
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Koskinen, Kari; Lyyra, Antti; Mallat, Niina; Tuunainen, Virpi

Trust and risky technologies: Aligning and coping with Tesla Autopilot

Published in:

Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS 2019)

Published: 01/01/2019

Document Version

Publisher's PDF, also known as Version of record

Published under the following license:

CC BY-NC-ND

Please cite the original version:

Koskinen, K., Lyyra, A., Mallat, N., & Tuunainen, V. (2019). Trust and risky technologies: Aligning and coping with Tesla Autopilot. In *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS 2019)* Hawaii International Conference on System Sciences. <https://hdl.handle.net/10125/60013>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Trust and risky technologies: Aligning and coping with Tesla Autopilot

Kari Koskinen
London School of
Economics
k.m.koskinen@lse.ac.uk

Antti Lyyra
London School of
Economics
a.k.lyyra@lse.ac.uk

Niina Mallat
Aalto University School
of Business
niina.mallat@aalto.fi

Virpi Kristiina Tuunainen
Aalto University School
of Business
virpi.tuunainen@aalto.fi

Abstract

Products are increasingly digitized, and they incorporate digital components, smart features and partial automation. Modern cars are a prime example of consumer-oriented automation; they sense the environment and perform specific driving tasks on the driver's behalf. The driving assistance and safety features provided by automation are under constant development, and as these features evolve, drivers experience and learn about their capabilities as they use them and develop their trust in automation in the light of new experiences and information. In this paper, we present a study on how trust in car automation unfolds as users gain experiences and information that conflicts with their expectations concerning the level of automation. We use Tesla Model S car as our case technology and explore how its users develop their trust and cope with issues with the novel automation technology. Our findings suggest important directions for future research of consumer-oriented automation and digitized products.

1. Introduction

Digital transformation is rapidly changing the products we use in our everyday lives. Our phones, computers, tablets and loudspeakers understand our talk, communicate back to us and have a surprisingly wide repertoire of skills to entertain or assist us in everyday tasks. Our vacuum cleaners need no steering but navigate around the house autonomously and return back to base for recharging before the battery runs out. Our cars can sense the traffic around them, assist us in braking, lane-keeping and collision detection and some cars may even park themselves. Even better, some of these products are continuously improved through regular software updates. But therein lies a trade-off. In the old world, when we bought mechanical products carefully manufactured,

tried and tested in different conditions, what we saw was often very close to what we got. With the new wave of digitized and connected products, we may buy into the future-looking visions and desires yet end up with unfinished products and continuous updates [1].

Digitized and connected products can be defined as products that are composed of both physical and digital components and connected to communication networks [2] [3]. The incorporation of computer-controlled functionality and connectivity enables the cost-efficient modification of product functionality remotely and over time without changing the physical embodiment of a product [2] [3]. Unlike the physical embodiments that tend to receive their final form by the end of manufacture, the computer-controlled functionality remains more open-ended. The reprogrammability of digital components allows for frequent change, and, this way, digitized and connected products are malleable and can remain constantly in the making [4] [5]. Digitisation and connectivity also provide means for gathering real time data on the use and behavior of a product, and this data can then be used to develop the product further.

From a buyers' perspective, this open-endedness means in practice that users are increasingly buying products whose functionality, utility and reliability remain to some extent unknown while changing over time. Users are therefore subjected to a degree of uncertainty and vulnerability, yet they are also expected to frequently revise their expectations concerning the product functionality as updates are released and new experiences and information emerge. The presence of uncertainty and vulnerability assumes that [6] trust plays a prominent role when users decide whether to rely or use certain functionalities.

Modern cars with autonomous driving features provide a prime example of digitized and connected products in the making [1] [2]. Although car manufacturers adhere to different development, deployment and marketing strategies, they converge at their aims to introduce more sophisticated autonomous

driving features to assist the driver or to replace the driver altogether.

Unlike airline pilots or other professionals, regular drivers do not receive training for using the automated driving features. Instead, they form their expectations based on marketing, test drives and any other available information before the purchase and then are expected to adjust the expectation based on experience and new information. Through experience, drivers may learn how to maneuver a car with its autonomous features, especially when encountering situations and issues and after any feature updates. Yet the misaligned expectations regarding the capabilities of driving automation may put the driver, passengers and other road traffic in danger [7]. Therefore, the role of trust in car automation can be viewed as accentuated.

Several studies have shown the importance of trust in the adoption and use of automation [6] [10] and autonomous driving technologies [8] [9]. While the earlier literature on human factors and automation has focused on task allocation [11] and the interplay between the operator and technology of automation from various perspectives in professional and controlled settings such as space and aviation industries [7], there is a little research on how consumers form trust in the autonomous driving technologies that are incomplete and under development.

In this paper, we study the formation of trust in the view of consumer-oriented automation and with reference to digitized and connected products which are constantly in the making using the Autopilot functionality of Tesla Model S as a case study. Empirical evidence is gathered from online discussions forums where Tesla drives share their experiences and discuss the capabilities and functioning of the Autopilot functionality.

To this, our research question can be summarized as follows: How does trust in automation unfold in the context of digitized and connected cars, which are constantly in the making?

The subsequent chapter reviews the related literature on trust providing conceptual framing. Chapter three presents the case Tesla as a case to study and describes collection and analysis procedures. Then, chapter four presents the empirical findings, and their implications are discussed in chapter five. The concluding remarks are in chapter six.

2. Literature review

This chapter outlines the conceptual framing that provides the backdrop upon which the formation of the trust in automation can be reflected.

2.1. Trust as irrational risk taking

Trust plays a prominent role in human action. We habitually extend trust and expect trustworthiness as we depend on the competence and good intentions of others [12]; we trust others to act on their commitments, yet misplaced trust may bring serious consequences or prove even fatal. The importance of trust in social interactions and material exchanges has attracted attention in various fields of research, ranging from sociology [13] [14] to organization studies [15] and from psychology [12] to human factors and automation studies [6]. This has resulted multiple definitions of trust, such as faith, expectations, predictability, familiarity, risk, reliance and dependence in association with their antecedents and consequences under different conditions.

Luhmann [13] postulates that trust is needed when embarking on ventures that depart from the continuity of the familiar. In this view, trust is intimately associated with uncertainty and possibility of unwanted outcomes. Decision-makers generally aim at positive outcomes by evaluating the degree of uncertainty and expectations regarding some future states and events [16], but events often unfold in an unforeseen manner. Although some specific risks can be quantified by weighing probabilities of positive and negative outcomes with their respective gains and losses, many future states and events remain fundamentally uncertain for all practical purposes; they transcend rational analyses given the complexity of environmental conditions, lack of information and limitations of our cognitive faculties [17]. However, the lack of purely calculative and rational analysis in decision-making does not prevent us from making choices. We routinely make decisions without appreciating all possible consequences [18].

To emphasize the connection between the trust and decision-making, Luhmann [13] [14] differentiates between non-choice (confidence) and choice (trust). Confidence belongs to the realm of situations and relationships a person is automatically part of but has no control over the future states and events. Trust, in turn, applies to situations where the future states and events remain fundamentally uncertain and risky, yet there is a real possibility to refrain from entering into a situation that puts the decision-maker into a vulnerable position, while simultaneously forgoing potential benefits and gains attainable by opting in.

When risks cannot be fully analyzed [17], trust black-boxes the impenetrable complexity and renders a foundation upon which decisions can be based and made [14] [13]. Therefore, trust, regardless of how it is formed or is constructed, allows us to act and make decisions. Trust indicates readiness for risk taking,

which is not necessarily rational in the traditional sense of the word.

2.2. Trust in technology and automation

A decision to depend on automation requires trust. The promise of automation is that a product or system performs tasks on the user's (operator) behalf; automation can be defined to refer "*the full or partial replacement of a function previously carried out by the human operator*" [11] (p. 287). This brings forward the functional comparability between the human and automated control of operations, and it also indicates that the scope and capabilities of automation are contingent upon tasks, technologies and context.

As the scope and capabilities vary, a user should be able to trust that the reliance on automation (agent) would produce a favorable outcome. This is formulated by Lee and See [6] as follows: "*[t]rust can be defined as the attitude that an agent will help achieve and individual's goals in a situation characterized by uncertainty and vulnerability*" (p. 51). Therefore, trust is viewed to precede a decision to rely on automation that puts the user potentially in the harm's way.

Occasionally, trust in automation is considered as equivalent to trust in humans, however, research on decision support systems shows that people respond differently to human and computer-based assistants [19]. Several differences also emerge from the ways how technologies are designed, developed and regulated and how they are marketed to consumers and discussed in user communities, bringing a variety of organizational, institutional and sociological aspects into consideration [20]. In addition, technologies are often black boxes; their internal functioning remains hidden from and inexplicable to their users, especially as technologies grow in complexity. Instead of being able to scrutinize the internal processes and underlying logic, the users are only able to experience the outcomes they produce [21]. Therefore, they come to evaluate the trustworthiness of some particular technology, its competence and good intentions [12], based on their experiences and the information they gather from the manufacturer and other users. All this affects the extent to which users trust in automation and how they may or may not put it to use.

Parasuraman and Riley [22] differentiate between the misuse, disuse and abuse of automation. Misuse refers to situations where excessive reliance is placed on automation without appropriate scrutiny, whereas disuse has to do with situations where automation becomes under-utilized or neglected. Abuse, in turn, refers to design practices where no sufficient attention is paid to how a product or system and its operator will

interact, leading potentially to substandard automation and interaction designs, which create environments and situations that feed to both the misuse and disuse of automation. Ideally, to avoid misuse and disuse, the expectations concerning the capabilities of automation and the actual capabilities of automation should be aligned.

2.3. Formation of trust

The relationship between the expectations and capabilities of automation and the formation of this relationship can be viewed through the lenses of alignment and coping.

Alignment [6] deals with the extent to which a person's trust in the automated system is aligned with its actual capabilities, that is, the alignment between the perceived and actual trustworthiness. Ideally, these two should be the same. However, this is not often the case leading to the situations that can be characterized by overtrust or distrust. Overtrust occurs when the level of trust exceeds the capabilities of the system, and distrust indicates a situation where the level of trust is below the system's capabilities. Overtrust is prone to lead use to misuse and overreliance, whereas distrust may lead to underutilization.

The alignment between the level of trust and system capabilities can be approached through the process of calibration [6], which assumes that users adjust the level of perceived trustworthiness in the light of new experiences and information. Functional and temporal specificity and resolution influences the process of calibration [6]. Resolution concerns how a change in the capabilities of automation reflects on the perceived level of trustworthiness. Overall, if resolution is low, a major change in the capabilities of automation may produce only minor change in trust. Furthermore, functional specificity refers to what extent the capabilities of specific automated functionalities are recognized. If a user is able to direct trust at specific functionalities and capabilities, functional specificity is high. In turn, if the level of trust reflects the system as a whole, functional specificity is low. Moreover, specificity refers also to temporality. High temporal specificity indicates instant reactions to changes in automated capabilities, whereas low temporal specificity signals a delay between the changes in the capabilities of automation and the subsequent adjustment of the levels of trust. Overall, calibration, resolution and specificity provide lenses through which the formation and alignment of trust can be approached.

Coping, in turn, shifts the focal point from the alignment and process of calibration to cognitive and behavioral efforts that users exert to overcome the

situations that are taxing or exceed the resources at their disposal [33]. This may happen when users are not able to calibrate and align their expectations towards automation yet are being left to deal with a problematic experiences and situations. Although a user's decision to exert or withdraw trust can be viewed as a consumer's risk-coping mechanisms [23], Pearlman and Schooler [24] offer a more fine-grained conceptualization. They describe three categories of coping mechanisms, which are activated depending on the situation and problematic experience. The first category refers to situations where coping is carried out by eliminating or modifying the conditions that give rise to the problematic experience, and this could be viewed as a decision to extend or withdraw trust. The second one takes the form of perceptually controlling the meaning of the experience to neutralize its problematic character. This can, for example, take the form of comparison in which one's experience is compared to those of others thus reducing the uniqueness of the experience and also the stress following from that experience. Alternatively, the problematic aspects of the experience can be selectively ignored or reasoned way [25]. The third category of mechanisms refers to the controlling of emotions to keep the stress and emotional consequences of the experience under control. Instead of eliminating the problematic experience or controlling its meaning, the third category acknowledges the problematic nature of the experience, and the coping takes the form of a person accepting the experience as something inevitable one has simply to live with. Stress therefore becomes something that cannot be avoided but can only be controlled while hoping that somehow rewards would follow. This is similar to Luhmann's [13] notion of confidence which characterizes the settings and situations one is automatically part of but has no control over the future states and events. Thereby, coping can take place through the elimination of a problematic experience, the removal of cognitive dissonances or accepting a situation as inevitable.

To conclude, the alignment of the perceived and actual trustworthiness [6] [25] along with the three mechanisms of coping [24] to overcome the lack of alignment offer sensitizing devices [26] with reference to which the formation of trust in automation can be examined.

3. Method

This study into the formation into trust in automation in the context of digitized and connected products is designed as an exploratory case study [27] [28]. We chose Tesla Model S car as our case study as

Tesla is often considered as one of the most prominent manufacturers of automated driving. The company markets the cars with the future-looking promises of fully autonomous driving as the capabilities of cars are gradually improved and updated.

The first Model S cars were delivered to customers in June 2012, and the car serves as a prime example of a digitized and connected product that incorporates consumer-oriented automation in the form of the Autopilot functionality [1]. Tesla's efforts to augment and automate driving are often subsumed under the label Autopilot, which, broadly speaking, is a general referent for a suite of functions that provide safety, driving assistance and self-driving features and functionalities; it consists of a set of capabilities each geared to serve a specific functional purpose. Currently, in terms of SAE International's Levels of Driving Automation for On-Road Vehicles, the most advanced feature of Tesla's Autopilot package (Autosteer (Beta)) reaches the level two on the scale of one to five, where one stands for driver assistance and five for full automation [29]. The company states that their long-term goal is to increase the level automation incrementally and to reach eventually the level of full automation.

Thereby, the Autopilot functionality is under continuous development. Over time, Tesla has developed and introduced new functionalities for automated driving and increased their capabilities. Updates to the software that controls the Autopilot functionalities are frequently released. In 2016, the company started equipping cars with a more capable set of sensors and computing hardware to pave the way towards more demanding scenarios of automated driving.

The Autopilot functionality and the utilization of frequent software updates render the Model S as a revelatory case, although it could also be considered as a representative case as it belongs to well-established product group of premium sedans [27].

3.1. Data collection

Data was collected from the Tesla Motors Club (TMC) discussion forum. TMC was formed in 2006 and was the first independent online Tesla community. Today it remains the largest community of Tesla users and enthusiasts. The forum features discussions dating from today to the early days of first Tesla releases. The structure of the forum and the discussions is well organized and covers a wide variety of topics. Many TMC posters own a Tesla and have first-hand, real experience of the technology.

For our study, the postings on the forum offered a rich, first-hand description from users who had already

formed an initial trust in the automation technology and were now reflecting upon the trustworthiness based on their own experiences and information from others shared in the forum. This enabled us to examine through narratives from the users how their trust in automation were manifested as new experiences and information emerged.

To narrow the data down to topic areas related to trust in the view of automated driving, we focused on discussion threads covering incidents and accidents. We first created a shortlist of 53 threads selected based on the topics of threads. Of these, we selected three threads for further examination, one of which discussed a crash reported in the news and two which were personal accounts of accidents a driver had experienced and reported on the forum. We decided to focus on threads that reflect everyday experiences and have a sufficiently large number of postings to obtain a rich set of data. We excluded discussions of the two published fatal accidents in USA and China as extreme cases.

Table 1. Description of the selected discussion threads

	Time period	Posts	Posters
Thread 1	23/12/16 - 25/1/17	100	50
Thread 2	16/5/16 - 5/8/17	358	89
Thread 3	13/1/16 - 9/3/16	593	162

Summary details of the threads are listed in Table 1 above. All three threads were initiated and mostly commented during 2016 they were thus comparable in terms of the maturity of technology discussed and public information about the technology and incidents available at the time.

3.2. Data analysis

We extracted the selected discussion threads from the discussion forum into offline documents with a script and then loaded them into Atlas.ti 8.0 software, which is a tool for qualitative data analysis.

For the two first threads (Thread 1 and Thread 2), we analyzed the postings with an interpretative clustering method in which text excerpts that have similar patterns or characteristics are grouped and then developed into conceptualizations [30]. During the first round of coding, we identified themes such as incidents that had happened, perceived causes for those incidents and thus causes for distrust, mechanisms to cope with the distrust, and development trajectories in the user trust toward automation technology. Within each coded quote we

also coded each poster separately to be able to analyze similarities and differences at the user level.

During the second round of coding, the coded quotes were compared, and the text documents were reread in order to check that the codes were coherently applied. At this time, some quotes were re-coded, some codes deleted, some re-named and some new quotations were included.

Thread 3 was used to confirm the results from the two first threads and to ensure the applicability of the generated codes to a similar thread.

4. Findings

This chapter outlines the content of the analyzed threads and reports the themes and concepts identified during the data analysis.

The first of the analyzed threads thread started with a Tesla driver describing an accident that he had been involved in: *"A week ago I was headed to work in my Model S and the AP [autopilot] pulled me into a truck driving by my side. I was driving into the sun so had the AP on to assist with my safety. Cruising along just fine until all the sudden I hear an impact and a crunch to my right. My car had pulled me into a Semi. Fortunately I had my hand on the wheel and was able to pull away from the truck before going completely under it."*

The second started with a poster quoting a news article which reported a recent Tesla accident. Arianna Simpson had been driving her Model S north from Los Angeles on I-5, cruising in the autopilot mode: *"All of a sudden the car ahead of me came to a halt. There was a decent amount of space so I figured that the car was going to brake as it is supposed to and didn't brake immediately. When it became apparent that the car was not slowing down at all, I slammed on the brakes but was probably still going 40 when I collided with the other car."* The news article reported that Tesla's log showed the autopilot mode had been disengaged at the time she manually hit the brake.

In both cases, the threads continued with different posters discussing the perceived causes of these accidents while also sharing similar experiences and reflecting how they felt about Autopilot after learning from these incidents and their own experiences.

The third thread reported a minor collision with another car after an Autopilot update and due to misaligned expectations: *"My car slowed as well. But when the car in front of me came to a complete stop (not a sudden emergency stop, but rather a gradual stop), I expected my car to do the same (as it had been doing previously). It didn't. I slammed on the brakes in that dreadful instance before I realized my car wouldn't stop in time, but I still hit the car in front of*

me. “ This was followed by a discussion on whether the autopilot can be trusted or not, many posters reminding others to be aware that *“The car doesn't drive itself, it's driver assistance, not self driving.”*”

The initial analysis confirmed our expectations of Tesla as a digital and connected product that undergoes continuous development and change. Also, it was clear from the threads that not the drivers' expectations concerning the capabilities of driving automation and the actual technological capabilities were not fully aligned. Many reported having encountered unexpected situations and issues that differed from their initial expectations. The drivers were constantly learning and forming their expectations and trust towards the technology as they used and learned from Autopilot. Also, many users also acknowledged the “beta” status of Autopilot and expected the autonomous features to improve over time.

4.1. Perceived causes for misalignment and distrust

We first analyzed perceived causes of these incidents to identify potential causes for misalignment and distrust. This gives important information for understanding how to improve the alignment and to overcome distrust. Our data indicated that the users were divided on why incidents happened, others blaming the driver and others the technology.

The human driver. Some posters saw the human drivers as root cause for all issues. In the end, they should be able to control the car in all circumstances. Further arguments calling for driver responsibility outlined that sufficient documentation and information is available for the drivers to be aware of the issues and to adequately prepare for them.

Several posters stated that the driver is responsible of the system at all times. The driver needs to pay attention to the system and environment and be prepared to take over at any time. This line of argument maintained that the driver is to blame for any and all incidents - not the system. *“Classic cruise control, TACC [Traffic-Aware Cruise Control], AutoPilot ... it does not matter - in all cases at all times the driver is responsible for the safe movement of the vehicle. The failure is with the driver.”*

Moreover, some argued that current limitations of the technology are documented well enough in Tesla user manual and/or is flagged in warnings when users turn on Autopilot. Drivers just need to be familiar of these to know how the technology actually performs and what is expected from them to prevent accidents. *“Her case is described in the manual very clearly: 'Warning: Traffic-Aware Cruise Control can not*

detect all objects and may not brake/decelerate for stationary vehicles, especially in situations when you are driving over 50 mph (80 km/h) and a vehicle you are following moves out of your driving path and a stationary vehicle or object, bicycle, or pedestrian is in front of you instead. [...]’ The problem is who would want to read the manual?”

Some also argued that the driver needs to be aware of “known” bugs and issues, the system and its design as well as be aware that the system is still “beta” and not working perfectly in all situations. The drivers were seen to be responsible to seek information of the system and take caution when using it. *“But if the technology is functioning as designed, whereas the design has known and acknowledged limitations, how is it a technology 'problem'? TACC, AEB, Autosteer have limitations, but I don't think it's correct to characterize those as 'flaws' or 'problems'. Those technologies are doing exactly what they're supposed to do.”*

The other camp countered the arguments regarding the drivers' responsibility and maintained that as the system performs well most of the time requiring little control, people learn to trust it too much. They wouldn't pay enough attention all the time to be prepared to take control in situations when the system suddenly does not behave as expected. *“Yeah, the fact that it works really well for long stretches of freeway, and then suddenly fails for a variety of reasons, is the real problem. It really does lull you into complacency. It is all very nice to get up on your high horse and say that people who don't monitor the road when AP is running are losers, stupid and deserve to die. Our brains, though, are wired to adapt to situations. When you drive 100 trouble free miles, it is human nature to get sloppy.”*

Some argued that the reasons behind human errors and misaligned expectations were linked to Tesla's marketing practices and the lack of driver training. In terms marketing, it was commented that Tesla provides too optimistic a picture of the car's performance in its marketing materials and during the sales process. In terms of training, some argued that Tesla would need to educate new drivers better about the capabilities, known limitations and issues so that the drivers would better informed and prepared to prevent accidents. *“The driver is 100 percent responsible for the vehicle, 100 percent of the time. If they don't understand this, they likely should not be driving any vehicle. In my opinion Tesla needs to do a much better job explaining this before delivery. Clearly some drivers will need additional training/education that outlines the limitations of the TACC/Autosteer before they start using it.”*

The technology. Many argued that the technology is not "good enough" or needs to improve. These comments critiqued the technology and the way it has been designed as a reason for the incidents. Some posters criticized some specific technical feature, such as Autopilot, whereas some included the Tesla brand in the conversation and argued that some features of Tesla technology are not as advanced as some other brands in the market. *"This was caused by the number 1 problem of autopilot. Not knowing whether or not it is engaged. Twice now I've been cruising on the freeway thinking AP was on when it wasn't. There are many events which cause AP to turn off once you turn it on. The only indication Tesla gives you that AP has turned off is a relatively quiet chime (and a very small icon that turns from blue to grey). This isn't enough of a warning that AP is no longer active. Tesla needs to address this now."* *"[...] and yes, for all the great stuff Tesla has today, it's got some catching up to do with others – at least from my POV having owned and experienced several vehicles with variants on the "automated safety assist" capabilities."*

However, there were several counter arguments in favor of the technology as well. It was maintained that regardless of the limitations the technology is safer than human drivers and reduces accidents; or it is just a driver assist feature and should be treated as such rather than semi-autonomous feature; or that the current state of technological development does not enable better functionality and therefore performs as well as it can shifting responsibility back to the human driver. *"The car is not supposed to drive itself. Do you own an AP equipped Tesla? The thinking that the car is really autonomous is a big problem. AP is really driver's assist and not autonomous. Read the docs."*

These causes for misalignment and distrust could be seen as largely external to the commenters themselves, i.e., in most cases, these were comments made on why a certain event had occurred and who was to blame. However, in the comments on how users dealt with these issues, the focus was more on users reflecting their own behavior and trust towards technology. In the following section we will look at how the different levels and resolutions of trust alignment resulted in particular coping behavior in relation to automated driving.

4.2. Formation of trust and coping behaviors

Many users shared their own experiences about incidents or narrow escape situations when using the autonomous driving functions of their Tesla cars: either Autopilot (AP) or Traffic-Aware Cruise Control (TACC) or both. While the technology is the same and the described incidents were similar, we were able to

distinguish three types of users whose perception of the trustworthiness of the technology developed into different directions after incidents. We named these groups *Trust*, *Low Trust* and *No Trust* and describe each of them in more detail below.

Group 1: Trust. The users in this group agreed that the technology is imperfect and does not always work flawlessly. Despite the incidents these users had encountered they still felt positive about using the automation technology and described ways in which they had learned to use the technology to prepare for and avoid the incidents. These users further maintained that they understood when and where the technology could fail and could therefore manage these situations. This had not always been the case from the start but rather an outcome from learning by using.

The quote below is illustrative of the users in this group. The poster first describes a situation where the automation may not function as expected, describes the poster's strategy for that situation which indicates confidence in being able to successfully manage the situation and gives an indication that this has been learned over time. Finally, the poster highlights satisfaction with the automation technology.

"I've had several instances where I've felt the need to brake myself. 100 percent of the time it has been where there is a stopped (or very slow) vehicle ahead and my speed is above 40mph. If my vehicle doesn't slow down once that vehicle appears on the display, I prepare to brake. Perhaps TACC would've kicked in but when it gets uncomfortably close, I step in just like I did when using dumb cruise control. My new way of driving: turn on TACC and Autosteer (if available), watch the flow of traffic, take over, repeat This has not lessened my AP enjoyment in the least. In fact, I see AP as providing the best of both worlds – autonomous driving when it is boring (start/stop traffic and long monotonous stretches) and self-drive the rest of the time."

It was clear that for this group of users their own experiences as well as those shared in the discussion forum had not diminished their trust in the technology. On the contrary, through experience and learning they had been able to develop preconceptions on how the technology may behave in different situations and then prepare their own actions as a strategy for those situations, thus growing more confident that they are in control of the situations and can cope with using the automation.

Over time, this group of users were able to calibrate and align their expectations of technological capabilities with sufficient functional specificity and resolution [6]. This group resorted to the second coping mechanism by Pearlin and Scouler [24] to

control the meaning of their experiences by developing an understanding of situations where the technology could fail. This way, they are able to plan their own actions for those situations, which gives them a sense of gaining control while reducing stress. Through the alignment of expectations and the control of the meaning of their experiences, users were able to maintain or even increase their trust in automation.

Group 2: Low trust. The users in this group agreed that the technology is imperfect and has its faults, and they had experienced similar issues as the first group. These users, however, were significantly less confident in understanding when and how the technology could behave unexpectedly and felt they needed to be more cautious when continuing to use the technology.

"I was one of the biggest fans and advocates of AP and TACC, and have seen it stop safely 98 percent of the time. But these highly reported accidents, together with many unreported others, and my own experience have lead me to now be more cautious and wary of the features, as there definitely are a few cases in which it just doesn't respond adequately." "Personally I view the Tesla autopilot [as] a darn good capable and impressive system but still one that is always looking for new ways to kill me or damage the car so it can't be fully trusted. I don't think that opinion will change until such time as FSD [full self-driving] has been implemented and has years of demonstrated success."

When compared with the first group, the key differentiating factors for group two were lower confidence in being able to foresee unexpected situations, demanding a higher level of alertness when using the technology, which appeared to lead to higher levels of stress. However, this group continued to use the automation technology while trying to manage the stress and uncertainty, but with lower level of trust than group one.

Unlike the group one, it appears that this group of users was not able to calibrate and align their expectations of technological capabilities with sufficient functional specificity and resolution [6]. Therefore, they appeared to resort to the third coping mechanism by Pearlin and Schooler [24] by seeking to control their emotions to keep the stress and emotional consequences of the experience under control. They had accepted the unexpected behavior of the automation as inevitable and controlled the resulting stress by highlighting the need for cautiousness and alertness on the road, yet they had not been able to develop an understanding of technological capabilities to reduce the stress by gaining the feeling of control.

Group 3: No trust. The users in this group reported experiences of incidents similar to the two other groups but differed in that their accounts

described the lowest level of trust. They expressed their disappointment with the technology as they learned from their own negative experiences and the information provided by others. They commented the faults in the technology and felt that it was very unpredictable. Some had stopped using autonomous features in particular situations or did not use them at all anymore or regretted purchasing them.

"My 'truck lust' experiences are much more frequent than that. So much so that I usually disengage AP until I pass the truck to avoid the stress." "AP is definitely not the safety feature I thought it was on purchasing and with 20:20 hindsight I'd not spec AP1 or AP2 for a current purchase - too many unknowns related to future performance"

It appears that the expectations regarding the trustworthiness of the technology were not aligned with the technological capabilities nor calibrated in the light of new information [6]. Users in this group did not feel they were in control in the same way as the users in the *trust* group, and they could not manage the stress as the users in the *low trust* group did. This group resorted to the first coping mechanism by Pearlin and Schooler [24] switching off automation to eliminate or modify the problematic conditions, thereby leading to disuse of automation.

5. Discussion

The purpose of this study was to explore how trust in automation unfolds and develops along different trajectories in the context of digitized and connected products and in situations where users gain new experiences and information that might be in conflict with their previous expectations. Through our analysis of users perceptions and responses, we were able to discover several interesting insights that contribute to the current knowledge on trust in autonomous vehicles and on digitized and connected products more broadly.

First, it is clear from our data that the autonomous car technology is still in the making and that users frequently encounter unexpected situations with the technology. The users are therefore constantly learning and adjusting their expectations and trust in automation as well as their ways of managing and coping with challenging situations. For digitized and connected products, this means a continuous need for the users to dynamically calibrate their trust both when they learn and discover new information on product functionality and when software updates change the behavior of a product. To the best of our knowledge, this dynamic formation of trust on automation has not been studied earlier and this paper proposes an important direction for future research on car automation and digitized and connected products.

Second, we identified different perceived causes for the misalignment of expectations and distrust. Our findings suggest that the causes are centered around the debate of human driver responsibility vs. technological readiness. Marketing and driver training were seen as means to increase drivers' knowledge on potential issues and as a way to increase their ability to prepare for and avoid them.

Third, we identified three different user groups where the users had very similar experiences on the same automation technology but for whom, based on the same new information, the trust in the technology had developed to different directions. While the users in the *no trust* group stopped using particular automation features of their cars, the *trust* and *low trust* groups both continued using the technology but expressed the differing levels of alignment and mechanisms of coping mechanisms while projecting differing levels of stress when discussing their views on the current and future use. While the literature on trust in technology has highlighted that system characteristics, such as functionality, reliability and helpfulness contribute to greater user trust in a system [30], our findings bring forward the individual differences in the perceptions and attitudes regarding the trustworthiness and readiness for risk-taking even in the context of the same technology.

Finally, our findings imply that by identifying origins of distrust and by working to align users' expectations with technological capabilities the challenges of using automation could be mitigated. With more training, the users in *low trust* and *no trust* groups could use automation more confidently within the boundaries of its affordances and constraints. Over time, the users may also be able to perceive improvements in the technology and reduction of the issues, and potentially even contribute to the changes to improve their trust. Also, even for imperfect systems, there are users who accept their limitations and are able to overcome initial distrust through the processes of calibration and coping. Our findings indicate a connection between trust formation and coping mechanisms and suggest that both streams of literature contribute to the theoretical base for trust in digitized and connected products.

5.1. Limitations and future research

One limitation in our research is that we used data from Tesla Motors Club online forum rather than data directly collected by researchers. We thus had little data about user demographics and characteristics, such as innovativeness, which could have an impact in trust formation. One direction for future research could be to extend the current research with multi-method data

by collecting either interview or survey data of Tesla users and including the measures of individual characteristics which could further explain trust in automation as well as its development over time.

Another limitation in our study was that the data covered two threads discussed at the same point in time. While this was justifiable in the exploratory phase of the research to exclude potential interference of changes over time, a longitudinal analysis in the second phase could provide a more insightful picture. Our data indicates that users expect the technology to improve over time, which may also affect their willingness to cope with issues. Interesting questions for future research are what happens if the technology does not develop as users expect and what strategies could be used to mitigate negative reactions.

Finally, our case example Tesla represents both a very innovative technology and a relatively expensive brand. Therefore, Tesla users are not a representative sample of a general population but rather users that are likely representatives of innovators or early adopters [32]. Thus, they may be more willing to accept issues and limitations of technology. Furthermore, the Tesla product development and marketing strategy is distinct compared to other brands and the company is actively making the autonomous features available for users while at the same time updating and improving these functionalities. Future research could extend the analysis to more varied user groups: users of other car brands and those who prefer to perform driving tasks manually, themselves. Lastly, the research should be extended to users of other types of digitized products.

6. Conclusions

Through the case Tesla Model S car, we studied how trust in automation unfolds and develops in situations where users gain new experiences and information that contradicts their expectations of the functionality of the automation. Our findings contribute to the existing knowledge on trust in automation technology and digitized and connected products by identifying three different groups of users who had developed three different trust trajectories and used different coping mechanisms for overcoming uncertainties and stress caused by issues with the automation. Our findings link the theories on trust formation and coping mechanisms and suggest that the skills and characteristics of users, their readiness to calibrate expectations under new experience and information and the mechanisms of coping rather than the solely technological capabilities of automation may explain the dynamic formation of trust. Our analysis of the causes of distrust further indicate that clear communication and training of users in the short

term and clearly presented improvements in the automation technology in the long term, with potential user involvement in the development, may significantly reduce user distrust and help them manage issues with digitized and connected products under continuous development. Key directions for future research include understanding the characteristics of users in the three groups in more detail and extending the analysis to wider user groups, to other car automation technologies, and to those who prefer manual driving and ultimately to users of other types of digitized and connected products.

References

- [1] A. K. Lyyra and K. M. Koskinen, "The Ambivalent Characteristics of Connected, Digitised Products: Case Tesla Model S,," *Nordic Contributions in IS Research*, p. 57–69, Aug 2016.
- [2] O. Henfridsson, L. Mathiassen and F. Svahn, "Managing technological change in the digital age: the role of architectural frames," *Journal of Information Technology*, vol. 29, p. 27–43, Jan 2014.
- [3] Y. Yoo, O. Henfridsson and K. Lyytinen, "Research Commentary—The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research," *Information Systems Research*, vol. 21, Dec 2010.
- [4] R. Garud, S. Jain and P. Tuertscher, "Incomplete by design and designing for incompleteness," *Design Requirements Engineering: A Ten-Year Perspective*, p. 137–156.
- [5] J. Kallinikos, A. Aaltonen and A. Marton, "The ambivalent ontology of digital artifacts," *MIS Quarterly*, vol. 37, pp. 357–370, Jun 2013.
- [6] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," *Human Factors*, vol. 46, no. 1, p. 50–80, 2004.
- [7] I. Y. Noy, D. Shinar and W. J. Horrey, "Automated driving: safety blind spots," *Safety science*, vol. 102, p. 68–78, 2018.
- [8] A. Shariff, J.-F. Bonnefon and I. Rahwan, "Psychological roadblocks to the adoption of self-driving vehicles," *Nature Human Behaviour*, vol. 1, no. 10, p. 694, 2017.
- [9] J. K. Choi and Y. G. Ji, "Investigating the importance of trust on adopting an autonomous vehicle," *International Journal of Human-Computer Interaction*, vol. 31, no. 10, p. 692–702, 2015.
- [10] P. A. Pavlou, "Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model," *International Journal of Electronic Commerce*, vol. 7, no. 3, p. 101–134, 2003.
- [11] R. Parasuraman, T. B. Sheridan and C. D. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, vol. 30, no. 3, p. 286–297, 2000.
- [12] K. Hawley, "Trust, distrust and commitment," *Nous*, vol. 48, no. 1, p. 1–20, 2014.
- [13] N. Luhmann, "Familiarity, Confidence, Trust: Problems and Alternatives," in *Trust: Making and Breaking Cooperative Relations*, Oxford, Basil Blackwell, 1998, p. 94–107.
- [14] N. Luhmann, *Trust and power*, John Wiley & Sons, 1979.
- [15] D. M. Rousseau, S. B. Sitkin, R. S. Burt and C. Camerer, "Not so different after all: A cross-discipline view of trust," *Academy of management review*, vol. 23, no. 3, p. 393–404, 1998.
- [16] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice," *Science*, vol. 211, no. 4481, p. 453–458, 1981.
- [17] H. A. Simon, *The sciences of the artificial*, MIT press, 1996.
- [18] D. Kahneman, *Thinking, fast and slow*, Macmillan, 2011.
- [19] P. Madhavan and D. A. Wiegmann, "Similarities and differences between human–human and human–automation trust: an integrative review," *Theoretical Issues in Ergonomics Science*, vol. 8, no. 4, p. 277–301, 2007.
- [20] J. Stilgoe, "Machine learning, social learning and the governance of self-driving cars," *Social studies of science*, vol. 48, no. 1, p. 25–56, 2018.
- [21] B. Nardi and J. Kallinikos, "Technology, Agency, and Community: The Case of Modding in World of Warcraft I," in *Industrial Informatics Design, Use and Innovation: Perspectives and Service*, IGI Global, 2010, p. 174–186.
- [22] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human factors*, vol. 39, no. 2, p. 230–253, 1997.
- [23] J. Cho, "The mechanism of trust and distrust formation and their relational outcomes," *Journal of Retailing*, vol. 82, no. 1, p. 25–35, 2006.
- [24] L. I. Pearlin and C. Schooler, "The structure of coping," *Journal of Health and Social Behavior*, p. 2–21, 1978.
- [25] L. Festinger, *A theory of cognitive dissonance.*, vol. 2, Stanford university press, 1962.
- [26] H. K. Klein and M. D. Myers, "A set of principles for conducting and evaluating interpretive field studies in information systems," *MIS quarterly*, p. 67–93, 1999.
- [27] R. K. Yin, *Case study research: Design and methods*, London and Singapore: Sage, 2009.
- [28] K. M. Eisenhardt, "Building theories from case study research," *Academy of management review*, vol. 14, no. 4, p. 532–550, 1989.
- [29] K. Habib, S. Ridella and J. Quandt, "Automatic vehicle control systems, Investigation into fatal accident involving Tesla Model S,," 2017.
- [30] M.B.Miles, A. Huberman, M. Huberman and M. Huberman, *Qualitative data analysis: An expanded sourcebook.*, Sage, 1994.
- [31] N. K. Lankton, D. H. McKnight, R. T. Wright and J. B. Thatcher, "Research note using expectation disconfirmation theory and polynomial modeling to understand trust in technology," *Information Systems Research*, vol. 27, no. 1, p. 197–213, 2016.
- [32] M. R. Everett, *Diffusion of innovations*, New York, 1995.
- [33] R. S. Lazarus and S. Folkman, *Stress, appraisal, and coping*, New York, NY: Springer, 1984.