I'm curious, I'm open to it, I test it, I trust it! A focus groups study to understand a-priori trust in automated buses

Jan C. Zoellick, Lina Kluy, Susanna Rössle, Janne Witte, Liane Schenk, Adelheid Kuhlmey, Stefan Blüher

Citation details
Zoellick JC, Kluy L, Rössle S, Witte J, Schenk L, Kuhlmey A, et al. I'm curious, I'm open to it, I test it, I trust it! A focus groups study to understand a-priori trust in automated buses. Transportation Research Part F: Traffic Psychology and Behaviour. [Online] Elsevier BV; 2021;81: 55–64.
DOI: 10.1016/j.trf.2021.05.016

I'm curious, I'm open to it, I test it, I trust it! A focus groups study to understand a-priori trust in automated buses

Jan C. Zoellick[a]*, Lina Kluy[b], Susanna Rössle[a,c], Janne Witte[a,d], Liane Schenk[a], Adelheid Kuhlmey[a], Stefan Blüher[a]

[a] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Sociology and Rehabilitation Science, Charitéplatz 1, 10117 Berlin, Germany

[b] Karlsruher Institute of Technology, Institut für Arbeitswissenschaft und Betriebsorganisation, Engler-Bunte-Ring 4, 76131 Karlsruhe, Germany

[c] Freie Universität Berlin, Institute of Philosophy, Habelschwerdter Allee 30, 14195 Berlin, Germany

[d] Humboldt-Universität zu Berlin, Institute of Social Sciences, Universitätsstr. 3b, 10117 Berlin, Germany

* Corresponding author.

E-Mail: jan.zoellick@charite.de

Abstract

Trust is regarded as one of the main predictors for adopting automated buses (ABs). However, theories about trust (development) in technology generally vary and an in-depths study about trust in ABs specifically is still outstanding. The present study fills this gap by presenting results from focus group interviews to trust (development) in shared automated buses prior to exposure. The objectives of this study are to contrast participants' naïve concepts of trust with theory and to identify underlying factors influencing a-priori trust in ABs. Results show that the *N* = 21 focus group participants use different strategies to familiarise themselves with the new technology of ABs, e.g., comparisons with familiar technologies, fundamental tendencies to approach or avoid, additional information seeking, or anthropomorphisation. These strategies largely support existing theories on trust (development) in technology. Differences between naïve interpretations of trust and its theoretical assumptions were found in focus group debates where more control over technology limited uncertainty and led to *more* trust. While theories suggest control and trust to be incompatible opposites, participants see control as a way to enhance trust. We provide starting points for further theory development and expansion and stress the importance of explanations in emerging technologies for trust and acceptance building.

Keywords: autonomous driving; trust in automated vehicles; automated buses; focus groups; theories on trust

1. Introduction

Automated buses (ABs) are an emerging technology for public transport being able to move more than 10 passengers simultaneously on programmed routes without a driver (Nordhoff et al., 2018; Zoellick, Kuhlmey, Schenk, Schindel, & Blüher, 2019b). Those ABs are “able to perform all driving functions under certain conditions” corresponding to level 4 or above of the classification by the US-American National Highway Traffic Safety Administration (NHTSA, 2017, p. 4). Fig. 1 depicts two examples of buses in use in pilot projects in Berlin, Germany (Zoellick, Kuhlmey, Schenk, Schindel, & Blüher, 2019a; Zoellick et al., 2019b). Embedded in smart mobility systems, ABs promise societal and environmental benefits, e.g. inclusive mobility (Pettigrew, 2017), reduced greenhouse gas emissions (Greenblatt & Saxena, 2015; Wadud, MacKenzie, & Leiby, 2016), and a positive change in urban spaces (Henderson & Spencer, 2016). Trust is regarded as one of the most important concepts in explaining the degree of acceptance and thereby the (non-)use of ABs (Choi & Ji, 2015; Kaur & Rampersad, 2018). Thus, trust is necessary for the ABs to fulfil their promises.

(A) (B)

Fig. 1. Electric, shared ABs at the Charité campuses. (A) EasyMile EZ10 at Charité Campus Mitte. (B) Navya Arma at Charité Virchow Klinikum. Adopted from “Assessing acceptance of electric automated vehicles after exposure in a realistic traffic environment” by J.C. Zoellick, A. Kuhlmey, L. Schenk, D. Schindel, and S. Blüher, 2019, *PLOS One*, *14*(5), p. 2.

However, little is known about trust in ABs. A first constraint is the dominance of quantitative questionnaire studies within this research field (Versteegh, 2019) that share the methodological limitation of focusing on breadth rather than depths of knowledge. Many of the qualitative studies on trust in automated mobility conceptualise the vehicles as privately owned passenger cars (Bazilinskyy, Kyriakidis, & de Winter, 2015; Buckley, Kaye, & Pradhan, 2018; Li, Blythe, Guo, & Namdeo, 2019). The few studies reporting

results on ABs use trust as one outcome among many (Nordhoff, de Winter, Payre, van Arem, & Happee, 2019; Nordhoff, Stapel, van Arem, & Happee, 2020; Zoellick et al., 2019a). In their interview study after a ride in an AB, Nordhoff et al. (2019) coded 6% of the material with the main category "trust" without sub-categories. Similarly, in the coding scheme of Nordhoff et al. (2020), trust is operationalised as a sub-category of perceived safety and discussed accordingly. Both of these studies collected their data on trust in ABs *after* the ride, not a-priori trust before the ride. Thus, to our knowledge little is known about the topic of a-priori trust in ABs.

The purpose of this paper is to investigate a-priori trust in ABs in depth using results from focus group participants (N = 21) and interpretative methods (Kuckartz, 2016). Their naïve conceptualisations of trust (development) are contrasted with theoretical deliberations based on academic literature. Findings from these contrasts expand theory and provide insights towards trust in newly induced automation processes generally.

In the following sections, we will first present the concept of human-machine trust, which differs from human-human trust, so that we can then discuss existing literature on trust in automated vehicles in more detail. This will provide the theoretical basis for our qualitative study on a-priori trust in automated buses.

1.1. Theory on trust in ABs

While authors generally agree that trust becomes relevant in human-machine interaction and subsequent use (Parasuraman & Riley, 1997; Wu, Zhao, Zhu, Tan, & Zheng, 2011), definitions vary in their emphasis on emotional (Plutchik, 2001), behavioural (Hergeth, Lorenz, Vilimek, & Krems, 2016), situational (Goto, 1996), or attitudinal aspects (Lee & See, 2004). Most commonly, trust is seen as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 54). In that sense, trust in machines such as ABs is conceptually closely related to interpersonal trust (Hoff & Bashir, 2015), but differs in its development and reference point (Madhavan & Wiegmann, 2007). Interpersonal trust is linked to (perceived) expertise of the other person and it typically increases from low default with experience; whereas trust in machines is linked to performance and it decreases from high default based on expectations of perfection with experience of errors (Madhavan & Wiegmann, 2007).

Trust in machines and its development is influenced by several factors. Hoff and Bashir (2015) proposed an empirically driven model that identified influencing factors on human

trust in automation. They suggest that there are four layers of trust (Fig. 2). The first layer consists of dispositional aspects such as demographic variables. Situational trust, the second layer, is characterised by the influence of environmental factors and the internal state of the operator. The third layer consists of knowledge and expectations prior to system interaction. Those three layers form initial expectations and the tenor of interaction. During interaction, trust evaluations can shift dynamically based on experiences with several aspects of the system (fourth layer). In contrast to other studies on trust in ABs, the focus of this article lays on the first three layers of a-priori trust as a baseline before interaction.

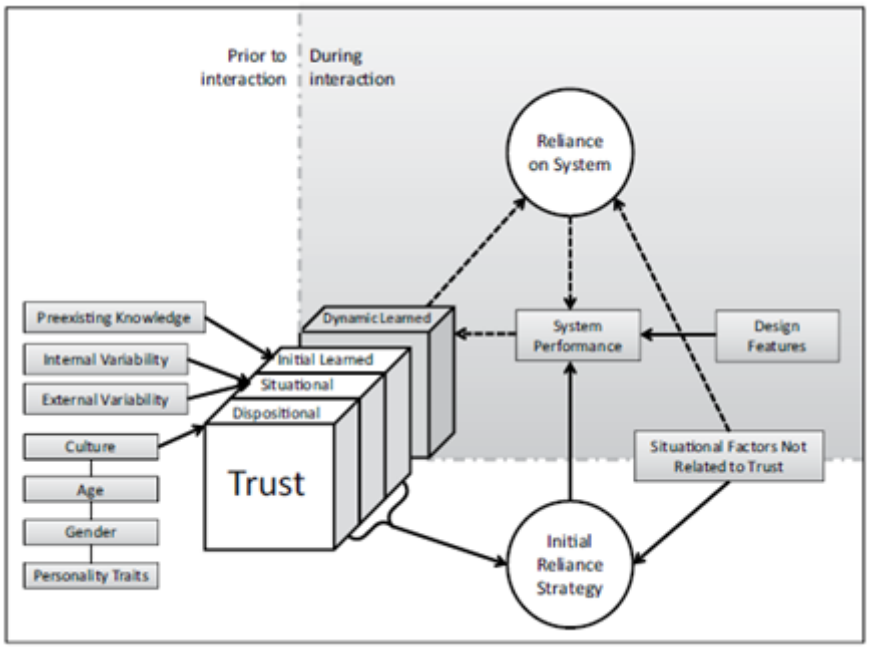

Fig. 2. Full model of factors that influence trust in automation. Adopted from “Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust” by Hoff & Bashir, 2015, *Human Factors*, *57*(3), p. 427.

1.2. Research questions

The aims of this article are to discover underlying factors that influence the development of a-priori trust in ABs, and to contrast naïve trust conceptualisations of potential users with theoretical assumptions found in academic literature. Correspondingly, the research questions are formulated as follows:

1. Which underlying factors influence the existence and development of a-priori trust (and mistrust) in ABs?

2. What are similarities and differences of theoretical assumptions about human-machine trust in comparison to naïve conceptualisations of a-priori trust in ABs by potential users?

2. Material and methods

2.1. Recruitment and participants

The present case study was conducted in advance of introducing four ABs at two campuses of a university clinic in Germany (Fig. 1), two Easymile EZ10 and two Navya Armas, being employed on round courses with 0.8 and 1.2 km length and designated stops (Zoellick et al., 2019b). Participants were recruited through notifications submitted via the intranet of the university clinic and offered a compensation of 25€ for participating. Forty-four clinic employees and university students responded to the call indicating their role in the hospital and their main campus for work or studies. We selected N = 21 participants (9 women) for the four focus groups based on their location and role to reflect the diversity of potential users within this case study. Characteristics of participants included in this article are presented in Table 1. These were collected via interviewer observations in the focus groups.

Table 1

Pseudonyms and descriptions of focus group participants in selected passages

Pseudonym	Age group	Mode of transport	Living conditions	Miscellaneous
B1m	*40s*	*Private car, annual ticket public transport; used to have a company car*	*Lives with family in the Brandenburg area*	*criticises prices of annual public transportation ticket*
C1f	*Mid-30s*	*Public transport user, previously car driver; owns rarely used car*	*Berlin suburb with husband and daughter*	*formerly excessive car driver in Bavaria*
D1f	*Mid-40s*	*Walking; public transport for long distances; owns 23-year-old car*	*Not disclosed*	*walks long distances, uses car only Sundays*
E1f	*End-30s*	*Cycling; former public transport user, because cycling was too dangerous*	*Not disclosed*	*enthusiastic cyclist (including shared bikes), has not driven a car for 16 years*
G2f	*Mid-50s*	*Private car, public transport (regularly)*	*Not disclosed*	*is keen on parking aids and technical development, wants better public transport clocking*
H2m	*30s*	*Private car exclusively*	*15-minute commute to work with nice route*	*Sports-affine*
I2f	*Mid-40s*	*Scooter; previously public transport that become fuller and chaotic*	*Not disclosed*	*Critical of corporations developing technology as business models*
J2m	*Mid-30s*	*Public transport; car rides are too stressful, but public transport is also increasingly stressful*	*Not disclosed*	*Fond of e-mobility and ride-sharing*
M3m	*Mid-40s*	*Private car and motorcycle enthusiast; public transport or cycling in the city*	*Not disclosed*	*-*
S4m	*End-40s*	*Private car; saves time compared to public transport*	*Not disclosed*	*Sceptical about e-mobility and low cost promises of automated public transport*
U4f	*Mid-30s*	*Bicycle; no driver's license*	*Not disclosed*	*Dreams of a car-free city with cheap and effective public transport*

Note. Pseudonyms consist of order of contribution (A to U; ascending alphabetically), focus group number (1 to 4), and sex (m = male; f = female). Participant characteristics were based on appearance and contributions during focus group discussions.

2.2. Interviewing procedure and analysis

All focus groups were conducted between February and March 2018 – prior to introducing the ABs – using the same semi-structured guideline (supplementary material) by the same team of two experienced researchers. Besides other topics such as use, safety, or mobility generally, trust was included in the guideline with the stimuli "to what extent do you trust the ABs?" and "what would increase your trust / mistrust?" after describing the ABs and presenting pictures of the vehicles taken from the manufacturers' homepages. In one focus group the topic of trust emerged from the participants themselves without the use of these stimuli. With this study design, we were able to extract expectations of participants towards ABs generally and in their application in this project particularly.

We analysed the qualitative material following qualitative content analysis by Kuckartz (2016) using deductive and inductive categorisation of text. Here, the basic unit of analysis is the so-called unit of meaning – a passage of the material that provides enough context to stand for itself without further explanation. Such a unit of meaning can be a sentence, but often times involves entire paragraphs or – particularly in the case of focus groups – a back-and-forth between participants themselves and the interviewer. With these characteristics, the same unit of meaning can be assigned to multiple categories (one-to-many classification) which is particularly prevalent in longer passages. As in other coding schemes, categories should be distinguishable and different from each other, but do not need to be exclusive since the same passages can be assigned to multiple categories. We refer to Kuckartz (2016, pp. 100-111) for a detailed description of the method and its iterative processes in category development and coding.

In sum, all focus groups together lasted 401 minutes (M = 100.25 min; SD = 14.55 min), and audio files were transcribed by an external transcription bureau into 74,887 words. First, all material was coded deductively based on the topics of the interview guideline as categories. One of those categories was trust and its passages were selected for further analysis. Second, inductive categories emerging from the material were formed. All categories were applied to the material again in an iterative process to clarify and distinguish them. For the present article, all units of meaning with the deductive category "trust" (4,799 words) were further analysed. Three experienced researchers analysed this subset of material to identify inductive categories independently and merged them into a shared set of categories in one meeting. This set was then used to classify the material independently again and further adapted in an iterative process supported by

the institute's interpretation circle with eight experienced researchers as an additional measure for quality control.

3. Results and discussion

3.1. Main categories of a-priori trust

The final set of categories for analysing the focus groups can be found in Table 2. Their specifications, relationships to trust (development), and interplay are described and interpreted below. Following the methodological approach of qualitative content analysis by Kuckartz (2016), the results section is structured based on passages and their content rather than the categories, because the same passage can be assigned to multiple categories simultaneously. Structuring based on categories would require multiple uses of extensive passages. Selected passages below are either particularly representative for the category or show the variance within a category. Results are illustrated with exemplifying quotes from the focus groups and contrasted with findings from literature.

Table 2

Final set of categories for qualitative analysis

Categories	Description	Frequency of occurrence
Comparisons of ABs to other (mobility) technologies	*Familiarisation through similarities and differences of the ABs to other (mobility) technologies.*	*12*
Curiosity vs. scepticism when approaching ABs	*Reactions of dealing with the unknown concept of ABs. Curiosity refers to inquisitiveness and tendencies to approach the ABs; scepticism refers to avoidance and retention.*	*12*
Maintaining of control over ABs	*Circumstances and features marked by the desire or exertion of control over (mobility) technology.*	*7*
Understanding the functionalities and processes of ABs	*The wish for more information and detailed knowledge about ABs and their role in developing trust.*	*6*
Ignorance and naivety as source of positive attitudes towards ABs	*Contrast to the wish for understanding. Ignorance refers to content with non-information; naivety refers to an unprejudiced, light-hearted view on ABs.*	*3*

Anthropomorphisation	*Attributing or desiring "human" features in the ABs.*	*5*
Habit and experience of other technology	*Description of past experiences with technology as well as habits that arose from the use of it. This includes the interplay between experience and habits.*	*7*
Perceived safety of the ABs	*Requirements that enhance a subjective assessment of protection from danger. Differentiated into the source of safety – human or machine.*	*25*
Reliability and error rate of the ABs	*Performance of a technical system in (repetitive) interaction with high (error-prone) or low (reliable) rates of malfunction.*	*9*

Note. Some categories depict a spectrum with opposing poles (e.g., curiosity vs. scepticism) or cover closely related concepts (e.g., ignorance and naivety). With this approach, we can capture, interpret, and discuss values and attitudes within categories instead of contrasting opposing categories.

3.2 Underlying factors of a-priori trust in ABs

3.2.1. Comparisons to other technologies and transmission of trust

Participants compared the (unknown) concept of ABs to known mobility technologies (e.g., subways, airplanes, cableway vehicles, lifts, or gondolas). This comparison can be seen as a strategy to familiarise themselves with ABs and thus form a basis of trust. The model by Hoff and Bashir (2015) suggests that there is initially learned trust that is based on the pre-existing knowledge of a system. The comparison to other technologies can therefore increase trust due to the knowledge gained. Passages 1 to 3 show a diversity of comparisons and a way of approaching the concept of ABs through familiarisation.

Passage 1:

M3m: I'm comparing this, for example, I'd be actually quite open-minded, without a driver even, because, for example, I'm comparing that right now, when I'm sitting in the back of the subway or something like that I have, yeah, I don't actively think about the driver and how he's feeling right now, or whether he is feeling well, or whether he can pay attention to things real quickly, emergencies or something like that or in emergency situations, I don't even know, I also have a certain basic trust in technology. Well, I'm not even aware that the subway itself with its driver in that sense still, I don't actively think about it, that's what I just noticed actually. Anyway…

Interviewer: A revelation!

M3m: Yes. Yes, because it's also spatially so far apart from each other when I enter

somewhere in the back, I don't notice it in that sense. At the most, when the S-Bahn or the U-Bahn (a/n: subway) passes me, that I still see the driver in the front. But otherwise I have a certain basic trust in it. Get in, drive, and get out. So I could also imagine it here.

Passage 2:

D1f: In a gondola, there is no one inside either – I'm just comparing it with a gondola – when I'm sitting up there, then the interaction is through a push button. Well, I'm not a pronounced gondola rider, but I go, well I do not, I don't ride it, but it's basically the same. The thing drives without even having a person in there. And you have to trust, that you can have contact immediately, like in a lift, if you get stuck, there is a button through which you can call someone.

Passage 3:

S4m: And the sense of security, I mean you fly on the plane and you don't even know if there is a pilot sitting at all in the cockpit or if automatic piloting is set and he is sleeping or, I don't know, drinking coffee, or being on the toilet and nobody's there. Anyhow, you don't know it. There is a closed door. So you're sitting in a plane and you don't know if someone has the control…

U4w: But I know that there is someone, who could intervene in an emergency.

S4m: You don't know that either. You don't see it.

U4f: Yes I do. I do.

The main topic of these comparisons is the subjective experience of being a passenger. Direct control over one's own movement is relinquished to either an invisible person whose presence becomes irrelevant (M3m and S4m), or to mere technology with a button for human communication (D1f). In all instances, this loss of control necessitates trust. Put differently, trust can only exist in a situation of uncertainty, vulnerability, and the absence of control (Donick, 2019; Lee & See, 2004). Particularly M3m (passage 1) presents an interesting case as his trust in the human U-Bahn driver is unknowingly transferred to technology making both interchangeable. Passages 1 and 3 are exemplary for how trust is often automatically generated (M3m) or reflexively denied (U4f) in routine activities such as a flight or a subway ride. This supports the view on trust as routines and norms about technology in general that are activated whenever engaging with specific artefacts instead of explicit deliberations about the particular piece of technology currently encountered (Wagner, 1994). Additionally, technology is not questioned as long as it performs as expected demonstrated by the U-Bahn ("*get in, drive, and get out*", M3m) or the plane's autopilot.

3.2.3. Approaching ABs: Naivety, curiosity and scepticism

The second way to approach ABs consisted of fundamental tendencies to approach and avoid, or put differently, of curiosity and scepticism. In passage 4, scepticism is ascribed to the potential user's attitudes towards the invention of the railway. Here, the dispositional level of trust (Hoff & Bashir, 2015) can explain which factors play a role in creating a positive or negative attitude towards ABs. Potential users of different age, gender, or cultural background are likely to develop different attitudes towards technology. Trust is therefore shaped by external factors that make it easier or harder to approach ABs.

Passage 4:

C1f: I think that, when the railway was introduced everyone made a big fuss out of it; they just didn't understand it. When the first cars came there were people that resisted and said I would never get into one and in the end they got driven to their own funeral and well, they didn't make that experience, but yeah… Moreover, there are always… especially with these kinds of cars (a/n: self-driving) I have the feeling that there will always be difficulties. People just can't accept it straight away.

In passage 4, scepticism towards new technologies as an anthropological constant is addressed and transferred to ABs. Technological innovations seem to be accompanied by an additional quality of trust necessary for adoption and acceptance. However, C1f does not share the sceptical attitude, but rather ironizes it (the first ride being the one to the own funeral). The trivialisation of scepticism can also be observed in passage 5 in which the values curiosity (E1f) and scepticism (B1m) clash.

Passage 5:

E1f: I believe that I am myself naïve enough to think that technology is like… functions better than individuals behind the wheel. Well, actually I don't really know anything about it, but I'm not worried, because I think someone will… the thing will learn and manage it. And one day I will be able to save my cab money.

B1m: This will probably happen as you said with a probability of 99.8 percent. But I mean probably due to my almost biblical age, as I said, one has witnessed events like for example Chernobyl and those are things that have – well at least for me – entrenched themselves deeply in my consciousness as something where because of malfunction technology got out of hand and eventually at a certain point became incontrollable – regardless of what you do. And of course, it's not a bit comparable, obviously, because this is a small bus that drives slowly around the place, but as I said just as an example. So finally, my subliminal fear is that this technology fails and this bus develops a malfunction that leads to people's harm – either outside the buses or inside the bus.

E1f: But I think it's madness that so many people trust in technology. So, technology is applied in operating rooms; everybody has a computer at home; everybody lets navigations systems navigate them. And then suddenly such a thing drives without a driver and everybody gets anxious that something bad happens! I think we have so much technology where we live and I personally do not understand this kind of…

By her own account, E1f does not have enough knowledge about ABs cognitively, but is affectively neither anxious nor afraid to use them. The result is unequivocal trust and the belief in superiority of technology over humans ("functions better than individuals behind the wheel", E1f). B1m, on the other hand, is cognitively convinced of the vehicles' safety ("99.8%"), but affectively anxious. This is illustrated by his comparison between the buses and Chernobyl as the maximum credible accident and reason for his scepticism towards autonomous technology. In the situation of uncertainty because of autonomous technology, E1f reacts "naïvely" and trusts unconditionally, while B1m wishes for a controlling authority. Even if B1m rates the security level cognitively relatively high his trust towards ABs is low because of affective concerns.

3.2.4. Maintaining control, perceived safety, and reliability of the ABs

The previous passage illustrated the tension between the concepts control and trust. If uncertainty is a necessary premise for the definition of trust (Donick, 2019, p. 11; Lee & See, 2004) then trust and control are incompatible counterparts. This wish for control becomes more distinct in the following passages dealing with the presence of a human operator (passage 6) and an openly accessible emergency button (passage 7).

Passage 6:

B1m: Why don't I just as in a plane put a human inside instead? This is also a human being who has a job who earns money. That doesn't even have to be, is not highly qualified work, but as I said a good and solid job. And for all I care he is just sitting at the wheel and looks around or babbles with the passengers and entertains them, but I know he is there and for my salvation, like, if something evil happens: he is there! In principle I would really like that, it'd be still autonomous, but, you know…

[…]

Interviewer: What I withheld so far is that in our test phase here at the campus for one year a so-called operator rides along, a person, who is inside the car.

B1m: Thank God!

B1m uses religious language ("salvation", "evil", "God") to describe his relief about not being solely dependent on technology. This underlines the strong desire for control, in this case manifested through the professional human operator who metaphorically

arrives as saviour and practically serves as contact person and mediator between technology and passenger.

Passage 7:

C1f: Well, if the car drives and we see a cat, is there a possibility to press stop somewhere, as a passenger, to by ourselves make the thing…

Interviewer: That exists, yes.

C1f: Yes, that's good. That's good. It is good, right?

D1f: Yes, I don't have a problem getting in there. I don't have a problem with that thing at all. I think it's rather funny.

Interviewer: What if it would not exist? What would you think about it then?

C1f: Let's put it that way: I would, I'll try it, because I think it is interesting. It's not my means of transportation, but I would definitely try it. But it will be fun-, well, it will be interesting. Well, it would be thrilling, I think. And that I have the possibility to control it – maybe I am a control person, could be – that comforts me a lot. I would first of all look where this button is.

D1f: This now worries me that there is a button with which everybody sitting inside can live out their own sensitivities. No.

C1f: No, well, only in an emergency.

D1f: Well, the emergency… If there is a cat, someone gets a fright, because he loves cats. But the thing already recognises the cat.

C1f: Yes, he recognises it.

D1f: But if everyone has access to this button, and has a problem, well, have fun with it. Then I am not riding along.

C1f: Oh ok.

D1f: No, if everybody has access to it, I don't like it so much.

C1f is relieved by the possibility to exercise procedural control over the autonomous machine and describes herself as a control person. D1f, on the other side, sees this kind of control through humans as a threat that lowers her willingness to ride the AV. This signifies a sceptical attitude towards the human (unprofessional, arbitrary, and affect-driven) element in the human-machine interaction. Accordingly, the question who should control whom – humans the technology or the technology humans – is answered quite differently. On the one hand, technology supposedly performs better than humans never being sleepy, drunk, or in a bad mood (M3m, passage 1; E1f, passage 5) becoming an ideal corrective for human inadequacy (Gehlen, 1978). On the other hand, uncontrolled

technology produces devastating results (B1m, passages 5 and 6). However, humans as control mechanisms seem also inadequate if responsibilities are not clearly assigned (passage 6).

The debate about the button in passage 6 resembles the theory of Madhavan and Wiegmann (2007) who argue that humans are initially met with scepticism until they have proven their trustworthiness. This is contrasted by passage 8, in which technology must prove its trustworthiness as well. Here, participants differ in their extent to grant a-priori trust. For H2m, trust is not granted a priori when dealing with ABs. Instead, trust has to be earned through experience. J2m on the other hand describes a generally high a-priori trust that is however based on conditions of safety. Particularly, different speeds are discussed and a categorical difference between slow ABs and fast ABs is introduced.

Passage 8:

Interviewer: Do you trust the vehicles?

G2f: Yes.

H2m: I think trust is something that has to be proven first, something you have to get used to. I would not go as far and say I'm getting into this vehicle and trust it right from the start. Instead, I would rather gradually gain trust by driving it more often.

J2m: Well, up to this *speed I'd trust it, because I'm just imagining with my naïveté that, if there is any mistake, they'll stop anyway, yeah. Well, that's just what I'm talking myself into believing. As such, up to the speed they are driving now, 20 or 45 even, it would not be an issue for me. Higher up, I would get careful, because that is a whole new ball game.*

G2f: No, absolutely, at 20 I would also get in and say, yes, I'm curious, I'm open to it, I test it, I trust it.

Contrary to theoretical conceptualisations of trust (Madhavan & Wiegmann, 2007), trust is not the default for H2m when dealing with technology. Instead, repeated interaction builds trust similar to how Madhavan and Wiegmann (2007) conceptualise interpersonal trust. In the rest of passage 9, "naïve" curiosity is embedded in cognitive restrictions about safety. The "whole new ball game" (J2m) presents a categorical change of perceived risk from acceptable to unbearable. Trust is therefore not expressed unconditionally, but in a clearly defined spectrum of conditions.

3.2.5. Understanding the black box and anthropomorphisation

The technology of ABs is increasingly attributed to be equipped with human agency (Brand, 2018), because it changed from being trivial and predictable towards complex

and connected configurations outside the realm of understanding. In its effectiveness, technology becomes opaque; like a black box with observable inputs and outputs of the machine, but hidden mechanisms in between (Latour & Venn, 2002). In passage 9, participants express the wish to understand the "black box" by visualising the actions of the machine. Participants humanise the process of AV decision making and thereby translate the machine's algorithms into human categories to experience the world. This approach can be seen in the context of the development of technology from trivial to non-trivial (Donick, 2019). If the technology is non-trivial, more human agency is credited and trust is defined by repetition and transparency about intentions. Therefore, the proposed separation of the concepts of trust towards machines and trust between humans is challenged (Madhavan & Wiegmann, 2007). Passages 5, 7, and 8 are consistent with the view of trust and control as opposites in situations of uncertainty and self-efficacy. A contradicting viewpoint is presented in passage 9, which expresses that participants are more willing to trust ABs the more they have control over them, e.g. in the form of information.

Passage 9:

I2f: Well, I could imagine running in its way, just to see what happens.

G2f: To test it?

J2m: Is he knocking me over?

I2f: Well, with a safe distance. Then I would like to know how they work exactly and to what they actually react. And that would probably at least strengthen my trust, if I understood it better.

Interviewer: Yeah, what else would increase trust? Understanding the behaviour of the vehicle, I'd say.

G2f: Just write more about how safe they are, that a vehicle has been used again, this and this many people have used it, it is usable around the clock. Well, that's something I'd say, oh great, sounds good, it works, so you can use it. And, yeah, people with prejudices might be convinced too. Yeah, if that many use it I can get in as well.

H2m: Maybe also reviews of rather older people who trust the vehicle. I feel like young people are more open to such things and will do it right away. And when I read that an 80-year-old elderly lady gets in there and writes 'great' and that everything went well, then I'd rather trust the whole thing.

J2m: Another point for visually perhaps, well, those information, that's all already kind of happened. However, if you're sitting in there live, maybe some kind of visualisation would be, when he drives, just projected into the windowpane, like a head-up display:

these are my lines and I'm driving there. So, it's not totally unexpected, I say in quotation marks, 'what he does'. And if there is a car parking, then there should be like a red circle around it, so that you simply project into the windowpane what he thinks, what the computer thinks. Well, and then one has a sense of security, I say, or suggesting it. Such a, somehow such a visualization of like, what he does.

G2f: Or a voice that says, well, in five minutes we reach the station, I don't know, Virchowweg, corner, or here, I don't know, maybe that's important in the beginning. So, you have elderly, right, that they have the feeling, yes, there is someone who is watching all of this somehow.

J2m: Exactly, well, that… Also, in live operation as well, when you're sitting inside, not only that you've read a newspaper before and that means, yes, there was an 80-year-old woman in there, she thought it was great. But that, I say, yes, I call it live operation again, it helps me when I see something that's visually orientated and that's why I would say that would help me too. When I can sit there and then there are two yellow lin-…, like, similar to a backup camera, there is sometimes, a picture, a … camera projected inside, that would be, yeah, stuff like that.

Additionally, passage 9 describes different trust-building mechanisms. Information about the functioning of the vehicle is discussed on different levels – in direct experience, when and why the AV stops, or through visualisations of the route or the AV's "thinking" (J2m). Information reduces uncertainty of the situation and trust is less necessary than without information. Therefore, the desire for information can be interpreted as desire for control that is, theory-driven, contrary to trust (Lee & See, 2004). However, participants see information as helpful in building *trust*. This contradiction can be resolved by interpreting the feeling described as trust rather as a feeling of *comfort*. Trust itself is a strategy to feel comfortable in a situation of uncertainty and to give up responsibility.

Furthermore, anthropomorphisation of technology is used as strategy to reduce discomfort and as a third way to approach the ABs. In the participants' imaginations, the AV is capable of thinking (J2m), perceiving through senses, and it should be able to speak (G2f). For the passengers this anthropomorphisation of technology creates (imagined) transparency or simplicity of the machine and thus ensures greater trust in the ABs. This is underpinned by other evidence on anthropomorphisation and trust (Verberne, Ham, & Midden, 2012, 2015).

Lastly, this passage draws a line to behaviour or function (of humans and machines) and its communication. The participants realise that, similar to a human brain, the processes of the intelligent, non-trivial technology are not easily traceable. J2m and G2f both desire communication of the AV to elicit certain functions and rationales. While humans use language, gestures, etc. to explain behaviour, machines need another way of

communicating their functions. Participants imagine ways for ABs to communicate and thus understand the ABs' "intentions". This shows similarities in human-to-human and human-to-machine interaction. The will to avoid mysterious errors and malfunctions through communication to gain more trust also resembles the theory that a-priori trust in technology is high and decreases rapidly if errors occur (Hoff & Bashir, 2015; Madhavan & Wiegmann, 2007).

3.3. Summary

Our analysis revealed that a-priori trust (development) in automated buses is characterised by several factors as presented in the focus group discussions answering our first research question. First, participants used several comparisons to other mobility technologies (e.g., gondola, lift, or subway) to grasp the concept of ABs and familiarise themselves with its characteristics. Second, participants differed in their fundamental tendencies to approach or avoid the concept of ABs demonstrating curiosity or scepticism. This tendency is shaped by pre-existing knowledge (e.g., Chernobyl) or situational aspects (e.g., speed of the ABs) supporting the model by Hoff and Bashir (2015). Some participants desire more understanding of the black box behind the ABs while others are satisfied with their nescience. In either case, anthropomorphisation is used to feel comfortable with the "learning", "thinking", or otherwise perceiving ABs. Lastly, participants differ in their desire for control exerted on the ABs either in the form of emergency stop buttons or human operators. Here, technology is seen by some to surpass human performance in traffic and thus they see ABs as a solution to human errors (e.g., tiredness or drunkenness) while others desire human control at all times.

Our second research question addressed the fit between theoretical conceptions of trust and naïve conceptions by participants. Here, the model of Hoff and Bashir (2015) described three layers of a-priori trust, namely dispositional factors, situational variance, and pre-existing knowledge. We found evidence for all three factors in the focus group discussions. However, theoretical deliberations suggest that people grant machines a-priori trust whilst expecting perfect functioning; when experiencing errors trust deteriorates (Madhavan & Wiegmann, 2007). Particularly the inductive categories "ignorance and naivety" and "curiosity vs. scepticism" suggest that for many of the participants this is the case. Even those initially critical of ABs because of traumatic error experiences (e.g., Chernobyl, B1m) are willing to cognitively grant a-priori trust in the functioning of ABs ("99.8%" safety, B1m). In contrast, for H2m trust is inextricably linked to experience, i.e., trust has to be earned regardless whether the agent is human or

mechanic. Thus, we demonstrate that the assumption of high trust granted a-priori (Madhavan & Wiegmann, 2007) does not apply for all potential users. Lastly, trust defined as an attitude towards another agent in an uncertain situation (Lee & See, 2004) suggests that trust and control are opposites. Participants in the focus groups were ambiguous about this assumption. Some suggested an increase of control to *enhance* their trust. This supports theories about the demand for explanations of automation (Janssen, Donker, Brumby, & Kun, 2019), but stands in contrast to theories about uncertainty (Donick, 2019; Lee & See, 2004). This contradiction could be explained by diverging definitions of trust between theory and laypeople. In this case, participants might consider trust (because of reflexive inaccessibility) as a positive feeling of comfort with the ABs, for which control and information are beneficial.

3.4. Strengths, limitations, and future research

The qualitative approach in this study enabled an in-depth understanding of preconceptions about trust in ABs, and factors relevant for trust development. Our results suggest that at least for some participants trust assessments might not be only partially cognitively accessible. This finding questions the informative value of attitudinal trust conceptions in quantitative measures. Those measures arguably provide superficial information on a cognitive, reflexive *kind of trust*. They however disregard inaccessible parts such as the fundamental tendencies curiosity or scepticism as well as the level of information necessary to trust or feel comfortable with ABs. We thus enrich the corpus of predominantly quantitative academic literature on trust in automated mobility (Versteegh, 2019) with a qualitative study on a-priori trust in ABs.

However, the results of this study need to be interpreted considering limitations. One limitation is the sampling bias due to self-selection: although participants represented a broad spectrum of potential users in this use case, they may be particularly interested in ABs. At the same time, they are all hospital-associated persons with a centre of life in (the outskirts of) a large city. The use case represented a very specific scenario before ABs were employed in a protected environment. Hence, the results might not be generalisable to other forms of automated vehicles or trust after experience.

For future research, a focus could lie on other influences on a-priori such as media coverage. Reports, particularly those about the two deadly accidents involving a pedestrian and a Volvo XC90 on 18 March 2018 in Arizona and involving a Tesla Model X and a concrete barrier on 23 March 2018 in California which became caesuras in automated driving in USA (Boudette, 2018; Griggs & Wakabayashi, 2018) possibly affect

trust in ABs through additional knowledge about their safety (layer 3 in Hoff & Bashir, 2015). Our focus groups were conducted before these accidents occurred and thus could not capture this effect on a-priori trust.

4. Conclusion

This study demonstrates that qualitative methods enable a critical reflection to approach the question of a-priori trust in ABs and produce promising results that support and extend theories on trust (development). Our results suggest that potential users can build trust towards a previously unknown concept through different trajectories. Here, previous experiences, fundamental tendencies to be curious or cautious towards new technologies, or the degree of understanding the black box influence the level of a-priori trust between potential users. Providing similarities with existing technologies helps to grasp the concept of ABs. Many of those mechanisms were supported by theories of trust in automation, but none of the models includes all factors. Thus, we see a need in revisions of trust theories to include, e.g., affective components of trust. On the one hand, trust seems to be a combination of control beliefs, information, and cognitive assessments; on the other hand, it apparently includes affective components as well as dispositions to approach or avoid new technologies.

Acknowledgements

We thank the members of the interpretation circle at our institute – particularly Arlett Wenzel – for providing valuable feedback about our interpretations that increased the quality of our analyses. We would also like to thank our project partners Berliner Verkehrsbetriebe (BVG) and the Berlin Senate Department for the Environment, Transport and Climate Protection.

This work was supported by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit) under Grant 16EM3157-2. The funders had no role in study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Declaration of interest: none.

References

Bazilinskyy, P., Kyriakidis, M., & de Winter, J. C. F. (2015). An international crowdsourcing study into people's statements on fully automated driving. *Procedia Manufacturing, 3*, 2534-2542. doi:10.1016/j.promfg.2015.07.540

Boudette, N. E. (2018, 31 March). Fatal Tesla crash raises new questions about autopilot system. *The New York Times.* Retrieved from https://www.nytimes.com/2018/03/31/business/tesla-crash-autopilot-musk.html

Brand, L. (2018). *Künstliche Tugend: Roboter als moralische Akteure*: Verlag Friedrich Pustet.

Buckley, L., Kaye, S.-A., & Pradhan, A. K. (2018). A qualitative examination of drivers' responses to partially automated vehicles. *Transportation Research: Part F, 56*, 167-175. doi:10.1016/j.trf.2018.04.012

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. doi:10.1080/10447318.2015.1070549

Donick, M. (2019). *Die Unschuld der Maschinen. Technikvertrauen in einer smarten Welt*. Wiesbaden: Springer.

Gehlen, A. (1978). *Der Mensch. Seine Natur und seine Stellung in der Welt* (12. ed.). Wiesbaden: Akademische Verlagsgesellschaft Athenaion.

Goto, S. G. (1996). To trust or not to trust: Situational and dispositional determinants. *Social Behavior and Personality: an international journal, 24*(2), 119-131. doi:10.2224/sbp.1996.24.2.119

Greenblatt, J. B., & Saxena, S. (2015). Autonomous taxis could greatly reduce greenhouse-gas emissions of US light-duty vehicles. *Nature Climate Change, 5*(9), 860-863. doi:10.1038/nclimate2685

Griggs, T., & Wakabayashi, D. (2018, 21 March). How a self-driving Uber killed a pedestrian in Arizona. Retrieved from https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html

Henderson, J., & Spencer, J. (2016). Autonomous vehicles and commercial real estate. *Cornell Real Estate Review, 14*(1), 44-55.

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, 58*(3), 509-519. doi:10.1177/0018720815625744

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407-434. doi:10.1177/0018720814547570

598 Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of
599 human-automation interaction. *International Journal of Human-Computer*
600 *Studies, 131*, 99-107. doi:10.1016/j.ijhcs.2019.05.006

601 Kaur, K., & Rampersad, G. (2018). Trust in driverless cars: Investigating key factors
602 influencing the adoption of driverless cars. *Journal of Engineering and*
603 *Technology Management, 48*, 87-96. doi:10.1016/j.jengtecman.2018.04.006

604 Kuckartz, U. (2016). *Qualitative Inhaltsanalyse. Methoden, Praxis,*
605 *Computerunterstützung* (3. ed.). Weinheim und Basel: Beltz Juventa.

606 Latour, B., & Venn, C. (2002). Morality and technology. *Theory, Culture & Society,*
607 *19*(5-6), 247-260. doi:10.1177/026327602761899246

608 Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance.
609 *Human Factors, 46*(1), 50-80. doi:10.1518/hfes.46.1.50_30392

610 Li, S., Blythe, P., Guo, W., & Namdeo, A. (2019). Investigation of older drivers'
611 requirements of the human-machine interaction in highly automated vehicles.
612 *Transportation Research: Part F, 62*, 546-563. doi:10.1016/j.trf.2019.02.009

613 Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-
614 human and human-automation trust: An integrative review. *Theoretical Issues*
615 *in Ergonomics Science, 8*(4), 277-301. doi:10.1080/14639220500337708

616 NHTSA. (2017). Automated driving systems 2.0: A vision for safety. *National Highway*
617 *Traffic Safety Administration & US Department of Transportation.* Retrieved
618 from https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-
619 ads2.0_090617_v9a_tag.pdf

620 Nordhoff, S., de Winter, J., Payre, W., van Arem, B., & Happee, R. (2019). What
621 impressions do users have after a ride in an automated shuttle? An interview
622 study. *Transportation Research: Part F, 63*, 252-269.
623 doi:10.1016/j.trf.2019.04.009

624 Nordhoff, S., de Winter, J. C. F., Madigan, R., Merat, N., van Arem, B., & Happee, R.
625 (2018). User acceptance of automated shuttles in Berlin-Schöneberg: A
626 questionnaire study. *Transportation Research: Part F, 58*, 843-854.
627 doi:10.1016/j.trf.2018.06.024

628 Nordhoff, S., Stapel, J., van Arem, B., & Happee, R. (2020). Passenger opinions of the
629 perceived safety and interaction with automated shuttles: A test ride study with
630 'hidden'safety steward. *Transportation Research: Part A, 138*, 508-524.
631 doi:10.1016/j.tra.2020.05.009

632 Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse,
633 abuse. *Human Factors, 39*(2), 230-253. doi:10.1518/001872097778543886

634 Pettigrew, S. (2017). Why public health should embrace the autonomous car.
635 *Australian & New Zealand Journal of Public Health, 41*(1), 5-7.
636 doi:10.1111/1753-6405.12588

637 Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary
638 roots, a fact that may explain their complexity and provide tools for clinical
639 practice. *American scientist, 89*(4), 344-350.

640 Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in smart systems: Sharing
641 driving goals and giving information to increase trustworthiness and
642 acceptability of smart systems in cars. *Human Factors, 54*(5), 799-810.
643 doi:10.1177/0018720812443825

644 Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2015). Trusting a virtual driver that
645 looks, acts, and thinks like you. *Human Factors, 57*(5), 895-909.
646 doi:10.1177/0018720815580749

647 Versteegh, M. L. J. (2019). *Trust in automated vehicles: A systematic review.* (Bachelor
648 of Science), University of Twente, Twente.

649 Wadud, Z., MacKenzie, D., & Leiby, P. (2016). Help or hindrance? The travel, energy
650 and carbon impacts of highly automated vehicles. *Transportation Research:*
651 *Part A, 86*, 1-18. doi:10.1016/j.tra.2015.12.001

652 Wagner, G. (1994). Vertrauen in Technik. *Zeitschrift für Soziologie, 23*(2), 145-157.
653 doi:10.1515/zfsoz-1994-0205

654 Wu, K., Zhao, Y., Zhu, Q., Tan, X., & Zheng, H. (2011). A meta-analysis of the impact
655 of trust on technology acceptance model: Investigation of moderating influence
656 of subject and context type. *International Journal of Information Management,*
657 *31*(6), 572-581. doi:10.1016/j.ijinfomgt.2011.03.004

658 Zoellick, J. C., Kuhlmey, A., Schenk, L., Schindel, D., & Blüher, S. (2019a). Amused,
659 accepted, and used? Attitudes and emotions towards automated vehicles, their
660 relationships, and predictive value for usage intention. *Transportation*
661 *Research: Part F, 65*, 68-78. doi:10.1016/j.trf.2019.07.009

662 Zoellick, J. C., Kuhlmey, A., Schenk, L., Schindel, D., & Blüher, S. (2019b). Assessing
663 acceptance of electric automated vehicles after exposure in a realistic traffic
664 environment. *PLOS ONE, 14*(5), e0215969. doi:10.1371/journal.pone.0215969

665