THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Designing for Appropriate Trust in Automated Vehicles

A Tentative Model of Trust Information Exchange and Gestalt

FREDRICK EKMAN

Department of Industrial and Materials Science
Division Design & Human Factors

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2020

Designing for Appropriate Trust in Automated Vehicles
-A Tentative Model of Trust Information Exchange and Gestalt

FREDRICK EKMAN

Technical report no IMS-2020-6

Department of Industrial and Materials Science
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Printed by Chalmers Reproservice

Gothenburg, Sweden 2020

Abstract

Automated vehicles (AVs) have become a popular area of research due to, among others, claims of increased traffic safety and user comfort. However, before a user can reap the benefits, they must first trust the AV. Trust in AVs has gained a greater interest in recent years due to being a prerequisite for user acceptance, adoption as well as important for good user experience. However, it is not about creating trust in AVs, as much as creating an appropriate level of trust in relation to the actual performance of the AV. However, little research has presented a systematic and holistic approach that may assist developers in the design process to understand what to primarily focus on and how, when developing AVs that assist users to generate an appropriate level of trust.

This thesis presents two mixed-method studies (Study I and II). The first study considers what factors affect users trust in the AV and is primarily based on a literature review as well as a complementary user study. The second study, a user study, is built upon Study I and uses a Wizard of Oz (WOz) approach with the purpose to understand how the behaviour of an AV affects users trust in a simulated but realistic context, including seven day-to-day traffic situations.

The results show that trust is primarily affected by information from and about the AV. Furthermore, results also show that trust in AVs have primarily four different phases, before the user's first physical interaction with the AV (i), during usage and whilst learning how the AV performs (ii), after the user has learned how the AV performs in a specific context (iii) and after the user has learned how the AV performs in a specific context but that context changes (iv). It was also found that driving behaviour affects the user's trust in the AV during usage and whilst learning how the AV performs. This was primarily due to how well the driving behaviour communicated intentions for the users' to be able to predict upcoming AV actions. The users' were also affected by the perceived benevolence of the AV, that is how respectful the driving behaviour was interpreted by the user. Finally, the results also showed that the user's trust in the AV also is affected by aspects relating to different traffic situations such as perceived task difficulty, perceived risk for oneself (and others) and how well the AV conformed to the user's expectations. Thus, it is not only how the AV performs but rather how the AV performs in relation to different traffic situations.

Finally, since design research not only considers how things are, but also how things ought to be, a tentative explanatory and prescriptive model was developed based on the results presented above. The model of *trust information exchange and gestalt* explains how information affecting user trust, travels from a *trust information sender* to a *trust information receiver* and highlights the important aspects for developers to consider designing for appropriate trust in AVs, such as the design space and related variables. The design variables are a) the message (the type and amount of information), b) the artefact (the AV, including communication channels and properties) and c) the information gestalt, which is based on the combination of signals communicated from the properties (and communication channels). In this case, the gestalt is what the user ultimately perceives; the combined result of all signals. Therefore, developers need to consider not only how individual signals are perceived and interpreted, but also how different signals are perceived and interpreted together, as a whole, an information gestalt.

Keywords: trust; automated vehicles (AVs); mixed method research; trust information; trust phases, driving behaviour, explanatory and prescriptive model, information gestalt.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my examiner and main supervisor professor MariAnne Karlsson, to whom I owe my accomplishments. Thank you MariAnne for your helpful and insightful comments. Secondly, I would like to thank my co-supervisors' Dr. Helena Strömberg and Dr. Lars-Ola Bligård for your invaluable advice and new perspectives that you have given me. Without these I would not have been where I am today.

I would also like to thank my colleague and friend, PhD candidate Mikael Johansson for all discussions that brings my understanding about Life, the Universe and Everything forward. Furthermore, to all my colleagues at the division for Design & Human Factors, thank you for being supportive, inspiring and the best colleagues' one could hope for.

Too my dear family and friends – I am very grateful to enjoy your support, unconditional love and affection, thank you.

Last but not least, Emelie, thank you for your support, understanding and encouragement in my endeavours.

Appended publications

Paper A

Ekman, F. Johansson, M. Sochor, J. (2017). *Creating appropriate trust in automated vehicle systems: A framework for HMI design.* **IEEE Transactions on Human-Machine Systems, 48 (1) 95-101.**

Contribution: Ekman and Johansson planned and conducted literature review and complementary user. Ekman and Johansson planned and conducted the analysis. Ekman and Johansson wrote the paper with guidance and feedback from Sochor.

Paper B

Ekman, F. Johansson, M. Bligård, LO., Karlsson, M., Strömberg, H. (2019). *Exploring automated vehicle driving styles as a source of trust information.* **Transportation Research Part F: Traffic Psychology and Behaviour, 65, 268-279.**

Contribution: Ekman, Johansson, Bligård, Karlsson and Strömberg planned the study. Ekman and Johansson conducted the study and data analysis with assistance from Bligård, Karlsson and Strömberg. Ekman and Johansson wrote most parts of the paper with guidance and feedback from Bligård, Karlsson and Strömberg. Karlsson and Strömberg wrote parts of the paper. Bligård and Karlsson conducted part of the statistical analysis.

Paper C

Trust in What? Exploring the interdependency between an Automated Vehicle's Driving Style and Traffic Situations. (2020). **Manuscript submitted to Transportation Research Part F: Traffic Psychology and Behaviour.**

Contribution: Ekman, Johansson, Bligård, Karlsson and Strömberg planned the study. Ekman and Johansson conducted the study and data analysis with assistance from Bligård, Karlsson and Strömberg. Ekman wrote the paper with guidance from Karlsson and feedback from the other authors.

Additional publications

Ekman, F. Johansson, M. Sochor, J. (2016). *To See or Not to See: The Effect of Object Recognition on Users' Trust in Automated Vehicles*. **Proceedings of the 9th Nordic Conference on Human-Computer Interaction, 1-4.**

Ekman, F. Johansson, M. Karlsson, M. (2018). *Understanding Trust in an AV-context: A Mixed Method Approach*. **Proc. 6th Humanist Conf, 13-14.**

Ekman, F. Johansson, M. Karlsson, M. (2018). *Designing Multi-modal Interaction–A Basic Operations Approach.* **Congress of the International Ergonomics Association, 43-53.**

Strömberg, H. Ekman, F. Bligård, LO. (2019). *Keeping a finger in the pie?*
Proceedings of the 31st European Conference on Cognitive Ergonomics, 118-126.

Johansson, M. Ekman, F. Karlsson, M. (2020). The Devil is in the Details - Trust Development During Initial Usage of an Automated Vehicle. **Manuscript submitted to Proceedings of the 7th Humanist Conference, Rhodes Island, Greece, 24-25 September 2020.**

Johansson, M. Ekman, F. Bligård, LO. Karlsson, M. Strömberg, H. (2020). Talking Automated Vehicles – Investigating users' understanding of an automated vehicle during initial usage. **Manuscript submitted to Behaviour & Information Technology.**

TABLE OF CONTENTS

1 INTRODUCTION

1.1 BACKGROUND

Automation may carry out functions previously conducted only by humans (Parasuraman, Sheridan, & Wickens, 2000). In the past, interaction with automation has primarily been designed for, and used, by expert users (pilots in aviation or operators in the process industry for example) but since automation has developed and matured, it has also become more available to novice users (Janssen, Donker, Brumby, & Kun, 2019). Automation is now readily available in areas such as education (Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013) (e.g. social- and educational robots) and transportation (automated vehicles) (Janssen et al., 2019).

The topic of automated vehicles (AVs) has come to be well-researched, due to claims of increased traffic safety and improved user comfort (Gold, Korber, Hohenberger, Lechner, & Bengler, 2015; Merat, Jamson, Lai, & Carsten, 2012; Naujoks, Mai, & Neukum, 2014; Payre, Cestac, & Delhomme, 2016). However, levels of safety and comfort are affected by the level of automation. Automated systems in cars may be defined by SAE's six levels of automation [LoA] (SAE, 2018), with LoA 0 defining a fully user-operated vehicle[1] (a conventional vehicle) and LoA 5 referring to a fully automated vehicle (no need for user involvement in the task of driving). LoAs 3 to 5 are levels of automation covering "automated driving systems" (ADS) which carry out the entire dynamic driving task (DDT) while engaged, with no need for user involvement other than as a fallback operation [LoA 3 only].

Nevertheless, before users can reap the benefits of AV use, they must first trust the vehicle. According to earlier studies in automation, trust is a precondition for the use of automated systems (Parasuraman & Riley, 1997); not only because it is essential to user acceptance (Ghazizadeh, Lee, & Boyle, 2012), but because it is also a prerequisite for a good user experience (Waytz, Heafner, & Epley, 2014).

However, user trust needs to be appropriate to the actual performance of the automated system. This ensures that the outcome of the user-automation interaction is as safe as possible (Lee & See, 2004). Too high a level of trust in automated system (relating to its performance) can lead to misuse, with users operating the automated system in unintended ways (Itoh, 2012; Parasuraman & Riley, 1997). This might lead to negative outcomes, and in worst case, accidents, resulting in injuries or even fatalities. On the other hand, if user trust in the automated system is too low, this may lead to disuse, with users choosing not to use it at all (Parasuraman & Riley, 1997) even though the automated system might conduct the driving task more safely than the user. Automation system knowledge and experience are important in aiding an appropriate level of trust. It is therefore important for users to understand an AV's limitations and constraints (Edelmann, Stümper, & Petzoldt, 2019). This allows them to generate an appropriate level of trust and minimise the possibility of disuse or, even worse, misuse.

Today, user trust in AVs is often studied in (AV) driving simulators but, since perceptions of risk, uncertainties and interdependencies are important aspects for trust (Mayer, Davis, & Schoorman, 1995; Rempel, Holmes, & Zanna, 1985), one might argue that driving simulators do not entail the perceived risks (Large, Burnett, Morris, Muthumani, & Matthias, 2018) and uncertainties fundamental to achieving a valid measurement of user trust in AVs. Furthermore, trust can be affected by different types of information sources (Lee & See,

[1] A user-operated vehicle is a conventional vehicle that uses a human driver to control the vehicle.

2004). For example, many studies focus on conveying trust relating to information from 'in-car displays' (Helldin, Falkman, Riveiro, & Davidsson, 2013; Seppelt & Lee, 2007; Stockert, Richardson, & Lienkamp, 2015), such as graphical user interfaces (GUIs) located in the vehicle cockpit. However, others have found that how the car behaves also affects user trust, either through conveying vehicle capability (Price, Venkatraman, Gibson, Lee, & Mutlu, 2016) or intentions of cooperation (Kauffmann, Naujoks, Winkler, & Kunde, 2018).
Thus, AV driving behaviour also seems a very important consideration when designing for appropriate trust in AVs. Therefore, an important next step in designing for an appropriate level of trust in AVs is more realistic studies that include more uncertainties and risk perceptions, whilst focusing on AV driving behaviour as a format for conveying different types of trust information.

However, further knowledge of trust is not only about conducting realistic studies. Designing for appropriate trust includes many aspects that must be considered why there is a need for an aid, assisting developers to design AVs that support users generating an appropriate level of trust. However, few studies have (to this author's knowledge) considered where and how developers should direct their attention when designing for appropriate trust in AVs. This involves identifying the design space and which related variables might be designed, to assist the user in generating an appropriate level of trust.

1.2 AIM AND QUESTION POSED

This research aims to identify the design space[2] and relevant design variables[3] which may be used by developers to enable users to generate an appropriate level of trust in AVs. However, to identify and propose relevant design variables for consideration, the following questions must be addressed:

RQ1a: What factors[4] affect user trust in an AV?
RQ1b: Which of these factors should be considered from a design perspective, so as to generate an appropriate level of trust?

The above two research questions, 1a and 1b, were the initial ones posed but, once the factors affecting trust had been identified, some of their specifics needed further investigation. These included the driving behaviour of the vehicle and driving behaviour relating to specific traffic situations. Thus, four more questions were posed, 2a/b and 3a/b, in order to understand how driving behaviour and contextual aspects affect user trust.

RQ2a: Does an AV's driving behaviour affect user trust in it?
RQ2b: If so, how does the AV's driving behaviour affect user trust?

and

[2] Design space is defined as the space in which design variables are located and in which the designer can operate by adjusting these variables, according to desired effects (such as increasing or decreasing user trust in an AV).

[3] Design variables are variables than can be deliberately designed to generate an effect of increasing or decreasing user trust in an AV. For example, which type of information the user is allowed to receive from the AV.

[4] Factors affecting trust, hereinafter called "trust factors", are factors affecting user trust in an AV before, during or after an interaction with it.

RQ3a: Are there any aspects of traffic situations (depending on the AV's driving behaviour) that affect user trust in the AV?
RQ3b: If so, how do traffic situations affect user trust in an AV?

1.3 THESIS OUTLINE

This thesis is organised as follows:

- **Chapter 1** provides an introduction to the topic, as well as the aim and research questions to be answered.
- **Chapter 2** presents the frame of reference upon which the research in this thesis is based.
- **Chapter 3** describes the research approach, including the author's theoretical and philosophical perspective and methodology used to answer the research questions.
- **Chapter 4** presents the main empirical results obtained and provides brief answers to each of the research questions.
- **Chapter 5** presents an explanatory and prescriptive model of how information affecting trust is communicated from developers of automated vehicles to the user, as well as highlight the design space and included design variables.
- **Chapter 6** presents empirical considerations, implications of the model presented and brings forward methodological issues encountered during the project.
- **Chapter 7** presents the conclusions and implications.

2 FRAME OF REFERENCE

The introduction highlighted the importance of generating an appropriate level of trust relating to the actual performance of an AV. Therefore, the frame of reference is dedicated to describing relevant theories regarding trust in general and factors affecting it. This chapter also describes other aspects that may indirectly affect user trust in automated systems.

2.1 THE FUNDAMENTALS OF TRUST

2.1.1 What is Trust?

Trust can be defined as:

> *"the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability"* (Lee & See, 2004).

Trust is an attitude held by a *trustor*[5] towards an agent. The agent might be either a human or a machine. The agent in which the trustor puts trust is hereinafter referred to as the *trustee*[6]. For a collaboration between a trustor and trustee to be initiated, the trustor needs an incentive, such as a goal (for example a banker helps a trustor to earn money by placing the trustor's money in funds on the stock market). Finally, there need to be risks or uncertainties and, hence, the possibility that the collaboration might fail (Mayer et al., 1995; McKnight & Chervany, 2000).

For the trustor, at the start of a collaboration with an unknown trustee, trust is based only on *beliefs* about the trustee (see Figure 1). These beliefs are generated by information on, and impressions of, the trustee. Through affective evaluation of this information, the trustor's belief may become an attitude of trust towards the trustee. When trust has been established, an intention to rely on the agent may grow. This, in turn, may become a behaviour; more specifically, *reliance*. Therefore, trust is an attitude, borne of a belief about the trustee and an intention to rely on them (Lee & See, 2004).

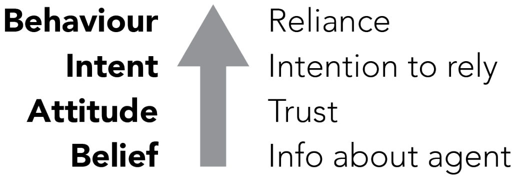

Figure 1 - From belief to behaviour in the context of trust.

In summary, trust is an attitude that might lead to a trustor's behaviour of relying on a trustee (a human or technological agent) and is, therefore, a key aspect in collaborative activities for which there is a goal; and especially when there is a risk of something going wrong.

2.1.2 Interpersonal Trust

Trust is often mentioned in the context of relationships as "interpersonal trust", with trustworthiness viewed as a desired quality for a well-functioning relationship (Rempel et al., 1985). Different aspects have been identified as affecting the trust formation process. More

[5] A person who forms trust in an agent.

[6] An agent such as a person or a machine in which trust is formed.

specifically, the trustee's capability or competence, benevolence and integrity (Mayer et al., 1995); (McKnight & Chervany, 2000) but also predictability (McKnight & Chervany, 2000). *Ability* or *competence* refers to how strongly a trustee has the power to achieve the trustor's goals. *Benevolence* is the trustor's expectation towards the trustee; that he or she is motivated to act in favour of the trustor. *Integrity* refers to the expectation of the trustee; that he or she keeps promises and tells the truth. Finally, *predictability* refers to the consistency of the trustee's actions; affording the trustor the ability to foresee future actions (McKnight & Chervany, 2000).

Therefore, in a collaboration including two people striving towards a common goal, it is highly important that the trustee is (in the eyes of the trustor) competent enough to help the trustor to reach his or her goal(s), shows benevolence towards them, has integrity, keeps promises, tells the truth and exhibits consistent behaviour over time.

2.2 TRUST IN AUTOMATION

Trust is important, not only to positive interpersonal relationships; it is also a key aspect in the user-automation interaction, if the trustor (hereinafter referred to as the user) is to accept the trustee (hereinafter referred to as the automated system or AV) (Ghazizadeh et al., 2012). Acceptance describes to what degree a user intends to use and adopt a system (Adell, 2010) e.g. an automated system. Trust in automation has similarities to interpersonal trust. The aspects affecting the trust formation process (ability/competence, benevolence, integrity and predictability) resemble the field of trust in automated systems and share three fundamental, corresponding trust aspects with it (see Section 2.2.1 – Fundamental Trust Aspects). However, other trust factors also affect trust (see Section 2.2.2 – Trust Factors). Furthermore, the user processes the information given by the automated system (and hence the automated system's characteristics, such as fundamental trust aspects and trust factors) using three different cognitive processes (see Section 2.2.3 – Processing Trust). Additional to the fundamental aspects of trust and trust factors, there are individual environmental aspects which must be considered; these also affect user trust in the automated system (see Section 2.2.4 – User, Automation & Context).

2.2.1 Fundamental Trust sources – Trust Formation

According to Lee and Moray (1992) and Lee and See (2004), performance, purpose and process are three sources of information from which the user draws relevant information about the goal-orientated characteristics of an automated system, to form and maintain an appropriate level of trust. Therefore, trust can be formed *"from a direct observation of system behaviour (performance), an understanding of the underlying mechanisms (process), or from the intended use of the system (purpose)"*(Lee & See, 2004, p. 67). If user trust is based on several information sources, it will be more stable than if it were based on only one (Lee & See, 2004).

Information on performance refers to the capability, reliability and predictability of an automated system and is similar to capability/competence in interpersonal trust. Information on performance takes into account the current and historical operation of the automated system and describes its capability to satisfy user goals. User trust is therefore affected by how well the automated system performs (Hoff & Bashir, 2015).

Information of purpose refers to the designer's intended use for the automated system and describes why the automated system was developed. Information of purpose is similar to benevolence in interpersonal trust but, since no current automated systems possess their own intentions, the term refers instead to the designer's intention for the automated system (Lee & See, 2004).

Information on process refers to the attributes of the automated system, rather than specific behaviours or actions. Information on process is similar to the interpersonal aspects of dependability and integrity (Lee & See, 2004)

2.2.2 Trust Factors

However, as indicated earlier, other factors also seem to affect users trust. One important aspect is helping the user create an approximate representation of the automated system functionality and capability. This allows the formation of a correct mental model for proper use of the automated system (Lee & See, 2004; Toffetti et al., 2009), thereby minimising the risk of not understanding the automated system's limitations (Saffarian, de Winter, & Happee, 2012). One way to help users understand automated system limitations is to train them in the automated system's functionality before and after first usage (Parasuraman, Sheridan, & Wickens, 2008; Saffarian et al., 2012; Toffetti et al., 2009).

Another way to assist users in understanding the automated system's capabilities and functions is to provide continuous, accurate feedback (Dekker & Woods, 2002; Thill, Hemeren, & Nilsson, 2014; Toffetti et al., 2009; Verberne, Ham, & Midden, 2012). This feedback may be divided into two types; action feedback and learning feedback (Stanton & Young, 2000). Action feedback is information provided directly after an action has been carried out and supports fast learning. Learning feedback is more detailed information about the performance, often provided during training. This leads to slower but more enduring skill knowledge (Banks & Stanton, 2016). A combination of these two different types of feedback is optimal for enduring skill knowledge as well as for a quick understanding of automated system capabilities. However, it is also important to present feedback promptly, clearly and non-intrusively (Saffarian et al., 2012). Feedback might be distracting to the user (Stanton & Young, 2000) and, if presented at the wrong time, could lead to distrust in the automated system (Saffarian et al., 2012).

There are also important aspects regarding what type of information is provided to the user and how much. Automated system transparency has also been identified as an important trust factor, as it may help users achieve a greater feeling of control by helping them predict how the automated system will behave (Verberne et al., 2012). One type of automated system transparency might be to show automated system uncertainty[7] (Beller, Heesen, & Vollrath, 2013; Jian, Bisantz, & Drury, 2000). Another might be presenting error information after an incident, to explain why it occurred and the extent to which overall automated system performance is affected (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003; Stanton & Young, 2000). Furthermore, earlier research has shown that "why & how" information is important, as it allows the user to better understand the intentions of automated system. In this case, "how" information describes how the automated system will solve a pending task and "why" information explains its actions (Koo et al., 2014).

However, users are different. Thus, it is important for user trust that non-critical automated system functions can be customised to correlate with user preferences (Merritt & Ilgen, 2008; Saffarian et al., 2012; Verberne et al., 2012). Another way to account for user needs is to

[7] System uncertainty – can be explained as showing system reliability or rather the lack thereof in order for users to understand that the automated system is operating at a reduced level of reliability.

design adaptable automated system that automatically adjust to the user's cognitive and physical preferences (Helldin et al., 2013). This may be achieved by such means as only showing users relevant information and thus lowering their mental workload (Saffarian et al., 2012).

It has also been found that how information is given may affect user trust. Automated systems that is designed to be more human-like by using anthropomorphic features (Hoff & Bashir, 2015; Waytz et al., 2014) may affect user trust. One experiment (Waytz et al., 2014) found that using anthropomorphic features (giving the AV a name, for example) increased participants' trust in the AV (Waytz et al., 2014). However, other research has claimed that anthropomorphic features in automated systems may have less of an effect on trust, owing to other aspects which may annul the trust-generating effect of anthropomorphism. These other aspects include: easy-to-understand information about the vehicle's awareness and actions; the performance and style of the vehicle's driving behaviour; and how appropriately the information provided by the automated system has been adapted to the situation (Aremyr, Jönsson, & Strömberg, 2019). Finally, the way the automated system is portrayed has been shown to affect user trust (Hoff & Bashir, 2015; Lee & See, 2004). For instance, if an automated system is portrayed as an expert to users, it may be perceived as more reliable than humans carrying out the same task (Madhavan & Wiegmann, 2007).

2.2.3 Processing Trust – The User's Cognitive Processes

According to Lee and See (2004), when information that affects trust is conveyed from an automated system (or information about the automated system is conveyed from elsewhere) to a user, it is processed by him/her through three cognitive processes. These processes are a) affective, b) analogical and c) analytic and are affected by the available information and how it is displayed. *The analogical* process involves connecting earlier, familiar experiences and using them to assess the trustworthiness of the automated system, based on similarities and differences. The *analytic* process undertakes rational assessment of the agent's trustworthiness, logically evaluating information based on the user's understanding of the automated system's functions. The *affective* process describes the emotional process; the feeling of trust. The affective process is the most fundamental and influential trust process in user behaviour, affecting both the analogical and analytic trust processes. It is also the least cognitively demanding process. The analogical process is used when information about the automated system is lacking and earlier experiences (with similar agents) are used to assess trustworthiness. By contrast, the analytic process involves logical argumentation regarding the automated system and trustworthiness and is used to evaluate information about the automated system (Lee & See, 2004). It is therefore important, when designing for an appropriate level of trust, to create an interaction that considers the cognitive processes.

2.3 USER, AUTOMATION AND CONTEXT

Thus, trust is affected by different factors conveyed by the automated system and processed through different cognitive processes. However, it is not only automated system that affects user trust. According to Hoff and Bashir (2015), two other elements affect user trust; the user him/herself and the environment in general. These dimensions correlate directly to three layers of trust; dispositional, situational and learned trust (Marsh & Dibben, 2003). *Dispositional* trust is the user's general tendency to trust automation, irrespective of automated system or context-specific attributes. Rather, this aspect examines such things as the user's age, gender, culture and personality traits. *Situational* trust includes two dimensions of variability; the internal and the external. The internal dimension relates to the user's self-confidence, expertise in the task at hand, mood and attentional capacity. The external dimension relates to situational aspects, such as workload, perceived risks, automated system complexity, type of automated system, task difficulty, organisational setting, perceived benefits and how the task is framed. *Learned* trust is trust based on current or previous interaction with the automated system. Previous experiences that have generated pre-existing knowledge and affect user trust are called *initially learned* trust. This includes trust affecting such aspects as attitudes/expectations, understanding of the automated system, experience with it and the automated system and/or brand's reputation. The other aspect is *dynamic learned* trust. This is trust generated when interacting with an automated system. The dynamic learned trust generated during the interaction might, in turn, generate a level of reliance on the automated system (Hoff & Bashir, 2015).

2.4 SUMMARY

To conclude:

- trust is an attitude held by a trustor towards a trustee, either human or machine. The trustor needs an incentive to collaborate (such as achieving a goal) as well as the possibility that the collaboration might fail.
- in the context of this thesis the trustor is the user of the AV and the trustee the automated system.
- the user's trust in automation is based on three primary sources of (trust) information from the automated system; performance, purpose and process information.
- the user processes this information through three cognitive processes; analytic, analogic and affective.
- the user's trust is affected not only by the automated system per se but also by user-related aspects such as dispositional, situational and learned trust. This includes everything from a user's cultural context, to what the user knows about the automated system and how the user perceives the context.

3 RESEARCH APPROACH

3.1 THEORETICAL & PHILOSOPHICAL PERSPECTIVE

All research is based on philosophical assumptions about the reality of our surrounding world; also known as *philosophical worldviews* (Creswell & Plano Clark, 2017, Guba & Lincoln, 1994). These worldviews shape not only how we view reality but also govern the processes of research. As a researcher, it is important to understand one's philosophical worldview in order to then justify one's practices. Therefore, the following section presents this author's theoretical and philosophical perspective and describes the methodology used in conducting his studies.

Personal Setting

With an educational background in industrial design engineering, this author's focus has been first and foremost on 1) understanding the design problem at hand, 2) the users encountering the problem and their needs and 3) the context in which user and problem are situated. Design problems do not appear in a vacuum but are situated in contexts. From an activity theory perspective,[8] this can be described in terms of a system including a user (subject) who carries out activities with intentionality and desire (object), using artefacts as mediating tools to interact with the objective world (Kaptelinin & Nardi, 2006, pp. 3-13). Furthermore, an artefact may only be understood by first capturing "….the context of human activity – by identifying the ways people use this artefact, the needs it serves and the history of its development" (Nardi, 1996, pp. 45-68). Thus, it is not possible to understand a design problem without first understanding the user, the activity and the context.

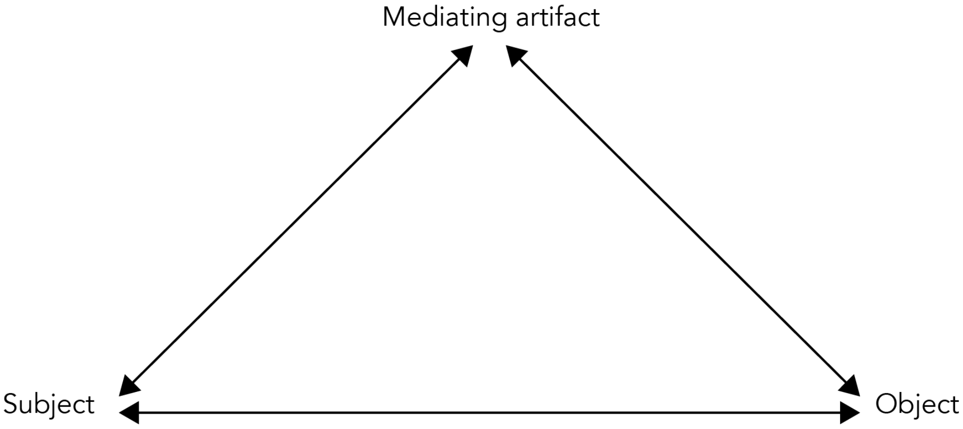

Figure 2 - The most common reformulation of Vygotsky's mode (Knutagård, 2002).

This author's research focuses on the design problem of how to generate an appropriate level of trust in automated systems, such as automated vehicles (AVs), from an individual user perspective (focusing on the user as they use the vehicle, the artefact). Furthermore, an activity theory perspective permits a focus not only upon user and AV but allows consideration of the context. In other words, the traffic environment in which the user's activity takes place, using an AV as a mediating tool to reach a destination (objective).

[8] Activity theory (Kaptelinin & Nardi, 2006) being the Division for Design & Human Factors' basis for design problems.

Philosophical Worldviews(s)

This author believes that an objective world exists, with or without our presence. However, that world is shaped and affected by our interpretation of it (cf. the ontology of a critical realist (Guba & Lincoln, 1994). Thus, complete understanding of objective reality can never truly be possible because we reshape our perspective on the world every day. This perspective is changed through new experiences and by acquiring new knowledge.

Trust is something that most people can relate to. Moreover, a person's trust in a trustee changes over time through new experiences and by acquiring new knowledge about the trustee. Therefore, user perception is the most important source of information in gaining a better understanding of trust in AVs (cf. constructivism (Creswell & Clark, 2017). Perception may be accessed through what users verbalise in, say, interviews and questionnaires. However, question-based methods are only one means of understanding what factors affect user trust in AVs. This author has adopted a pragmatic approach (cf. pragmatism (Creswell & Plano Clark, 2017), choosing the most relevant methods to answer the research questions in the best way possible. This means using mixed-methods research to gain as nuanced an image of trust as possible and address the design problem of incorrectly calibrated trust relating to actual AV performance.

3.2 METHODOLOGY

Figure 3 describes the research process. Two studies (Study I and Study II) were conducted for this thesis. Each had different purposes and outcomes, but both have contributed to answering the research questions. The studies are reported in Papers A, B and C.

Studies I and II were exploratory in nature and designed to use mixed methods. They were conducted in sequence (see Figure 3), with the planning and completion of Study II built upon the results of Study I.

Research questions 1a and 1b:

> **RQ1a: What factors affect user trust in an AV?**
> **RQ1b: Which of these factors should be considered from a design perspective, so as to generate an appropriate level of trust?**

were answered by Study I (see Figure 3).

Study II further explored what directly AV-related trust factors affect user trust, by looking at the effect of driving behaviour. It also examined what contextual aspects in traffic situations affect user trust in AVs. Contextual aspects, such as risks (Hoff & Bashir, 2015), seem to be an important consideration when studying user trust in AVs. Therefore, two further bipartite research questions were formulated:

> **RQ2a: Does an AV's driving behaviour affect user trust in it?**
> **RQ2b: If so, how does the AV's driving behaviour affect user trust?**

and

> **RQ3a: Are there any aspects of traffic situations (depending on the AV's driving behaviour) that affect user trust in the AV?**
> **RQ3b: If so, how do traffic situations affect user trust in an AV?**

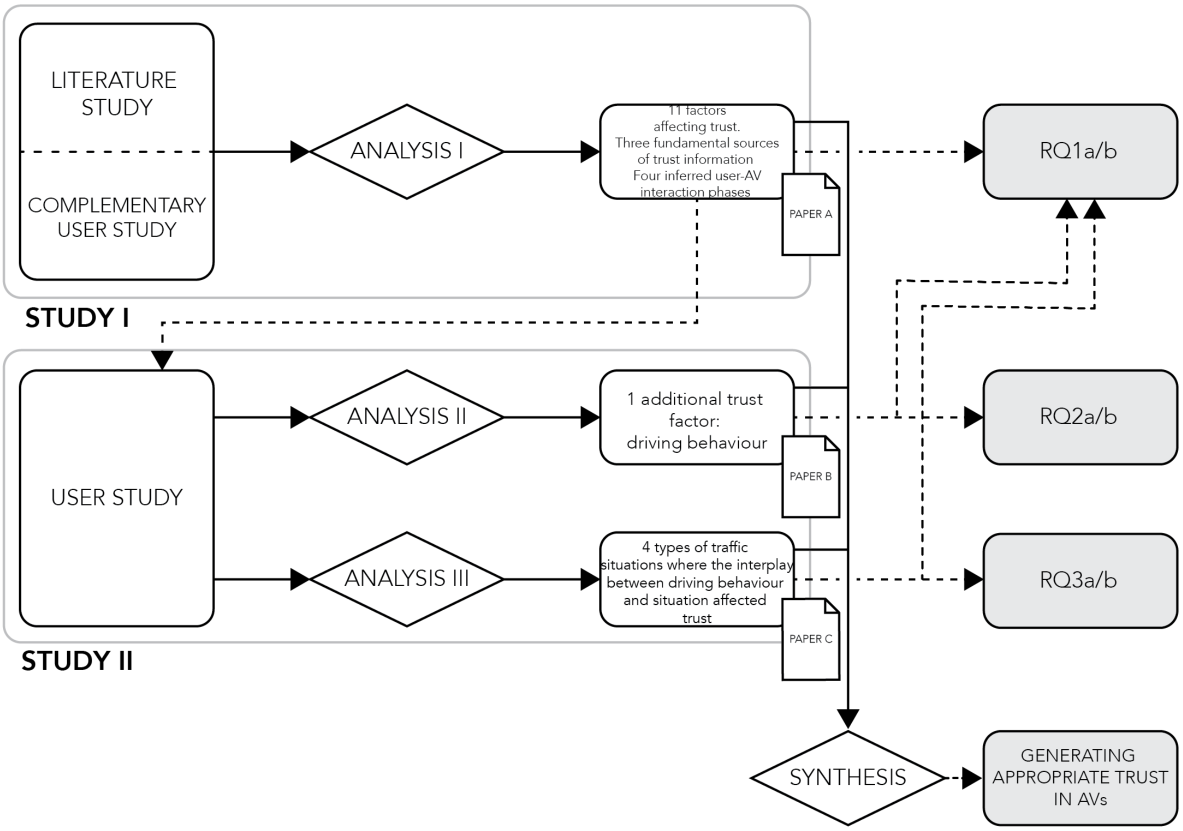
LITERATURE STUDY
COMPLEMENTARY USER STUDY
ANALYSIS I
11 factors affecting trust.
Three fundamental sources of trust information
Four inferred user-AV interaction phases
PAPER A
RQ1a/b
STUDY I
USER STUDY
ANALYSIS II
1 additional trust factor:
driving behaviour
PAPER B
RQ2a/b
ANALYSIS III
4 types of traffic situations where the interplay between driving behaviour and situation affected trust
PAPER C
RQ3a/b
STUDY II
SYNTHESIS
GENERATING APPROPRIATE TRUST IN AVs

Figure 3 – Organisation of research.

3.3 STUDY I

3.3.1 Method

The aim of Study I (see also Paper A) was to investigate how an appropriate level of user trust in an AV can be generated. This was achieved primarily by conducting a literature review to identify which factors have been found to affect user trust in AVs and what events[9] take place in the user-AV interaction to generate the most appropriate level of trust.

The literature study was based on a grounded-theory literature review method, comprising a five-stage process (Wolfswinkel, Furtmueller, & Wilderom, 2017). The two main areas of interest in the literature review were trust and human-machine interaction (HMI).

A complementary user study was conducted to provide context-specific input on how and when trust factors affect user trust, as well as confirming the relevance of the events in the user-AV interaction. The study involved nine participants (five males and four females), aged between 23-55; these subjects had held a driving licence for between 5 and 37 years. Data was collected via semi-structured interviews during and after the participants had driven a semi-automated vehicle (Level 2 SAE) on a stretch of road with low-density traffic. The participants were also observed when driving, to identify events taking place in the user-AV interaction. This provided more in-depth information on which events most needed trust-affecting factors to help the user generate an appropriate level of trust.

[9] Events are defined as touchpoints; points of interaction between user and AV where trust-affecting factors can be brought in, to assist the user in generating an appropriate level of trust in the AV.

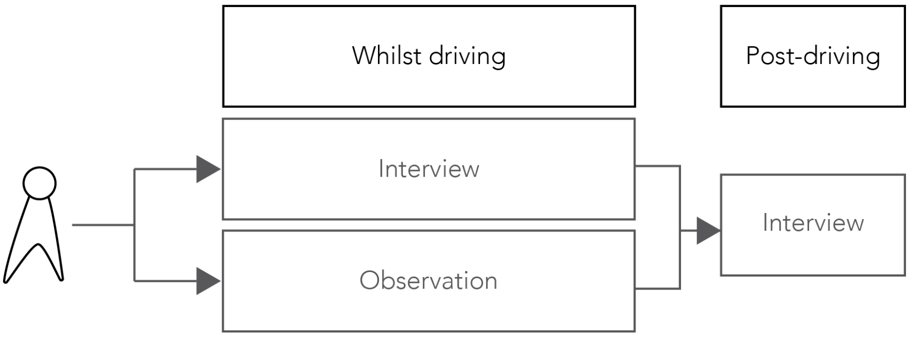

Figure 4 - Method design including two phases: whilst driving and post-driving.

3.3.2 Analysis

The analysis in Study I was accomplished in three steps: (i) analysing the information extracted from the literature review, focusing on what trust factors are in play, plus important events taking place between user and AV, and (ii) analysing the interview data with, and observations of, participants driving the semi-AV, focusing on what factors (identified from the literature) were present. Also, identifying relevant events in the interaction between user and AV and (iii) comparing the data provided in step (i) and (ii), looking for similarities and/or discrepancies and finally compiling them.

3.4 STUDY II

3.4.1 Method

Study II's aim was to investigate whether and how the vehicle's driving behaviour affects the user trust in the AV (see also Paper B) during the interaction with an AV, and how driving behaviour affects user trust in the AV in everyday traffic situations (see also Paper C).

An experiment with a Wizard of Oz (WOz) approach was set up to investigate how the driving behaviour of an AV (acceleration, braking and lane placement etc.) affects user trust, plus how the driving behaviour affects user trust in the AV in different traffic situations. This approach involved a standard car being remodelled to be perceived and experienced as a fully automated vehicle. However, it was actually operated by a "wizard" driver via secondary controls (steering wheel, accelerator and brake pedals, plus gear selector) sitting in the back seat. The wizard simulated two different driving behaviours, "Defensive" and "Aggressive".

Eighteen participants (ten male and eight female) between the ages of 20 and 55 years experienced the AV on a test course. Each participant underwent two test runs; experiencing one of the two driving styles' in each test run. These test runs comprised seven different realistic traffic situations designed specifically for the test.

The mixed-methods design was used to allow parallel extraction of both quantitative and qualitative data, so that the different datasets could be combined and compared with each other during the analysis (cf. Creswell & Plano Clark, 2017). The mixed-methods design helped extract information regarding (i) which factors, (ii) when and (iii) how the factors affected user trust in the AV. Therefore, data on perceived trust was collected in two different phases using a combination of methods (see Figure 7). In Part 1 of the peri-trial phase, a momentaneous trust assessment was introduced to collect data during participants' interaction with the AV during seven different traffic situations. Part 2 took place directly after each test run, to collect data on participants "overall" trust in the AV (via a trust questionnaire) and to allow participants to chart how their trust in the AV changed during the test run (via a trust curve). The peri-trial phase was then iterated once more to allow each participant to experience both driving behaviours. The post-trial phase was included to allow participants to compare both driving behaviours (experienced during both test runs). To stimulate the participants, the trust curve was introduced as a mediating tool. This helped participants further reflect on and discuss their levels of trust in the AV in specific situations, plus their overall trust.

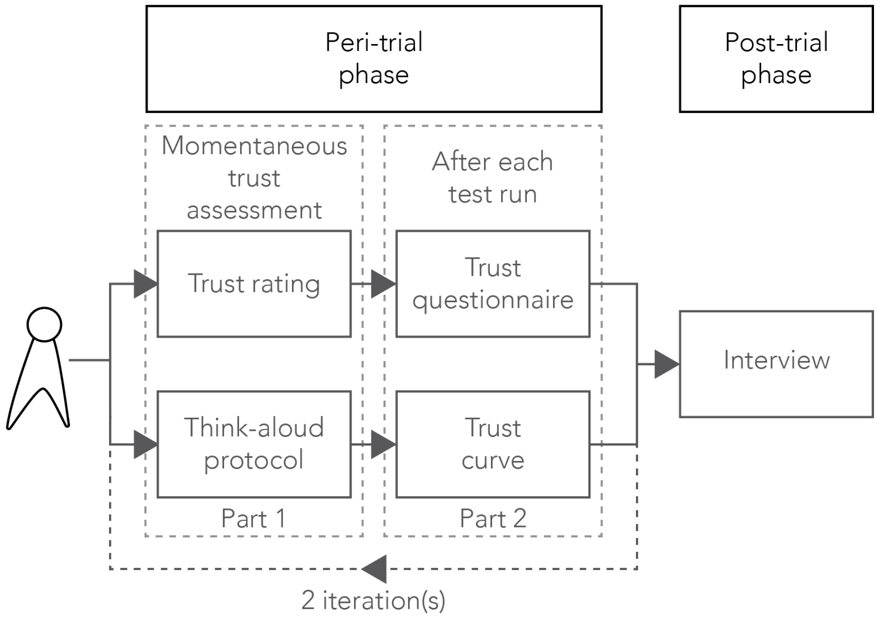

Figure 5 - Convergent mixed-methods design, including methods.

3.4.2 Analysis

The analysis of Study II was divided into two parts.

The first analysis (see also Paper B) focused on how the AV driving behaviour affected user trust and included data collected via trust ratings, think-aloud protocols, the trust questionnaire, and post-trial interviews. For the momentaneous trust ratings, a median value for each trust rating (given for the seven situations faced during the test runs) was calculated for each participant and driving behaviour. For the trust questionnaire, the participants' degree of agreement with eight different items was calculated for each driving behaviour and the results compared. A Wilcoxon signed-rank test (cf. Siegel & Castellan, 1988) was also used, to determine any statistical differences between participants' momentaneous ratings of trust and their trust questionnaire scores for the two driving behaviours.

The data from the think-aloud protocols, trust curve explanations (peri-trial phase) and post-trial interviews were analysed using an iterative thematic analysis (cf. Braun & Clarke, 2006). The questions guiding the analysis were 1) what factors explain users'/drivers' trust in the AV? and 2) what factors explain users'/drivers' trust in the respective driving behaviours of the AV? The transcripts were coded according to elements that were deemed relevant.

The second analysis (see also Paper C) focused on how the AV's driving behaviour affected user trust in various everyday traffic situations. The analysis was based on data collected from trust ratings, think-aloud protocols and post-trial interviews. Differences in trust ratings between the "Defensive" and "Aggressive" driving behaviours in the seven traffic situations were calculated. The difference in trust was determined by comparing each participant's trust score in each of the seven traffic situations. The trust curves, drawn by the participants after each test run, were analysed. These were annotated with a (+) for positive tangencies and a (-) for negative tangencies in each participant's curves relating to the respective traffic situations. The number of positive and negative annotations for each situation and driving behaviour were then summarised. Finally, an analysis of the think-aloud data and post-trial interviews was conducted, using a targeted search of participants' statements relating to each traffic situation. The statements for each situation were then analysed focusing on known contextual aspects affecting trust, such as perceived risks and task difficulty (cf. Hoff and Bashir, 2015), plus unknown contextual trust aspects.

3.5 SYNTHESIS

The results of Study I and Study II were synthesised by identifying trust-affecting factors relating to the AV and context, plus any interdependencies between the elements and how users perceive these factors and the interdependence between AV and context. This in order to create a model that could explain and predict how trust in AVs are generated and therefore assisting developers in designing for appropriate trust. The approach might be compared to doing a puzzle; trying to find the right pieces and combining them to generate a full image of trust in AVs (in context of which AV and user operate). Hence, two main questions were posed, to guide the synthesis in the direction of this aim. These guiding questions were:

What design variables[10] *are relevant when designing for an appropriate level of trust in AVs?*

And

How can a design space,[11] *in which users trust AVs, be designed for and illustrated, to accommodate an understanding of what elements to consider and prioritise when developing AVs?*

[10] Design variables are variables than can be deliberately designed to generate an effect of increasing or decreasing user trust in an AV. For example, which type of information the user is allowed to receive from the AV.

[11] Design space is defined as the space in which design variables are located and in which the designer can operate by adjusting these variables, according to desired effects (such as increasing or decreasing user trust in an AV).

4 FINDINGS

4.1 STUDY I

The aim of Study I was to investigate how an appropriate level of user trust in an AV can be generated during user-AV interactions necessary to achieve user goal(s) (see also Paper A). This was achieved by identifying factors affecting user trust in automation in general and in AVs specifically.

4.1.1 Results

The literature study revealed several different but theoretically related factors, from several areas of research, affecting user trust in AVs[12]. The factors were structured into two main groups depending on if the trust affecting factor referred to the user or the AV. Factors referring to the AV were clustered into *Information about personality, Information on system capabilities* and *Willingness to accommodate to the user.* Factors referring to the user were clustered into *Information to support user understanding*. The literature study also identified the importance of time and dynamics regarding trust.

The complementary user study confirmed the importance of different factors in different usage phases. The complementary user study also proposed usage phases important to consider regarding user trust in AVs.

Trust Factors

Information about personality *i.e. how the AV is perceived*. Earlier research has shown that anthropomorphism i.e. making the system more "human-like" by, say, giving the system a name, gender and voice (Hoff & Bashir, 2015; Waytz et al., 2014) has been shown to have an effect on user trust. Furthermore, by designing an automated system that is perceived as an expert system i.e. portraying the system as competent may also affect user trust (Hoff & Bashir, 2015; Lee & See, 2004), since users may tend to trust automation more than humans carrying out the same task (Madhavan & Wiegmann, 2007).

Information on system capabilities *i.e. the ability of the* system. Earlier research has identified feedback as being important for user trust, that is, continuous system output, ideally available to all the users' senses (Dekker & Woods, 2002; Lee & See, 2004; Thill et al., 2014; Toffetti et al., 2009; Verberne et al., 2012). Furthermore, uncertainty information, that is, showing system degradation, such as sensors not fully functioning but still within acceptable boundaries (Beller et al., 2013; Jian, Bisantz, & Drury, 1998) has also been identified to be important. Other types of information that has been shown to be important is "why & how" information, which involves the system presenting information on upcoming actions. "How" information explains how the system will solve a pending task, while "why" information provides an explanation for its actions. The combination of "why" and "how" information may lead to the user maintaining responsibility for controlling the AV, if the AV is semi-automated (Koo et al., 2014). Finally, error information provided after an error or incident (to explain why it happened and the extent to which the overall system is affected) (Dzindolet et al., 2003; Stanton & Young, 2000) may increase user trust in the AV (Dzindolet et al., 2003).

Accommodating disposition *i.e. how adaptable an automated system is to the needs of the user.* An adaptable system, that is, a system adaptable to users' psychological and physiological states relating to the current situation (Helldin et al., 2013) and to user

[12] These identified factors affecting trust is also presented in Frame of reference.

preferences may be positive regarding trust. Users often trust high-level complex automation less than low-level automation and, thus, may need a system that adapts to user needs regarding levels of control. This may, therefore, improve safety and efficiency (Hoff & Bashir, 2015). Furthermore, customization may also be important i.e. opportunities for the user to adjust non-critical system functions, in order to customise the system to personal preferences (Saffarian et al., 2012; Verberne et al., 2012) (and obtain individually-adapted information). Different individuals need different interventions to correctly calibrate their trust (Merritt & Ilgen, 2008) to the actual performance of the automation. Finally, common goals i.e. that the user perceives the automated system as sharing the user's goals, also seem to be important which could be done by aligning the system's purpose with that of the user. This might be achieved by the system proposing goals, which the user can then accept or decline (Davidsson & Alm, 2009; Lee & See, 2004).

Information to support user understanding *i.e. assist user in understanding system functionality and capability.* Factors referring to the user involves primarily users' mental model of the system. It is important that a user of an AV has a correct mental model of the automated system. Therefore, assisting the user in creating an approximate representation of system functionality and capability helps users understand how to use the system correctly (Lee & See, 2004; Toffetti et al., 2009). One way could be through training. Conducting training before and after system usage, in order to improve users' system knowledge (Parasuraman et al., 2008; Saffarian et al., 2012; Toffetti et al., 2009) may assist users to form an appropriate level of trust in AVs.

Thus, it seems as information about the AV, from the AV, and that the AV accommodates to the user is key to fully understand and accept the AV and therefore generate an appropriate level of trust during the interaction with AVs.

Four phases

From the literature study as well as the complementary user study was concluded that trust is a dynamic concept, changing over time. The aforementioned trust factors have primarily been used to affect users trust *during* the interaction with an AV, whilst the user is learning how the AV works, thus the learning phase. However, from the literature study was also identified that trust is not only affected *during* the interaction with an AV. According to the theory of reasoned action (TRA) created by Ajzen and Fishbein (1980); (Fishbein & Ajzen, 1975) and presented in Lee and See (2004), user trust in automation is initially based on a belief about a trustee arising from information communicated via reputation and gossip. This affects user trust-formation before the user has even interacted with the automated system. Thus, there is an important trust-affecting phase *before* the user has even encountered, much less interacted with, the AV. The phase before the user have encountered and interacted with the AV is hereineafter denoted as the pre-use phase.

Furthermore, later in the user-AV interaction, when a user understands how the AV operates in a specific context (such as on a specific route), trust is based mostly on dependability. It is not as important for the user that the automated system shows intentions; rather, it is more important for the user to receive feedback about the automation's performance (Lee & See, 2004). For instance, a user may fully understand how the AV operates on a specific route to and from work but may not understand this if a different route to work was taken with, say, more traffic. The change in context is an important consideration since automation performance may vary due to the external environment. So, if the user does not fully understand how a change of context might affect the AV's performance, it could lead to misuse and disuse of the automation (Hoff & Bashir, 2015). This highlights another two important phases to consider; a performance phase when the user fully understands the AV capabilities and limitations and nothing changes. However, a change in context *after* the user has learned how the AV works in a given context (a specific route), the user may need to re-learn AV capabilities and limitations in the new context i.e the user is once again entering a (re)-learning phase. The performance- to relearning phase may loop back and forth as soon as a new context, that the user is unfamiliar with, presents itself.

Finally, the *amount* of information is an important factor affecting user trust, since a more transparent system may increase a user's feeling of control by helping predict automation behaviour (Verberne et al., 2012). Thus, more information of AV performance and AV limitations may be extra important before the user fully understands the AV as well as when a context changes to something the user is not familiar with, throwing the user back to a re-learning phase.

Study I, reported in Paper A, contributes to the field of user trust and interaction with AVs, firstly by consolidating trust information and by identifying and highlighting different phases for AV developers to consider when designing for appropriate trust in AVs. Secondly, the results show that trust needs to be viewed holistically, including not just the AV but the user-AV interaction, temporality and various types of trust factors.

4.2 STUDY II

The aim of Study II was, firstly, to investigate whether and how the vehicle’s driving behaviour affects the user trust in the AV (see also Paper B) during interaction with an AV (cf. learning phase identified in Study I) and, secondly, to investigate how the AV’s driving behaviour affects the user trust in the AV during everyday traffic situations (see also Paper C).

4.2.1 Results

Driving behaviour

The results of Study II show that participant trust in the AV was generally high. However, the “Defensive” driving behaviour was perceived as more trustworthy than the “Aggressive” driving behaviour, receiving a momentaneous trust rating median of 6 (on a 7-step scale) compared to 5 (p<0.01) for the “Aggressive” driving behaviour. Similar results were shown in the trust questionnaire ($M_{\text{“Def”}}$=6 vs $M_{\text{“Agg”}}$ = 5.5; p<0.01), “Defensive” being perceived as more trustworthy. This shows that driving behaviour affected user trust. The main explanation for “Defensive” being perceived as more trustworthy was primarily that it was perceived as more predictable than the “Aggressive” driving. It was perceived as more predictable, primarily because the “Defensive” actions were taken earlier and more calmly, whilst the “Aggressive” actions were found to be more sudden and unpredictable and therefore perceived as less trustworthy. However, the participants also found that the AV (through its driving behaviour) showed its intentions by coming to a halt for a pedestrian waiting to cross the road, which was perceived as showing benevolent behaviour towards the pedestrian. Thus, driving behaviour is an important factor to consider, not just because it affects user trust through greater or lesser predictability; it could also be used to convey intentions and benevolence.

Driving behaviour & Traffic Situations

The results in Study II also show that participant trust was affected not only by the AV’s driving behaviour per se; perceptions of the AV’s trustworthiness were also affected by aspects relating to different traffic situations and the AV’s driving behaviour relating to them. This included task difficulty (perceived ease of a task), perceived risks and how well the AV conformed to user expectations of how a traffic situation should be conducted. Sometimes, this affected the participant’s trust more than could be accounted for by the AVs driving behaviour alone.

Perceived task difficulty. In situations with low perceived task difficulty, participant trust was affected by the driving behaviour to a lesser degree. One explanation might be that, in situations with perceived low task difficulty, there was nothing that highlighted the actual capabilities of the AV because the corrective driving actions needed from the AV were minor and few in number. Therefore, it was difficult for the user to understand the AV’s actual capabilities and limitations and build an appropriate level of trust in it.

Perceived risk. Perceived risk to oneself and others also affected participant trust in the AV. Perceived risk to oneself affected participant trust in the AV to a greater extent in low visibility (little information provided by the environment) when there was difficulty predicting what would happen next. Inability to obtain sufficient information about what would happen next in the environment affected user trust because the perception of risk increased. Other aspects affecting user trust included the AV initiating an action in a perceived risk-filled situation without the user knowing why. In these situations, participant trust dropped and neither of the driving behaviours could compensate for it. Rather, the feelings of risk were amplified. Perceived risk to others also affected participant trust to a large extent but not as much as did perceived personal risk. Perceived risk to others seemed to affect participant trust to a large extent, since an accident involving vulnerable road users (VRUs, like pedestrians or cyclists) could lead to severe injuries to those individuals (compared to, for instance, hitting an object such as a signpost). Overall, the perceived risk was higher when there were humans involved in the traffic situation. Therefore, driving behaviour was important to the participants and needed to be well-adapted to the situation. Examples included encountering a traffic situation involving VRUs, when participants perceived the AV (through its driving behaviour) as more or less benevolent, risk-aware (keeping distance from VRUs) and respectful (coming to halt before a VRU crossed a zebra crossing). Thus, if a driving behaviour was benevolent, risk-aware and respectful towards VRUs, participant trust in the AV increased.

Conforming to expectations. How the AV's driving behaviour conformed to user expectations of how situations should be conducted also moderately affected participant trust in the AV. The focus of participants' attention was on how well the AV conformed to the unwritten rules of deceleration and lane positioning. The "Defensive" driving behaviour was generally perceived as best conforming to user expectations concerning deceleration and lane positioning. For example, the "Defensive" behaviour meant slowing down earlier and taking wider turns on roundabouts; this matched participants' expectations of how an AV should negotiate traffic situations. Therefore, the "Defensive" driving behaviour was perceived as more trustworthy.

The results show the importance of adapting driving behaviour to different traffic situations, such as low visibility, and that perceived contextual aspects such as perceived risks, task difficulty and conforming to user expectations are important considerations for assisting the user in forming an appropriate level of trust in the AV.

Study II contributes primarily to the field of trust and interaction with AVs by presenting driving behaviour as a trust-affecting factor, in that it directly affects user trust. However, it seems that driving behaviour can convey different types of trust information, such as performance information i.e. by showing predictability, as well as purpose information i.e. by showing benevolence. Study II also contributes by identifying (user-) perceived contextual aspects affecting user trust in the AV, such as perceived task difficulty and perceived risks.

Thus, these results have implications for a) how to design driving behaviours that assist users in generating an appropriate level of trust relating to the actual performance of the AV and b) how a driving behaviour needs to be designed for different traffic situations.

4.3 ADDRESSING THE RESEARCH QUESTIONS

Six research questions were posed. The aim was to provide answers that in turn, could assist in identifying a design space and relevant variables for developers to consider in enabling users to generate an appropriate level of trust in AVs. The first research questions (1a/b) support this quest by identifying relevant factors affecting trust. The second research questions (2a/b) further explore other factors which may affect user trust, such as driving behaviour. Finally, the third research questions (3a/b) consider the contextual influence on user trust in AVs.

RQ1a: What factors affect user trust in an AV?
RQ1b: Which of these factors should be considered from a design perspective, so as to generate an appropriate level of trust?

The results of Study I identified four clusters of factors affecting users trust in AV: *Information about personality, Information on system capabilities, Accommodating disposition and Information to support user understanding*. Thus, it seems as primarily information from and about the AV is key in affecting users trust. That information from and about a trustee, e.g. an automated system such as an AV, is important for user trust has been identified in earlier research. As described in Frame of Reference, according to Lee and See (2004) and Lee and Moray (1992) performance, purpose and process information are the general basis of trust in automation and important sources from which the user draws relevant information in order to form trust. These factors refer to fundamental constructs such as predictability, reliability, capability, benevolence, faith, dependability and integrity on the part of the AV, and to information on which users base their trust. Hence, one might argue that described trust-affecting factors identified in Study I are consciously or unconsciously based on one or more information sources and that Study I therefore identified and confirmed a number of factors that could be described in terms of performance, purpose and process information. Accordingly, user trust in automation is affected primarily by three fundamental sources of trust information (performance, purpose and process information) during user interaction with an automated system.

However, from a design perspective it is also important to consider the dynamic aspect of trust changing over time. Before the first physical interaction with an AV, i.e. during the pre-use phase, user trust may be based on beliefs about the AV's reputation, by such means as word of mouth.

During the interaction with an AV, i.e. during the learning phase, user trust is affected by information from and about the AV and primarily information about the AV's; performance, purpose and process information provided directly by the AV and interpreted by the user, as well as by other factors such as how "expert-like" the system is perceived to be (Study I). Study II also identified driving behaviour as a trust factor affecting participant trust in general. According to the participants, trust was affected primarily by the vehicle's behaviour conveying predictability, intentions and benevolence, of which predictability and benevolence relates to two of the fundamental sources of trust information, specifically performance and purpose. This further confirms the relevance of the fundamental sources of trust information for users trust in AVs. This also shows the need to consider how AV driving behaviour should be designed to convey an appropriate level of trust to AV users.

Once the user has fully learned how the AV operates and makes its decisions in a specific context, it is more important to then present intentions which inform the user about the AV's performance. This is because later in the user-AV interaction, trust is likely to be based on how dependable the AV is perceived to be (cf. performance phase, Study I). If the context changes and the user experiences the AV in a new context, it is important that the user understands how the AV's performance may have been affected by the change (Hoff & Bashir, 2015). Therefore, users may need more information about the AV's capabilities in the new context and thus re-learning AV capability and limitations (cf. re-learning phase).

Thus, from a design perspective and based on the factors identified in the literature study and in Studies I and II, the relevant trust factors that need consideration are primarily performance, purpose and process information, as proposed by Lee and Moray (1992) and Lee and See (2004). Furthermore, these trust factors need to be considered during four different usage phases; pre-use, learning and performance and/or re-learning phase. Furthermore, it also seems as the driving behaviour is a factor affecting trust during the learning phase, primarily through communicating AV predictability, showing intentions as well as benevolence (see answer to RQ2a/b). Where predictability and benevolence are related to performance and purpose information and thus further supporting that the relevant trust factors that needs to be considered are performance, purpose and process information. Finally, it also seems as there are important factors affecting trust related to the context, that is in traffic situations, such as: (i) perceived task difficulty, (ii) perceived risk to oneself and (iii) to others (VRUs, for example) and (iv) how well the AV conformed to the user's expectations of how a situation should be handled (see answer to RQ3a/b).

RQ2a: Does an AV's driving behaviour affect user trust in it?
RQ2b: If so, how does the AV's driving behaviour affect user trust?

According to the results of Study II, user trust is affected by the AV's driving behaviour. This is because there were different effects on trust according to which driving behaviour the user was experiencing (represented by the "Aggressive" or "Defensive" styles). "Defensive" driving behaviour was generally perceived as instilling more trust than its "Aggressive" counterpart. Trust in the AV was increased when the AV was perceived to: (i) show predictability by decelerating in good time before a traffic situation, (ii) show its intentions by braking whilst in proximity of VRUs (which was interpreted by the participant as the AV having detected the other road user) and (iii) show benevolence by keeping a good distance from VRUs such as pedestrians and cyclists. Thus, overall, the "Defensive" driving behaviour was perceived as more trustworthy than the "Aggressive" one.

RQ3a: Are there any aspects of traffic situations (depending on the AV's driving behaviour) that affect user trust in the AV?
RQ3b: If so, how do traffic situations affect user trust in an AV?

The results of Study II show that users' trust is affected, not only by the AV's driving behaviour, but also by perceptions of how the AV's handling of traffic situations affects user trust. The four aspects identified as affecting user trust were: (i) perceived task difficulty, (ii) perceived risk to oneself and (iii) to others (VRUs, for example) and (iv) how well the AV conformed to the user's expectations of how a situation should be handled.

Situations with perceived low task difficulty only affected participant trust in the AV to a small degree, since it was problematic for participants to fully understand the actual capabilities and limitations of the AV in these situations. They therefore had difficulty building an appropriate level of trust in the AV. On the other hand, perceived risk to oneself affected participant trust to a large extent, due to low situational visibility making it difficult for the user to get sufficient information (from the traffic situation) and predict what would happen next. The participants were, therefore, unable to understand whether the AV had sufficient capability to handle the “unknown” future. Perceived risk to others also greatly affected participant trust, but less than “risk to oneself”. The rationale was that an accident involving a VRU could lead to severe injuries to that individual. Finally, if the AV did not conform to user expectations of how a traffic situation should be negotiated (in terms of deceleration and lane positioning), participant trust decreased, albeit only moderately.

5 GENERATING APPROPRIATE TRUST IN AUTOMATED VEHICLES

This author’s beliefs on approaching design problems from a systemic perspective, plus his aim to identify the design space and relevant design variables (to help developers enable users to generate an appropriate level of trust in AVs), have guided his presentation of a thesis that is descriptive but which also explains and prescribes a perspective on trust. Prescribing solutions to any given problem is a fundamental aspect of design. According to Simon (1996), *“The natural sciences are concerned with how things are…Design, on the other hand, is concerned with how things ought to be, with devising artifacts to attain goals”* (Simon, 1996, p. 114). Thus, the following section is explanatory and prescriptive in nature, describing this author’s view on how to apply the findings of Chapter 4 by illustrating the design space and related design variables.

Based on previous research and the results of Study I and Study II trust seems to primarily be formed by information from and about the AV. To this end, the author has adopted and adapted Shannon and Weaver’s model of communication. This was initially developed to understand how information is communicated between a sender and a receiver, for instance within the context of tele-communication (Shannon & Weaver, 1949). It was later also used to show how an artefact (such as a product), through its design, may be a carrier of information (Monö, 1997). The rationale of using primarily Monö’s model (but also the Shannon-Weaver model), as a foundation arises from trust being based on information about and from a trustee. Both the Shannon-Weaver and Monö models include a sender, or source, that produces a message that is sent to the receiver, or target, via a transmitter. However, the transmitted message may be affected by disturbances (i.e. noise) before reaching its target.

The proposed model, the ‘*Model of trust information exchange and gestalt*’ (see figure 7), explains how trust information in exchanged from a developer i.e. trust information sender, of an AV to an AV user i.e. the trust information receiver.

As mentioned in the results of Study I (see chapter 4. Findings) there are four different phases in the trust formation processes; the pre-use phase, the learning phase and the performance and/or re-learning phase. This means that AV developers must not only consider the AV per se as an artefact generating appropriate trust but to also consider e.g. how the manufacturer of the AV is portrayed through ads or how AV developers portray new AV models through e.g. commercials. Therefore, the model of trust information exchange and gestalt may be used to understand how an appropriate level of trust in AVs may be generated through other artefacts than the AV. Thus, the model may be used in different ways through the use of different artefacts in different trust formation phases. However, the focus of the work presented in this thesis is the learning phase i.e. while the user is still learning how the AV operates - thus, learning how the AV operates during usage - including the capabilities and limitations of the AV.

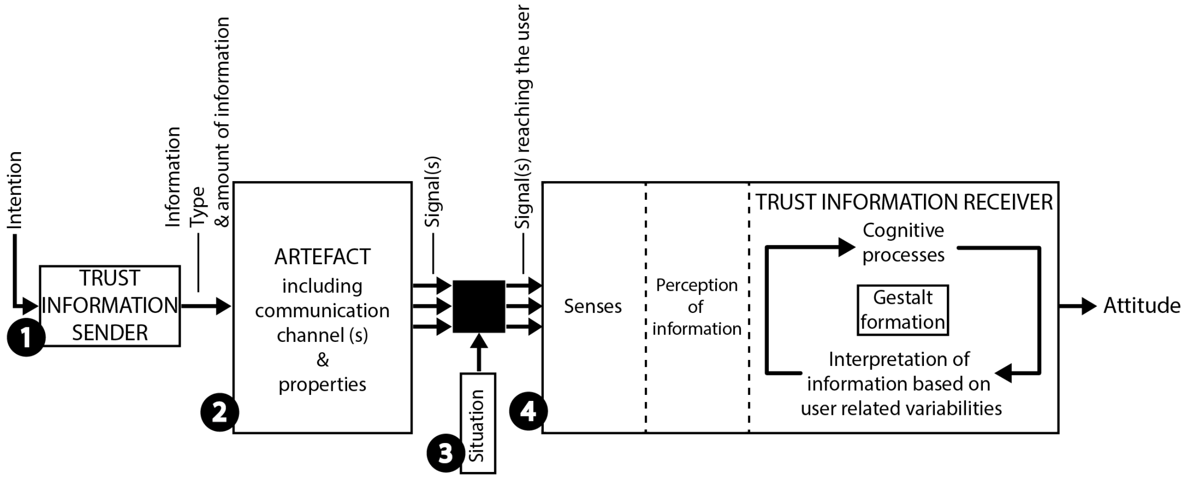

Figure 7 – Model of trust information exchange and gestalt via an artefact, primarily based on and adapted from Monö's model (1997).

Intention, trust information sender & artefact

Monö's model considers companies or developers as possible sources of a message. According to the model, the trust information created, starts first as an intention from the developers and then travels from the developer (trust information sender, see 1, figure 7) to a trust information receiver (see 4, figure 7) via a transmitter, which could be an artefact such as an AV (see 2, figure 7).

The message contains the information needed for the user to create an appropriate level of trust in relation to the actual capabilities of the AV and should therefore contain *information type* (such as information on predictability) and *amount of information* (e.g. giving predictability information before each intersection). However, an AV constitutes a complicated technological system with many sub-systems, several of which communicate information to the user. Therefore, this author considers it necessary to include "communication channels", as integral elements of the AV. This is an important consideration when generating appropriate trust in AVs, since the type and amount of information may be communicated using different communication channels. These might include displays, such as graphical user interfaces, but also an AV's driving behaviour (see results of Study II). Furthermore, the information channels have specified properties. For instance, a display can be formatted as a graphical user interface containing x, y and z functions doing x_2, y_2 and z_2 and driving behaviour could be based on a specific driving style that contain x, y and z properties: accelerating x_2, deaccelerating y_2 and turning z_2 and so on. The communication channel and each individual function (and sub-function) of that communication channel will send out signals to the user; small pieces of information that are interpreted by the user (see Signal(s) figure 7).

Signals

According to Monö (1997), a signal may be defined as an action and might be a directive given in a specific situation. For example, an AV giving indications to the user (behind the wheel) by flashing the right turn indicator (as a way of communicating its intentions and soon-to-be-executed actions). However, as mentioned earlier, trust information may be communicated through different communication channels. For example, the user might receive signals from either indicator flashing as well as from the AV slowly beginning to turn (driving behaviour). Both these signals reach the user, who then makes a combined interpretation of them.

(Traffic) situations affect perception & interpretation
The signals reaching the user may be further affected by how the AV "behaves" in traffic situations in which the user and AV are situated. Monö (1997) describes disturbances as aspects which hinder the design (and thus its intended message) from being perceived correctly (or rather as intended by the trust information sender) by a user. Disturbances might be contextual factors such as noisy traffic, an obstructed view, tiredness and so on. Thus, based on the results of Study II, which showed how contextual aspects relating to traffic situations affected the perception and interpretation of the AV, developers need to consider how the context in which an AV operates affects the perception of signals of the AV. However, this is a complex task due to a mixture of information signals from the AV, the traffic situation and other contextual aspects. It might, therefore, be viewed as a "black box" in which a lot of information coincides and interacts (see black square, figure 7).

The trust information receiver & information gestalt
The signals reaches the receiver's sensory organs and is then perceived and interpreted, based on three different cognitive processes. These are the analytic, analogic and affective processes (Lee & See, 2004), plus specific, user-related variabilities (Hoff & Bashir, 2015; Marsh & Dibben, 2003). However, here the user (i.e. the trust information receiver) interprets all signals communicated from the AV (as well as information given from the traffic situations) together, as a whole. This combination of multiple signals generates an information gestalt (see 4, figure 7). According to Monö (1997), "gestalt" can be defined as *"an arrangement of parts which appears and functions as a whole that is more than the sum of its parts"* (Monö, 1997, p. 33). In other words, individual attributes/functions/properties form signals in the artefact which interact with each other and are perceived and interpreted by the trust information receiver, not as isolated factors, but as a whole, as a gestalt. It is therefore important to consider the information gestalt since AV users seem to interpret an AV's trustworthiness based not only on different signals communicated from different channels but primarily on the interpretation of the AV's information gestalt. The gestalt in turn, is the basis for the user's attitude towards the artefact (AV), that is the user's level of trust.

Thus, **the model of trust information exchange and gestalt** highlights the design space and related design variables which need to be considered when designing for appropriate trust (see Figure 8). These have been identified as *the message*, in other words, the information type: performance, purpose and process information (Lee & See, 2004) as well as how much information to give. *The transmitter* (artefact) means the AV and related communication channel(s) (such as displays & driving behaviour), including properties (such as acceleration), through which little bits of trust-affecting information (signals) may be communicated to the user. The final and perhaps most important variable is the gestalt; how the sum of all signals is perceived as a whole by the user (trust information receiver). In addition, it is important to consider the purpose of the design, i.e. the intention to assist users in generating an appropriate level of trust and for this intention to be clear to everyone involved in the design process. This is considered a precondition and the basis for creating a coherent message that is to be perceived and interpreted coherently by the user (trust information receiver).

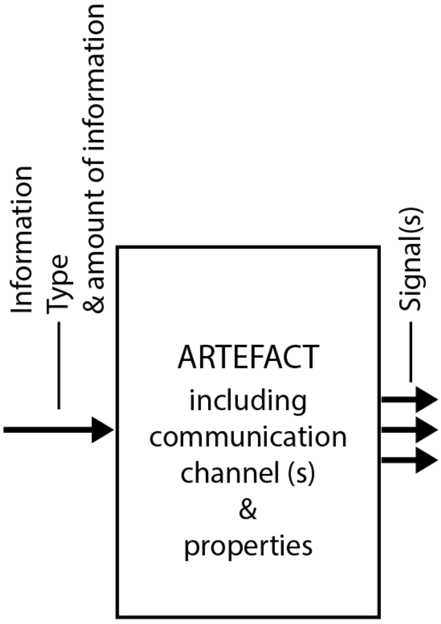

Figure 8 - Design space and included design variables that can be used by developers to design for appropriate trust.

The model of trust information exchange and gestalt contributes to the field of user trust in, and interaction with, AVs by explaining how trust-information is communicated from developers via an artefact (e.g. an AV) and its communication channels (such as in-car displays and/or driving behaviour). Also, by explaining how the sum of trust information (information gestalt) may be perceived differently from individual signals. Furthermore, the model prescribes a new perspective for developers to consider, focusing on the following design space and related variables: type and amount of information, communication channels (including properties) for communicating trust information and evaluating how choice of information type and amount plus communication channels (including properties) are perceived together. This constitutes the information gestalt.

6 DISCUSSION

6.1 EMPIRICAL CONSIDERATIONS

The following paragraphs reflect on (with reference to other related research) the results attained in Studies I and II.

The results of Study I show that trust in automation is primarily affected by automation-related information: *Information about (AV) personality, Information on system capabilities, Accommodating disposition and Information to support user understanding*. Thus, it seems as information from and about the AV is key in affecting users trust. These identified factors can, as argued earlier, be described in terms of three (fundamental) sources of information; performance, purpose and process information. Performance information refers to the perceived capability of the automation and what it does to achieve a trustor's goal. Purpose information describes the intended use of the automation and whether it is used within the realm of the designer's intentions. Process information refers to how the automation makes decisions and whether it is suited to assisting users in reaching their goal. This is consistent with the results of Study II showing that (primarily) performance information (such as predictability) was important to participant trust in the AV. So, even though the trust dimensions as presented by Lee and Moray (1992) and Lee and See (2004) are primarily identified in the area of (general) automation, and not in the context of AVs specifically, they are equally important when designing for appropriate trust in AVs. This is further supported by Lee et al. (2016) who, by applying the trust dimensions, identified design aspects that also affected user trust formation in the area of AVs. Thus, it seems that when designing for appropriate trust, the fundamental sources of trust information also need to be considered in context of automated vehicles.

The results of Study II show that the driving behaviour affected user trust in AVs and further that the "Defensive" driving behaviour was considered more trustworthy than did the "Aggressive" driving behaviour primarily due to being more predictable. The perception of higher predictability was the result of the AV showing its intentions earlier, e.g. starting to slow down earlier, less aggressive acceleration, or earlier and therefore more clearly, positioning itself in the lane before a turn. The results are supported by findings of Bellem, Thiel, Schrauf, and Krems (2018) who studied the effect of acceleration, deacceleration and jerk movements on users' experience of comfort and found that users prefer low acceleration as well as early motion feedback in lane changes. Thus, early and calmly showing intentions before a situation are recommended for a driving behaviour that is intended to increase users' trust.

However, based on the results of the second analysis in Study II, there are contextual aspects related to traffic situations which affect how AV driving behaviour is perceived, such as perceived risks, perceived task difficulty and how well the AV conforms to expectations. Thus, a driving behaviour should not be designed with specific properties without considering the traffic situation in which it will operate. Therefore, it is important to consider how driving properties are interpreted in general but to also consider how driving properties are interpreted by users in specific traffic situations. Thus, designing specific driving properties for different traffic situations may be highly important for the feeling of comfort (Bellem et al., 2018), a correlating aspect to trust (Siebert, Oehl, Höger, & Pfister, 2013). Other results have also acknowledged the importance of context to user trust in AVs. These include the findings of Frison, Wintersberger, Liu, and Riener (2019) who found that the environments greatly affect user acceptance and experience of the AV. In particular, highways and rural roads were perceived as more trustworthy than urban areas, possibly due to the former being perceived as more complex and risk-filled. These findings highlight the importance of considering not only AV driving behaviour regarding trust, but that context (the situation and environment) also needs to be considered. Furthermore, the design of a driving behaviour needs to be broken down into individual properties (such as acceleration) and individual manoeuvres in order to understand how each manoeuvre is perceived by users from a trust perspective. Thus, the breakdown of the driving behaviour is twofold. The first part considers the design of the *general driving behaviour properties* (such as acceleration, deacceleration, lane changes, distance to objects and so on), while the second considers *specific individual manoeuvres* and how these are perceived in different realistic traffic situations, in order to generate an appropriate level of trust.

6.2 THE MODEL OF TRUST INFORMATION EXCHANGE AND GESTALT

Chapter 5 presented the model of trust information exchange and gestalt. The model shows how an intention held by an AV developer (to assist users in generating appropriate trust) is turned into a message, consisting of a type of trust-affecting information and an amount of trust information. The trust-affecting information may then be distributed through different communication channels as specific properties (such as driving behaviour properties or in-car display properties) and sent to the user as signals. These signals (meaning properties generating signals which affect trust) are interpreted as a whole by users; a gestalt. Therefore, it is highly important to consider an AV's gestalt when designing for appropriate trust. This ensures the coherence of all information given to the user and that the whole generates an appropriate level of trust.

This is further supported in a recently published framework by Domeyer, Lee, and Toyoda (2020). Their *automation-incidental user communication* framework also considers communication (of information) in terms of a message being sent (or not) to the user as the influencing aspect that affects trust, perception of risk and acceptance even though it focuses primarily on incidental users. In other words, road users who, by incident, are "affected" by the AV technology. For example, a pedestrian who needs to communicate with an AV so as to cross a road. According to Domeyer et al. (2020), it is important to consider information provided by different aspects of the AV as a whole, in order to enhance the interaction. This further emphasises the importance of focusing, not only on a single signal affecting trust at a given time, but of considering how all signals together, affect the user's trust.

However, whereas Domeyer et al. (2020) suggest that driving behaviour (for example) is a signal, the model of trust information exchange and gestalt proposed in this thesis treats driving behaviour as a communication channel imbedded into the artefact. The communication channels send signals to the user through specific driving properties which becomes a signal that is interpreted by the user. All properties (from all communication channels, such as flashing lights on in-car displays, or acceleration from driving behaviour) constitute the signals which jointly form the information gestalt, which users ultimately interpret and use as the basis for their trust.

So how should a developer go about verifying that all signals jointly from a gestalt are consistent with the intentions of generating appropriate trust? An approach to evaluating both individual signals and the combination of all signals together (a gestalt) is a four-step approach to "seeing mental states" (Becchio, Koul, Ansuini, Bertone, & Cavallo, 2018) in humans, and was proposed for the area of AVs by Domeyer et al. (2020). The four-step approach might be a favourable procedure, since it is based on human perception and may assist in designing the interpretability of an agent's (an AV's) intentions and states (Domeyer et al., 2020). Thus, applying this approach would make it possible to identify whether users perceive a signal (and gestalt) from the AV (or not) and whether the signal is correctly perceived, relative to the intentions set by the developers.

6.2.1 Greater Demands On Interdisciplinary Collaboration

Considering additional design variables, such as incorporating more communication channels (like driving behaviour) opens up new possibilities. However, it also places greater demands on designers to generate a unified, coherent, safe and easy (to interpret) "expression" of the AV (Strömberg, Bligård, & Karlsson, 2019). Furthermore, managing the increased need for a coherent information gestalt may also create greater demand for interdisciplinary collaboration in vehicle development. This is because different company departments need to consider how their respective system design (such as in-car displays) will be perceived and interpreted in terms of trust. Also because the design will relate to other system designs in generating a coherent perception and thus a trustworthy AV gestalt. Therefore, it is important for different departments and thus different competences (such as those working with HMI/UX, driving behaviour algorithms, safety and various other aspects) to have the same view on how the intention (of generating appropriate trust) should be communicated through the design variables and what the information gestalt will ultimately be. To do this, one may start in the end, with the gestalt. This can be done by using metaphors, such as viewing the intended gestalt as a horse (as proposed by Flemisch et al. (2003). Research has shown that using metaphors early in the design process may aid members of a design project to create a joint understanding of a conceptual vision (Strömberg, Pettersson, & Ju, 2020).

6.3 METHODOLOGICAL ISSUES

Having an activity theory perspective has assisted in understanding the dynamic relationship between user, AV and context and has contributed to the research design, in both Study I and Study II. In other words, it has aided the incorporation of as realistic an environment as possible, so as to include aspects fundamental to trust formation, such as perceived risks. If simulator studies had been conducted instead, the validity of the trust measurements would have decreased, due to low (or non-existent) perception of risks and uncertainties. However, the ecological validity might still be discussed. For example, although the experiment in Study II was created to simulate a realistic scenario, there are indications of users having a relatively high level of initial trust (see results of Study II). This could at least in part be explained by having a setting which incorporated an enclosed test facility and having a test leader and operator (wizard driver) in the AV. However, since the experiment compared two design concepts ("Aggressive" and "Defensive" driving behaviours) the relative difference should still be rather similar to that found in a naturalistic context, even though trust levels may then have been generally lower. In addition the experiment was conducted using a wizard driver. This was a professional driver, but nonetheless a human driver and, thus, the consistency of the wizard driver's actions may not have been as good as or as consistent as an AV system. This, in turn, may have affected the results but due to technical limitations, a WOz approach was the only viable option for evaluating the effect of driving behaviours in a fully automated vehicle.

This licentiate thesis also contributes to the area of trust in AVs by showing the importance of using mixed-methods research so as to compare and combine datasets from different methods (Creswell, 2014) and gain results that are as nuanced and reliable as possible. Nevertheless, the sample size and qualitative nature of the data in both Study I and Study II limit the generalisability of the results (Creswell, 2014). Thus, to fully understand the relationship between user, AV and traffic situations regarding trust, more studies are needed. Ideally, these should have larger sample sizes and preferably be longitudinal and naturalistic in nature.

7 CONCLUSION AND IMPLICATIONS

This thesis presents the findings of two studies, Study I and Study II.

The findings from Study I identified and confirmed a number of factors that could be described in terms of performance, purpose and process information as presented by Lee and See (2004). Furthermore, Study I identified four important phases that needs to be considered regarding trust; pre-use, learning and performance and/or re-learning phase. Finally, Study I also identified that the ‘amount of information’ is important for user trust. In other words, how much information the user receives from and about the AV is important to user trust.

Study II identified driving behaviour as a factor affecting trust in AVs, with a “Defensive” driving style was in generally considered more trustworthy than was an “Aggressive” one. Study II also identified four aspects relating to different traffic situations and the AV’s driving behaviour relating to them. For example, perceived task difficulty, perceived risks (for oneself and others) and how well the AV conformed to the users’ expectation of how a traffic situation should be conducted. Thus, trust was affected, not only by driving behaviour but also how the AV interacted, relative to the traffic situation. This revealed an interdependence between AV and situation which must be considered when designing for appropriate trust in AVs. Therefore, it is important for AV developers to consider not only the driving behaviour but also the interdependence between AV behaviour and traffic situations.

This thesis also presents a model based on a synthesis of the findings presented in Study I and Study II, “the model of trust information exchange and gestalt,” primarily adapted from Monö’s model of a product’s communicative functions (Monö, 1997) which, in turn, builds on Shannon and Weaver’s classic communication model). The model of *trust information exchange and gestalt* shows how information affecting user trust is conveyed from developers, starting with an intention, to the user of an AV, that in turn forms trust. Moreover, the model identifies a possible design space and related variables which AV developers should consider when designing for appropriate trust in AVs. The variables are a) the message (the type and amount of information), b) the artefact (the AV, including communication channels and properties) and c) the information gestalt, which is based on the combination of signals communicated from the properties (and communication channels). In this case, the gestalt is what the user ultimately perceives; the combined result of all signals.

As an example, suppose a developer wants to design an AV to assist users in generating an appropriate level of trust in reference to the AV’s actual performance. The developer knows that user trust increases if the AV is perceived as benevolent towards the user. Therefore, the message is designed for benevolence, with the AV acting benevolently every time the user and the AV approach a pedestrian wanting to cross the road. Thus, the developers decide to use the communication channel of driving behaviour to communicate benevolent behaviour by designing driving properties. These might be soft deacceleration (starting to slow down x meters before pedestrian at a rate of y meters per second) as well as coming to a complete halt five meters before the pedestrian. The signals the AV sends (soft deacceleration as a pedestrian is approached and coming to complete halt before them) is then perceived by the user as a respectful gesture towards the pedestrian and, in turn, interpreted as a benevolent act making user trust increase. However, if one of the properties is changed (such as *not* coming to a complete halt five meters before a pedestrian waiting to cross the road, the user’s interpretation of the AV may be different. Thus, the gestalt, the combination of signals, properties and communication channels, needs to correspond to each other, so that the gestalt is interpreted uniformly. The above example also shows the importance of considering the situation in which the AV will operate. For example, without considering the pedestrian wanting to cross the road, the design of the properties might have been different and, therefore, also the perception and interpretation of the gestalt.

Thus, the implication is that AV developers need to test and evaluate relevant communication channels for use in communicating trust-affecting information and to clearly define properties which generate signals that are interpreted by the user as intended. Furthermore, to fully understand how the gestalt of the AV is perceived and interpreted by future customers, testing and evaluation should primarily involve novice users and realistic environments, including real traffic situations.

REFERENCES

Adell, E. (2010). *Acceptance of driver support systems.* Paper presented at the Proceedings of the European conference on human centred design for intelligent transport systems.

Ajzen, H., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior.

Aremyr, E., Jönsson, M., & Strömberg, H. (2019). *Anthropomorphism: An Investigation of Its Effect on Trust in Human-Machine Interfaces for Highly Automated Vehicles*, Cham.

Banks, V. A., & Stanton, N. A. J. A. e. (2016). Keep the driver in control: Automating automobiles of the future. *53*, 389-395. Retrieved from https://ac.els-cdn.com/S0003687015300247/1-s2.0-S0003687015300247-main.pdf?_tid=76919f79-c9b1-4252-9d5b-90338650b4b9&acdnat=1542191776_2a29c20a74db686c23245d81622fdb84

Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of life reviews, 24*, 67-80.

Bellem, H., Thiel, B., Schrauf, M., & Krems, J. F. (2018). Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation research part F: traffic psychology and behaviour, 55*, 90-100.

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: an approach using automation uncertainty. *Hum Factors, 55*(6), 1130-1141. doi:10.1177/0018720813482327

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*: Sage publications.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*: Sage publications.

Davidsson, S., & Alm, H. (2009). *Applying the "Team Player" Approach on Car Design*, Berlin, Heidelberg.

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work, 4*(4), 240-244.

Domeyer, J. E., Lee, J. D., & Toyoda, H. (2020). Vehicle Automation–Other Road User Communication and Coordination: Theory and Mechanisms. *IEEE Access, 8*, 19860-19872.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*(6), 697-718. doi:10.1016/s1071-5819(03)00038-7

Edelmann, A., Stümper, S., & Petzoldt, T. (2019). *Specific Feedback Matters-The Role of Specific Feedback in the Development of Trust in Automated Driving Systems.* Paper presented at the 2019 IEEE Intelligent Vehicles Symposium (IV).

Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behaviour: An introduction to theory and research *Intention and Behavior: An Introduction to Theory and Research*.

Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., & Schutte, P. C. (2003). The H-Metaphor as a guideline for vehicle automation and interaction.

Frison, A.-K., Wintersberger, P., Liu, T., & Riener, A. (2019). *Why do you like to drive automated? a context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces.* Paper presented at the Proceedings of the 24th International Conference on Intelligent User Interfaces.

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

Helldin, T., Falkman, G., Riveiro, M., & Davidsson, S. (2013). Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. 210-217. doi:10.1145/2516540.2516554

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

Janssen, C., Donker, S., Brumby, D., & Kun, A. (2019). History and Future of Human-Automation Interaction. *International Journal of Human-Computer Studies*. doi:10.1016/j.ijhcs.2019.05.006

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (1998). Towards an empirically determined scale of trust in computerized systems: Distinguishing concepts and types of trust. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Vols 1 and 2*, 501-505.

Kaptelinin, V., & Nardi, B. A. (2006). *Acting with technology: Activity theory and interaction design*: MIT press.

Kauffmann, N., Naujoks, F., Winkler, F., & Kunde, W. (2018). *Learning the "Language" of Road Users - How Shall a Self-driving Car Convey Its Intention to Cooperate to Other Human Drivers?*, Cham.

Knutagård, H. (2002). *Introduktion till verksamhetsteori*: Studentlitteratur AB.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2014). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM), 9*(4), 269-275. doi:10.1007/s12008-014-0227-2

Large, D. R., Burnett, G., Morris, A., Muthumani, A., & Matthias, R. (2018). *A Longitudinal Simulator Study to Explore Drivers' Behaviour During Highly-Automated Driving*, Cham.

Lee, J., Kim, N., Imm, C., Kim, B., Yi, K., & Kim, J. (2016). A Question of Trust: An Ethnographic Study of Automated Cars on Real Roads. *8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (Automotiveui 2016)*, 201-208. doi:10.1145/3003715.3005405

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

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors, 46*(1), 50-80. doi:DOI 10.1518/hfes.46.1.50.30392

Madhavan, P., & Wiegmann, D. A. (2007). Effects of information source, pedigree, and reliability on operator interaction with decision support systems. *Human factors, 49*(5), 773-785.

Marsh, S., & Dibben, M. (2003). The role of trust in information science and technology. *37*(1), 465-498.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review, 20*(3), 709-734. doi:Doi 10.2307/258792

McKnight, D. H., & Chervany, N. L. (2000). What is trust? A conceptual analysis and an interdisciplinary model. *AMCIS 2000 Proceedings*, 382.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: dispositional and history-based trust in human-automation interactions. *Hum Factors, 50*(2), 194-210. doi:10.1518/001872008X288574

Monö, R. G. (1997). *Design for product understanding: The aesthetics of design from a semiotic approach*: Liber.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J.-J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning, 1*(209-0015), 13.

Nardi, B. A. (1996). *Context and consciousness: activity theory and human-computer interaction*: mit Press.

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

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

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160. doi:10.1518/155534308x284417

Price, M. A., Venkatraman, V., Gibson, M., Lee, J., & Mutlu, B. (2016). *Psychophysics of Trust in Vehicle Control Algorithms* (0148-7191). Retrieved from

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in Close Relationships. *Journal of Personality and Social Psychology, 49*(1), 95-112. doi:Doi 10.1037/0022-3514.49.1.95

SAE. (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. In: SAE International.

Saffarian, M., de Winter, J. C. F., & Happee, R. (2012). Automated Driving: Human-Factors Issues and Design Solutions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 56*(1), 2296-2300. doi:10.1177/1071181312561483

Seppelt, B. D., & Lee, J. D. (2007). Making adaptive cruise control (ACC) limits visible. *International Journal of Human-Computer Studies, 65*(3), 192-205. doi:https://doi.org/10.1016/j.ijhcs.2006.10.001

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of communication. *Urbana: University of Illinois Press*.

Siebert, F. W., Oehl, M., Höger, R., & Pfister, H.-R. (2013). *Discomfort in automated driving–the disco-scale.* Paper presented at the International Conference on Human-Computer Interaction.

Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences. (2nd editon).

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.): MIT press.

Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science, 1*(4), 315-331. doi:10.1080/14639220052399131

Stockert, S., Richardson, N. T., & Lienkamp, M. (2015). Driving in an increasingly automated world - approaches to improve the driver-automation interaction. *6th International Conference on Applied Human Factors and Ergonomics (Ahfe 2015) and the Affiliated Conferences, Ahfe 2015, 3*, 2889-2896. doi:10.1016/j.promfg.2015.07.797

Strömberg, H., Bligård, L.-O., & Karlsson, M. (2019). *HMI of Autonomous Vehicles - More Than Meets the Eye*, Cham.

Strömberg, H., Pettersson, I., & Ju, W. (2020). Enacting metaphors to explore relations and interactions with automated driving systems. *Design Studies, 67*, 77-101. doi:https://doi.org/10.1016/j.destud.2019.12.001

Thill, S., Hemeren, P. E., & Nilsson, M. (2014). *The apparent intelligence of a system as a factor in situation awareness.* Paper presented at the 2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA).

Toffetti, A., Wilschut, E. S., Martens, M. H., Schieben, A., Rambaldini, A., Merat, N., & Flemisch, F. (2009). CityMobil. *Transportation Research Record: Journal of the Transportation Research Board, 2110*(1), 1-8. doi:10.3141/2110-01

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

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113-117. doi:10.1016/j.jesp.2014.01.005

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2017). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems, 22*(1), 45-55. doi:10.1057/ejis.2011.51