per group and compared to the subsequent mode by using the formula described earlier to calculate significant trust changes (see Appendix C). A one-way repeated measures ANOVA was conducted to determine whether there was a statistically significant difference in *Number of Detected Trust Changes* between the 8 *Algorithms,* applied to the same dataset (see Appendix D). For this analysis, the data of the slider was used.

Results

Number of Detected Trust Changes between Algorithms

There were no outliers as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. The data was normally distributed for each algorithm, except the 0.5s algorithm (p =.044), as assessed by boxplot and Shapiro-Wilk test (p >.05), respectively. The assumption of sphericity was not met, as assessed by Mauchly's test of sphericity, $\chi2(2) = 130.99$, $p < .001$. Epsilon (ε) was 0.226.

The application of the 8 algorithms elicited statistically significant changes in the number of detected trust changes, $F(1.581, 14.229) = 9.423$, $p = .004$ (see Table 5).

Table 5

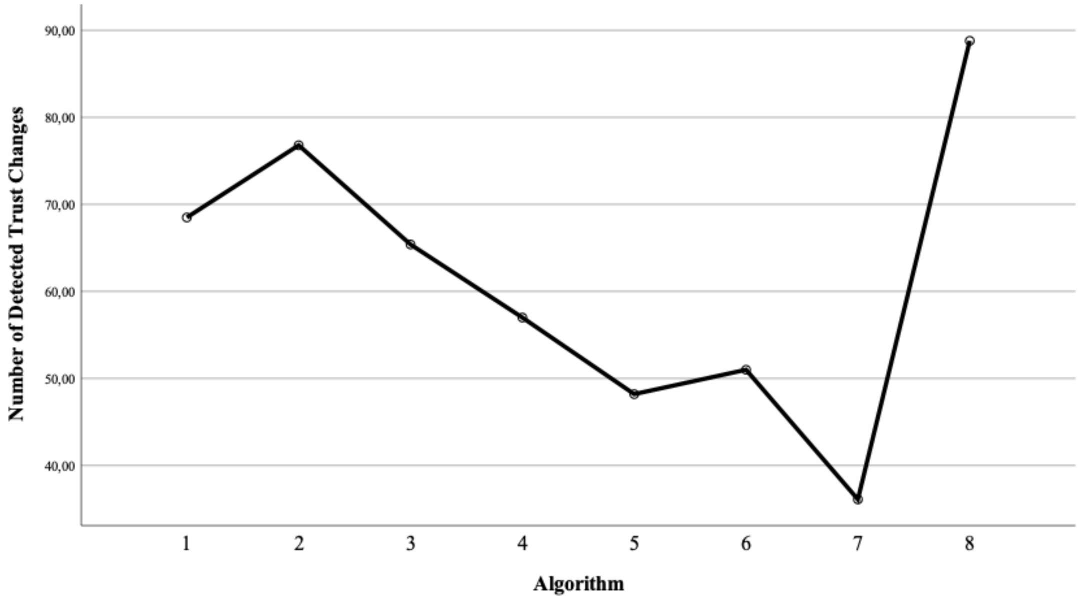
The effect of algorithms on the number of detected trust changes

Algorithm	Mean (M)	Standard Deviation (SD)
0.1s	68.5	53.61
0.5s	76.8	60.68
1s	65.40	50.53
2s	57.00	42.89
3s	48.20	34.66
4s	51.00	37.57
5s	36.10	22.97
FiFo	88.80	69.56

Post-hoc analysis with a Bonferroni adjustment revealed that the number of detected trust changes was not significantly different between any two means. Therefore, there was not one algorithm that showed significantly more detected changes than another by direct comparison. However, the application of algorithm 8 led to more detected trust changes than any other algorithm (see Figure 9).

Figure 9

Means of Number of Detected Trust Changes per Algorithm



Note. This figure shows how many trust changes were detected by each algorithm within the fixed speed dataset. All algorithms were applied to identical data.

Therefore, the null hypothesis can be rejected. The alternative hypothesis that the application of *Algorithms* has an effect on the *Number of Detected Trust Changes* can be accepted, but we cannot accept the alternative hypothesis that algorithm 8 specifically leads to significantly more changes detected than other variations.

Discussion

Experiment 1 tested eight algorithms to determine which would result in the highest number of detected trust changes. It was hypothesized that one of the algorithms would result in significantly more detected trust changes, i.e., would be more versatile than the others.

The results of Experiment 1 did not result in one algorithm showing significantly more detected changes than any others. In simulators where a high level of noise can be expected due to sudden movements not caused by participants or otherwise caused noise, a mode algorithm with a medium timeframe such as 0.2s could be chosen. Nevertheless, algorithm 8 was able to detect the highest number of changes.

In case technical noise is eliminated, the FiFo algorithm may be better suited to capture all fluctuations of trust. Technical noise can be eliminated, for instance by excluding values that occur for a duration that would not be possible for a human participant. An example of this would be a value of 1 that occurs in one millisecond preceded and followed by values of 60. In that case, it would not be possible for the user to move the slider fast enough in that duration (correct values would decrease, such as 60, 55, 40, 12, 1, then increase again which would take several milliseconds).

Due to the fact that algorithm 8 has a higher temporal sensitivity and that the algorithm used to induce speed changes in real time for an adaptive driving simulator, a temporal delay may cause speed changes to feel unpredictable to participants. Since algorithm 8 detected the highest number of trust changes and was able to detect trust changes immediately, it seems the most appropriate algorithm to apply to the adaptive driving simulator. Therefore, all further analyses below that used adaptive automation used algorithm 8 as a basis for detecting trust changes.

Experiment 2

The goal of Experiment 2 was to assess how trust may increase after experiencing adaptive automation for participants with initially low-levels of trust compared to a fixed-speed simulation. For this purpose, Experiment 1 was used as the control condition. Pre-and post-test questionnaires were used to compare trust between speed conditions, and the average amount of trust collected by the slider throughout the experiment was compared. We expected that the post-test questionnaire results of the adaptive speed condition would be higher after the adaptive speed condition compared to before the experiment and compared to the fixed speed condition.

Additionally, we aimed to validate the slider values by correlating the scores with verbal trust measures in both experiments. It was hypothesized that verbal trust indications would show high positive correlations with the slider trust data. We expected that the average trust during the adaptive speed condition would be higher compared to the fixed speed condition.

Lastly, the goal was to validate the trust data collected by the slider by collecting verbal trust data throughout the experiment and correlating these scores. We expected a high positive correlation.

Methods

Participants

For the adaptive driving simulation, there were 10 participants consisting of 6 males and 4 females. 6 participants were students. The remaining 4 subjects were of various professions. The minimum age was 18 and the maximum was 25 (M = 21.2, SD = 2.49). Additionally, there were participants from Germany (n = 8), the Netherlands, (n = 1), and Lithuania (n = 1). The driving experience ranged from 0.5 to 8 years (M = 2.9, SD = 2.17). They also indicated their current prevalence of driving activity with the options of every day (n = 4), twice per week (n = 3), every few months (n = 2), and never (n =1).

For the data analysis the data of both experiments were combined. In total there were 20 participants consisting of 13 males, and 7 females. 15 participants were students. The remaining 5 subjects were of various professions. The minimum age was 18 and the maximum was 27 (M = 21.3, SD = 2.41). Additionally, there were participants from Germany (n = 12), the Netherlands, (n = 5), Cyprus (n = 1), Moldova (n = 1), and Lithuania (n = 1). The driving experience ranged from 0.5 to 10 years (M = 3.35, SD = 2.37). They also indicated their current prevalence of driving activity with the options of every day (n = 6), twice per week (n = 5), once per week (n = 2), once per month (n = 2), every few months (n = 2), and never (n =3). The requirements and the sampling method were identical for both experiments.

Materials

Speed Changes

The recorded trust changes using algorithm 8 were used to induce speed changes (see Appendix E). For every increase in trust, the speed would immediately increase, and for every decrease in trust the speed would immediately reduce by 5 km/h. The minimum speed of the vehicle was identical to the speed change increments, in this case 5 km/h. The maximum speed was set to 40 km/h. The remaining materials were identical to Experiment 1.

Task

In Experiment 2, a between-subject design, another group of participants experienced an adaptive speed driving simulation based on the most versatile, (in this case most detected changes) algorithm as tested in Experiment 1. The tasks and procedure of Experiment 2 were identical to Experiment 1. However, participants only started at a speed of 20km/h. After the algorithm used showed a significant trust change, the speed of the vehicle changed by 5km/h. The speed change took place immediately after the trust change.

Data Analysis

Changes in Trust Scores based on Time and Speed Condition

Before the data were analyzed, pre- and post-trust scores were calculated. A two-way mixed ANOVA was used to assess the effect of *Time (pre-test, post-test)* and *Speed Condition (adaptive speed, fixed speed)* on the questionnaire *Trust Scores* (see Appendix F).

Average Slider Scores based on Speed Condition

To prepare the data for the third analysis, the mean of the slider values was calculated per participant. An independent-samples t-test was run to determine if there were differences in the *Average Slider Scores* during the experiment between the adaptive and fixed speed simulations.

Verbal Trust Scores and Slider Trust Scores correlation

Prior to analyzing the data, the physical trust scores needed to be prepared for comparisons with the verbal measures. For every time point at which the verbal measure was collected, the mean of the slider values +/- 5 seconds were calculated (see Appendix C). Furthermore, the slider values were transformed to the same scale as the verbal scores. A Spearman's rank-order correlation was run to assess the relationship between *Verbal Trust Scores* and *Slider Trust Scores* (see Appendix G).

Results

Changes in Trust Scores based on Time and Speed Condition

There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. Pre-and post-trust scores were normally distributed, as assessed by Normal Q-Q Plot. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance (p > .05). There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices (p =.500). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi 2(2) = .00$, $p < .001$.

There was no statistically significant interaction between the *Speed Condition* and *Time* on *Trust Scores*, $F(1, 18) = 0.638$, p =.435, η_p^2 =.034. The main effect of *Time* showed a statistically significant difference in mean trust at the different time points, $F(1, 18) = 8.84$, p =.008, η_p^2=.329, but no significant main effect of *Speed Condition* on *Trust Scores* was found $F(1, 18) = 1.916$, p =.183, partial η_p^2 =.096. There was a significant mean increase in *Trust Scores* pre-test (M = 3.34, SD =.59) to post-test (M = 4.06, SD = 1.08), of .72, $F(1, 18) = 8.938$,

p =.008, $\eta_p{}^2$ =.329. In other words, *Time* was found to have a significant effect on *Trust Scores*, but the effect occurred regardless of the *Speed Condition*.

Average Slider Scores based on Speed Condition

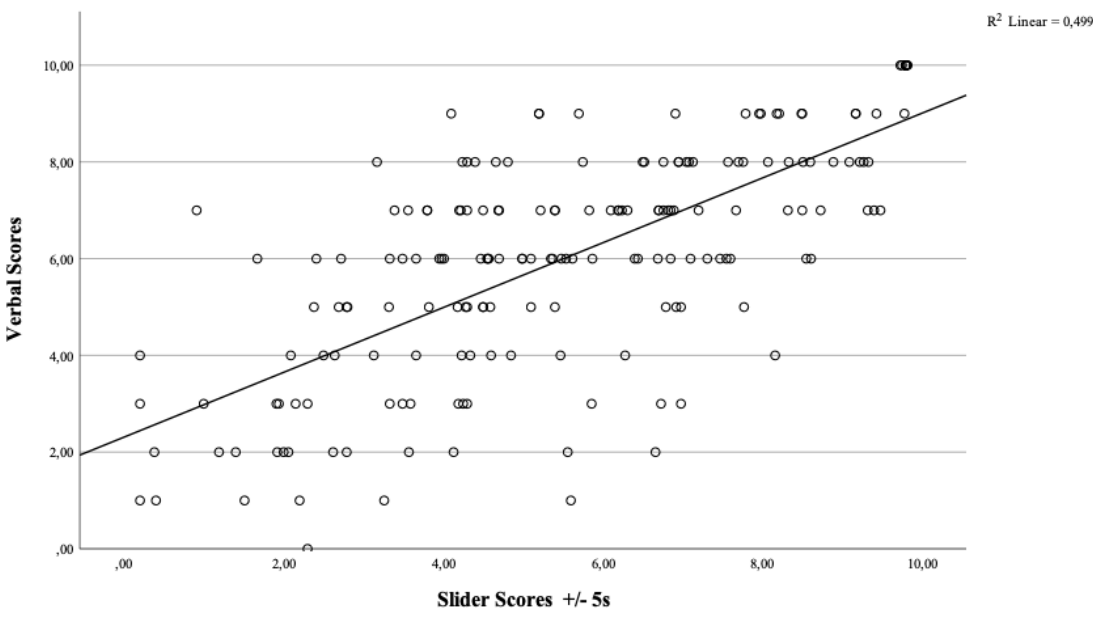
There were no outliers in the data, as assessed by inspection of a boxplot. *Average Slider Scores* for each level of *Speed Condition* were normally distributed, as assessed by Shapiro-Wilk's test (p > .05), and there was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .219). No statistically significant difference was found between the *Average Slider Scores* of the adaptive *Speed Condition* compared to the fixed speed *Speed Condition* M = -11.61, 95% CI [- 29.05 to 5.83], p =.089.

Verbal Trust Scores and Slider Trust Scores correlation

A visual inspection of the scatterplot showed the relationship between the two variables to be monotonic (see Figure 10). There was a strong positive correlation between verbal and physical trust scores, rs = .706, p < .001.

Figure 10

Scatterplot of Verbal and Slider Scores



Discussion

Post-Test Trust Levels

The hypothesis for Experiment 2 was that the post-test trust questionnaire results would be significantly higher in the adaptive speed simulation compared to pre-test questionnaire scores and compared to the fixed speed simulation. We observed a significant increase in trust scores over time, regardless of the type of simulation. However, we found no statistically significant difference in average slider scores between the adaptive speed condition and the fixed speed condition. This suggests that the type of simulation (i.e., whether the vehicle's speed was fixed or adaptive) did not have a significant effect on trust ratings. There are multiple possible explanations for why the results were not significant.

Previous research on trust in autonomous vehicles has found that trust can increase over time as users become more familiar with the system (Clement et al., 2022). This would explain why trust increased in all conditions. However, this could also explain why there was no observable difference between speed conditions. Within a single session of driving simulation, familiarity would have increased trust to the point that differences between speed conditions might not have had a significant impact.

Another possible explanation is that the driving style of the vehicle may have been perceived as aggressive by participants in both speed conditions Driving style can be defined as how the act of driving is conducted (Eboli et al., 2017). A defensive driving style involves smooth and fluid vehicle movements, i.e., predictable driving behavior. An aggressive driving style refers to the opposite, for instance, rapid acceleration and deacceleration without cause, and not breaking in curves.

Based on these definitions, the behavior of the vehicle in both speed conditions would be classified as aggressive, which is why trust may have not increased more than due to familiarity. The vehicle does not deaccelerate before curves and accelerates after, and it does not decrease speed when driving up or down a mountain. The speed changes in the adaptive condition are not gradual. A defensive driving style has been found to promote participant trust, while an aggressive driving style can decrease it (Ekman et al., 2019).

An alternative explanation for the aggressive driving style is that the behavior of the vehicle does not replicate the behavior that drivers of manual vehicles are used to. Trust can be established through comfortability (Hartwich et al., 2018). AVs should aim to replicate manual driving by slowing down before approaching curves and accelerating after, gradual increases and decreases of speed, etc.

The current simulator drives at a constant speed, except for speed changes caused by changes in slider scores. Therefore, the missing realism and familiarity of how participants would drive themselves may have made the vehicle less trustworthy (Hartwich et al., 2018). Especially the absence of gradual increases and decreases in speed is a concern in how participants rate the trustworthiness of AVs (Dikmen & Burns, 2016).

The average speed of the adaptive simulator was 40km/h, while that of the fixed simulator was 20km/h. Sagberg et al. (2015) reported that participants indicated higher speeds as one of the main concerns for predicting accidents in other vehicles. Therefore, the adaptive simulated vehicle may have been perceived to be less safe due to an average higher speed compared to the fixed speed simulated vehicle.

Validity of slider

Additionally, we predicted that during the experiments, verbal and slider trust ratings would correlate strongly positively. This would suggest that the same concept, namely trust, is measured by both measurement tools. The assumption that there was a strong positive correlation between verbal and slider trust scores was accepted as confirmation that the slider could indeed measure the concept of trust.

General Discussion and Conclusion

The purpose of this study was to develop an algorithm that would process trust data collected using a slider, and then use this algorithm in adaptive automated driving simulators to determine whether trust ratings changed for participants with low levels of trust. We also compared trust data collected using the slider to verbal trust indications to assess whether they both measured the concept of trust. To achieve these goals, we conducted two experiments:

In Experiment 1 eight algorithms were tested to determine the most versatile algorithm i.e., the algorithm that detected the highest number of changes despite individual sliding differences. In Experiment 2 the most versatile algorithm from Experiment 1 was used to induce speed changes in an adaptive speed simulation. The trust ratings of the adaptive speed simulation were compared to those in a fixed speed simulation. Lastly, the relationship between the slider data and verbal measures throughout the simulation was compared.

Limitations and future research

There are several limitations to our study that should be noted. First, the slider used in this study had maximum points of 0% trust and 100% trust. Remaining at these points for a long time did not result in additional speed changes. If a participant continued to not trust the vehicle, further reducing speed could improve their overall trust throughout the experiment. Future research could include further changes of speed if participants remain at the maximum points for a longer time.

Additionally, the simulated driving in our study was not dynamic, with no slowing down before curves or acceleration after. This may not accurately reflect real-world driving conditions, which could affect trust ratings. The behavior of real automated vehicles includes dynamic decreases and increases in speed based on the route structure. Further research may include natural speed changes that adapt to the road.

Another limitation is that the speed changes in our study were not gradual, with an immediate change of 5km/h. This may not accurately reflect how speed changes in real-world situations where those changes are more gradual. We also did not include a randomized speed change condition. This would have allowed us to compare how participants respond to speed changes in general rather than just in the adaptive speed simulation. Furthermore, our study did not include event-based measures of trust, such as comparing trust before, during, and after sharp curves or steep declines. This may have limited our understanding of how trust changes in response to specific events during driving.

Another limitation is that our study was a simulation, and it is possible that trust ratings in a real AV might differ from those in the simulation. The slider itself would need to be revised for real driving scenarios to avoid slider movements through bumps in the road and other disturbances. Finally, our study had a relatively small sample size, which may have limited the generalizability of our findings. Overall, these limitations should be considered when interpreting the results of our study.

Conclusion

The main hypothesis that trust would increase after experiencing adaptive automation could not be confirmed in this study. Furthermore, we were not able to find one algorithm that was significantly better at detecting more trust changes than another. However, the highest number of detected changes and the absence of delays result in the highest feasibility of the FiFo algorithm to compare data.

The strong positive correlation of the verbal-and slider trust data suggests that the slider does indeed measure the concept of trust. This finding indicates that the slider is a valid and feasible tool for measuring trust data in real-time and continuously and poses an alternative to measures such as questionnaires. The main advantage of using the slider in driving simulator studies is that trust can be measured before, during, and after events. In the context of adaptive automation, the findings of this study provide a basis for adaptive automation based on trust.

While the use of the slider and the FiFo algorithm are appropriate measures for the use of adaptive driving simulation, the experiment set-up and speed change parameters should be refined in future studies. Based on the limitations and literature review of this study, we suggest building a more naturalistic driving simulator that reduces its speed before curves, accelerates after and has gradual speed changes. These changes could reduce participants perceiving the behavior of the vehicle as aggressive.

Furthermore, we advise that the maximum points of the slider lead to further speed changes when participants remain on them for a longer time. Since improvements in trust can be expected due to familiarity alone, multiple sessions or longer simulation times could be tested. This is to determine if the presence of an adaptive simulator can significantly increase trust compared to a fixed speed simulator. Lastly, events could be included in future studies to assess how participants' trust responses change based on speed condition.

Overall, we found that the use of the slider is feasible for implementation in driving simulator studies to measure trust continuously. The FiFo algorithm can detect trust changes with a high temporal sensitivity and no delays, while we also observed that the slider is a valid measure to capture the concept of trust. Future studies should focus on refining the above-mentioned parameters to assess the effects of adaptive driving simulations.

References

Calhoun, G. (2022). Adaptable (Not Adaptive) Automation: Forefront of Human–Automation Teaming. *Human Factors*, *64*(2), 269-277. https://doi.org/10.1177/00187208211037457

Clement, P., Veledar, O., Könczöl, C., Danzinger, H., Posch, M., Eichberger, A., & Macher, G. (2022). Enhancing Acceptance and Trust in Automated Driving through Virtual Experience on a Driving Simulator. *Energies, 15, 781*, *15*(3), 781. https://doi.org/10.3390/EN15030781

de Visser, E., & Parasuraman, R. (2011). Adaptive Aiding of Human-Robot Teaming: Effects of Imperfect Automation on Performance, Trust, and Workload. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231. https://doi.org/10.1177/1555343411410160

Dikmen, M., & Burns, C. M. (2016). Autonomous Driving in the Real World: Experiences with Tesla Autopilot and Summon. *Automotive UI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, *2016-January*, 225–228. https://doi.org/10.1145/3003715.3005465

Duarte, F., & Ratti, C. (2018). The Impact of Autonomous Vehicles on Cities: A Review. *Journal of Urban Technology*, *25*(4), 3–18. https://doi.org/10.1080/10630732.2018.1493883

Eboli, L., Mazzulla, G., & Pungillo, G. (2017). How to define the accident risk level of car drivers by combining objective and subjective measures of driving style. *Transportation Research Part F: Traffic Psychology and Behaviour*, *49*, 29–38. https://doi.org/10.1016/J.TRF.2017.06.004

Ekman, F., Johansson, M., Bligård, L. O., Karlsson, M. A., & Strömberg, H. (2019). Exploring automated vehicle driving styles as a source of trust information. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 268–279. https://doi.org/10.1016/J.TRF.2019.07.026

Hartwich, F., Beggiato, M., & Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving – effects of drivers’ age and driving style familiarity. *Ergonomics*, *61*(8), 1017-1032. https://doi.org/10.1080/00140139.2018.1441448

Hoff, K. A., & Bashir, M. (2014). Trust in Automation: Integrating Empirical Evidence on Factors that Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoff, K. A., & Bashir, M. (2013). A Theoretical Model for Trust in Automated Systems. *CHI ’13 Extended Abstract on Human Factors in Computing Systems*, 115–120. https://doi.org/10.1145/2468356.2468378

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2010). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International journal of cognitive ergonomics*, *4*(1), 53-71. https://doi.org/10.1207/S15327566IJCE0401_04

Kalayci, T. E., Kalayci, E. G., Lechner, G., Neuhuber, N., Spitzer, M., Westermeier, E., & Stocker, A. (2021). Triangulated investigation of trust in automated driving: Challenges and solution approaches for data integration. *Journal of Industrial Information Integration*, *21*, 100186. https://doi.org/10.1016/J.JII.2020.100186

Kidwell, B., Calhoun, G. L., Ruff, H. A., & Parasuraman, R. (2012). Adaptable and Adaptive Automation for Supervisory Control of Multiple Autonomous Vehicles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 428–432. https://doi.org/10.1177/1071181312561096

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/HFES.46.1.50_30392

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Raats, K., Fors, V., & Pink, S. (2020). Trusting autonomous vehicles: An interdisciplinary approach. *Transportation Research Interdisciplinary Perspectives*, *7*, 100201. https://doi.org/10.1016/J.TRIP.2020.100201

SAE. (2018). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. *SAE International, 3016*, 1-35. https://doi.org/10.4271/J3016_201806

Sagberg, F., Selpi, S., Piccinini, G. F. B., & Engström, J. A. (2015). A Review of Research on Driving Styles and Road Safety. *Human Factors*, *57*(7), 1248–1275. https://doi.org/10.1177/0018720815591313

Verberne, F. M., Ham, J., & Midden, C. J. (2012). Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(5), 799–810. https://doi.org/10.1177/0018720812443825

Walker, F. (2021). To trust or not to trust?: Assessment and calibration of driver trust in automated vehicles. [Doctoral dissertation, University of Twente]. https://doi.org/10.3990/1.9789055842766

Walker, F., Dey, D., Martens, M., Pfleging, B., Eggen, B., & Terken, J. (2019). Feeling-of-safety slider: Measuring pedestrian willingness to cross roads in field interactions with vehicles. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-6). https://doi.org/10.1145/3290607.3312880

Walker, F., Wang, J., Martens, M. H., & Verwey, W. B. (2019). Gaze behaviour and Electrodermal Activity: Objective measures of drivers' trust in automated vehicles Integrated Cooperative Automated Vehicles. *Transportation Research Part F: Traffic Psychology and Behaviour, 64*, 401-402. https://doi.org/10.1016/j.trf.2019.05.021

Wang, J., Zhang, L., Huang, Y., & Zhao, J. (2020). Safety of Autonomous Vehicles. *Journal of Advanced Transportation*, *2020*, 1-13. https://doi.org/10.1155/2020/8867757

Appendices

Appendix A: Pre-test questionnaire English and German

Automated vehicle behavior: a driving simulator study

The aim of this study is to simulate automated vehicle behavior and look at users' trust ratings during the simulation. Only participants with a specific trust rating can participate in this study. Your initial trust in self-driving cars will be measured through this questionnaire. Filling out the questionnaire will not take you more than 2 minutes. If your trust score is below or above a predetermined value, your data will be deleted immediately. Conversely, if your trust score is in line with research requirements, the experimenter will contact you to schedule an appointment for the driving simulator study. Importantly, you will only receive study credit if you are invited to participate in the actual study.

Informed Consent

I confirm that:

- o I do not suffer from any neurologic, psychiatric, or psychological condition.
- o I do not have any color vision deficits and I do not wear glasses (does not include contact lenses).
- o I do not suffer from motion sickness.
- o I have never driven or been a passenger in an automated vehicle before (for example, Tesla with Autopilot)
- o I have a driver's license.
- o I understand that I can terminate my participation immediately if I feel uncomfortable or unwell.
- o I participate in a sober state and of my own free will.

Demographic Questionnaire

Name (to match your trust scores with further data collection - you may also choose a nickname or number):

____

Age:

____

Gender:

- o Female

- Male
- Prefer not to say
- Other

Nationality:

____

Years of driving experience:

____

How often do you drive on European roads?

- Never
- Once per month
- Once per week
- Twice per week
- Every day

Handedness:

- Left
- Right

Pre-test questionnaire

Through this brief questionnaire, we would like to measure your attitudes toward self-driving cars. We understand that your knowledge of self-driving cars might be limited, so please answer based on your ideas and expectations. Please respond as truthfully as possible, and keep in mind that there is no "correct" answer.

1. I am cautious about self-driving cars

1– 2 – 3 – 4 – 5 – 6 – 7

2. Self-driving cars are reliable

1– 2 – 3 – 4 – 5 – 6 – 7

3. I would entrust my car to self-driving functions for lane keeping, lane changing, alerts following object recognition, etc.

1– 2 – 3 – 4 – 5 – 6 – 7

4. I can count on self-driving cars

1– 2 – 3 – 4 – 5 – 6 – 7

5. Self-driving cars can have harmful consequences

1 – 2 – 3 – 4 – 5 – 6 – 7

6. I trust self-driving cars

1– 2 – 3 – 4 – 5 – 6 – 7

7. I assume that self-driving cars will work properly

1 – 2 – 3 – 4 – 5 – 6 – 7

Contact

In case of any questions, you may contact m.m.kowalski@student.utwente.nl

Please provide your email address. You will be contacted by the researcher to schedule an appointment for the simulator study in case your trust score is suited for this experiment.

____

Thank you!

Thank you for your participation.

Verhalten von automatisierten Fahrzeugen: Eine Fahrsimulator Studie

Das Ziel dieser Studie ist es das Verhalten von automatisierten Fahrzeugen zu simulieren und das Vertrauenslevel von Nutzern während der Simulation zu untersuchen. Nur Teilnehmer mit einem bestimmten Level an Vertrauen können an dieser Studie teilnehmen. In diesem Fragebogen wird Ihr ursprüngliches Level an Vertrauen gemessen. Das Ausfüllen dieses Fragebogens sollte nicht mehr als 2 Minuten dauern. Wenn Ihr Vertrauenslevel unter oder über einem bestimmten Referenzwert liegt, werden Ihre Daten umgehend gelöscht. Sollte Ihr Level an Vertrauen zu den Anforderungen dieser Studie passen, werden Sie kontaktiert, um einen Termin für die Fahrsimulatorstudie an der University of Twente zu vereinbaren. Bitte beachten Sie, dass study credits nur dann vergeben werden können, wenn Sie an der Simulatorstudie teilgenommen haben.

Einverständniserklärung

Ich bestätige, dass:

- ich nicht an neurologischen, psychiatrischen oder psychologischen Konditionen leide.
- ich nicht farbenblind bin und keine Brille trage (Kontaktlinsen sind nicht mit inbegriffen).
- ich nicht an Übelkeit bei starken Bewegungen leide (z.B. Seekrankheit).
- ich nie ein selbstfahrendes Auto gefahren oder Beifahrer in einem selbstfahrenden Auto gewesen bin (z.B. Tesla mit Autopiloten).
- ich einen Führerschein habe.
- ich verstehe, dass ich meine Teilnahme jederzeit beenden kann, sobald ich mich unwohl fühle.
- ich nüchtern bin und aus freiem Willen teilnehme.

Demographischer Fragebogen

Name (um das Vertrauenslevel den weiteren Daten zuzuordnen - es kann auch ein Spitzname oder eine Nummer gewählt werden):

____

Alter:

____

Geschlecht:

- weiblich
- männlich
- keine Angabe

- Anders

Nationalität:

____

Jahre an Fahrerfahrung:

____

Wie oft fahren Sie auf europäischen Straßen?

- Nie
- Einmal im Monat
- Einmal die Woche
- Zweimal die Woche
- Jeden Tag

Händigkeit:

- Links
- Rechts

Pre-Test Fragebogen

Durch diesen kurzen Fragebogen wird Ihre Einstellung gegenüber selbstfahrenden Fahrzeugen gemessen. Obwohl Ihr Wissen über selbstfahrende Autos eventuell limitiert ist basieren Sie Ihre Antworten bitte auf Ihre Ideen und Erwartungen. Bitte antworten Sie auf alle Fragen so ehrlich wie möglich und beachten Sie, dass es keine "korrekte" Antwort gibt.

1. Ich bin vorsichtig mit selbstfahrenden Autos.

1– 2 – 3 – 4 – 5 – 6 – 7

2. Selbstfahrende Autos sind zuverlässig.

1– 2 – 3 – 4 – 5 – 6 – 7

3. Ich würde meinem Auto selstfahrende Funktionen wie Spurführung, Spurwechsel, Warnungen zur Objekterkennung, etc. überlassen.

1– 2 – 3 – 4 – 5 – 6 – 7

4. Ich kann auf selbstfahrende Autos zählen.

1– 2 – 3 – 4 – 5 – 6 – 7

5. Selbstfahrende Autos können schädliche Folgen haben.

1 – 2 – 3 – 4 – 5 – 6 – 7

6. Ich vertraue selbstfahrenden Autos.

1– 2 – 3 – 4 – 5 – 6 – 7

7. Ich nehme an, dass selbstfahrende Autos ordnungsgemäß funktionieren.

1 – 2 – 3 – 4 – 5 – 6 – 7

Kontakt

Bitte geben Sie Ihre E-Mail-Adresse an. Sie werden kontaktiert, um einen Termin für die Simulatorstudie zu vereinbaren, sollte Ihr Vertrauenslevel für das Experiment geeignet sein.

____

Vielen Dank!

Dankeschön für Ihre Teilnahme.

Appendix B: Post-test questionnaire English and German

Post-Test Questionnaire

Name (to match your trust scores with further data collection - you may also choose a nickname or number):

____

1. I was cautious about the self-driving car

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

2. The self-driving car was reliable

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

3. I would entrust my car to the tested self-driving functions (for example, lane-keeping)

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

4. I could count on the self-driving car

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

5. This self-driving car can have harmful consequences

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

6. I trusted the self-driving car

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

7. The self-driving car worked properly

Not at all 1– 2 – 3 – 4 – 5 – 6 – 7 extremely

Have you noticed any speed changes during the driving simulation?

- o Yes
- o No

Do you have any additional comments?

Why did you move the slider during the experiment?

Please enter your Sona participant-ID if applicable to receive the credit:

Post-Test Fragebogen

um das Vertrauenslevel den weiteren Daten zuzuordnen - es kann auch ein Spitzname oder eine Nummer gewählt werden: ______________

1. Ich war vorsichtig mit dem selbstfahrenden Auto.

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

2. Das selbstfahrende Auto war zuverlässig.

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

3. Ich würde meinem Auto die getesteten selbstfahrenden Funktionen überlassen (z.B. Spurführung).

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

4. Ich konnte auf das selbstfahrende Auto zählen.

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

5. Das selbstfahrende Auto kann schädliche Folgen haben.

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

6. Ich habe dem selbstfahrenden Auto vertraut.

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

7. Das selbstfahrende Auto hat ordnungsgemäß funktioniert.

Überhaupt nicht 1– 2 – 3 – 4 – 5 – 6 – 7 Extrem

Haben Sie Geschwindigkeitsveränderungen während der Fahrsimulation bemerkt?

- o Ja
- o Nein

Haben Sie zusätzliche Kommentare?

Warum haben Sie den Schieberegler während des Experiments bewegt?

Bitte füllen Sie Ihre SONA-ID aus um die study credits zu erhalten, falls dies für Sie zutrifft.

Appendix C: Script used to extract speed changes per algorithm

```
library(tidyverse)
########## Helper functions ############
#' @param v Vector of values
#' @return most common element in `v`
get_mode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

#' prepare a raw data set for further analyses
#'
#' removes "Alert" rows, renames the "Trust Data" column as
"trust" and coerces
#' time since start column (Duration) into a numeric value
(time) while removing
#' all other columns.
#'
#' @param raw_data 'raw' data as read from the csv file, e.g.
`readr::read_csv("filename.csv")`
#' @return tibble with columns `trust` and `time`.
prep_data <- function(raw_data) {
  data <- raw_data %>%

    # filter out 'alert' rows
    filter(Time != "Alert Occured") %>%

    # create new columns, and drop any columns that aren't
mentioned
    transmute(
      # rename trust column
      trust = `Trust Data`,

      # remove 's' from Duration and coerce it to numeric
      time = Duration %>% stringr::str_remove_all("s") %>%
as.numeric()
    )

  # return the new data
  data
}

#' get the number of significant trust changes for a data set,
for a given group
#' size and threshold.
```

```
#'
#' @param data prepared data as returned from `prep_data`
#' @param group_size size of 'groups' to calculate mode for
(defaults to 10)
#' @param threshold threshold value for what is considered a
significant change,
#'     calculated as `threshold * mode(t-1)`
#' @return number of significant change points
get_change_points <- function(data,
                              group_size = 10,
                              threshold = 0.1) {

  change_data <- data %>%
    # assign 'group' numbers using some clever integer
division
    mutate(group = 0:(n() - 1) %/% group_size) %>%

    # tell the tidy functions to great each 'group' as a
separate batch
    group_by(group) %>%

    # calculates mode per group/batch, and returns a single
row for each group
    summarize(mode = get_mode(trust)) %>%

    # determine if each row is significantly different from
the previous
    mutate(change = abs(mode - lag(mode)) > (threshold *
lag(mode)))

  # return the sum of the `change` column we obtained above.
This works because
  sum(change_data$change, na.rm = TRUE)
}

#' get the mean trust scores for a data set at given time(s)
#' returns the mean trust score for the `n` measurements
before `time`(s).
#'
#' @param data prepared data as returned from `prep_data`
#' @param times time or vector of times in seconds
#'     to get mean scores for.
#' @param n number of measurements to use (defaults to 200).
get_trust_scores <- function(data, times, n = 200) {

  # create an empty vector to hold scores
```

```
  scores <- numeric()

  # loop over the times, calculate score for each time.
  # note: using `.time` to distinguish from the `time` column
in the data.
  for (.time in times) {
    .score <- data %>%
      # use only rows that occured before `.time`
      filter(time < .time) %>%

      # use the last n rows, these will be the most recent
      tail(n) %>%

      # pull out the 'trust' column
      pull(trust) %>%

      # take the mean of the trust values, ignoring any
missing values
      mean(na.rm = TRUE)

    # attach the score for each time on the end of the scores
vector
    scores <- c(scores, .score)
  }

  # return the scores vector
  scores
}

# Preprocess the data
data <- prep_data(raw_data)

# Obtain number of change points
# Just manually call the function for each group size; example
change_points_N10 <- get_change_points(data, 10)
change_points_N25 <- get_change_points(data, 25)

# Get average trust score at given times example
trust_scores <- get_trust_scores(data, c(15, 30, 45), 200)
```

Appendix D: Script used for Analysis 1 in SPSS

```
#Descriptive Statistics and identifying outliers and normality
DATASET ACTIVATE DataSet1.

EXAMINE VARIABLES=M1 M2 M3 M4 M5 M6 M7 M8
  /PLOT BOXPLOT STEMLEAF NPPLOT
  /COMPARE VARIABLES
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

#One-way repeated measures ANOVA and bonferroni post-hoc test
+ Mauchlys test for sphericity

GLM M1 M2 M3 M4 M5 M6 M7 M8
  /WSFACTOR=Methods 8 Polynomial
  /MEASURE=Changes_detected
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(Methods) TYPE=LINE ERRORBAR=NO
MEANREFERENCE=NO YAXIS=AUTO
  /EMMEANS=TABLES(Methods) COMPARE ADJ(BONFERRONI)
  /PRINT=DESCRIPTIVE ETASQ
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=Methods.
```

Appendix E: Code of how algorithm 8 was implemented and how speed changes are induced

```
  //This part checks if there is a significant trust change in order to call the speed change
method
    private void checkTrust()
    {
        // Debug.Log("trust at first " + TrustData + "& prevVal at first" + prevVal);
        if (prevVal == 0 || TrustData == 0 || TrustData > slider_max || TrustData < slider_min)
//this is to ignore any hardware noise.                    Noise filtering
        {
            prevVal = TrustData;

            firstTimerValue = TrustData;

            Debug.Log("trust " + TrustData + "& prevVal " + prevVal);
        }
        else
        {

            if (TrustData != prevVal)

 // calculate the margin & change in trust, if change in trust > margin then we call
speedChange()

                if (!flag)

                {
                    margin = percent_change_inTrust * firstTimerValue;
is trust change significant?

                }

                float changeinTrust = Math.Abs(TrustData - firstTimerValue);

                if (Time.time != 0)
                {

                    if (changeinTrust >= margin)
                    {

                        NewprevVal = prevVal; // NewprevVal holds the previous value of prevVal

                        flag = false;
                        functionHasBeenCalled = false; // Ensuring that Fifo stops when theres a speed
                        change call.
                    }

                    else
```

```
                {
                    flag = false;
                }
            }
            else prevVal = TrustData;

            if (!flag)
            {
                prevVal = TrustData;

            }
        }
    }
}
/*

 * This function is responsible for changing the speed
 */
private void speedChange()
{

    if (TrustData > firstTimerValue) //if there is a positive significant trust change speed
increases
    {
        currentSpeed = currentSpeed + speed_step;
        outputModule.speedChange();
        Debug.Log("Speed Increased");
        Invoke("clearList", 0.2f); //calls the clearList()
        firstTimerValue = TrustData;
        Invoke("removeFirstEntry", FirstTimerDuration);
    }
    Else //if there is a negative significant trust change speed decreases
    {
            currentSpeed = currentSpeed - speed_step;
            outputModule.speedChange();
            Debug.Log("Speed Decreased");
            Invoke("clearList", 0.2f);//calls the clearList()
            firstTimerValue = TrustData;
            Invoke("removeFirstEntry", FirstTimerDuration);
    }
    if (currentSpeed <= 5f) //if the speed is at 5, it will not decrease more
    {
        currentSpeed = 5;
    }
    if (currentSpeed >= 40.0f) //if the speed is at 40 it will not increase more
    {
        currentSpeed = 40;
    }
```

```
    }
  /* This method is called from update, it reads all the trust values from the slider,
   * Checks if it is in the margin or not, then calls the CheckTrust methods & calls the FiFo
method.
   *
   */
    private void ReadTrustValue()
    {
        TrustDataRaw = float.Parse(testvalue); //the value from arduino in float

        TrustDataRawList.Add(TrustDataRaw); //adds the raw trustdata in this list

        {
            TrustDataRawList.Clear(); //clears the list
        }
        if (!FixedSpeed)
        {
            oneInoneOut();
        }
        if (!FixedSpeed)
        {
            {
                anotherMargin = percent_change_inTrust * firstTimerValue;
                float anotherChnageinTrust = Math.Abs(item - firstTimerValue);

                if (anotherChnageinTrust >= anotherMargin)
                {
                    checkTrust();
                }
            }
        }
    }
    /*
   * This method starts removing first entry from a list after firsttimerduration seconds, Fifo
   * for every element that is added to the TrustDataMost list.
   *
   * Concept of one in one out (FiFo)
   */
    private void oneInoneOut()
    {
        if (functionHasBeenCalled == true)
        {
            removeFirstEntry();
        }
    }

    /*
     *This methods clears the list. It is called after logging an alert of significant change,
meaning the list is cleared after significant change was detected
     */
    private void clearList()
```

```
    {
      TrustDataMostList.Clear();
      functionHasBeenCalled = false;
        Debug.Log("cleared");
      Debug.Log("first timer value after clear " + firstTimerValue);
    }
    /*This method removes the first element in a list called by Fifo (first element is removed)
     * and if there is a significant change holds the first element of that time

    private void removeFirstEntry()
    {
       Debug.Log("removing first element");
      TrustDataMostList.RemoveAt(0);
       timerStartValue = TrustDataMostList[0];

     if(timerStartValue == 0) {
          for (int i = 1; timerStartValue > 0; i++)
          {
            timerStartValue = TrustDataMostList[i];
           }
      }
      if (Time.time != 0 )
      { //if there is a significant change hold the first entry and a new list is started
         if (timerStartValue != 0) {
            if (TrustData > timerStartValue || TrustData < timerStartValue) //if turst > then first
value or < then
            {
               firstTimerValue = timerStartValue; //the firstTimerValue is the timerStartValue

            }
         }
      }
      functionHasBeenCalled = true;
    }
}
```

Appendix F: Script used for Analysis 2 in SPSS

```
#Descriptives

EXAMINE VARIABLES=Pre Post BY Condition
  /PLOT BOXPLOT NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

#Two-way mixed ANOVA with Bonferroni + outlier identification
with residuals + levenes test of equality of error variances

GLM Pre Post BY Condition
  /WSFACTOR=time 2 Polynomial
  /MEASURE=trust
  /METHOD=SSTYPE(3)
  /SAVE=SRESID
  /PLOT=PROFILE(time*Condition) TYPE=LINE ERRORBAR=NO
MEANREFERENCE=NO YAXIS=AUTO
  /EMMEANS=TABLES(Condition) COMPARE ADJ(BONFERRONI)
  /EMMEANS=TABLES(time) COMPARE ADJ(BONFERRONI)
  /EMMEANS=TABLES(Condition*time)
  /PRINT=DESCRIPTIVE ETASQ HOMOGENEITY
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=time
  /DESIGN=Condition.

#checking for normality using residuals

PPLOT
  /VARIABLES=SRE_1 SRE_2
  /NOLOG
  /NOSTANDARDIZE
  /TYPE=Q-Q
  /FRACTION=BLOM
  /TIES=MEAN
  /DIST=NORMAL.
```

Appendix G: Script used for Analysis 3 in SPSS

```
#Descriptive Statistics

DATASET ACTIVATE DataSet1.
EXAMINE VARIABLES=Average_Trust BY Condition
  /PLOT BOXPLOT NPPLOT
  /COMPARE GROUPS
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

#T-test

T-TEST GROUPS=Condition('Adaptive' 'Fixed')
  /MISSING=ANALYSIS
  /VARIABLES=Average_Trust
  /ES DISPLAY(TRUE)
  /CRITERIA=CI(.95).
```