Applied Ergonomics 66 (2018) 18–31

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Introduction matters: Manipulating trust in automation and reliance in automated driving

Moritz Körber*, Eva Baseler, Klaus Bengler

Technical University of Munich, Boltzmannstraße 15, D - 85747, Garching, Germany

A R T I C L E I N F O

Article history:
Received 26 February 2017
Received in revised form
28 June 2017
Accepted 19 July 2017
Available online 12 August 2017

Keywords:
Automated driving
Trust in automation
Reliance

A B S T R A C T

Trust in automation is a key determinant for the adoption of automated systems and their appropriate use. Therefore, it constitutes an essential research area for the introduction of automated vehicles to road traffic. In this study, we investigated the influence of trust promoting (*Trust promoted* group) and trust lowering (*Trust lowered* group) introductory information on reported trust, reliance behavior and take-over performance. Forty participants encountered three situations in a 17-min highway drive in a conditionally automated vehicle (SAE Level 3). Situation 1 and Situation 3 were non-critical situations where a take-over was optional. Situation 2 represented a critical situation where a take-over was necessary to avoid a collision. A non-driving-related task (NDRT) was presented between the situations to record the allocation of visual attention. Participants reporting a higher trust level spent less time looking at the road or instrument cluster and more time looking at the NDRT. The manipulation of introductory information resulted in medium differences in reported trust and influenced participants' reliance behavior. Participants of the *Trust promoted* group looked less at the road or instrument cluster and more at the NDRT. The odds of participants of the *Trust promoted* group to overrule the automated driving system in the non-critical situations were 3.65 times (Situation 1) to 5 times (Situation 3) higher. In Situation 2, the *Trust promoted* group's mean take-over time was extended by 1154 ms and the mean minimum time-to-collision was 933 ms shorter. Six participants from the *Trust promoted* group compared to no participant of the *Trust lowered* group collided with the obstacle. The results demonstrate that the individual trust level influences how much drivers monitor the environment while performing an NDRT. Introductory information influences this trust level, reliance on an automated driving system, and if a critical take-over situation can be successfully solved.

1. Introduction

The introduction of automated vehicles to road traffic is motivated by several predicted, beneficial outcomes (Maurer et al., 2015; Stanton and Young, 1998) such as mitigating the negative effects of routine drives on drivers' health and compensating the negative effects of the predicted increase in passenger traffic by increasing traffic efficiency (Payre et al., 2014; Roberts et al., 2011). Beyond that, although advancements in passive and active safety technologies have led to a significant reduction in road accidents (Choi and Ji, 2015), European data, for example, show that 26,000 road fatalities were still reported in the European Union in 2015 (European Commission, 2016). It is assumed that fully automating the driver's tasks will reduce human error, such as speeding or distraction, and, thereby, the number of fatalities further still. However, these claimed benefits may only occur if automated vehicles are successfully implemented into road traffic and trust in this technology is a vital precondition for this. Ghazizadeh et al. (2012) stated in their *Automation Acceptance Model* that trust is a crucial contributor to an individual's acceptability of automation technology and several previous studies have empirically shown that trust is a key determinant for reliance on automated systems (Bailey and Scerbo, 2007; Muir and Moray, 1996), adoption of automation (Gefen et al., 2003; Lee and Moray, 1994), and the intention to use autonomous vehicles (Choi and Ji, 2015). In other words, “operators tend to use automation that they trust while rejecting automation that they do not” (Pop et al., 2015). Multiple research disciplines focus on trust, and there are several models with multiple dimensions of trust that more or less overlap. Based

* Corresponding author.
E-mail addresses: moritz.koerber@tum.de (M. Körber), eva.baseler@tum.de (E. Baseler), bengler@tum.de (K. Bengler).

http://dx.doi.org/10.1016/j.apergo.2017.07.006
0003-6870/

on Mayer et al. (1995) and Lee and See (2004), we define trust in automation as "*the attitude of a user to be willing to be vulnerable to the actions of an automation based on the expectation that it will perform a particular action important to the user, irrespective of the ability to monitor or to intervene.*" This definition implies that trust is a multidimensional construct that is based on relevant characteristics of the automated system (e.g., reliability, predictability) and the trustor himself (e.g., propensity to trust). Driving automation represents a novel and complex technology and, contrary to flight automation in aviation, its users will not be experts who have a deep understanding of its functionality and principles (Körber and Bengler, 2014). Thus, its use represents a situation of uncertainty and vulnerability in which the user entrusts his well-being to the automated driving system (Lee and See, 2004; Walker et al., 2016). However, trusting a system is not a binary all-or-none decision. The conceptualization of trust in a system rather refers to trusting and reliance as a graded process, with the degree of trust being dynamic and situational (Hoff and Bashir, 2015; Lee and See, 2004). This degree does not have to exclusively concern a system as a whole but may be specific to particular functions of it. For example, in a study on a supervisory process control task, trust was distinct to the specific automatic controller (Lee and Moray, 1994) and an automation failure did not cause trust to decline in the remaining similar, but independent automatic controllers (Lee and Moray, 1992; Muir and Moray, 1996). In contrast, Keller and Rice (2009) found that when a completely reliable aid was presented with an unreliable aid, operators tended to rate both aids the same in a global, system-wide trust rating rather than treating them as different systems with different reliabilities (component-specific trust). The exact degree of functional specificity is probably moderated by an operator's experience with the system, its complexity, the information presented to the operator and their goals in operating the system (Lee and See, 2004). Indeed, this reflects the conceptualization of trust as a mainly affective response with influences by analytic and analogical processes. Since the future driving automation users will not be a homogenous group of trained experts, potential variability in driver characteristics such as trust can lead to very different outcomes, especially in time-critical situations like take-over situations (Creaser and Fitch, 2015; Körber et al., 2016a,b). For example, in June 2016, the first fatal accident caused by a self-driving car occurred. The driver completely relied on the Autopilot function of his Tesla Model S, which misinterpreted a white tractor-trailer crossing the highway against a bright sky for a road sign overhead (The Guardian, 2016). Accordingly, one of the two published Research Needs Statements regarding human factors research on automated vehicles by the Transportation Research Board (TRB) pertains to the misuse and abuse of automated vehicles (Creaser and Fitch, 2015). Thus, to ensure a safe introduction of automated vehicles to road traffic, it is crucial to take trust into account. Otherwise, the benefits of driving automation could vanish even if the system is used.

1.1. The role of trust in automated driving

Contrary to autonomous vehicles, conditionally automated driving still requires a driver. In SAE Level 3 (SAE International, 2016), drivers are included in the control loop as a fallback level and will be required to take over manual control at system limits. Beyond that, it is also possible that the system might require a transfer to Level 2 if the conditions necessitate it (Creaser and Fitch, 2015). An automated driving system will allow the driver to take his eyes off the road and engage in NDRTs and driving simulator studies show that drivers are willing to do so, possibly increasing the demand of a take-over situation (Carsten et al., 2012; Llaneras et al., 2013; Radlmayr et al., 2014). This act of reliance is only performed if the driver trusts driving automation enough to completely hand over the driving task. However, trust predicts not only whether an automated system is used but also how it is used: Parasuraman and Riley (1997) categorized the interaction with automation into four styles which can be linked to operator's trust in automation. Among *Use*, *Disuse*, *Abuse,* the authors highlight the negative effects of *Misuse*, inappropriate over-trust when the operator's trust exceeds the automated system's capabilities. Mahr and Müller (2011) stated in their model that too much trust leads to over-reliance on automation and therefore to a risk adaption, i.e. the driver takes risks he would not have taken without an automated system. Operators then tend to be vulnerable to monitoring failures (Bagheri and Jamieson, 2004; Bailey and Scerbo, 2007) and tend to exhibit longer reaction times (Beller, Heesen, and Vollrath, 2013; Helldin, Falkman, Riveiro, and Davidsson, 2013) or poorer reaction quality in critical events (McGuirl and Sarter, 2006; de Waard, van der Hulst, Hoedemaeker, and Brookhuis, 1999). Hence, not only a minimum level but an appropriate level of trust is crucial: The operator has to know the capabilities of an automated system and should monitor it adequately when it is close to the limits of its capability (Carlson et al., 2014). Otherwise, the consequences are unexpected situations in which the driver may not able to react in time.

The take-over of vehicle control can be critical if the automated driving system is operated in an unfamiliar, unexpected or unstructured environment, situation or condition, because then the situation's demand may exceed the capacity for reacting since such situations have an increased demand (Shinar et al., 2005; Wagner and Koopman, 2015). For example, Payre et al. (2016) found a higher take-over time with increasing trust in an emergency situation if training was insufficient. Consistent across different levels of automation, inappropriate levels of trust lead to extended reaction times or poorer reaction quality in hazardous situations (Abe, Itoh, and Tanaka, 2002; McGuirl and Sarter, 2006; Parasuraman and Riley, 1997). The causal mechanism could lie in participants' monitoring strategy: Muir and Moray (1996) as well as Bagheri and Jamieson (2004) found a decrease in monitoring with increasing trust. Hergeth et al. (2016) also reported a negative correlation between participant's trust in automation and the extent of monitoring of a highly automated driving system during the engagement with an NDRT. Accordingly, better-calibrated trust, achieved by the display of system confidence or reliability (Dzindolet et al., 2003; McGuirl and Sarter, 2006), led to faster braking responses in a study by Seppelt and Lee (2007). Beller et al. (2013) showed that the presentation of information on an automated system's uncertainty improves situation awareness, improves a driver's mental model of the automated driving system, increases trust, and leads to an increased time to collision in the event of an automation failure. Drivers in the study of Helldin et al. (2013), who were informed of the automated system's uncertainty, were better prepared in take-over situations while, on average, spending more time doing other activities. This study aims to investigate the relationship between trust, reliance behavior and take-over performance in conditional automated driving in greater detail. Firstly, we investigate if trust and reliance behavior can be manipulated by prior information (see next section). Secondly, trust is one factor, besides others such as self-confidence, that influences reliance (Lee and Moray, 1992; Lee and See, 2004), but it does not completely determine it because intentions and attitudes do not completely determine behavior (Ajzen and Fishbein, 1980; Meyer, 2004). Therefore, we use explicit behavioral measures in the form of reliance behavior as well as implicit measures in the form of eye tracking. Furthermore, the study aims to investigate the relationship between reported trust and the safety-critical outcome take-over performance. The following relationships are expected: