This is a pre-print version of

Körber, M., Prasch, L., & Bengler, K. (in press). Why do I have to drive now? Post hoc explanations of take-over requests. *Human Factors: The Journal of the Human Factors and Ergonomics Society*

Why do I have to drive now? Post Hoc Explanations of Take-Over Requests

Moritz Körber[a], Lorenz Prasch[a], & Klaus Bengler[a]

[a]Chair of Ergonomics, Technical University of Munich, Munich, Germany

Correspondence concerning this article should be addressed to Moritz Körber, Chair of Ergonomics, Technical University of Munich, Boltzmannstraße 15, D – 85747 Garching, Tel: +49 89 289 15376, Email: koerber@lfe.mw.tum.de

Manuscript category: Extended Multi-Phase Study

Exact word count (not including title page, abstract, tables, biographies, references, key points): 6274

Acknowledgements: The authors thank Joachim Vandekerckhove and Torrin M. Liddell for their statistical advice. We also thank Jonas Schmidtler for his comments on this manuscript.

Abstract

Objective: It was investigated whether providing an explanation for a take-over request in automated driving influences trust in automation and acceptance.

Background: Take-over requests will be recurring events in conditionally automated driving which could undermine trust as well as acceptance and, therefore, the successful introduction of automated vehicles.

Method: Forty participants were equally assigned to either an experimental group provided with an explanation of the reason for a take-over request or a control group without explanations. In a simulator drive, both groups experienced three take-over scenarios that varied in the obviousness of their causation. Participants rated their acceptance before and after the drive and rated their trust before and after each take-over situation.

Results: All participants rated acceptance on the same high level before and after the drive, independent of the condition. Control group's trust ratings remained unchanged by take-over requests in all situations, but the experimental group showed decreased trust after experiencing a take-over caused by roadworks. Participants provided with explanation felt stronger that they had understood the system and the reasons for the take-overs.

Conclusion: A take-over request did not lower trust or acceptance. Providing an explanation for a take-over request had no impact on trust or acceptance, but increased the perceived understanding of the system.

Application: The results provide insights into users' perception of automated vehicles, take-over situations and a fundament for future interface design for automated vehicles.

Keywords: Technology acceptance, Trust in automation, Human-automation interaction, Automated driving, Take-over request

Précis: In this study, we investigated the effect of an explanation of the reason for a take-over request on trust and acceptance of driving automation. An experimental group provided with explanations and a control group given no explanations experienced three TORs that varied in the obviousness of the reason for the take-over.

Introduction

Advances in passive and active safety technologies have led to a remarkable increase in traffic efficiency and safety (Kühn & Hannawald, 2016). Automated vehicles are currently being introduced to the consumer market, with the intention to provide an even higher standard (Watzenig & Horn, 2017). However, societal goals do not necessarily coincide with a driver's personal goals (Adell, Várhelyi, & Nilsson, 2014b). Consequently, previous research that accompanied the introduction of Advanced Driver Assistance Systems (ADAS) has shown that to guarantee a successful introduction of a new technology it is necessary to evaluate its deployment not only from a technological perspective but also from a driver-centered perspective (Bengler et al., 2014; Regan, Horberry, & Stevens, 2014). Whereas excellent system performance may be sufficient from a technical point of view, a system's functionality must be known, understood, believed in, and valued by the driver in order for it to be accepted and used (Adell et al., 2014b; van der Laan, Heino, & de Waard, 1997). An unsystematic introduction without a driver-centric approach may give rise to issues such as information overload, over-reliance, or negative behavioral adaptation to the technology (Broughton & Baughan, 2002; Mahr & Müller, 2011; Parasuraman & Riley, 1997). This can lead to low acceptance or even disuse of the new system after its introduction despite all the possible benefits (Lee & Seppelt, 2012).

Acceptance represents a multidimensional attitude that results out of the fulfillment of the user's individual needs and requirements. It consists of an affective as well as a rational-cognitive (e.g., perceived usefulness) component and is an antecedent of the intention to buy and to use a system (Adell et al., 2014b; van der Laan et al., 1997; Schade & Baum, 2007). We define acceptance as an attitude and follow Adell's (2009) definition of acceptance as "the degree to which an individual intends to use a system and, when available, to incorporate the system in his/her driving" (p. 31). Acceptance is closely related to actual usage of a system because, as described in the Theory of Planned Behavior (Ajzen, 1991), attitudes influence the intention to use a system and, thereby, actual behavior. Based on this theory, the Technology Acceptance Model (TAM; Venkatesh, Morris, Davis, & Davis, 2003) has successfully explained the adoption of driver assistance systems or automated vehicles in several studies (Choi & Ji, 2015; Ghazizadeh, Peng, Lee, & Boyle, 2012; Meschtscherjakov, Wilfinger, Scherndl, & Tscheligi, 2009).

The introduction of driving automation will only generate the claimed benefits if the technology is accepted by the drivers and used appropriately (Najm, Stearns, Howarth, Koopmann, & Hitz, 2006). Contrary to manual driving, in conditionally automated driving (Level 3 in SAE, 2016), the driver is removed from the driving task and a driving automation operates the vehicle. The driver merely acts as a fallback level and has to take over vehicle control at system limits. This concept of vehicle control represents a novelty for the majority of the driving population, which is why acceptance is not guaranteed and has to be investigated (Payre, Cestac, & Delhomme, 2014).

Trust as a necessary precondition of acceptance

Given the close relationship between trust in automation and reliance on it (Bailey & Scerbo, 2007; Körber, Baseler, & Bengler, 2018), it seems reasonable to include trust in an acceptance framework. Indeed, previous research has shown that trust is a key determinant for the adoption of new technologies (Gefen, Karahanna, & Straub, 2003), the adoption of automation (Lee & Moray, 1992, 1994; Parasuraman & Riley, 1997), and the intention to use autonomous vehicles (Choi & Ji, 2015). The incremental value of investigating trust in studies on acceptance has been successfully shown by several studies such as on an on-board monitoring system (Ghazizadeh, Peng et al., 2012), on ADAS (Trübswetter & Bengler, 2013) and on the reliance on and intention to use automated vehicles (Choi & Ji, 2015). Consequently, trust in automation as a determinant of acceptance of automation has been included in in Arndt's model of acceptance of ADAS (2011) and in the Automation Acceptance Model (AAM) of Ghazizadeh, Lee, and Boyle (2012). In the AAM, trust partially mediates the effect of the operator's beliefs and external variables on perceived usefulness and perceived ease of use, but also has a direct effect on the behavioral intention to use an automation. Hence, trust in automation is a necessary condition that has to be fulfilled before acceptance may arise. Put simply, "operators tend to use automation that they trust while rejecting automation that they do not" (Pop, Shrewsbury, & Durso, 2015, p. 1). Therefore, it is necessary to include an assessment of trust in automation in a study on acceptance of automation.

Increasing trust and acceptance by providing explanations

Operator and automation are not isolated entities but act as a joint system, i.e. as a team (Bengler, Zimmermann, Bortot, Kienle, & Damböck, 2012). Therefore, a driving automation cannot be considered in isolation from its users and must be designed following a human-centered approach to perform in conjunction with the human interacting with it (Billings, 1997; Christofferson & Woods, 2002; Sheridan & Parasuraman, 2005). In comparison to ADAS, a driving automation represents a more sophisticated automated system, an increase in autonomy and authority (Parasuraman, Sheridan, & Wickens, 2000). While a status icon alone may be sufficient for a less complex function such as a lane departure warning system, it may no longer be sufficient to support effective coordination with more complex machine agents like a driving automation, which require more coordination (Norman, 1990; Sarter, 2008). Coordination needs an adequate model of the automation's intentions and actions. In order to design automated systems as "cooperative partners rather than as mysterious and obstinate black boxes" (Christofferson & Woods, 2002, p. 4), they should act neither capriciously nor unobservably (Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004; Lee & Seppelt, 2009).

However, feedback alone is not enough; the interactions have to be as comprehensible for the driver as possible to create a common ground and, thereby, to ensure the construction of a correct mental model (Clark & Brennan, 1991). Drivers of automated vehicles will not be experts but laypersons who do not possess complete in-depth knowledge of the automation and must at first build themselves a mental model of its functioning (Walker, Stanton, & Salmon, 2016). A user generally builds his mental model based on the information provided by the system or interactions with it (Naujoks & Totzke, 2014). Hence, to ensure trust in driving automation, it is crucial to provide the driver with obvious and comprehensible information on its intentions, state, capacity, and upcoming actions to help them to understand and make it predictable. Otherwise, the increase in autonomy and authority creates an intransparent black box where users cannot comprehend or retrace the actions (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Verberne, Ham, & Midden, 2012).

Automation failures result in a drop of trust in the automated system (Lee & See, 2004), however, as Lewandowsky, Mundy, and Tan (2000) concluded, this drop represents more than a simple perception of whether an automation failure occurred since the failure's impact depends on its predictability rather than on its magnitude. A drop in trust in ADAS only follows if problems were omitted in a description of the system given beforehand (Beggiato & Krems, 2013) or if the failures were inconsistent with the perceived design of the system or occurred unpredictably (Lees & Lee, 2007). The attitude toward an automated system is, therefore, not purely based on performance (Lewandowsky et al., 2000). Even if the system exhibits high performance, a discrepancy between the operator's expectations and the system's behavior, i.e. a large gulf of evaluation (Norman, 2013), can erode trust (Lee & See, 2004). If operators had prior knowledge of the magnitude of the failure (Riley, 1996) or if the failure was predictable or if its cause was comprehensible (Dzindolet et al., 2003), a decrease in trust did not occur. Accordingly, Gold et al. (2015), as well as Hergeth et al. (2015), observed a slight increase in trust after the experience of a take-over request (TOR) since the automation worked as described beforehand. Dimensions such a predictability, understanding or transparency have been proposed as a basis for trust in automation (Hoff & Bashir, 2015; Lee & See, 2004), which has been empirically shown in several studies (Choi & Ji, 2015; Muir & Moray, 1996; Seong & Bisantz, 2008). For example, Beller, Heesen, and Vollrath (2013) presented the uncertainty of an automation in an interface which led to better knowledge of fallibility and, in consequence, to higher trust ratings and increased acceptance. Users rated an adaptive cruise control system that took over the driving task as more trustworthy and acceptable when it provided information on this action (Verberne et al., 2012). Forster, Naujoks, and Neukum (2017) found that the provision of auditory explanations of the automation's actions led to higher reported trust.

Besides these aforementioned cognitive aspects, Adell, Várhelyi, and Nilsson (2014a) suggested investigating the emotional reactions of the driver such as irritation or stress in research on user acceptance. Beaudry and Pinsonneault (2010) already showed that anxiety is negatively related to the use of information technology. Individuals tend to search for or create explanations for unpleasant events afterward if no immediate reason can be deduced from the environment or prior knowledge, referred to as retrospective control (Thompson, 1981). Since unexpected TORs are rather stressful situations (Maule & Hockey, 2012), providing an explanation after the TOR might alleviate the negative affective reaction and promote a feeling of control. Accordingly, Koo et al. (2015) reported that providing information yielding reasons for the actions of an auto-brake function created the least anxiety, highest trust and was preferred by the drivers. Hence, avoiding negative emotions is essential in guaranteeing user acceptance.

In this study, we explicitly focus on take-over situations. We investigate if providing an explanation for a take-over request increases system transparency and understanding and, in doing so, also increases trust in automation as well as acceptance of the automation. We expect that an explanation should avoid a decrease in trust and acceptance when a take-over situation occurs because it guarantees the construction of an appropriate mental model by helping to bridge the gulf of evaluation (Norman, 2013), enabling a driver to learn when a take-over situation is to be expected and how to react appropriately (Larsson, Kircher, & Andersson Hultgren, 2014). The created predictability and comprehension of the situation should mitigate the negative impact of a TOR on trust (Riley, 1996). An explanation also helps to avoid automation surprises (Sarter, Woods, & Billings, 1997) and negative emotional reactions, caused by unexpected situations, which are known to reduce acceptance.

Depending on the situation, providing information can, however, also be counterproductive. Whereas an explanation beforehand is often not possible due to technological limits (e.g., sensor range; Gold & Bengler, 2014), a presentation simultaneous with the TOR might overload information processing capacity and may result in a delayed reaction (Walch, Lange, Baumann, & Weber, 2015; Wickens, 2002). Besides a possible objective detrimental effect, subjective ratings of real-time feedback appear to be more negative as well. Koo et al. (2015) reported that the participants felt subjectively overstrained if too much information was presented during the automatic brake maneuver. Similarly, Roberts, Ghazizadeh, and Lee (2012) compared the acceptance of real-time with post-drive driving performance feedback. Drivers rated real-time feedback as more obtrusive, less useful and less easy to use. To provide the explanation without a loss in performance and appraisal but still linked to the situation at hand, we suggest presenting the explanation directly after regaining vehicle control and stabilizing the vehicle, i.e. after the situation was solved and when workload is at a sufficiently low level. To increase the generalizability of the results, we investigate the provision of explanations in situations with varying obviousness.

Pre-study

We conducted an online pre-study to evaluate if the chosen take-over situations were comprehensible and whether they differ in their obviousness of the reason for the take-over. In this survey, a total of 36 participants, 20 (55 %) male, 16 (45 %) female, between the ages of 18 and 51 (M = 25.60, SD = 6.30), watched videos of three different take-over scenarios (duration between 14 and 29 seconds, filmed in ego-perspective). The three scenarios, which we expected to vary in their obviousness, have been (a) GPS data missing (*GPS*; low obviousness), (b) Missing lane markings (*Missing lines*; medium obviousness), and (c) *Roadworks* (high obviousness). The videos were presented in a resolution of 680 × 400 pixels. The TOR signal was a sharp sinusoidal tone (3000 Hz) and a blinking hands-on icon (Appendix D) and was presented nine seconds prior to a theoretical take-over. After every video participants answered the following three questions on a five-point rating scale from *not at all* (1) to *very much* (5):

- "I think this TOR was a system failure."
- "It is obvious to me, why the TOR was triggered."
- "I would have wished for an explanation, why this TOR was triggered."

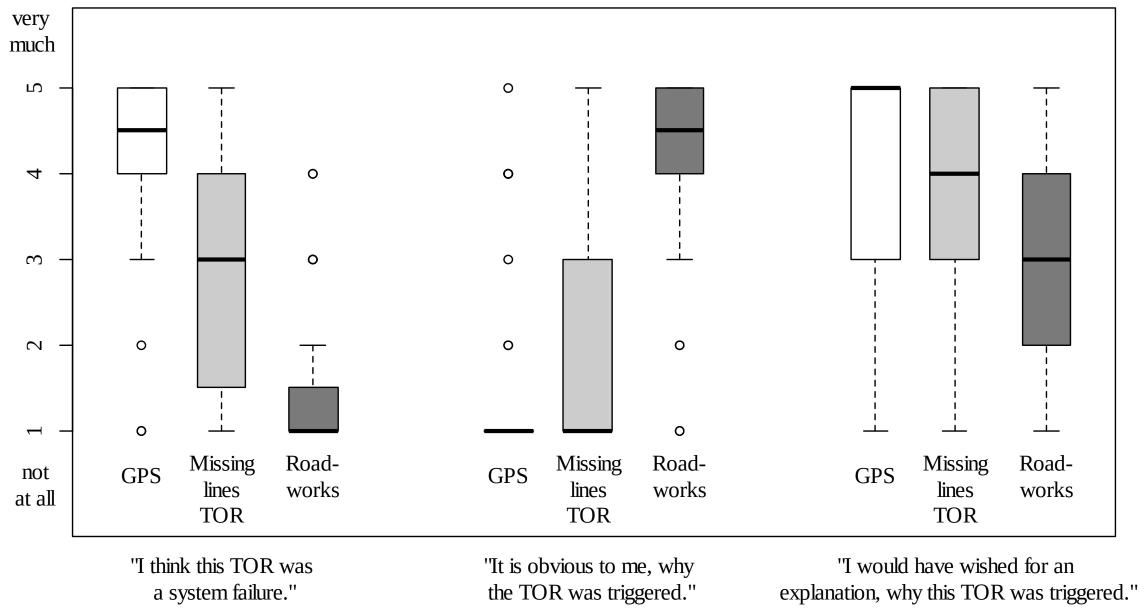

Figure 1: Reported answers on the videos by question and TOR.

The results, illustrated in Figure 1, show that the scenarios tend to differ for all three questions. In addition, the participants could elaborate as to why they thought the TOR was triggered. No participant could name the correct reason for the system limit for *GPS*, 35 % answered correctly for *Missing lines*, and 78 % could name the correct reason in *Roadworks*. The results of this pre-study are described in further detail in Prasch and Tretter (2016).

Main Study

Experimental Design and Scenarios

In the main study, we used a 2 × 3 mixed design. The factor *Explanation* (between-subjects) consisted of a control (*Control*) and an experimental group (*Explanation*). We assigned the participants equally and randomly to both groups. The *Explanation* group was provided with an explanation of the reason for the TOR after each take-over situation. This explanation was absent in the control group. The explanations conveyed the external reasons for the TOR as well as the internal implications for the system (Koo et al., 2015; Lombrozo, 2006). Every explanation had the same structure and wording with the only difference being the respective cause and effect: „*The take-over request was triggered because of [cause]. Due to [effect], driving in highly automated mode can temporarily not be continued.*" They were recorded by a female voice actor in a natural manner and friendly tone as recommended by Broadbent, Stafford, and MacDonald (2009). The explanations were presented on the mock-up speaker system at 68 dB 14 seconds after the presentation of the TOR. At the same time as the audio, a flashing icon was displayed in the head-up display (HUD) indicating the presence of an explanation. The participants of both groups carried out a non-driving-related task (NDRT), the Surrogate Reference Task (ISO 14198:2012, 2012), while driving in conditional automated mode (Level 3; SAE International, 2016). The factor *Scenario* (within-subjects) represented three take-over scenarios that each participant experienced in the course of the experimental drive (Figure 2 to 4): (a) *GPS*, (b) *Missing lines*, and (c) *Roadworks*. The scenarios were chosen to correspond to realistic take-over situations in automated driving (Aeberhard et al., 2015) and varied in their obviousness of the reason of the take-over – as tested in the pre-study. The scenario *GPS* represented a TOR caused by missing GPS data. Conditionally highly automated driving requires highly precise map data that is not available for every section of highways yet (Aeberhard et al., 2015). If this data is missing for the current section of the road, a TOR is emitted. In this scenario, no visible cue for the reason of the take-over was present. The scenario

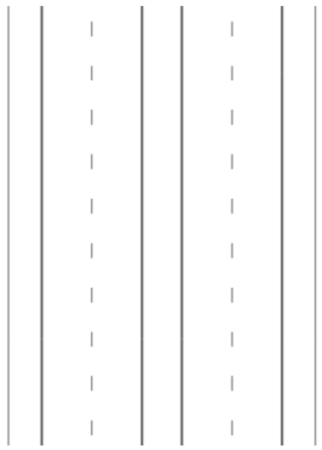

Figure 2: Schematic visualization of Scenario GPS.

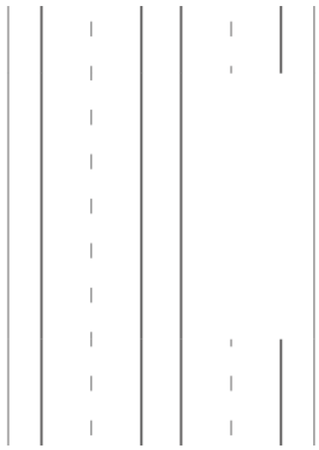

Figure 3: Schematic visualization of Scenario Missing lines.

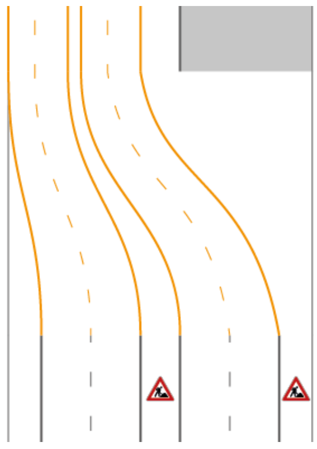

Figure 4: Schematic visualization of Scenario Roadworks.

Missing lines represented a highway section where the right lane markings were missing (Figure 3). Without lane markings, it is impossible for the vehicle to detect its exact position on the lane and a TOR has to be emitted. This scenario contained a visible cue for the reason of the take over in form of the missing lane markings. The scenario *Roadworks* (Figure 4) represented roadworks on the participant's lane, which required bypassing on an alternative lane. In such an unpredictable situation and without map data, conditionally automated driving becomes unavailable and a TOR is emitted. In this scenario, the reason for the TOR (roadworks) was directly visible to the driver. Every scenario was exactly 1000 meters (30 s at a speed of 120 km/h) long and started with a TOR nine seconds before the irregularity in the environment/the cue for the reason of the TOR (disappearing lane markings or yellow, swerving lanes in the roadworks scenario) appeared. This time budget corresponds to the time taken for a non-critical take-over process for the great majority of participants (Eriksson & Stanton, 2017). It was thereby ensured that all situations were experienced as non-critical to avoid a confounding influence of criticality. No other traffic was present during the TOR. After every scenario, the automation became available again, which was indicated by an icon in the instrument cluster. The order of the scenarios was permutated using a Latin square. In each situation, the NDRT was presented three times for 60 seconds, while the first presentation was interrupted by the TOR. In addition, to reduce the predictability of the TOR, the driving time prior to the TOR (ranging from 2.50 to 7.50 minutes) was manipulated by implementing up to two NDRT phases (Figure 5), also permutated according to Latin square.

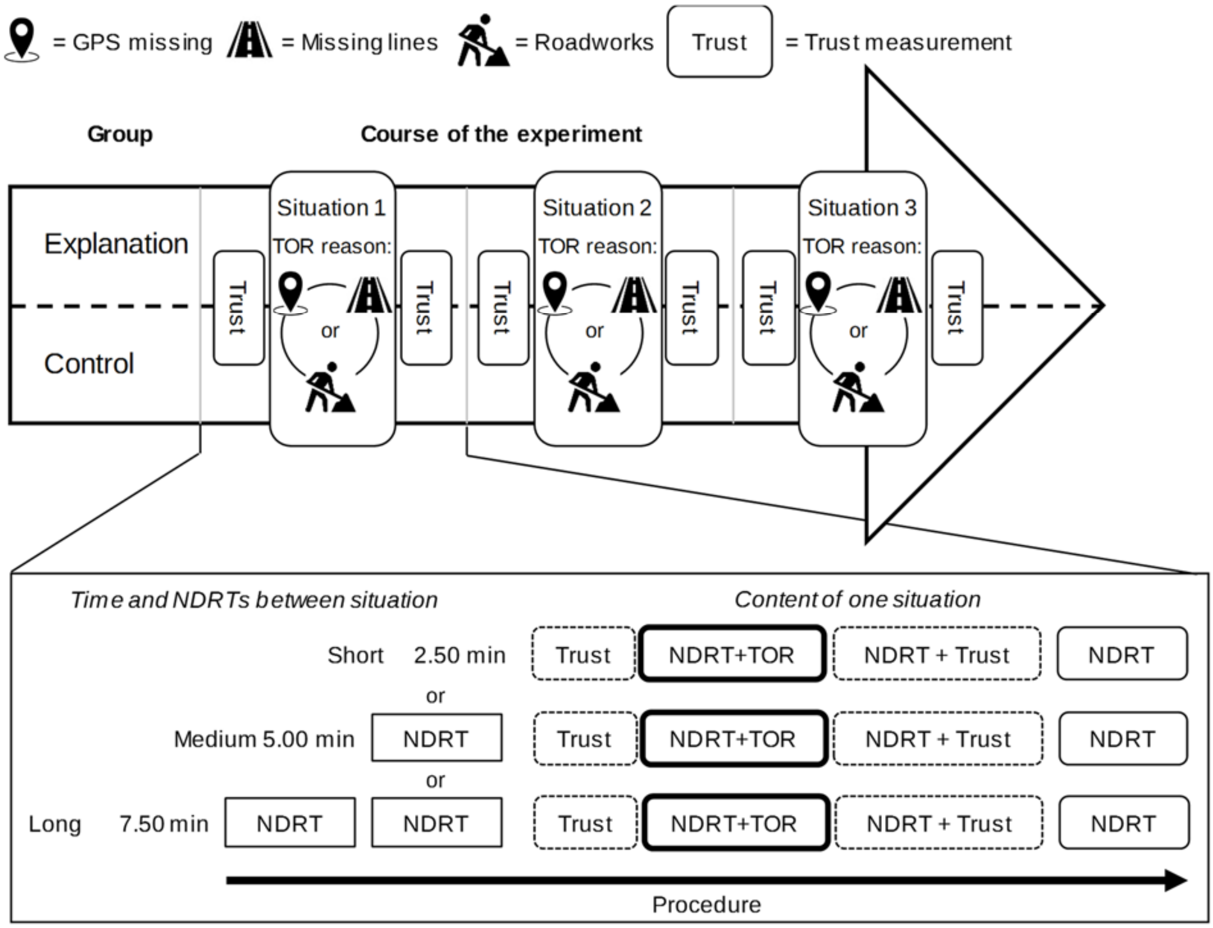

Figure 5: Procedure of the experiment.

Hence, a participant encountered three situations that each included one of the scenarios (*GPS*, *Missing lines*, *Roadworks*) with the time between the scenarios being either short/medium/long. Trust was measured before and after each of the three TORs. The design and procedure of this study was critically evaluated by the institute's interdisciplinary internal ethical review entity.

Instructions and experimental track

The experiment was carried out in a driving simulator on a two-lane freeway. Prior to the experimental drive, the participants received a written introduction to the automation which explained the functionality of the automation and its interface (e.g., icons). The participants were instructed that the automation is capable of executing lateral as well as longitudinal control without the need to monitor it. The automation can not solve every situation and the driver is requested to take-over vehicle control within a sufficient time budget in this case. The participants then performed an approximately five-minute familiarization drive. In this drive, they were prompted to steer and to brake manually, to turn on the automation and to observe the automation carrying out vehicle control. They also engaged in the

NDRT and experienced a single TOR. The drive came to an end when the participants indicated that they felt comfortable using the driving simulator. The following experimental drive was a single drive of approximately 30 minutes and contained three TORs in the aforementioned scenarios. Following previous studies (Gold, Körber, Lechner, & Bengler, 2016; Körber, Gold, Lechner, & Bengler, 2016), each TOR was represented by a blinking hands-on icon in the HUD and a sharp double earcon (3000 Hz at 74 dB) via the mockup speaker system with a time budget of nine seconds.

Sample

A total of $n = 40$ participants, 20 (50 %) female and 20 (50 %) male, took part in the study. The participants were between the ages of 21 and 30 ($M = 25.20$ years, $SD = 2.60$). All of them were students or employees at the Technical University of Munich. Possession of a valid driver's license was required for participation (mean duration of possession $M = 7.40$ years, $SD = 2.30$). Participants completed an informed consent form and acknowledged their voluntary participation and consent with a signature. Twenty-four (60 %) participants had already taken part in at least one driving simulator study. Annual mileage and acquaintance with automated driving are shown in Table 1. No participant reported an impairment relevant for driving. Participation was rewarded with candies. The three participants with the best performance in the NDRT were rewarded with vouchers for an online store worth 20, 30 and 50 Euros.

	Annual mileage in kilometers				Acquaintance with automated driving
	< 5,000	5,001–20,000	20,001–50,000	> 50,001	Median
Control	10	7	2	1	3
Explanation	8	7	5	0	2
Total	18	14	7	1	2.5

Table 1: Participant's annual mileage and reported acquaintance with automated driving (on a rating scale from 1 (lowest) to 5 (highest)).

Apparatus and Measures

Driving Simulator and Driving automation

The study was conducted in a static driving simulator equipped with a BMW 6-Series mock-up. Seven projectors provided a front view of approximately 180 °, side and rear mirrors, and a mockup of a HUD. The implemented driving automation performed on SAE Level 3, conditional automation (SAE International, 2016). The participants were asked to attend to the NDRT whenever it was present. The automation could be toggled via a button on the steering wheel and was also shut off by steering or braking input. The participants were instructed to switch on the automation whenever it was available. Its status was displayed via an icon in the top center of the instrument cluster.

Non-driving-related Task (NDRT)

While driving, participants had to perform an NDRT, the Surrogate Reference Task (SuRT; ISO 14198, 2012), which is a visual-manual-demanding task that simulates real life situations in which drivers are strongly engaged in an NDRT during conditional automated driving. In this task, the participants were presented a scatter of 50 white circles (diameter 40 px) in 18 columns and 15 rows on a black background. A single, larger circle (diameter 47 px) randomly implemented in this scatter represented the target stimulus. The participants' task was to find that larger circle and to highlight the respective column out of a total of six selectable columns. The task was presented for 60 seconds every 2.50 minutes on a 14 " Lenovo ThinkVision monitor at a resolution of 1366 × 768 pixels mounted on the center console and operated via an external numeric keypad next to the gear lever. To increase their motivation, participants were informed that their performance was being tracked and the best three participants would be rewarded with vouchers.

Acceptance Questionnaire

Following previous studies on the acceptance of ADAS (Adell, Várhelyi, & Hjälmdahl, 2008; Törnros, Nilsson, Östlund, & Kircher, 2002), we measured acceptance of the driving automation using a questionnaire by van der Laan et al. (1997). It represents a semantic differential consisting of two scales, *usefulness* and *satisfaction*, each containing nine bipolar items (e.g., useful–useless) that are rated on five-point rating scales from −2 to 2. The questionnaire was presented before and after the experimental drive via GoogleForms.

Trust Questionnaire

Trust in automation was measured with a single item, which has been shown as valid in previous studies (Beller et al., 2013; Brown & Galster, 2004; Hergeth, Lorenz, Vilimek, & Krems, 2016). The participants were prompted via an intercom system to rate their trust on a scale from 0 to 100 ("On a scale from 0 to 100, how much do you trust the system?") after each engagement in the NDRT. We analyzed only the trust ratings reported directly before and after each take-over.

Understanding of the Take-Over Request

To assess if the explanation of the TOR had an effect on the predictability and understanding of the automation, we presented four statements, which could be answered on a rating scale. Participants could rate how much they felt safe during the take-over, how much they felt that they understood the system, and how much they would like to know more about the system.

Procedure

After they had been welcomed by the experimenter, the participants received the instructions and filled out a questionnaire on demographic data. Next, participants started the familiarization drive and practiced the NDRT. Afterward, the participants filled out the Van Der Laan questionnaire for the first of two times. Subsequently, the experimental drive started. Upon completion, the same questionnaire was filled out for the second time and the participants were interviewed with regards to their experience of the scenarios. At the end, the participants were debriefed and the reward for participation was given.

Data Analysis

We relied on Bayesian parameter estimation to quantify the uncertainty in the parameter estimates and to obtain a full probability distribution on the resulting credible interval (Kruschke, 2015). For hypothesis testing, we relied on Bayes Factors (BF; Rouder, Speckman, Sun, Morey, & Iverson, 2009), which represent the ratio of the probability of the data given a null model to the probability of the data given an alternative model and thus quantifies whether the data are more compatible with a null model or an alternative (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2015). A BF, therefore, directly quantifies evidence as a likelihood ratio and also, contrary to a *p* value, is able to obtain evidence for a null hypothesis as it can distinguish between uninformative results and results supporting the null hypothesis (Dienes, 2014). A BF_{10} of 3, for example, states that the data is 3 times more likely in the alternative model than in the null model. If it equals 1, both models predict the data equally well or the data are uninformative for a decision. Lee and Wagenmakers (2013) interpret a BF_{10} 1–3 as anecdotal evidence, 3–10 as moderate evidence and > 10 as strong evidence. The data analysis was carried out by the BayesFactor package (Morey & Rouder, 2015) and scripts by Kruschke (2015) implemented in the statistical computer software R (R Core Team, 2016) and JAGS (Plummer, 2003). A Cauchy distribution with $r = 1/\sqrt{2}$ was chosen as the prior distribution for the effect size δ of the alternative model in the Bayesian *t* test. This weakly informative prior was chosen as a trade-off between results that are completely determined by data and the expectation of a small to medium effect size and represents an anchor point in psychological research (Schönbrodt et al., 2015). With this prior, a *p* value of $p = 0.05$ in an independent samples *t* test with $t(40) = 2.021$ corresponds to a $BF_{10} = 1.49$. We estimated the descriptive parameters with a normal prior and uninformative priors for its parameters ($\mu \sim N(\bar{x}, 1/(100 \cdot \sigma^2)$; $\sigma \sim U(\sigma/1000, \sigma \cdot 100)$).

Results

Acceptance

We compared both scales of the questionnaire between the experimental group (with explanations) and the control group (without explanations) as well as within each group before and after the experimental drive. The descriptive statistics for the scale *satisfaction* are reported in Tables 2 and 3. With regards to the reports of *satisfaction*, we found no difference between the groups before ($BF_{10} = 0.36$) and after the experimental drive ($BF_{10} = 0.42$). There was also moderate evidence that the ratings did not change within the control group before and after the experiment ($BF_{10} = 0.23$). Data were inconclusive whether a slight decrease in the *Explanation* group occurred ($BF_{10} = 0.74$). The results are visualized in Figure 6.

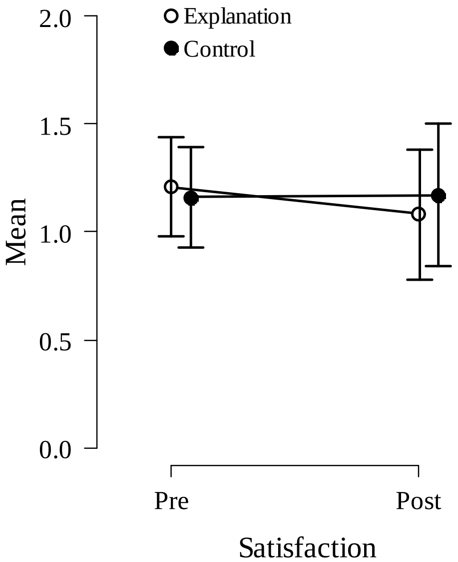

Figure 6: Difference before and after the experimental drive on the scale satisfaction by group; error bars = 95 % HDI.

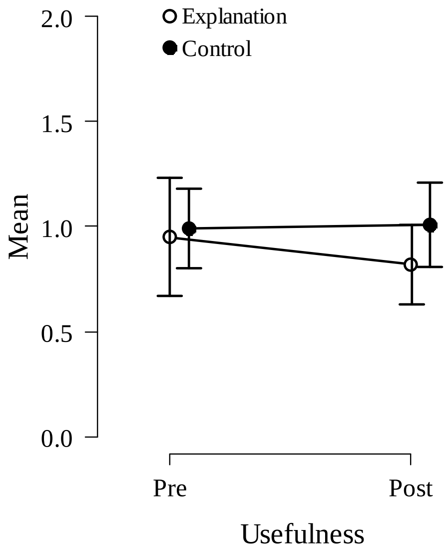

Figure 7: Difference before and after the experimental drive on the scale usefulness by group; error bars = 95 % HDI.

	Group	*N*	*M*	*SD*	HDI LL	HDI UL	α	*d*	BF_{10}
Satisfaction Pre-Exp	Control	20	1.16	0.49	0.93	1.38	0.62	0.05 [−0.58, 0.70]	0.36
	Explanation	19	1.21	0.49	0.98	1.44	0.68		
Satisfaction Post-Exp	Control	20	1.17	0.73	0.84	1.50	0.88	0.18 [−0.44, 0.85]	0.42
	Explanation	20	1.08	0.67	0.78	1.38	0.83		

Table 2: Sample description of the scores on the scale *satisfaction*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

Group	*d*	BF_{10}
Control	−0.02 [−0.43, 0.38]	0.23
Explanation	0.33 [−0.10, 0.78]	0.74

Table 3: Difference pre–post take-over situation on the scale *satisfaction*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

To investigate the interaction between the conditions and the time of measurement we conducted an ANOVA conceptualized as a hierarchical linear mixed model in which the levels are clustered within each factor, following the approach of Rouder, Morey, Verhagen, Swagman, and Wagenmakers (2016). Here, the effect of *group* and *point of measurement* are expressed in the effect size d_i where each factor gets a shared prior for its levels. Consistent to the prior the prior width for the expected range of effect sizes was set to $r = 0.5$ (medium), which correspondents to the prior width of $r = 1/\sqrt{2}$ for the Bayesian *t* test (Wagenmakers et al., 2017). *Participant* was included as a random factor. An ANOVA showed no interaction effect between *group* and *point of measurement* ($BF_{10} = 0.08$; Table 4).

Model	BF_{10}
Group	0.44
Point of Measurement	0.33
Group + Point of Measurement	0.15
Group + Point of Measurement + Group × Point of Measurement	0.08

Table 4: ANOVA for the scores of the scale *satisfaction* with the factors *group* and *point of measurement*; BF indicates comparison to a null model without any factors.

The data on the ratings of usefulness showed no difference between the groups before (BF_{10} = 0.39) and after the experiment (BF_{10} = 0.59) and also no change within a group ($BF_{10\ Control}$ = 0.25, $BF_{10\ Explanation}$ = 0.45; Tables 5 and 6; Figure 7). An ANOVA indicated no interaction effect between group and point of measurement (BF_{10} = 0.11; Table 7).

	Group	*N*	*M*	*SD*	HDI		α	*d*	BF_{10}
					LL	UL			
Usefulness	Control	20	0.99	0.60	0.80	1.20	0.69	0.14 [−0.49, 0.80]	0.39
Pre-Exp	Explanation	19	0.95	0.44	0.67	1.23	0.84		
Usefulness	Control	20	1.01	0.46	0.81	1.21	0.34	0.32 [−0.31, 1.03]	0.59
Post-Exp	Explanation	20	0.82	0.43	0.63	1.01	0.61		

Table 5: Sample description of the scores on the scale *usefulness*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

Group	*d*	BF_{10}
Control	−0.07 [−0.48, 0.34]	0.25
Explanation	0.24 [−0.18, 0.68]	0.45

Table 6: Difference pre–post take-over situation on the scale *usefulness*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

Model	BF_{10}
Group	0.59
Point of Measurement	0.33
Group + Point of Measurement	0.19
Group + Point of Measurement + Group × Point of Measurement	0.11

Table 7: ANOVA for the scores of the scale *usefulness* with the factors *group* and *point of measurement*; BF indicates comparison to a null model without any factors.

Trust Ratings

The participants reported their subjective trust on one item with a rating scale from 0 to 100 before (*Pre*) and after (*Post*) the experience of a scenario. The results are visualized in Figure 8.

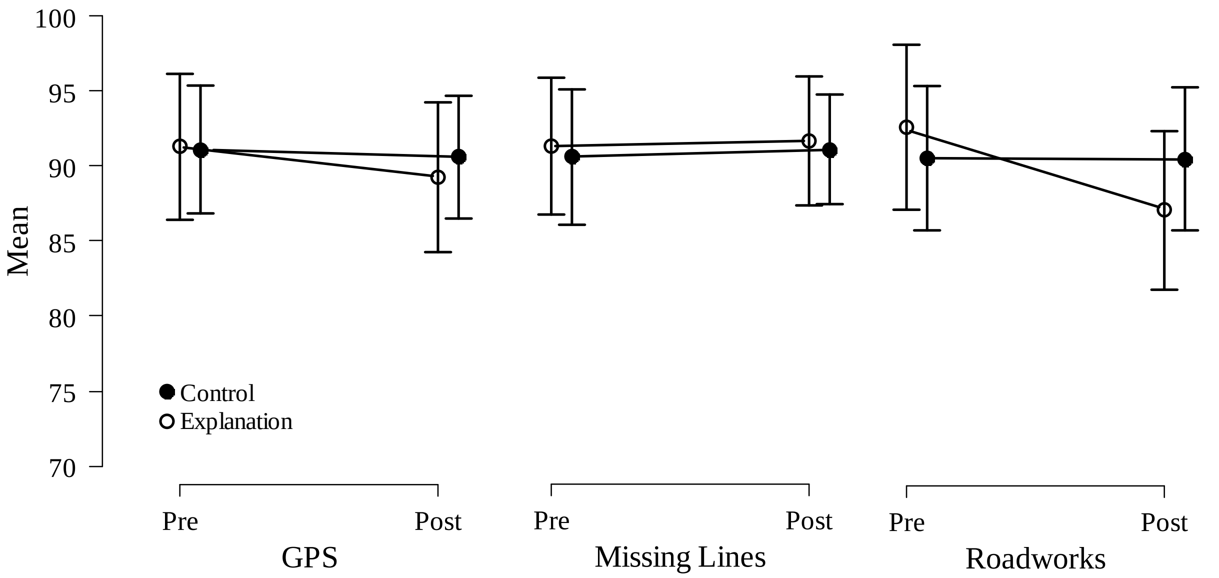

Figure 8: Differences in trust scores before and after the scenarios by *group* and by *scenario*; error bars = 95 % HDI.

Scenario a) GPS

We found no difference between the groups before (BF_{10} = 0.31) and after the experimental drive (BF_{10} = 0.33; Table 8) as well as no change within the groups ($BF_{10\ Control}$ = 0.25, $BF_{10\ Explanation}$ = 0.37; Table 9).

	Group	*N*	*M*	*SD*	HDI LL	HDI UL	*d*	BF_{10}
Pre-Scenario	Control	20	91.08	9.71	86.80	95.41	−0.02 [−0.57, 0.53]	0.31
	Explanation	20	91.29	10.43	86.70	96.14		
Post-Scenario	Control	20	90.59	9.03	86.52	94.64	0.11 [−0.43, 0.67]	0.33
	Explanation	20	89.28	11.35	84.06	94.26		

Table 8: Sample description of the trust scores in the scenario *GPS*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

Group	*d*	BF_{10}
Control	0.07 [−0.33, 0.48]	0.25
Explanation	0.20 [−0.21, 0.62]	0.37

Table 9: Difference pre–post take-over situation in the scenario *GPS*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

Scenario b) Missing Lines

The data also showed no difference between groups before (BF_{10} = 0.32) and after (BF_{10} = 0.31) the experimental drive (Table 10) as well as no change within the groups ($BF_{10\ Control}$ = 0.28, $BF_{10\ Explanation}$ = 0.24; Table 11).

	Group	*N*	*M*	*SD*	HDI LL	HDI UL	*d*	BF_{10}
Pre-Scenario	Control	20	90.60	10.29	86.06	95.19	−0.06 [−0.61, 0.49]	0.32
	Explanation	20	91.32	10.05	86.85	95.88		
Post-Scenario	Control	20	91.10	8.10	87.47	94.68	−0.05 [−0.61, 0.49]	0.31
	Explanation	20	91.65	9.34	87.56	95.94		

Table 10: Sample description of the trust scores; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; d = Cohen's *d*.

Group	*d*	BF_{10}
Control	−0.13 [−0.54, 0.28]	0.28
Explanation	−0.04 [−0.44, 0.37]	0.24

Table 11: Difference pre–post take-over situation in the scenario *Missing lines*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

Scenario c) Roadworks

We found no difference between the groups before (BF_{10} = 0.36) and after the scenario (BF_{10} = 0.46) as well as no change within the control group (BF_{10} = 0.23; Table 12 and 13). However, we found substantial evidence for a decrease in trust within the *Explanation* group of $\Delta = 5.54$ score points (5.98 %; *d* = 0.60 [0.13, 1.08]; Table 13).

	Group	*N*	*M*	*SD*	HDI LL	HDI UL	*d*	BF_{10}
Pre-Scenario	Control	20	90.65	10.54	85.85	95.36	−0.15 [−0.72, 0.39]	0.36
	Explanation	20	92.70	12.15	87.26	98.18		
Post-Scenario	Control	20	90.58	10.76	85.81	95.42	0.26 [−0.30, 0.83]	0.46
	Explanation	20	87.16	11.95	81.65	92.46		

Table 12: Sample description of the trust scores; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's d.

Group	*d*	BF_{10}
Control	0.01 [−0.40, 0.41]	0.23
Explanation	0.60 [0.13, 1.08]	6.56

Table 13: Difference pre–post take-over situation in the scenario *Roadworks*; HDI = 95 % highest density interval; LL = lower limit; UL = upper limit; *d* = Cohen's *d*.

We carried out an ANOVA to evaluate the evidence for an interaction effect. Data yielded no interaction effect in the scenarios *GPS* (BF_{10} = 0.06) and *Missing lines* (BF_{10} = 0.04), but moderate support for an interaction of *Group* and *Point of Measurement* in *Roadworks* (BF_{10} = 2.64); this is consistent to the analysis in Table 13. Table 14 lists the results.

Model	GPS BF_{10}	Missing Lines BF_{10}	Roadworks BF_{10}
Group	0.44	0.46	0.45
Point of Measurement	0.38	0.25	2.01
Group + Point of Measurement	0.17	0.12	0.95
Group + Point of Measurement + Group × Point of Measurement	0.06	0.04	2.64

Table 14: ANOVA with the factors *group* and *point of measurement*; BF indicates comparison to a null model without any factors.

Independent of the scenario, we investigated if the trust level changed in the course of the experiment. The data points in Figure 9 represent the mean of the pre and post situation trust rating. We found moderate evidence for an increase in course of the experiment (BF_{10} = 3.89) and moderate evidence that this effect was independent of *group* ($BF_{10\ \text{Interaction}}$ = 0.46; Table 15).

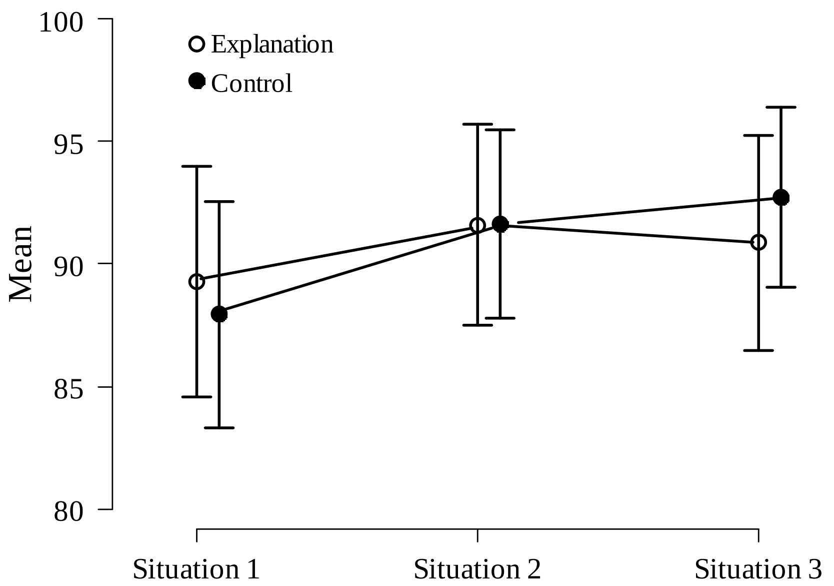

Figure 9: Development of the trust score in course of the experiment by group; error bars = 95 % HDI.

Model	BF_{10}
Group	0.40
Point of Measurement	3.89
Group + Point of Measurement	1.62
Group + Point of Measurement + Group × Point of Measurement	0.46

Table 15: ANOVA with the factors *group* and *point of measurement*; BF indicates comparison to a null model without any factors.

In an explanatory analysis, we compared the difference in the trust ratings between the rating before and after the TOR for participants who experienced Roadworks as their first, second, or third situation. While there was no difference in the trust ratings if the participants experienced Roadworks as their first situation (M_{Δ} = 2.17, BF_{10} = 0.42), the difference was already larger if it was the second situation (M_{Δ} = 3.57, BF_{10} = 1.50), and large if it was their last situation (MΔ = 10.29, BF_{10} = 4.50, d = 1.24). This trend was not observable in the control group ($BF_{10\ Situation\ 1}$ = 0.60, $BF_{10\ Situation\ 2}$ = 0.71, $BF_{10\ Situation\ 3}$ = 1.41). However, the sample sizes (n = 7) for these calculations are too small to conduct reliable and valid inferential statistical methods.

Understanding of the Take-Over Request

After the experiment, we asked the participants to rate four statements on their experience with the take-over situations. We used an ordinal probit model for parameter estimation, which assumes an underlying normal distributed metric variable that is mapped to the empiric ordinal values via response thresholds (Liddell & Kruschke, 2015). There was no evidence for a difference in the ratings of questions 1 and 4. However, the participants in the *Explanation* group felt stronger that it was clear why they had to take over (BF_{10} = 149.10) and that they had understood the system (BF_{10} = 14.71; Table 16).

	Control		Explanation		Δ*M*	BF_{10}
	Md	*M* (*SD*)	*Md*	*M* (*SD*)		
"During the take-over I always felt safe."	4	3.81 (0.63)	3	3.26 (1.44)	−0.55 [−1.65, 0.47]	0.93
"It was always clear to me why I had to take over."	3	2.00 (0.97)	4	3.80 (3.28)	1.80 [−0.07, 4.31]	149.10
"I feel that I have understood the system."	3	2.41 (0.67)	4	3.16 (1.37)	0.76 [−0.25, 2.07]	14.71
"I would like to know more about the system limits."	4	3.85 (0.99)	3	3.45 (1.58)	−0.41 [−1.64, 0.82]	0.26

Table 16: Descriptive results of the four questions after the experimental drive; *Md* = Median; *Mo* = Mode; *N* = 20.

Discussion

In this study, we investigated the effect of providing an explanation of the reason for a take-over request (TOR) on trust and acceptance of driving automation. An experimental group provided with an explanation of the reason for an occurred TOR and a control group given no explanations experienced three take-over situations that varied in the obviousness of the reason for the take-over.

Both groups indicated in the questionnaire prior to the experimental drive that they were satisfied with the system and found it useful. This appraisal did not change by experiencing the three take-over situations. Consistent with previous findings (Gold et al., 2015), it seems that participants do not view a TOR, as implemented in this study, as a threatening malfunction but rather as a legitimate warning of a system that is working correctly. In general, trust ratings increased slightly from experiencing the first take-over to experiencing the last take-over, independent of the condition. This increase in trust with increasing system experience and no experience of negatively evaluated events has been also been reported in similar studies (Beggiato, Pereira, Petzoldt, & Krems, 2015; Hergeth et al., 2016). Accordingly, a take-over situation did not influence the trust rating and we found no difference between both groups in the scenarios *GPS* and *Missing lines*. However, we found persuading evidence for a decrease in trust in the explanation group in the *Roadworks* scenario. An imaginable reason for this might be that the explanations led to a different evaluation of the automation's competence. The provided explanations might have conveyed the image of a more complex and competent system in contrast to the system in the control group which merely experienced performed lateral and longitudinal control. Therefore, it may be surprising for the participants of the explanation group that *Roadworks*, the most obvious reason for the TOR, could not be solved by the driving automation. A similar finding was observed by Madhavan, Wiegmann, and Lacson (2006) who observed that automation errors in easy trials led to greater mistrust than errors in difficult trials. Even small errors of an automated system affect trust more than a large error if the error was unexpected (Muir & Moray, 1996) and trust erodes if the system does not behave as expected even if it shows high performance (Lee & See, 2004). Since the assessment of the automation's competence requires some experience with the system and some exposure to the explanations, the effect should be the most pronounced in the last situation. Following this line of thought, we compared the how much the trust ratings changed by experiencing the TOR for participants who experienced *Roadworks* as either their first, second, or third situation in an explanatory, descriptive analysis. There was no change in the trust ratings if the participants experienced *Roadworks* as their first situation, but a large decrease occurred if it was their third situation. We did not observe this trend in the control group. Each of the three scenarios was implemented with a non-critical take-over time budget of nine seconds. While the road continued as a straight lane after the TOR in the scenarios *GPS* and *Missing lines*, *Roadworks* was the only scenario that required steering after the nine seconds to follow the alternative lane on the construction site (see Figure 4). Therefore, a miscalibration of trust might weight stronger than in the other scenarios and this might be the reason why a TOR might have a different influence on trust in this scenario.

Nevertheless, all scenarios were easily solvable. The participants might therefore not have seen the explanations as overly helpful since no problem occurred that may be explained to ease the mind. The lack of consequences and real risk in simulator drive might have alleviated the need for explanations as well. That being said, the explanations could have a stronger effect if the situations are more critical or more confusing. Lastly, the interaction with the automation was very short and limited to longitudinal and lateral control. Drivers might be more in need of transparency and explanation in more complex situations such as an overtaking maneuver, crossroads, or entering a highway. The results also have to be interpreted in light of the fact that both acceptance and trust, were on a very high level right from the beginning although the automation's functioning and limitations have been explained in a neutral and accurate way prior to the experiment. A possible reason for this fact may be that the study was conducted at a technical university with the majority of the participants being students. The affinity for and trust in technology may generally be on a very high level in such a sample. We, therefore, recommend repeating the study with a sample that has a lower affinity for technology and less experience with automated driving.

In their rating of their understanding of the TOR, the explanation group felt stronger than the control group that they had understood the system and that the reason for the take-over was clear to them. Hence, albeit the explanations had no systematic effect on trust and acceptance, the increase in transparency by the explanations seems to have been successful. Future studies should explicitly investigate whether this subjective increase indeed reflects an improvement in the constructed mental model. For example, drivers should then be able to predict a TOR in a novel situation with higher accuracy. Furthermore, behavioral measures such as take-over time or gaze behavior may also function as a behavioral indicator of system understanding since reaction times to expected events are lower (Larsson et al., 2014; Martens, 2004).

Limitations and future work

The study was conducted in a driving simulator to ensure that each participant experienced exactly the same scenarios. It is possible that the participants may have reported differently due to the lack of risk in a simulator, especially regarding their perceived safety during the take-over situations. Hence, providing an explanation could have a greater effect in a naturalistic drive. That being said, Eriksson and Stanton (2017) have shown that participants' behavior and subjective ratings did not substantially differ between an naturalistic automated on-road drive and a high fidelity simulator. We recruited a gender-balanced sample, but at the same time, mostly students from a technical university aged between 21 and 30 years took part. This led to a homogenous sample regarding affinity to technology, prior knowledge, as well as experiences, and trust in automation (Körber et al., 2016). Recent research has revealed moderating covariates such as age that may influence the attitudes toward automated driving (Hohenberger, Spörrle, & Welpe, 2016; Körber & Bengler, 2014; Payre et al., 2014). To increase the external validity of the results, we, therefore, recommend investigating attitudes toward automated driving with different demographics in future studies.

Key Points:

- Providing a post hoc explanation for a take-over request had small to no impact on trust or acceptance of a driving automation
- Providing a post hoc explanation increased the perceived understanding of the system and of the reason for a take-over request

References

Adell, E. (2009). *Driver experience and acceptance of driver support systems: A case of speed adaptation*. *Bulletin (Traffic & roads, Department of Technology and Society, Lund University): Vol. 251*. Lund: Lund University.

Adell, E., Várhelyi, A., & Hjälmdahl, M. (2008). Auditory and haptic systems for in-car speed management – A comparative real life study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(6), 445–458. https://doi.org/10.1016/j.trf.2008.04.003

Adell, E., Várhelyi, A., & Nilsson, L. (2014a). Modelling acceptance of driver assistance systems: Application of the unified theory of acceptance and use of technology. In M. A. Regan, T. Horberry, & A. Stevens (Eds.), *Human factors in road and rail transport. Driver acceptance of new technology. Theory, measurement and optimisation* (pp. 23–34). Farnham, Surrey, England, UK: Ashgate Publishing Ltd.

Adell, E., Várhelyi, A., & Nilsson, L. (2014b). The definition of acceptance and acceptability. In M. A. Regan, T. Horberry, & A. Stevens (Eds.), *Human factors in road and rail transport. Driver acceptance of new technology. Theory, measurement and optimisation* (pp. 11–21). Farnham, Surrey, England, UK: Ashgate Publishing Ltd.

Aeberhard, M., Rauch, S., Bahram, M., Tanzmeister, G., Thomas, J., Pilat, Y.,. . . Kaempchen, N. (2015). Experience, results and lessons learned from automated driving on Germany's highways. *IEEE Intelligent Transportation Systems Magazine*, *7*(1), 42–57. https://doi.org/10.1109/MITS.2014.2360306

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Arndt, S. (2011). *Evaluierung der Akzeptanz von Fahrerassistenzsystemen: Modell zum Kaufverhalten von Endkunden* [Evaluation of acceptance of driver assistance systems: A model of end customers' purchasing behavior]. *Verkehrspsychologie*. Wiesbaden: VS Verl. für Sozialwiss.

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321–348. https://doi.org/10.1080/14639220500535301

Beaudry, A., & Pinsonneault, A. (2010). The other side of acceptance: Studying the direct and indirect effects of emotions on information technology use. *MIS Quarterly*, *34*(4), 689–710.

Beggiato, M., & Krems, J. F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour*, *18*, 47–57. https://doi.org/10.1016/j.trf.2012.12.006

Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. F. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *35*, 75–84. https://doi.org/10.1016/j.trf.2015.10.005

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: An approach using automation uncertainty. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(6), 1130–1141. https://doi.org/10.1177/0018720813482327

Bengler, K., Dietmayer, K., Färber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 6–22. https://doi.org/10.1109/MITS.2014.2336271

Bengler, K., Zimmermann, M., Bortot, D., Kienle, M., & Damböck, D. (2012). Interaction principles for cooperative human-machine systems. *it - Information Technology*, *54*(4), 157–164. https://doi.org/10.1524/itit.2012.0680

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*. Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Broadbent, E., Stafford, R., & MacDonald, B. (2009). Acceptance of healthcare robots for the older population: Review and future directions. *International Journal of Social Robotics*, *1*(4), 319–330. https://doi.org/10.1007/s12369-009-0030-6

Broughton, J., & Baughan, C. (2002). The effectiveness of antilock braking systems in reducing accidents in Great Britain. *Accident Analysis & Prevention*, *34*(3), 347–355. https://doi.org/10.1016/s0001-4575(01)00032-x

Brown, R. D., & Galster, S. M. (2004). Effects of reliable and unreliable automation on subjective measures of mental workload, situation awareness, trust and confidence in a dynamic flight task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2004* (pp. 147–151).

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, *31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Christofferson, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research: Vol. 2. Advances in human performance and cognitive engineering research. Automation* (1st ed., Vol. 2, pp. 1–12). Amsterdam, Boston: JAI. https://doi.org/10.1016/S1479-3601(02)02003-9

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington: American Psychological Association. https://doi.org/10.1037/10096-006

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(4), 689–705. https://doi.org/10.1177/0018720816685832

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *2017 IEEE Intelligent Vehicles Symposium (IV)* (pp. 365–372). IEEE. https://doi.org/10.1109/IVS.2017.7995746

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51–90.

Ghazizadeh, M., Peng, Y., Lee, J. D., & Boyle, L. N. (2012). Augmenting the Technology Acceptance Model with trust: Commercial drivers' attitudes towards monitoring and feedback. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 2286–2290. https://doi.org/10.1177/1071181312561481

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49. https://doi.org/10.1007/s10111-011-0194-3

Gold, C., & Bengler, K. (2014). Taking over control from highly automated vehicles. In N. A. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.): *Vol. [8]. Advances in human factors and ergonomics 2014, Advances in human aspects of transportation, Part II* (pp. 64–69). Louisville, KY: AHFE Conference.

Gold, C., Körber, M., Hohenberger, C., Lechner, D., & Bengler, K. (2015). Trust in automation – Before and after the experience of take-over scenarios in a highly automated vehicle. *Procedia Manufacturing*, *3*, 3025–3032. https://doi.org/10.1016/j.promfg.2015.07.847

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Hergeth, S., Lorenz, L., Krems, J. F., & Toenert, L. (2015). Effects of take-over requests and cultural background on automation trust in highly automated driving. In *Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 331–337).

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 509–519. https://doi.org/10.1177/0018720815625744

Hoff, K. A., & Bashir, M. (2015). 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

Hohenberger, C., Spörrle, M., & Welpe, I. M. (2016). How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups. *Transportation Research Part A: Policy and Practice*, *94*, 374–385. https://doi.org/10.1016/j.tra.2016.09.022

ISO 14198. (2012). *PD ISO/TS 14198:2012 - Road vehicles - Ergonomic aspects of transport information and control systems - Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems*: BSI.

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(06), 91–95. https://doi.org/10.1109/MIS.2004.74

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*(4), 269–275. https://doi.org/10.1007/s12008-014-0227-2

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/j.apergo.2017.07.006

Körber, M., & Bengler, K. (2014). Potential individual differences regarding automation effects in automated driving. In C. S. G. González, C. C. Ordóñez, & H. Fardoun (Eds.), *Interacción 2014:*

Proceedings of the XV International Conference on Human Computer Interaction (pp. 152–158). New York, NY, USA: ACM. https://doi.org/10.1145/2662253.2662275

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19–32. https://doi.org/10.1016/j.trf.2016.03.002

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2. ed.). San Diego, CA: Academic Press.

Kühn, M., & Hannawald, L. (2016). Driver assistance and road safety. In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), *Handbook of driver assistance systems. Basic information, components and systems for active safety and comfort* (pp. 69–90). Cham: Springer Reference.

van der Laan, J. D., Heino, A., & de Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*(1), 1–10. https://doi.org/10.1016/S0968-090X(96)00025-3

Larsson, A. F.L., Kircher, K., & Andersson Hultgren, J. (2014). Learning from experience: Familiarity with ACC and responding to a cut-in situation in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 229–237. https://doi.org/10.1016/j.trf.2014.05.008

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, *40*(1), 153–184. https://doi.org/10.1006/ijhc.1994.1007

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

Lee, J. D., & Seppelt, B. D. (2009). Human factors in automation design. In S. Y. Nof (Ed.), *Springer Handbook of Automation* (pp. 417–436). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-78831-7_25

Lee, J. D., & Seppelt, B. D. (2012). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1615–1642). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118131350.ch59

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

Lees, M. N., & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics*, *50*(8), 1264–1286. https://doi.org/10.1080/00140130701318749

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123. https://doi.org/10.1037/1076-898X.6.2.104

Liddell, T. M., & Kruschke, J. K. (2015). Analyzing Ordinal Data: Support for a Bayesian Approach. *SSRN Electronic Journal*. Advance online publication. https://doi.org/10.2139/ssrn.2692323

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(2), 241–256. https://doi.org/10.1518/001872006777724408

Mahr, A., & Müller, C. (2011). A schema of possible negative effects of advanced driver assistant systems. In *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 116–121). Iowa City, Iowa: The University of Iowa, Public Policy Center.

Martens, M. H. (2004). Stimuli fixation and manual response as a function of expectancies. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(3), 410–423. https://doi.org/10.1518/hfes.46.3.410.50402

Maule, A. J., & Hockey, G. R. J. (2012). State, Stress, and Time Pressure. In O. Svenson & A. J. Maule (Eds.), *Time Pressure and Stress in Human Judgment and Decision Making* (1st ed., pp. 83–101). Springer.

Meschtscherjakov, A., Wilfinger, D., Scherndl, T., & Tscheligi, M. (2009). Acceptance of future persuasive in-car interfaces towards a more economic driving behaviour. In A. Schmidt (Ed.), *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 81–88). New York, NY: ACM. https://doi.org/10.1145/1620509.1620526

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*. Retrieved from https://CRAN.R-project.org/package=BayesFactor

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

Najm, W. G., Stearns, M. D., Howarth, H., Koopmann, J., & Hitz, J. (2006). *Evaluation of an automotive rear-end collision avoidance system* (No. DOT HS 810 569). Washington, DC.

Naujoks, F., & Totzke, I. (2014). Behavioral adaptation caused by predictive warning systems – The case of congestion tail warnings. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 49–61. https://doi.org/10.1016/j.trf.2014.06.010

Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *327*(1241), 585–593. https://doi.org/10.1098/rstb.1990.0101

Norman, D. A. (2013). *The design of everyday things*. New York, NY: Basic Books.

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

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Payre, W., Cestac, J., & Delhomme, P. (2014). Intention to use a fully automated car: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 252–263. https://doi.org/10.1016/j.trf.2014.04.009

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Kurt Hornik, Friedrich Leisch, & Achim Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003).* Vienna, Austria.

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual differences in the calibration of trust in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(4), 545–556. https://doi.org/10.1177/0018720814564422

Prasch, L., & Tretter, S. (2016). Ich bin dann mal raus. Die Bedeutung der Erklärbarkeit von Übernahmeaufforderungen für das Vertrauen in hochautomatisierte Fahrsysteme [I'm out then. The importance of explainability of take-over requests for trust in highly automated driving systems]. In

S. Franken, U. Schroeder, & T. Kuhlen (Eds.), *Mensch und Computer 2016 – Kurzbeiträge.* Aachen: Gesellschaft für Informatik e.V. https://doi.org/10.18420/muc2016-mci-0270

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from https://www.R-project.org

Regan, M. A., Horberry, T., & Stevens, A. (Eds.). (2014). *Driver acceptance of new technology: Theory, measurement and optimisation*. *Human factors in road and rail transport*. Farnham, Surrey, England, UK: Ashgate Publishing Ltd.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Human factors in transportation. Automation and human performance. Theory and applications* (pp. 19–35). Mahwah, NJ: Erlbaum.

Roberts, S. C., Ghazizadeh, M., & Lee, J. D. (2012). Warn me now or inform me later: Drivers' acceptance of real-time and post-drive distraction mitigation systems. *International Journal of Human-Computer Studies*, *70*(12), 967–979. https://doi.org/10.1016/j.ijhcs.2012.08.002

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2016). Bayesian analysis of factorial designs. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000057

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

SAE International (2016, September 30). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicle*. (Standard, J3016_201609).

Sarter, N. B. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 506–510. https://doi.org/10.1518/001872008X312233

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed., pp. 1926–1943). New York: Wiley.

Schade, J., & Baum, M. (2007). Reactance or acceptance? Reactions towards the introduction of road pricing. *Transportation Research Part A: Policy and Practice*, *41*(1), 41–48. https://doi.org/10.1016/j.tra.2006.05.008

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000061

Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, *38*(7-8), 608–625. https://doi.org/10.1016/j.ergon.2008.01.007

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics*, *1*(1), 89–129. https://doi.org/10.1518/155723405783703082

Thompson, S. C. (1981). Will it hurt less if I can control it? A complex answer to a simple question. *Psychological Bulletin*, *90*(1), 89–101.

Törnros, T., Nilsson, L., Östlund, J., & Kircher, A. (2002). Effects of ACC on driver behaviour, workload and acceptance in relation to minimum time headway. In *9th world congress on Intelligent Transport Systems.* Chicago: ITS America.

Trübswetter, N., & Bengler, K. (2013). Why should I use ADAS? Advanced driver assistance systems and the elderly: Knowledge, experience and usage barriers. In *Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 495–501). Iowa City, Iowa: The University of Iowa, Public Policy Center.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478.

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (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

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J.,. . . Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*. Advance online publication. https://doi.org/10.3758/s13423-017-1323-7

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: Investigating the feasibility of car-driver handover assistance. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications AutomotiveUI '15* (pp. 11–18). New York, NY: ACM. https://doi.org/10.1145/2799250.2799268

Walker, G. H., Stanton, N. A., & Salmon, P. (2016). Trust in vehicle technology. *International Journal of Vehicle Design*, *70*(2), 157. https://doi.org/10.1504/IJVD.2016.074419

Watzenig, D., & Horn, M. (2017). Introduction to automated driving. In D. Watzenig & M. Horn (Eds.), *Automated driving. Safer and more efficient future driving* (pp. 3–16). Cham, s.l.: Springer International Publishing.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Biographies

Dipl.-Psych. Moritz Körber

Dipl.-Psych. Moritz Körber is a graduate research associate working with Professor Dr. phil. Klaus Bengler at the Chair of Ergonomics at the Technical University of Munich. In 2012, he earned his diploma (German equivalent to a master's degree) in psychology and business at the University of Regensburg. His thesis topic was ethical leadership and its influence on employees' challenging citizenship behavior. After working on several User Experience projects, his primary research interests shifted to vigilance, automated driving and methodology.

Lorenz Prasch, MSc

Lorenz Prasch earned his BSc in cognitive science at the University of Tübingen and his master's degree in human factors engineering at the Technical University of Munich with a focus on system ergonomics and interaction design. In his master's thesis at the Chair of Ergonomics, he conceptualized, implemented, and evaluated the influence of post-hoc explanations of automation behavior on the users' trust, acceptance and perceived understanding of a highly automated driving system. Since October 2016, Lorenz Prasch is working at the Chair of Ergonomics as a research associate and continues to study the field of cooperation of highly automated vehicles.

Prof. Dr. phil. Klaus Bengler

Klaus Bengler graduated in psychology at the University of Regensburg in 1991 and received his PhD in 1995 in cooperation with BMW Group at the Institute of Psychology (supervisor: Prof. Dr. Zimmer). After his PhD, he was actively working on topics such as software ergonomics and evaluation of human-machine interfaces. He investigated the influence of additional tasks on driving performance in several studies within the EMMIS EU project and in contract with BMW Group. Multifunctional steering wheels, touch screens and ACC-functionality are examples for the topics of these investigations. In 1997 he joined the BMW Group. Several projects granted him the opportunity to gather experience in experimental design and different kinds of driving simulators as well as field trials. At BMW Group, he was responsible for the HMI project of the MOTIV program, a national follow-up on the program of PROMETHEUS. Within BMW Group Research and Technology, he was responsible for projects on HMI research and leader of the usability lab. Since May 2009 he is the head of the Chair of Ergonomics at the Technical University of Munich which is active in research areas like digital human modelling, human robot cooperation, driver assistance, HMI design and human reliability. He is leading the German Standardization Group (FAKRA) AK-10 “Mensch als Fahrzeugführer” and is an active member of ISO TC22 SC13 WG8 “Road vehicles - Ergonomic aspects of transport information and control systems” as well as a member of VDI working group “Menschliche Zuverlässigkeit”.