THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

User trust here and now but not necessarily there and then

A Design Perspective on Appropriate Trust in Automated Vehicles (AVs)

FREDRICK EKMAN

Department of Industrial and Materials Science
Division Design & Human Factors

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

User trust here and now but not necessarily there and then
-A Design Perspective on Appropriate Trust in Automated Vehicles (AVs)

Fredrick Ekman

ISBN: 978-91-7905-798-5
Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5264
ISSN 0346-718X

Published and distributed by
Department of Industrial and Materials Science
Division Design & Human Factors
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Telephone + 46 (0)31-772 1000

Cover by Mikael Johansson

Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Abstract

Automation may carry out functions previously conducted only by humans. In the past, interaction with automation was primarily designed for, and used by, users with special training (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 users who have no additional training on automation such as users of automated vehicles (AVs). However, before we can reap the benefits of AV use, users must first trust the vehicles. According to earlier studies on trust in automation (TiA), user trust is a precondition for the use of automated systems not only because it is essential to user acceptance, but also because it is a prerequisite for a good user experience. Furthermore, that user trust is appropriate in relation to the actual performance of the AV, that is, user trust is calibrated to the capabilities and limitations of the AV. Otherwise, it may lead to misuse or disuse of the AV.

The issue of how to *design for appropriate user trust* was approached from a user-centred design perspective based on earlier TiA theories and was addressed in four user studies using mixed-method research designs. The four studies involved three types of AVs; an automated car, an automated public transport bus as well as an automated delivery bot for last-mile deliveries (LMD) of parcels. The users' ranged from ordinary car drivers, bus drivers, public transport commuters and logistic personnel.

The findings show that user trust in the AVs was primarily affected by information relating to the performance of the AV. That is factors such as, how predictable, reliable and capable the AV was perceived to be conducting for instance a task, as well as how appropriate the behaviour of the AV was perceived to be for conducting the task and whether or not the user understood why the AV behaved as it did when conducting the task. Secondly, it was also found that contextual aspects influenced user trust in AVs. This primarily related to the users' perception of risk for oneself and others as well as perceptions of task difficulty. That is, user trust was affected by the perception of risk for oneself but also by the possible risks the AV could impose on other e.g. road users. The perception of task difficulty influenced user trust in situations when a task was perceived as (too) easy, the user could not judge the trustworthiness of the AV or when the AV increased the task difficulty for the user thus adding to negative outcomes. Therefore, AV-related trust factors and contextual aspects are important to consider when designing for appropriate user trust in different types of AVs operating in different domains.

However, from a more in-depth cross-study analysis and consequent synthesis it was found that when designing for appropriate user trust the earlier mentioned factors and aspects should be considered but should not be the focus. They are effects, that is the user's interpretation of information originating from the behaviour of the AV in a particular context which in turn are the consequence of the following design variables: **(I) The Who** i.e. the AV, **(II) What** the AV does, **(III)** by **What Means** the AV does something, **(IV) When** the AV does something, **(V) Why** the AV does something and **(VI) Where** the AV does something, as well as the interplay between them. Furthermore, it was found that user trust is affected by the interdependency between **(II) What** the AV does and **(VI) Where** the AV does something; this was always assessed together by the user in turn affecting user trust. From these findings a tentative Framework of Trust Analysis & Design was developed. The framework can be used as a 'tool-for-thought' and accounts for the activity conducted by the AV, the context as well as their interdependence that ultimately affect user trust.

Keywords: Automated Vehicles (AV); Appropriate User Trust; User-Centred Design Perspective; Trust Factors; Contextual Aspects; Information; Design Variables, Trust Framework

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my main supervisor Professor MariAnne Karlsson, to whom I owe my accomplishments. Thank you MariAnne for your helpful and insightful comments. What sets you apart is not just your scientific competence but also your devotion to your PhD students, always wanting the best for them and pushing them to achieve the very best possible – even though this takes so much of your time. It is difficult to adequately quantify your impact. I can only say that if I ever become half as competent as you are, I would regard my contribution to the research fraternity as a truly remarkable achievement. Thank you for being such a source of inspiration!

Secondly, I would like to thank both my unofficial and official co-supervisors Dr. Helena Strömberg and Dr. Lars-Ola Bligård for the invaluable advice and new perspectives that you have given me. Helena, thank you for your emotional support and for always trying to make me understand that I actually know 'stuff' and have important contributions to make. Not an easy feat due to my highly self-critical and stubborn disposition. I would also like to thank you for your valuable input on my ideas and for helping me figure out ways forward. Lars-Ola, thank you so much for your valuable input and knowledge of systems theory and for all the interesting, inspiring and fun discussions we have had over the past five years. You have both been fundamental to my development as a researcher. Thank you!

Furthermore, I would like to thank my colleague and dear friend Dr. Mikael Johansson for making this journey so truly motivating and enjoyable. It has been a luxury few have been granted. Thank you for all the interesting and inspiring discussions we have had and for the ones still to come. Really looking forward to what our future as (official) researchers will bring. Thank you all my past and present colleagues at the Division for Design & Human Factors, thank you for being supportive and inspiring and for being the best colleagues one could hope for during this process. Without you I would not have made it. A special thank you to Dr. Jana Sochor for seeing the promise beneath the rough surface and for giving me the confidence to pursue a career in research. For that I am eternally thankful.

Thank you to everyone at Volvo Car Corporation, Volvo Buses and HUGO AB who have freely given time, expertise, and the opportunity to conduct field research. I would like to especially thank all the participants who have given their own time to provide me with both inspiration and invaluable feedback. Thank you, Ilya, for your respectful, competent and sometimes last-minute language checks over the years!

Finally, my dear family and friends, I am very grateful to you for the privilege of enjoying your support, unconditional love and affection, thank you so much. Last but not least, my biggest gratitude goes to Emelie, thank you for your support, understanding and encouragement during this sometimes rough period, and to my son Karl, for being my biggest inspiration and for putting a smile on my face on a daily basis.

Appended publications

Paper A
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 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 B
Ekman, F., Johansson, M., Bligård, LO., Karlsson, M. & Strömberg, H. (2021). *Trust in What? Exploring the Interdependency Between an Automated Vehicle's Driving Style and Traffic Situations.* Transportation Research Part F: Traffic Psychology and Behaviour, 76, 59-71.
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.

Paper C
Ekman, F., Karlsson M. & Bligård, LO. *Aspects Influencing Users' Trust in an Automated Delivery Bot: A Pilot Study.* Submitted to Special Issue 'Trust in Automated Vehicles' (Frontiers in Psychology).
Contribution: Ekman planned and conducted the study with feedback from Bligård. Ekman wrote the paper. Karlsson wrote parts of the paper with feedback from Ekman and Bligård.

Paper D
Johansson, M. Ekman, F. Karlsson, M. Strömberg, H. & Jonsson, J. (2022) *ADAS at work: assessing professional bus drivers' experience and acceptance of a narrow navigation system*. Cognition, Technology & Work, 24, 625-639.
Contribution: Johansson planned the study with assistance from the other authors. Johansson, Ekman and Jonsson conducted the study and data analysis was carried out by Johansson and Ekman. Johansson wrote most parts of the paper with guidance and feedback from the other authors. Karlsson wrote parts of the paper with feedback from the other authors.

Paper E
Johansson, M. Ekman, F. Karlsson, M. Strömberg, H. Jonsson, J & Faleke, M. (2022). *Automation as an enabler: Passengers' experience of travelling with a full-length automated bus and expectations of future public transport system.* Transportation Research Arena (TRA).
Contribution: Johansson and Ekman planned the study with input from the other authors. Johansson, Ekman, Karlsson, Strömberg and Jonsson were all involved in conducting the study. Johansson and Ekman carried out the analysis. Johansson wrote the paper with 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.* In Proceedings of the 9th Nordic Conference on Human-Computer Interaction (p. 42), ACM, 1-4.

Ekman, F., Johansson, M., & Sochor, J. (2018). *Creating Appropriate Trust in Automated Vehicle Systems: A Framework for HMI Design.* IEEE Transactions on Human-Machine Systems, 48(1), 95-101.

Ekman, F., Johansson, M., & Karlsson, M. (2018). *Understanding Trust in an AV-context: A Mixed Method Approach.* In Proceedings of the 6th Humanist Conference, The Hague, Netherlands, 13-14 June 2018, 13-14.

Johansson, M., & Ekman, F. (2021). *The Devil is in the Details - Trust Development During Initial Usage of an Automated Vehicle.* 7th Humanist Conference. Rhodes Island, Greece, 1-6.

Johansson, M., Ekman, F., Karlsson, M., Strömberg, H., & Bligård, L. (2021). *Talking Automated Vehicles: Exploring Users' Understanding of an Automated Vehicle During Initial Usage.* 3rd International Conference on HCI in Mobility, Transport, and Automotive Systems, MobiTAS 2021, held as part of the 23rd International Conference, HCI International 2020, 262-272.

MariAnne Karlsson, I.C., Ekman, F., Johansson, M. (2019). *Designing Multi-modal Interaction – A Basic Operations Approach.* Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). IEA 2018. Advances in Intelligent Systems and Computing, vol. 827, Springer, 43-53.

Strömberg, H., Ekman, F., Bligård, L-O., & Johansson, M. (2019). *Keeping a finger in the pie - Exploring different collaborative interactions with autonomous vehicles.* In Proceedings of the 31st European Conference on Cognitive Ergonomics, 118-126.

Strömberg, H., Ramos, É., Karlsson, M., Johansson, M., Ekman, F., Bligård, L., & Bergstad, C. (2021). *A future without drivers? Comparing users', urban planners' and developers' assumptions, hopes, and concerns about autonomous vehicles.* European Transport Research Review, 13(1), 1-12.

Johansson, M., Ekman, F., Strömberg, H., Karlsson, M., & Bligård, L. (2021). *Capable and considerate: Exploring the assigned attributes of an automated vehicle.* Transportation Research Interdisciplinary Perspectives, 10, 1-10.

Abbrevations

AV – **Automated Vehicle(s)** refers in this thesis to automated car, bus and delivery bot (for automated delivery of parcels).

ADS – **Automated Driving System** i.e. the automated system in the AV that executes all driving-related functions, in other words the dynamic driving task (SAE, 2021)

ADAS – **Automated Driving Assistance System** includes a broad range of automated features, e.g. providing warning and/or lane-keeping assist systems (LKA) etc. It is therefore a broad and imprecise term (SAE, 2021). However, when referred to in this thesis it refers to an automated system that can dock a full-length public transport bus by itself at public transport bus stops.

OD – **Operative Domain** i.e. the natural domain for that specific AV, e.g. the context in which user and AV are situated. Includes physical, digital, professional and social characteristics which are referred to as contextual aspects – that may or may not affect user trust in the AV.

ODD – **Operational Design Domain** i.e. a specific domain such as a. highway, in which the AV can or cannot operate automatically (SAE, 2021)

DDT – **Dynamic Driving Task** i.e. all real-time operations and functions needed for a vehicle to operate properly in on-road traffic (excluding strategic functions such as selection of destination) (SAE, 2021)

TABLE OF CONTENTS

1. INTRODUCTION

The first chapter – Introduction – welcomes the reader to the subject of the thesis, trust in automated vehicles (AV), and presents the aim and research questions that contribute to the aim.

1.1. BACKGROUND

Automation may carry out functions previously conducted only by humans (Parasuraman et al., 2000). In the past, interaction with automation was primarily designed for, and used by, users with special training (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 users who have no additional training on automation (Janssen et al., 2019). Automation is now readily available in areas such as education (Mubin et al., 2013) (e.g. social and educational robots) and transportation (automated vehicles) (Janssen et al., 2019).

The concept of road vehicle automation, using automated vehicle systems (AVS), was introduced as far back as the end of the 1930s at the New York World's Fair. The launch of the DARPAS Grand Challenge in 2002, which focused on developing completely automated road vehicles that could travel the distance from Los Angeles to Las Vegas without any user intervention (Behringer et al., 2004) marked the fourth wave of AVS (Shladover, 2018). AVS are today researched, planned, developed, tested and/or used (to different extents) in several different operative domains and within different types of automated vehicles (AVs); within private motoring such as automated cars (e.g. Endsley, 2017; Stockert et al., 2015), public transport such as automated buses (e.g. Hagenzieker et al., 2020; Lundgren et al., 2020), and within the logistics chain, for instance automated delivery bots (e.g. Boysen et al., 2021; Sonneberg et al., 2019).

Interest in the subject of AVs has developed 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) in the operative domain of private and public transport, and to reduce personnel costs in last-mile deliveries of parcels and goods (Sonneberg et al., 2019) in the operative domain of logistics.

However, before we can reap the benefits of AV use, users must first trust the vehicles no matter domain. According to earlier studies on trust in automation (TiA), user trust is a precondition for the use of automated systems (Parasuraman & Riley, 1997); not only because it is essential to user acceptance (Ghazizadeh et al., 2012), but also because it is a prerequisite for a good user experience (Waytz et al., 2014). Even though user trust is important for use, acceptance and good user experience it is even more important that user trust in AVs is appropriate to the actual performance of the system in order to mitigate misuse and/or disuse (Itoh, 2012; Parasuraman & Riley, 1997). It is vital that users understand the limitations of the system in relation to the use context[1].

[1] **Use context -** hereinafter defined as including physical, digital, organizational and social characteristics which are referred to as contextual aspects (that may or may not affect user trust in the AV).

Most earlier research on user trust in AVs has primarily studied user trust in automated cars. Given the present trend to introduce AVs in different operative domains (OD) [2], is it possible to treat user trust the same in all domains? Is it possible to treat AVs of different types and ODs the same when designing for appropriate user trust? Will the user trust them based on the same characteristics whether he or she is a user of an automated car or a passenger in an automated public transport bus, or are there differences?

Secondly, much of the research on user trust is based on, or conducted in relation to, psychological tradition and theories, for instance Muir (1987), Mayer et al. (1995) or Kraus (2020) to name a few. However, if the aim is to 'design for appropriate user trust' in AVs then a design perspective is crucial. Without a design perspective, one is only describing and/or explaining without prescribing any solutions that can be understood and used by developers of AVs (and automation in general) and researchers alike, to design AVs that assist users in creating appropriate levels of trust during use.

Thus, the overall aim of the thesis is to contribute to AV design by developing further knowledge on how to design for appropriate user trust, in different types of AVs within different ODs.

[2] **Operative Domain (OD)** – Hereinafter defined as a specific use context, for which an AV has been purposefully developed to operate in e.g. a bus operates within the traffic system and a delivery bot within the logistic system.

1.2. RESEARCH QUESTIONS

To be able to contribute to the aim, three research questions were posed for the PhD project:

RQ1: What trust factors affect user trust in AVs in different ODs, during use?
The first question was formulated to assist in identifying what affects user trust in different AVs and ODs. Trust-affecting factors are defined as factors that impact trust, hereinafter called "trust factors". They are factors related to the AV that impact user trust in an AV during use.

RQ2: What contextual aspects affect user trust in AVs in different ODs?
The second question was formulated to assist in identifying what contextual aspects relating to the OD affect user trust. Contextual aspects are defined as factors related to the context in which the AV operates and that, in one way or another, affect user trust in the AV during use.

The combination of the first and second question was used as a way of understanding what fundamentally affects user trust from a design perspective.

RQ3: What is important to consider from a design perspective so as to be able to design for appropriate user trust?
The third question was formulated to be able to identify any exceptions to factors and/or contextual aspects relevant for designing for appropriate user trust.

1.3. DELIMITATIONS

Throughout the thesis, the focus will be on user trust affected by the AV itself (trust factors) and/or the OD (contextual aspects) in which the user and AV operate. The focus will be on user trust as affected during use. However, 'use' refers to short-term use, that is to say no longitudinal studies have been conducted.

A user's general tendency to trust, whether due to age, gender, culture and/or personality traits, will not be considered. Neither will a user's mood, confidence, area of expertise and/or attentional capacity.

Finally, using the terms 'factors' and 'aspects' should not be interpreted as if user trust has been analysed using any statistical technique (e.g. factor analysis or similar) even though these terms might create those kinds of expectations. The choice of terminology, that is to say factors and aspects, stems from earlier research within the area of user trust in automation. Thus, the same terminology will be used throughout the thesis.

1.4. THESIS OUTLINE

This thesis is organised as follows:

Chapter 1 provides an introduction to the subject, and details 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. This includes an overview of the methods used in four empirical studies, and a description of the analysis which included three stages; within-study, cross-study, and finally a cross-synthesis.
Chapter 4 presents the main empirical results obtained.
Chapter 5 presents the results of the within- and cross-study analyses and answers RQ1-3. In addition, chapter five presents the results of the cross-study synthesis: a Tentative Framework of Trust Analysis and Design.
Chapter 6 examines what to consider when designing for appropriate user trust and discusses methodological considerations encountered during the project.
Chapter 7 presents the conclusions.

2. FRAME OF REFERENCE

The second chapter of the thesis – Frame of Reference – presents the theories which have been used as a lens through which the empirical data extracted from the four empirical studies presented in the dissertation thesis has been viewed. First, the chapter introduces the reader to the origin of the theories and ends, via related work, with presenting my own reflections on the topic. The intention is to allow the reader to receive a reduced version of the body of knowledge from interpersonal trust via trust in automation (TiA) to the latest empirical findings of trust in AVs, hereinafter referred to as TiAV.

2.1. INTRODUCTION TO TRUST

2.1.1. 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 specifically the trustee's[3] ability 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[4] 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.

2.1.2. The Fundamentals of 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* 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*. 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 and/or uncertainties, hence the possibility that the collaboration might fail (Mayer et al., 1995; McKnight & Chervany, 2000).

For the trustor such as a person wanting to invest money in funds (trustor) at the start of a collaboration with an unknown banker (trustee), trust is based only on *beliefs* generated by information on, and impressions of, the banker. For instance, the person wanting to invest has received information from friends that one specific banker shows good investment performance and when meeting the banker, the banker gives the impression of being knowledgeable, which in turn creates trust based on beliefs. Through affective evaluation of this information, the investor's belief may become an attitude of trust towards the banker. When trust has been established, an intention to rely on the banker may grow. This, in turn,

---

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

[4] A person who forms trust in an agent

may become a behaviour; more specifically, *reliance*. Therefore, trust is an attitude, borne out of a belief about the trustee and an intention to rely on them (Lee & See, 2004).

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; especially when there are uncertainties about the outcome.

2.2. TRUST IN AUTOMATION (TiA)

Trust is important, not only to positive interpersonal relationships; it is also a key aspect in user-automation interaction if the trustor (hereinafter referred to as the user) is to accept the trustee (hereinafter referred to as the automated system) (Ghazizadeh et al., 2012). According to Adell (2010) acceptance describes the degree to which a user intends to use and adopt a system[5].

However, it is not as much about creating trust as it is about creating <u>an appropriate level of user trust</u>. 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 an 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 in turn lead to negative outcomes and in the 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 the system at all (Parasuraman & Riley, 1997) even though the automated system might conduct the task at hand more safely and efficiently than the user him- or herself.

Automation system knowledge and experience are important in aiding an appropriate level of trust. Therefore, it is important for users to understand an automated system's limitations and constraints (Edelmann et al., 2019) by allowing users of automation to observe the actual behaviours of the automated system (Muir, 1987). This is especially important since users without special training and without a sufficient understanding of the capabilities and limitations of an automated system, may view an automated system as more capable than it really is, which in turn may increase the risk of over-trust and thus also misuse (Lee & See, 2004). The risk of over-trust and consequently misuse may also be further increased by developers of automated systems misinforming their non-expert customers through exaggeration of automation system capabilities – also known as autonowashing (Dixon, 2020). An example of autonowashing in the area of automated cars is for instance Tesla who labelled a level 2[6] (SAE, 2021) automated car as "self-driving" although the car provides only a support feature, for example <u>supporting</u> the driver in maintaining position within the traffic lane.

Furthermore, trust in automation shares similarities with interpersonal trust. The aspects affecting the trust formation process (i.e. ability/competence, benevolence, integrity and predictability) resemble the field of trust in automated systems and share three fundamental, corresponding trust dimensions (see section 2.2.1. The Automation). Apart from the corresponding dimensions there are also other elements and contextual aspects that influence user trust (see section 2.2.2. The user, Automation & Context) as well as other factors (see

[5] This is one of several definitions of acceptance, but this is the one I use in my dissertation project.

[6] Level 2 (SAE) is defined as a support feature that is only capable of a limited amount of object and event detection and responses. This in turn means that the user must constantly supervise the automated system to be able to recognize and respond to events that the AV cannot handle.

section 2.2.3. Other Trust Affecting Factors). They all affect user trust in the automated system.

2.2.1. The Automation

According to Lee and Moray (1992) and Lee and See (2004), performance, purpose and process are three dimensions (adapted from interpersonal trust) of information from which a user draws relevant information about the goal-orientated characteristics of an automated system, to form and maintain an appropriate level of trust that in turn can lead to reliance on the automation. 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 dimensions, it will be more stable than if it were based on only one (Lee & See, 2004).

Information on performance refers to what the automation does and relates to the capability, reliability and predictability of an automated system and is similar to capability/competence in interpersonal trust. Information on performance considers the current and historical operation of the automated system, it describes the system's ability to satisfy user goals and is task- and situation- dependent (Lee & See, 2004). 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. However, purpose also refers to the degree to which the designer shows a positive orientation (via the automation) towards the user of the automation, for example accounting for the user's needs (Lee & See, 2004).

Information on process refers to how the automation operates and the appropriateness of the algorithms in any situation to assist in reaching the user's goal. Information on process is similar to the interpersonal aspects of dependability and integrity and shares similarities with understandability, that is to say the degree to which the user can understand the underlying mechanisms of why the automation behaves as it does in the current situation, as well as if that behaviour is favourable for reaching the user's goals(s) (Lee & See, 2004; Sheridan, 1992).

Allowing the user to receive the information related to performance, purpose and process can assist the user in forming a correct mental model of what the automated system should be used for and what the automated system can do, which benefits an appropriate level of trust in the automation (Lee & See, 2004; Toffetti et al., 2009), thereby minimising the risk of not understanding the automated system's limitations (Saffarian et al., 2012) which in turn also decreases the potential consequences of misuse and disuse. As mentioned earlier, trust is an attitude that may lead to reliance. This process can be supported by the dimensions of information relating to the automation (information about the automation) on different levels of detail and is part of a feedback loop. However, it is not only the automation itself that affects user trust but also the context and the user him- or herself (see Figure 1).

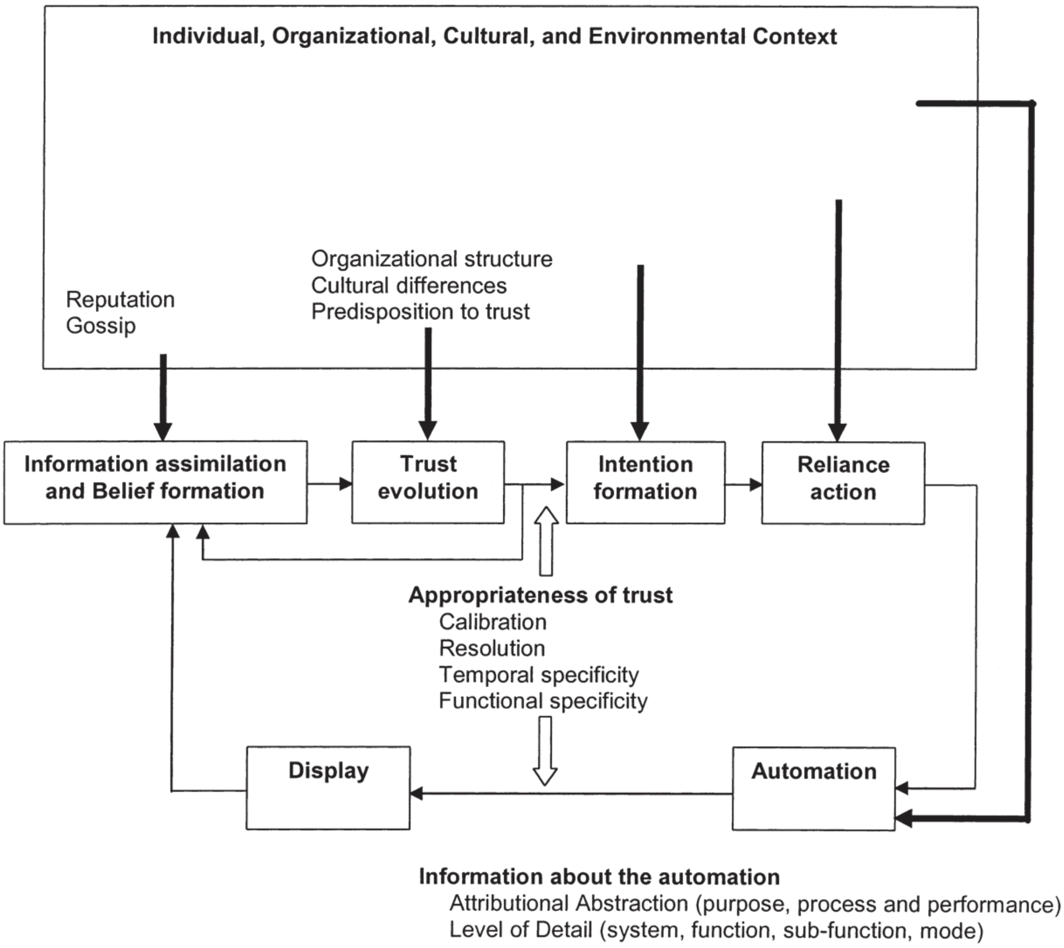

Figure 1 - Lee & See's (2004) conceptual model explaining how a belief via trust may lead to reliance and affected by contextual aspects as well as by information about the automation.

2.2.2. The User, Automation & Context

As mentioned, in addition to the automation, two other elements affect user trust: the user him/herself and the context, for example use environment and/or organizational setting (Hoff & Bashir, 2015; Lee & See, 2004; Marsh & Dibben, 2003; Muir, 1987). All three elements correlate directly to three layers of trust: dispositional, situational and learned trust (Hoff & Bashir, 2015; Marsh & Dibben, 2003) (see Figure 2).

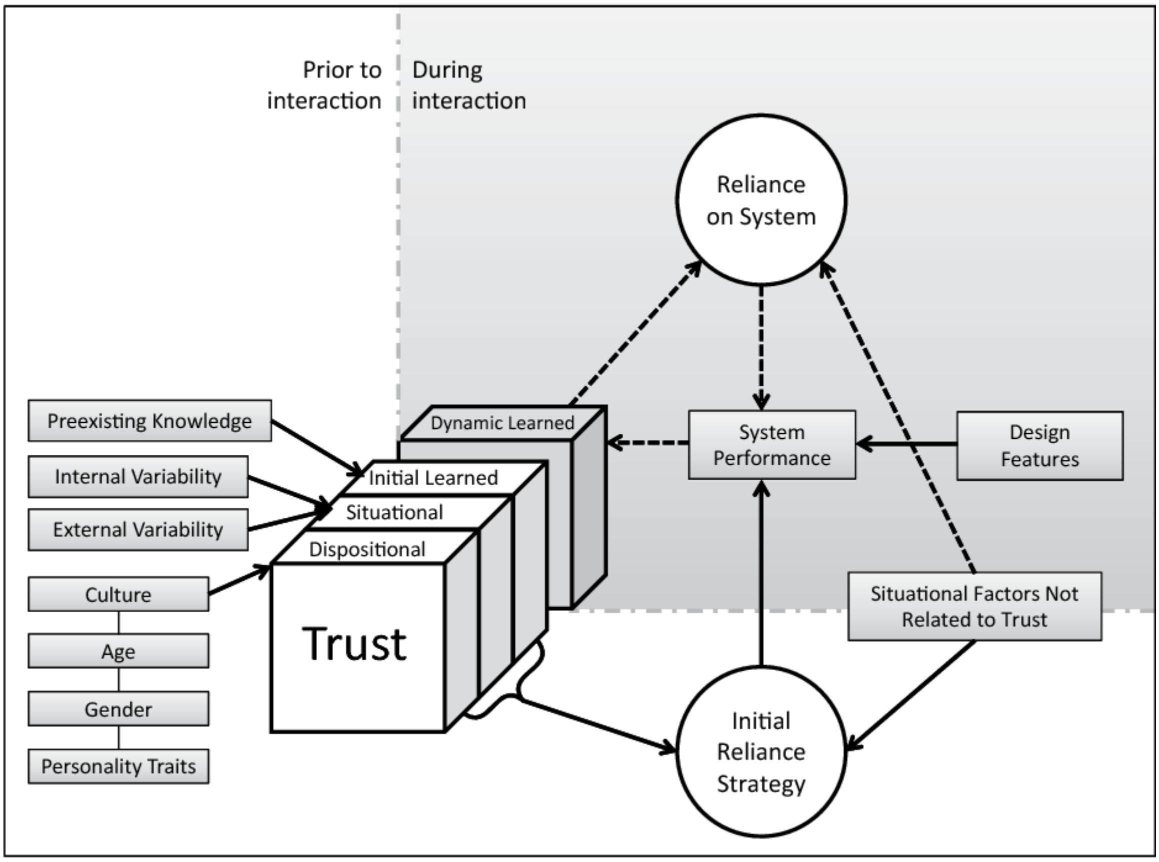

Figure 2 – Hoff & Bashir's (2015) three-layered conceptual model.

Dispositional trust is the user's general tendency to trust automation, irrespective of automated system or context-specific attributes. Rather, this aspect examines influencing factors such as the user's age, gender, culture and personality traits.

Learned trust is trust based on current or previous interaction with the automated system. Previous experiences and pre-existing knowledge affecting user trust are called *initially learned* trust. Initially learned trust creates the reliance strategy, that is to say how one intends to use the automated system, prior to an interaction. This includes trust-affecting aspects such as attitudes/expectations, understanding of and experience with the automated system and the automated system's and/or brand's reputation. The other aspect is *dynamic learned* trust. Dynamic learned trust is formed during an interaction with an automated system and primarily relates to the performance of the automation. Dynamic learned trust generated during the interaction might in turn generate a level of reliance on the automated system (Hoff & Bashir, 2015). Although dynamic learned trust is partly generated via experience it is also formed via *situational trust* (Marsh & Dibben, 2003) where situational trust is created in relation to one specific situation but since an interaction can entail several situations, situational trust affects dynamic learned trust that is formed during an interaction.

Situational trust is user trust created in relation to one specific situation (e.g. a traffic situation) at one moment in time (Marsh & Dibben, 2003) and 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 in a specific situation at one moment in time. The external dimension relates to contextual aspects which according to Hoff and Bashir (2015) include workload, perceived risks, automated system complexity, type of automated system, task difficulty, organisational setting, perceived benefits and how the task is framed in one situation.

Perceived risk and/or uncertainties is perhaps the most prevalent contextual aspect in TiA literature to affect user trust and is defined as a combination of the probability of negative outcomes and the degree of seriousness of the outcome (Numan, 1998). Risk is so important for user trust that Li et al.(2019) even suggest that without perceptions of risk, no matter the level, there is no trust.

However, there are other contextual aspects possibly affecting user trust as well. One of these is system complexity. This is defined as the quantity, in other words the number of items within a system, the variety of different items in the system and finally, the interconnections (cause and effect connections) between the items within the system (Cummings et al., 2010). In the case of automation such as AVs, one could argue that it could for instance relate to the number of other road users (quantity) and the different types of road user such as pedestrians and cyclists (variety) and to the degree to which they affect each other (cause and effect connections). This in turn may affect user trust.

Furthermore, the perception of how difficult a task is has also been shown to affect user trust. For instance, when a task is perceived as difficult, users may be more prone to believe in and use the automation (Schwark et al., 2010). Perceived workload has also been identified as a contextual aspect affecting user trust in automation (Hoff & Bashir, 2015) and can be defined as demands of work in terms of difficulty, complexity and time pressure (Popkin, 1999). It has been found that workload may affect the positive relationship between user trust in automation and automation use, where higher perceived workload increased user reliance on automation (Biros et al., 2004).

Moreover, perceived benefits – the advantage or profit you can get from something such as use of automation to assist in reaching a goal – affects user trust. It has been found that there are positive relationship between the two (Park et al., 2019) Finally, the organizational setting in which in interaction takes place has also been found to influence user trust in automation (Hoff & Bashir, 2015).

2.2.3. Other Trust Affecting Factors

Additionally, trust in automation has been found to be affected by: training in the automated system's functionality before and after first usage (Parasuraman et al., 2008; Saffarian et al., 2012; Toffetti et al., 2009), and/or by providing continuous and accurate feedback (Dekker & Woods, 2002; Thill et al., 2014; Toffetti et al., 2009; Verberne et al., 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 automated system's 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 argued as 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, it could lead to distrust in the automated system (Saffarian et al., 2012).

System transparency has also shown to affect trust 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 (Beller et al., 2013; Jian et al., 2000). System uncertainty can be explained as showing system reliability (or rather the lack thereof) for users to understand that the automated system is operating at a reduced level of reliability. Another type of transparency 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 et al., 2003; Stanton & Young, 2000).

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). Portraying an automated system using human-like features (see anthropomorphism) has also shown to affect user trust (Hoff & Bashir, 2015).

Thus, in summary:
(i) Trust is an attitude held by a trustor towards a trustee (can be either man or machine) such that the trustee will and can assist the trustor to achieve a goal in a situation where there are potential risks. Trust can be described as an attitude, created from a belief that in turn may transform into an intention to rely on a trustee and potentially end in reliance which can be defined as a behaviour by the trustor towards the trustee.
(ii) Trust in automation is (primarily) affected by three dimensions of information: performance, purpose and process information, that is to say information communicated to the user regarding how capable the automation is at satisfying a user's goal(s), information so that the user understands the intended use of the automation, and information allowing the user to understand how the automation makes decisions and to what degree the behaviour is appropriate for any given situation.
(iii) User trust is not only affected by the automation but also from the user's tendency to trust automation based for instance on cultural background and/or age etc. (dispositional trust), and;
(iv) The context in which user and automation interact seem to affect user trust. This since the context may include different aspects such as perceived risk, task difficulty, system complexity, workload, benefits and the organizational setting itself that may influence trust and reliance. Thus, the context is important to consider.
(v) System transparency seems to affect trust by for instance assisting users in achieving a greater feeling of control by helping the user to more easily predict how the automated system will behave.
(vi) The way the system is characterized has an impact on trust, such as designing/implementing anthropomorphic features.
(vii) Finally, mitigating mechanisms have also been identified as affecting trust, more specifically training (understanding of how to operate automation, that is to say the functionality of the automated system), and continuous and accurate feedback which can be divided into two types: learning and action feedback.

2.3. TRUST IN AUTOMATED VEHICLES (TiAV)

One application area of TiA that has increased in interest is trust in automated vehicles (TiAV)[7]. There is interest for example in identifying user tendency to trust AVs and why, as well as in solutions to mitigate primarily misuse but also disuse in order to minimize the risk of incidents when AVs are introduced into the market. However, more particularly there has been an increasing focus on empirically evaluating different TiA models.

2.3.1. Information affecting TiAV

One example of a study researching the effects of information on user trust in AVs is Hergeth (2016) who in a driver simulator study investigated the effects of the three dimensions performance, purpose and process as described by Lee and See (2004) on user trust in automation. He found that purpose and performance information (communicated via displays) had no significant effect on participants' trust in the automation during use, in other words on dynamic learned trust. However, process information had a positive effect on trust in automation. Based on other findings it seems that many of the antecedents relating to performance, purpose and process have had different effects on user trust in relation to TiAV (Koo et al., 2014; Körber et al., 2018; Noah & Walker, 2017).

Performance information, for instance, by giving introductory information affecting user perception of how reliable an automated system is, has been shown to affect user trust (Körber et al., 2018). In addition, the provision of uncertainty information, that is information related to how reliable the automated system is (Noah & Walker, 2017) during an interaction with an AV has been shown to affect TiAV (Beller et al., 2013).

Purpose information, such as information about why the automated system was developed and what it is to be used for, has also been shown to affect user trust. In a study conducted by Beggiato, Pereira, et al. (2015) participants received information via the owners' manuals about the AV's functionality before use, which in turn affected user trust – but not enough to provide a long-lasting effect in terms of assisting the user of the AV in maintaining a level of appropriate trust in the AV during use.

Process information, that is to say information about why automation acts in a certain way as well as the appropriateness of that behaviour in relation to the current situation, has been shown to decrease trust simply because the information has been received. This is due to a mismatch between user perception of the amount of competence it takes to (i) explain why the AV could not perform a task, and (ii) to actually perform the task. In other words, if the AV could explain why it could not perform the task, the AV must surely be capable enough to perform the task (Korber et al., 2018). On the other hand, providing information on *why* an action is conducted by the AV has, according to Koo et al. (2014), been shown to decrease anxiety and increase trust. Therefore it seems as if the action of providing the information is perceived as including more capability on the part of the AV than the task itself which may lead to a decrease in trust. If not, trust may increase. Thus, it seems that there are discrepancies between what effects different types of information have on TiAV.

[7] Trust in automated vehicles will hereinafter use the acronym 'TiAV' which is not an established acronym for trust in automated vehicles but only used in this thesis as a way of simplifying the task of reading, by clearly separating TiA and TiAV.

2.3.2. Learned Trust

One of the more ambitious recent research endeavours is the dissertation by Kraus (2020) who aimed to investigate the proposed dynamic formation and calibration of trust based on the theories put forward by Lee and See (2004) and Hoff and Bashir (2015) respectively. He did so with a series of empirical studies on trust in automated vehicles carried out in a driving simulator. The end-result was a Three Stages of Trust framework that describes trust formation in a specific automated system in three stages: **(i)** propensity to trust automation is formed and is based on personality, disposition and earlier knowledge of technology and automation, **(ii)** trust prior to a first interaction i.e. *initial learned trust*, is based on a propensity to trust and information available before automation use. From here expectations are created for the specific automated system, and **(iii)** the last stage, during an interaction with automation, that is to say *dynamic learned trust*, a user's expectations of the system is compared to the actual automated system behaviour and information given via displays. Thus, *initial learned trust* is partly based on information available before automation use, which in turn creates expectations regarding the specific system, while *dynamic learned trust* relates to expectations that are continuously compared to the automated system's behaviour as well as information from displays. The findings by Kraus further support but also develop the learned trust layer as described by Hoff and Bashir (2015).

2.3.3. Situational Trust

It has also been shown that the context in which user and AV are situated also affects user trust in the AV. A study by Frison et al. (2019) found that during automated highway driving the participants trusted the AV; on rural roads the driving style of the AV was more important but during the urban scenario the high complexity of the environment decreased trust to a major extent. Thus, it seems that both the type of road and traffic volume affect user trust in AVs to a great extent. Frison et al. (2019) further describe that the participants in their study felt most safe on a rural road with light traffic but had trust issues when they encountered the urban area with moderate traffic. Lastly, Walker (2021) also identified the effect of context on user trust in AVs. More specifically, he found that dynamic learned trust formed through real-life experiences of different traffic situations led to improvement in situational trust calibration even for traffic situations the participant had not yet encountered. Walker refers to this phenomenon as dynamic learned trust "generalization", which in turn shows that trust calibration is possible without users even needing to encounter all traffic situations to form trust.

2.4. REFLECTIONS

It seems that the general theories of TiA, as described by Lee and See (2004) and Hoff and Bashir (2015), have explanatory power also in the area of TiAV. The research on TiAV and more specifically initial, dynamic learned and situational trust, has been shown to be important to understand TiAV too.

However, there are limitations. In the vast majority of cases simulator studies have been used. As described by Raats et al. (2019), these studies are *based on a hypothesis that consists of fixed response options and decontextualized interactions; to test reliance in a system and how it develops during use"*. This statement is further supported by Li et al. (2019) who claim that without any perceived risk, trust measurements may be invalid, and further that it is often reliability that is being measured and not trust. Secondly, since simulators are rather rigid in terms of possible and unforeseen contextual effects upon the AV, as described by (Muir, 1987), the more stable the environment is, the more predictable it is and therefore user trust is inversely related to the amount of disturbance imposed by the environment upon the machine (whether it may be traffic situations or organizational influences), something a simulator study somewhat lacks. Therefore, it seems highly important to include the actual OD of AVs, at least as a way of creating user expectations of a realistic context in which (harmless) unforeseen things can happen in order for users to experience the need or willingness to trust the AV and to have ecological validity, that is to say to what degree measurements in a research setting are representative of the real world (Chang et al., 2022).

Furthermore, the bulk of research on TiAVs deals with automated cars, with less focus on different types of AVs and even less on comparisons between types of users, transport modalities and ODs, such as trust in automated full-length public buses and/or automated delivery bots[8]. However, the question is whether the TiA theories apply also for these other types of AVs. This is important since the relationship and interaction between a passenger in an automated full-length public bus and a driver of an automated car or user of an automated delivery bot is completely different. In an automated car or bus you interact with the AV from an internal point of view and with an automated delivery bot from an external point of view. The relationship with an automated car may be more personal than with an automated bus. Secondly, these types of AVs also have different purposes, such as transporting goods versus transporting people and/or being used within a work context versus being used as a tool for private use (automated car). Nevertheless, I would argue that it is also important to start with the same premises if trying to understand user trust in different AVs, that is to say starting with the same theoretical background, approach and perspective. Otherwise it is difficult to be able to create a unified perspective of user trust in AVs. If, indeed, they share any similarities in terms of user trust.

Finally, little of the current research in user trust in AVs has been conducted by design researchers and 'research for design' focusing on the user. Instead, a majority of user trust is based on, or conducted within, the tradition of psychology and similar approaches. This can lead to a less actionable foundation for someone who wants to design for appropriate user trust. Design concerns 'conception and realisation of new things' that are useful (Archer et al.,

[8] An automated delivery bot is a type of autonomous unmanned ground vehicle (AUGV). An AUGV is defined as a vehicle that requires minimal to no operator interference and that uses system mapping, GPS, stereo vision, sensors and edge computing to navigate itself from one location to another (Mishra & Das, 2019). One type of AUGV is micro-freight robots or bots for short to be used for example in the logistics chain, more specifically for last-mile deliveries (LMD) often the most inefficient and most expensive part of the logistics chain (Sindi & Woodman, 2020; Sonneberg et al., 2019).

2005). Designers are good at resolving badly defined problems, they adopt solution-focused strategies and according to Dorst (2011) humans tend to solve problems via deduction and/or induction whilst 'design' is more concerned about using abduction as a scientific approach. This allows the design researcher to pursue answers in an iterative manner by continuously going back and forth between real-life observations and matching these observations with earlier theory until a final theory (final conclusion) is created, a solution that can be useful.

Therefore, design is not only about searching for 'truth' to the same extent as for instance the natural sciences but even more so it is about creating value for someone or something as described by the statistician George E.P. Box: *"Essentially, all models are wrong, but some are useful"* (Box & Draper, 2007, p. 414). A quote that simplistically describes the underpinnings of design. For design the usefulness of something is sometimes more important than the complete truth. Therefore, based on the complex nature of designing for appropriate user trust, a design perspective can be useful since it is pragmatic and therefore can, if needed, take a different approach to the same problem with less concern for following a predetermined route to reach a solution.

Thus, the scope of this thesis has been on exploring what trust factors and contextual aspects affect user trust in different AV types, in their respective OD using as realistic study setups as possible and to understand, from a design perspective, how to help developers and researchers to develop AVs that assist users to create appropriate user trust, ultimately mitigating misuse and/or disuse.

3. RESEARCH APPROACH

The third chapter of the thesis – Research Approach – describes the author's theoretical and philosophical perspective as well as the research process. The main purpose of the research approach was to support the dissertation project in addressing the aim and assist in answering the research questions. The theoretical and philosophical perspective is described to allow the reader to get acquainted with my background and worldview(s), to communicate the rationale behind the research process and to contextualize the research. The research process is described to explain which research methods have been used for data collection and analysis as well as why.

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 not only shape how we view reality, they also govern the processes of research. As a researcher, it is important to understand one's philosophical worldview in order to justify one's practices. Therefore, the following section presents my theoretical and philosophical perspective and describes the methodology used.

3.1.1. Personal Setting

With an educational background in industrial design engineering, my 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. A design problem cannot be understood without first understanding the user, the activity and the context in which the activity is to be conducted.

My research focuses on the design problem of how to generate an appropriate level of trust in automated systems, such as automated vehicles (AVs). Furthermore, a design perspective stresses a focus not only on user and AV, but to also consider the context including physical, digital, organizational and social characteristics. In other words, the contextual setting such as the physical environment in which the user's activity takes place (referred to as OD), using an AV as a tool to reach a goal. Thus, the interplay between user, AV and context was my point of departure in the dissertation project.

Finally, the dissertation project and the thesis support 'research for design', in other words the research contributes to outcomes such as: *"…conceptual frameworks, guiding philosophies, design implications arising from the investigation of people and contexts, and design implications arising from the analysis of designed artifacts" (Forlizzi et al., 2009, p. 2892).* Thus, the dissertation not only intends to be descriptive, for example only depicting user trust but also prescribes solutions to support developers and researchers alike, to design AVs that assist users in creating appropriate levels of trust.

3.1.2. Philosophical Worldviews(s)

I believe that an objective world exists, with or without our presence, but also that the world is shaped and affected by our interpretation of the same (cf. the ontology of a critical realist Guba & Lincoln, 1994). Thus, a complete understanding of objective reality can never truly be possible as we reshape our perspective on the world every day. Our perspective changes through new experiences and through the acquisition of new knowledge.

Furthermore, trust is something that most people can relate to (at least as regards interpersonal trust). A person's trust in a trustee may change 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).

As of today, the most common way of assessing users' perception is via what users verbalise, for example in interviews and questionnaires. However, question-based methods are only one means of understanding what factors affect user trust in AVs. I have therefore 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. According to Kohn et al. (2021) it is important to make use of different data collection methods when studying trust for reliable and valid results. Therefore, mixed-methods research has been the foundational approach of the dissertation project. This includes different questions-based methods and to some extent objective measures (behavioural measures), primarily to gain as nuanced an image of trust as possible and to address the issue of being able to design for appropriate user trust relating to the actual AV performance in the context in which user and AV operate.

3.1.3. Research Setting

As important as being transparent with one's personal setting and worldviews is, it is equally important to describe the setting in which one's research has been conducted. The dissertation project presented in this thesis is based on four different studies of which all were done in collaboration with other stakeholders.

Study I

The first study (Study I), presented in ***Paper A and Paper B***, was part of the 'HaTRIC' project (https://www.saferresearch.com/projects/hmi-autonomous-vehicles-traffic-hatric). The project had the overall purpose of identifying specific vehicle characteristics that may affect user trust, users' understanding as well as users' overall experience of automated vehicles. The involved stakeholders included: Volvo Car Corporation, VTI (Swedish National Road and Transport Research Institute) and Chalmers University of Technology. The project included several studies of which one is presented in this thesis (Study I). The primary focus of Study I was what effect an automated car's driving behaviour has on user trust in the AV, and secondly, how users understand AVs. The study was conducted in cooperation with another PhD student (whose research focused on users' understanding of AVs).

Study II

The second study (Study II), presented in ***Paper C***, was part of the 'Carbon Neutral Urban Logistics' project (https://www.energimyndigheten.se/forskning-och-innovation/projektdatabas/sokresultat/?projectid=30587). The overall aim in the project was to develop an autonomous ground vehicle delivery system (referred to as automated delivery bot in this thesis) to better understand the environmental impact of urban logistics as well as the implications on society and infrastructure of shifting towards autonomous solutions. The involved stakeholders included: HUGO Delivery AB, Chalmers University of Technology, University of Gothenburg, HSB, Chalmersfastigheter, Akademiska Hus, Johanneberg Science Park, Ernst Rosen, Trafikkontoret, City of Göteborg and Johanneberg Science Park. The Carbon Neutral Urban Logistics project involved several studies, one of which is presented in this thesis (Study II). The focus of Study II was to evaluate how logistics personnel experienced and trusted a bot as a tool for last-mile deliveries (LMD) of parcels.

Study III and Study IV

The third and fourth studies (Studies III and IV), presented in ***Paper D*** and ***Paper E***, were part of the national 'KRABAT' project (https://www.drivesweden.net/en/node/796). The overall mission of the project was to push the development towards a shift in the Swedish transport system through the use of self-driving, electric and shared vehicles. Due to the type of project (i.e. a national level project), many stakeholders were involved. The project was divided into six sub-projects of which one relates to the two studies presented here. The purpose of both studies was to evaluate how users understand, accept, experience and trust an automated system, although Study III focused on how an automated system that docks the bus at bus stops was trusted by the bus driver and in Study IV, the focus was on public transport passengers and their experience, acceptance of and trust in a fully automated public transport bus.

3.2. RESEARCH PROCESS

3.2.1. Overall Methodology

The four studies (Studies I-IV) which make up the dissertation project were completed between 2017 and 2021. The studies differed from each other in character. Study I was an experiment in a controlled yet realistic environment (including simulated traffic situations) in which the independent variables were controlled for (driving behaviour including specific driving properties, see Papers A and B) and which included pre-determined groups based on age and gender. Study II was a field study conducted in the participants' work context where no consideration to gender or age distribution was made (see Paper C) as the participants were an already naturally formed group consisting of logistics personnel (cf. convenience sample, Creswell (2014)). Study III and Study IV were experiments in a controlled environment. Study III included a convenience sample where the participants were bus drivers who were interested in participating in the study (Paper D). Study IV included pre-determined groups based on age, gender and geographic affiliation (cf. small city versus big city, see Paper E).

As described earlier I believe that no design problem can be addressed without accounting for the context. This is further supported in the area of TiA since: (i) user trust is inversely related to the amount of disturbance posed by the environment on the machine (Muir, 1987) and; (ii) without any perceptions of risk (related to the interdependency between AV and context) a valid trust results may not be fully established (Li et al., 2019). Therefore, the studies were conducted in as realistic a context as possible, for instance by simulating realistic elements often present in the actual OD of AVs, such as everyday traffic situations, or by evaluating AVs in their actual OD.

However, it is equally important that the AV itself is perceived as credible and realistic so that the results are ecologically valid. However, since much of the AV technology, especially for the higher levels of automation such as L3-L5 (SAE, 2021) is still under development, 'Wizard of Oz' (WOz) approaches had to be used to (partly or fully) simulate the technology. The WOz approach basically involves having a human impersonating a machine without the participant knowing about it (Kelley, 1984). The WOz approach is a useful option early in the development phase, when there are still technological limitations (Habibovic et al., 2016) and has been used to evaluate user interfaces, driving behaviour and secondary tasks (Habibovic et al., 2016; Müller et al., 2019). The WOz approach was used to different degrees in the dissertation project.

Study I included regular car drivers experiencing a fully automated car, L5[9] (SAE, 2021) on a test track with simulated everyday traffic situations such as a pedestrian crossing the street. The fully automated car was completely simulated using a WOz approach and used no automation.

Study II included logistics personnel working with the logistics service at a university campus. They experienced a fully automated delivery bot L5 (SAE, 2021) operating within their logistics system. The intentions were to use an automated delivery bot that could be defined as having an automation level of L3, in other words that it could operate completely by itself under normal operation. Only in the event of technical issues would an operator be ready to respond remotely and take over the controls (SAE, 2021). However due to technical

[9] **SAE level 5** refers to the highest level of automation (Full Driving Automation) and is defined as an ADS (in an AV) that can perform the entirety of the DDT without being ODD dependent or needing a fallback-ready user (SAE, 2021).

issues the automated delivery bot had to be completely simulated (hence using the WOz approach) by developers remotely controlling the automated delivery bot.

Study III included professional bus drivers experiencing an automated bus L3[10] (SAE, 2021) and a system for automatically docking at bus stops in a naturalistic traffic setting. In this case no simulations were used.

In Study IV, frequent users of public transport experienced an automated bus on a test track including everyday traffic situations. The automated bus was pre-programmed to follow a route and, for example, to stop at intersections and stop to let people get on and/or off at bus stops. The traffic situations included several extras who acted and adapted to the bus in as natural a way as possible so that the impression was that the bus stopped for them (even though the bus lacked any possibility to adapt to other road users). Therefore, the automated bus was actually L3 (SAE, 2021) but perceived as L5 in that specific setting (see Table 1).

Table 1 – Summarizes the type of study conducted including type of user, type of AV, OD, actual automation level and perceived automation level.

Study No. (Year)	Characteristics	User & AV	OD	Actual Automation Level	Perceived Automation Level
Study I (2017)	*Experimental User Study*	*Regular car drivers & Automated car*	*Test-track including bidirectional rural road, urban streets and seven simulated traffic situations*	*SAE L0(WOz)*	*Perceived as SAE L5*
Study II (2019-20)	*Field Study*	*Logistic personnel & Automated Delivery Bot*	*Last-Mile Logistics at a university campus*	*SAE L0(WOz)*	*Perceived as SAE L5*
Study III (2021)	*Experimental User Study*	*Professional bus drivers & Automated bus*	*Real traffic conditions in an industrial area with five bus stops*	*SAE L3*	*Perceived as L3*
Study IV (2021)	*Experimental User Study*	*Public transport passengers & Automated bus*	*Test-track including bidirectional urban streets and simulated traffic situations*	*SAE L3*	*Perceived as SAE L5 due to OD*

[10] **SAE level 3** refers to mid-level automation (Conditional Driving Automation) and is defined as an ADS (in an AV) that can perform the DDT in specific ODD and that relies on a fallback-ready user (if something were to happen) (SAE, 2021).

3.2.2. Data Collection and Analysis

The data collection and analysis were conducted in three stages (see Figure 3). The first stage included data collection via the four studies conducted. The first stage also included a within-study analysis, in other words each study was separately analysed in relation to the separate aims of the studies.

The second stage included a cross-study analysis which allowed me to go back to the data collected in each study and compare the data against one another and against the frame of reference.

Finally, the third stage included a cross-study synthesis, which was conducted to interpret and synthesize the empirical findings from the within-study analysis and the cross-study analysis and then compared against the frame of reference. The synthesis led to the tentative framework of trust analysis and design.

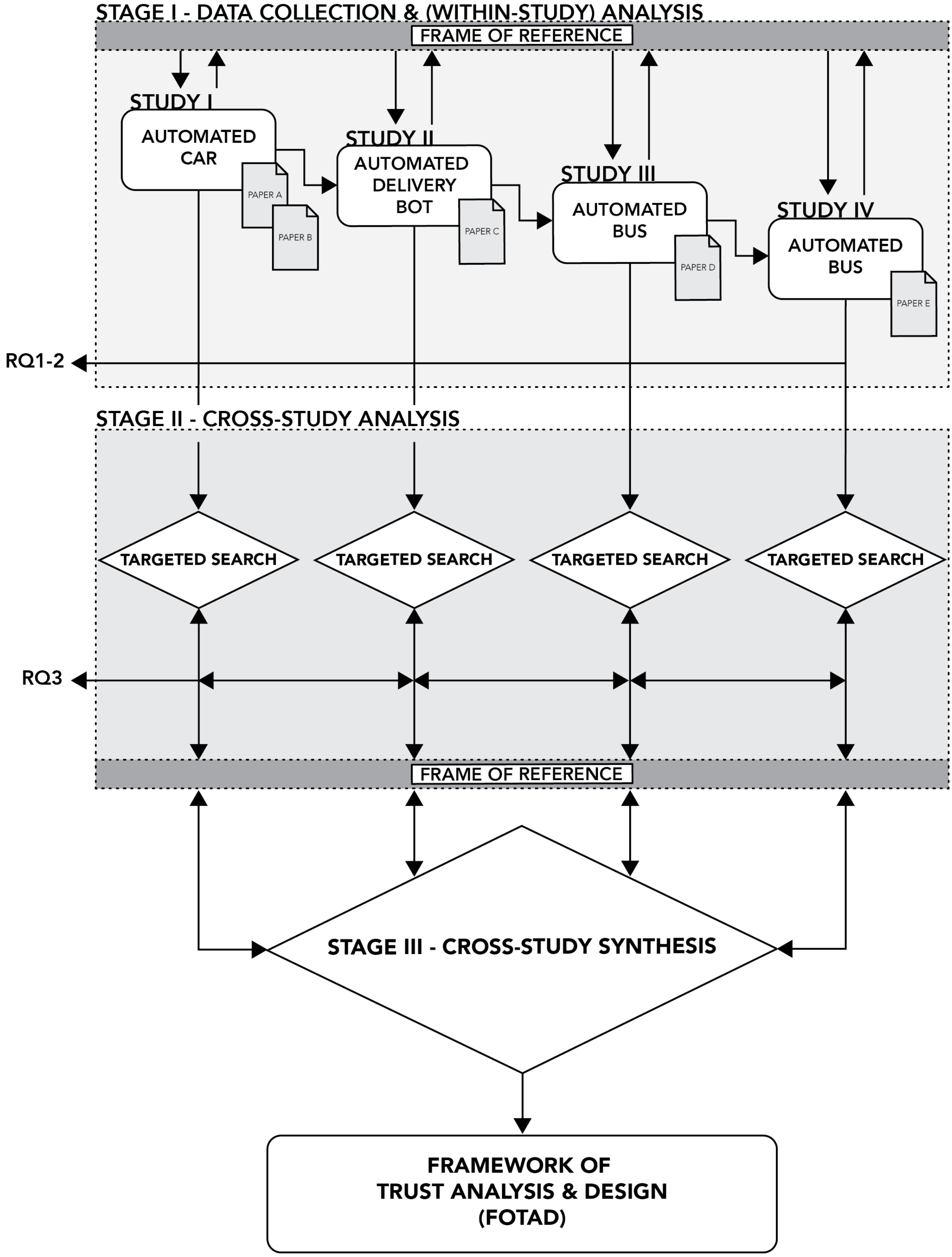

Figure 3 - Research Process

Stage I - Data Collection & (Within-Study) Analysis

Data Collection

As mentioned earlier, the overall nature of the research process has been exploratory in the sense that results from one study informed the direction of the next in terms of purpose and aim. This in turn determined which types of methods for data collection and analysis were to be used.

Furthermore, all studies have used Mixed Method Research Designs to be able to collect and analyse different data sets separately and then either compare and/or relate them to each other in a 'side-by-side' comparison to either confirm or disconfirm the data (cf. Creswell, 2014).

The choice of data collection methods has primarily included interviews and questionnaires. As the methods were considered useful methods for measuring trust, they were possible to implement in mixed method designs and they suited the context in which the studies took place.

Table 2 - Detailed description of data collection methods used in the four conducted studies.

Study No.	Data collection Methods
Study I	*Trust Questionnaire* *Semi-structured interviews* *Momentaneous trust assessment* *Trust curve* *Think-aloud protocol*
Study II	*Trust Questionnaire* *The Questionnaire on Contextual Aspects* *Semi-structured interviews*
Study III	*Trust questionnaire* *The Questionnaire on Contextual Aspects* *Semi-structured interviews* *Behavioural measures i.e. (hands-on-steering wheel)*
Study IV	*Trust questionnaire* *The Questionnaire on Contextual Aspects* *Semi-structured interviews*

Two different questionnaires were developed and used.

The first questionnaire – *Trust Questionnaire* – had its theoretical basis in the questionnaire created by Jian et al. (2000) and incorporated antecedents to performance, purpose and process information as described by Lee and See (2004) (see Appendix A for items).

The second questionnaire – *The Questionnaire on Contextual Aspects* – was based on the aspects identified by Hoff and Bashir (2015) and identified aspects related to the context that may affect user trust in automation. The second questionnaire was developed for Study II but was used throughout (Studies II-IV) (see Appendix B for items). The questionnaires were implemented primarily after a user-AV interaction within a specific OD had taken place

(Studies II-IV) (see papers A, B, D and E) but were adapted and used both before and after the interaction in Study II (see Paper C).

Two other data collection methods were also used to be able to capture momentaneous trust during a user-AV interaction (momentaneous trust assessment) either by prompting the participant seconds after a traffic situation had taken place or continuously via a think-aloud protocol (Charters, 2003) where the participants were allowed throughout the interaction to verbally explain their level of trust and what affected it (see Papers A and B for details). The two data collection methods were only used in Study I. In Study II the participants were not interacting with the AV (automated delivery bot) continuously; in Study III the participants drove a full-length bus on public roads and could not be disturbed while driving and finally, in Study IV, since it was not feasible to prompt all passengers and to allow them to think-out load whilst riding in the bus together neither method was used.

Another method that was only used in Study I was the '*Trust Curve'* developed by me and another PhD student. It was adapted from the UX-curve developed by Kujala et al. (2011) to assist users in reporting their experience with a product over time. The Trust Curve was used in a similar fashion as the UX curve but instead of measuring experience, the trust curve measured how user trust changed depending on the everyday traffic situations to which the users were exposed. The Trust Curve was re-introduced post-study to act as a 'mediating tool' (Karlsson, 1996) to stimulate the participants to further recall, reflect on and discuss the levels of trust in the AV during specific traffic situations (see Papers A and B for details).

The most frequently used method throughout all four studies was semi-structured interviews (see Papers A, B, C, D and E). Interviews were considered essential as the aim of the studies was to understand not only what but also how factors relating to the AV affected user trust; not only what but also how contextual aspects relating to the environment in which user and AV are situated affected user trust in the AV; and whether there are differences in trust between AVs and ODs and why these differences exist. Interviews were also important to discover new factors and aspects that may have impacted user trust, and not merely verify or reject what is already known from earlier research.

Only one behavioural measure was used and only in one study, Study III. In this study the users (i.e. bus drivers) interacted with the vehicle and the in-vehicle interfaces via normal controls (e.g. steering wheel, gas-and brake pedal) but also via a control panel from which they could activate the ADAS system while driving. The behavioural measure was a complement to questionnaires and interviews and was used by observing (via video data) users' behaviour in terms of their hands-on-steering wheel, while transitioning the control of the dynamic driving task (DDT) and while the AV was docking, to measure trust in the AV; in other words changes in hand position, for example if the users were hesitant to let go of the steering wheel. These are methods that have been suggested to be useful for studying user trust in AVs (e.g. Yu et al., 2021).

Within-Study Data analysis

A within-study data analysis was conducted in relation to the specific aims of each individual study. Study I had for example a twofold aim: (i) to investigate if and how the vehicle's driving behaviour affects user trust in AVs (an automated car) and, in particular, how this is expressed by users, and; (ii) to investigate whether and how user trust in AVs is affected by the AV's driving behaviour in relation to the characteristics of different everyday traffic situations. Therefore, Study I included not one but two data analyses which resulted in two different papers (see Paper A and Paper B). Studies II-IV on the other hand only included one within-study data analysis based on their respective aims.

Subjective & Objective Data Analysis
First and foremost, the analysis of the subjective data provided by the questionnaires was analysed using basic descriptive statistics, in other words classifying and summarizing numerical data based on the current sample (cf. Hinkle et al., 1988). For example, median trust scores were calculated for each driving behaviour (cf. momentaneous trust assessment Study I – Paper A). However, during the analysis of Study I, I became interested in including statistical methods such as nonparametric tests (due to only using ordinal scales) to identify possible correlations between trust in the AV and contextual aspects (see Studies II-IV). Therefore, I developed the "The Questionnaire on Contextual Aspects". The primary motivators were a desire to learn new ways of interacting with the data and to show the results in an easily digested way for the reader.

Furthermore, the subjective data elicited from the think-aloud protocols (Study I) and interviews (Studies I-IV) was analysed using an iterative thematic analysis as described by Braun and Clarke (2006). The excerpts extracted from the transcriptions were coded by two researchers (myself and another PhD student) to be able to cross-check to determine coding consistency between the two researchers (cf. Intercoder agreement Creswell, 2014). If at any time there was a mismatch between the researchers' interpretation of an excerpt and the code given to that excerpt, the coding was discussed until full consensus was achieved. Intercoder agreement was done on all studies except Study II. Due to time restrictions within the project conducted during the pandemic I had to code the material by myself.

The Trust Curve data (only used in Study I) was analysed by annotating positive (+) and negative (-) tangencies for each of the participant's curves. The number of positive and negative annotations for each situation and driving style was then summarised.

A targeted search of the participants' statements on user trust relating to traffic situations was also conducted for each traffic situation in Study I. Identified statements for each traffic situation was analysed and compared to previously established contextual aspects affecting trust (Hoff & Bashir, 2015) as well as unknown aspects.

In terms of the objective data collected in Study III, video recordings of behaviour changes in terms of hand positions were manually annotated by myself and another PhD student. This since changes in hand position; the position of hands on or off steering wheel being a behaviour commonly used to measure trust (e.g., Yu et al. 2021). The annotation scheme for hand position included five states: (1) both hands on steering wheel, (2) right hand only on steering wheel, (3) left hand only on steering wheel, (4) no hands on steering wheel, and (5) hands hovering (close to steering wheel without touching it and with palm(s) facing it).

For a summary of data analysis method see Table 3.

Table 3 - Overview of Data Analysis Methods used in Study I-IV.

Study No.	Data Analysis Methods
Study I	*Iterative Thematic Analysis* *Trust Curve Tangency Annotations* *'Targeted Search'* *Descriptive Statistics*
Study II	*Iterative Thematic Analysis* *Descriptive Statistics*
Study III	*Iterative Thematic Analysis* *Descriptive Statistics* *Annotations of behavioural changes*
Study IV	*Iterative Thematic Analysis* *Descriptive Statistics*

The results from the within-study analysis provided answers to RQ1 and RQ2, that is to say which trust-affecting factors and contextual aspects respectively were identified as affecting the participants' trust in each study. However, even though the within-study analysis provided indications, the answer to RQ3 required further analysis. Therefore, the interview material in each study was revisited and compared against one another via a cross-study analysis to be able to provide a more elaborate answer to RQ3.

Stage II - Cross-Study Analysis

The purpose of the cross-study analysis was to further investigate what was previously indicated in the interview material during the within-study analysis. These indications had been continuously thought about and even sporadically discussed with other researchers in terms of "it seems as if it is not only x, y, z affecting trust but also…" which in turn generated responses and discussions from which new ideas and thoughts emerged. Since this process was primarily internal and implicit (i.e. not actively analysing and synthesizing the material but rather thinking about it) and also involved discussion with other researchers, it can be viewed as an implicit analysis and synthesis process taking place in parallel with the planned and hands-on research practices conducted throughout the dissertation project.

What became evident in this process was that even though the identified factors and aspects are important to consider when designing for appropriate user trust, they are effects of something else, more in particular of how the user interprets what the AV does and where the AV does something. The user's experience and interpretation of these components and the interdependency and interplay between them were what actually affected user trust. Thus, the conclusion was that what creates the user's experience and ultimately forms user trust in an AV consists of a number of fundamental and underlying components and the interdependency and interplay between these components.

Hereby the cross-study analysis answered RQ3, that is to say what it is important to consider from a design perspective, so as to be able to design for appropriate user trust in terms of a number of fundamental design variables that support design for appropriate user trust.

Stage III – Cross-Study Synthesis

Even though the answer to RQ3 explained a number of fundamental design variables that can support when designing for appropriate user trust, it did not provide any guidance on how to practically go about it. However, my design perspective prompted me to not only describe a phenomenon but to try to prescribe solutions that may support in the design for appropriate user trust. Therefore, a cross-study synthesis was conducted. This enabled me to address the overarching aim of the thesis – how to design for appropriate user trust.

The cross-study synthesis 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 the OD of which an AV and user operate). The framework was developed in an iterative process by organizing and re-organizing the findings and continuously relating them to the frame of reference as well as to other research when necessary. The end-result is a tentative framework which serves as a 'tool-for-thought' to support developers and researchers alike when designing for appropriate user trust. A description of the proposed framework – The (Tentative) Framework of Trust Analysis and Design – can be found in Chapter 5.

4. EMPIRICAL STUDIES

The fourth chapter of the thesis – Empirical Studies – describes the four studies conducted in order to answer the research questions. Finally, the chapter describes the outcome of the within-study analysis. In other words, the main takeaways from each study.

4.1. STUDY I

4.1.1. Aim

The aim of Study I was twofold: (i) to investigate whether and how the AV's driving behaviour affects user trust in a fully automated AV, i.e. a fully automated car (Level 5 SAE) (see also Paper A) during interaction with an AV, and (ii) to investigate how driving behaviour affects user trust in the AV in everyday traffic situations (see also Paper B).

4.1.2. Method

To answer these questions an experiment with a Wizard of Oz (WOz) approach was set up. This approach involved a standard car being remodelled to be perceived and experienced as a fully automated vehicle (see Figure 4). 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, here referred to as "Defensive" and "Aggressive".

Figure 4 - Wizard of OZ (WOZ) operated car. To be perceived as fully automated SAE L5.

Eighteen participants (ten male and eight female) between the ages of 20 and 55 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.

A mixed-methods design was used 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 user trust was collected in two different phases. In Part 1 of the peri-trial phase, a momentaneous trust assessment scale 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 (through the Trust Questionnaire) and to allow participants to chart how their trust in the AV changed during the test run (via the 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 and explain the experienced difference between the two in terms of trust. To stimulate the participants, the trust curve was reintroduced as a mediating tool. This helped participants further reflect on and discuss both their levels of trust in the AV in specific situations as well as their overall trust.

4.1.3. Analysis

The analysis of Study I was divided into two parts.

The first analysis (see also Paper A) focused on how the AV's 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 i) what factors explain users'/drivers' trust in the AV? and ii) 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 B) 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 sum total of positive and negative annotations for each situation and driving behaviour was 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), as well as unknown contextual trust aspects.

4.1.4. Findings

Trust in Automated Car

The findings of Study I show that participant trust in the automated car 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 $M_{\text{“Agg”}} = 5.5$; $p<0.01$), with "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 style. It was perceived as more predictable primarily because the "Defensive" actions were taken earlier and more calmly, while the "Aggressive" actions were found to be more sudden and unpredictable. Interestingly, in some traffic situations the 'Aggressive' driving style was considered more trustworthy, for example in traffic situation #4 – Stopping for pedestrian at pedestrian crossing. The 'Aggressive' driving style was perceived as more trustworthy mainly because the automated car came to a full stop before the pedestrian crossing. The full stop was interpreted as a benevolent action towards the pedestrian, which in turn increased the participants' trust in the AV. Thus, driving behaviour is an important factor to consider, not just because it affects user trust through greater or lesser predictability but because it could also be used to convey intentions and benevolence.

Trust Affected by Contextual Aspects

The findings in Study II also show that participant trust was affected not only by the automated car's driving behaviour per se; perceptions of the automated car's trustworthiness were also affected by aspects relating to different traffic situations and the interdependence between the automated car's driving behaviour and the situation. These aspects included perceived task difficulty, perceived risks and how well the automated car conformed to user expectations of how a traffic situation should be negotiated. Sometimes, this affected the participant's trust more than could be accounted for by the automated car's driving behaviour alone.

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

Perceived risk to oneself and others. Perceived risk to oneself and others also affected participant trust in the automated car. Perceived risk to oneself affected participant trust in the automated car to a greater extent when in low visibility (little information provided by the environment), when the participant had difficulties 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 automated car 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 this. Rather, the feelings of risk were amplified. Perceived risk to others seemed to affect participant trust to a large extent, since their perception was that an accident involving vulnerable road users (VRUs, such as pedestrians or cyclists) could lead to severe injuries to those individuals (compared for instance to 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 automated car (through its driving behaviour) as more or less benevolent, risk-aware (keeping a safe distance from VRUs) and respectful (coming to halt before a VRU crossed a pedestrian crossing). Thus, if a driving behaviour was benevolent, risk-aware and respectful towards VRUs, participant trust in the automated car increased.

Conforming to expectations. The way the automated car's driving behaviour conformed to user expectations of how situations should be conducted, moderately affected participant trust in the automated car. The focus of participants' attention was on how well the automated car 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 automated car should negotiate traffic situations. Therefore, the "Defensive" driving behaviour was perceived as more trustworthy. The findings 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 automated car.

Main Takeaways:

- *Predictable (AV) driving behaviour increased user trust.*
- *Experienced benevolent driving behaviour towards other road users increased user trust.*
- *An action initiated by the car in a perceived risk-filled situation without the user knowing why, decreased user trust.*
- *Perceived risk to others was amplified when there were other humans involved in the traffic situation but if the driving behaviour was well-adapted to the situation user trust in the automated car increased.*
- *Perceived risks for oneself, i.e. little information given from the environment made it harder for participants to predict upcoming events which in turn increased their perception of risk and affected user trust in the automated car to a greater extent.*
- *When there was low perceived task difficulty, user trust was less affected. In other words, when the user experienced the DDT was easy, trust was affected to a lesser degree.*
- *Conforming to expectations, i.e. the way driving behaviour conformed to user expectations of how situations should be conducted, also affected user trust (to a moderate degree).*
- *User trust in the automated car was fundamentally affected by the perceived interplay between why the car conducts an action/task, how the action/task is conducted and when the action/task is conducted in relation to the OD (context), e.g. a traffic situation.*

4.2. STUDY II

4.2.1. Aim

The aim of Study II was to evaluate how logistics personnel experienced and trusted an automated delivery bot as a tool for last-mile deliveries (LMD) of parcels. Of particular interest was an investigation into how contextual aspects, that is to say factors relating to the physical environment and/or specific situations, affected a user trust in the bot (see also Paper C).

4.2.2. Method

The study involved logistic personnel at the logistic department at the University campus. The participants were nine males and three females (P1 – P12) with a mean age of 58 years (SD = 2.3). One group of four (two males/two females) worked at the Transport Central (TC) – organizing incoming deliveries and further distributing parcels to five different departments scattered throughout the University campus. The second group, logistics personnel working at the different departments (seven males/one female), received parcels from the TC and then reorganized and further distributed parcels to the end-recipients. The study used a WOz approach due to technical issues with the bot, with the intention to simulate an L5 automation, by remotely controlling the automated delivery bot. The automated delivery bot (see Figure 5) delivered primarily smaller parcels from TC to the five different departments across the University campus and back again, twice a day over a period of three days. The bot also delivered parcels between departments.

Figure 5 - Automated delivery bot (Photo: HUGO AB)

Data on users' experiences was collected in two phases – a *pre-interaction phase* and a *post-interaction phase*. More specifically, two questionnaires were distributed, a *Trust Questionnaire* and a *Questionnaire on Contextual Aspects*, and semi-structured interviews were conducted both before and after the participants had interacted with the bot.

The interview guide included questions about what benefits and issues the participants saw with using a bot for LMD, their trust in the bot, their general attitude towards the solution of a bot for LMD as well as their perception of how useful using a bot for LMD is and could be.

4.2.3. Analysis

The analysis of Study II was conducted in three steps. First a comparison was conducted between participants' responses to the *Trust Questionnaires,* distributed during the pre- and post-interaction phases, to identify any differences in the participants' expected versus experienced trust in the bot. Secondly, a non-parametric test, Spearman's rho, was used to determine any association between trust in the bot and contextual aspects. Third, a thematic analysis was performed on the qualitative data collected in the interviews. The thematic analysis was carried out to identify, analyse, and report patterns in the data (cf. Braun & Clarke, 2006), more specifically to identify what issues the participants registered in their day-to-day work using a bot for LMD, primarily focusing on benefits and issues with the bot, trust in the bot and usefulness of the bot for LMD. Finally, the data collected from questionnaires and interviews respectively was compared to see if there were any similarities or differences between the data sets.

4.2.4. Findings

Trust in Delivery Bot

The findings show that the participants' experienced trust in the bot in the post-interaction phase was lower than their expected trust in the pre-interaction phase, which had to do with the bot not conforming to the participants' expectations (see Figure 6).

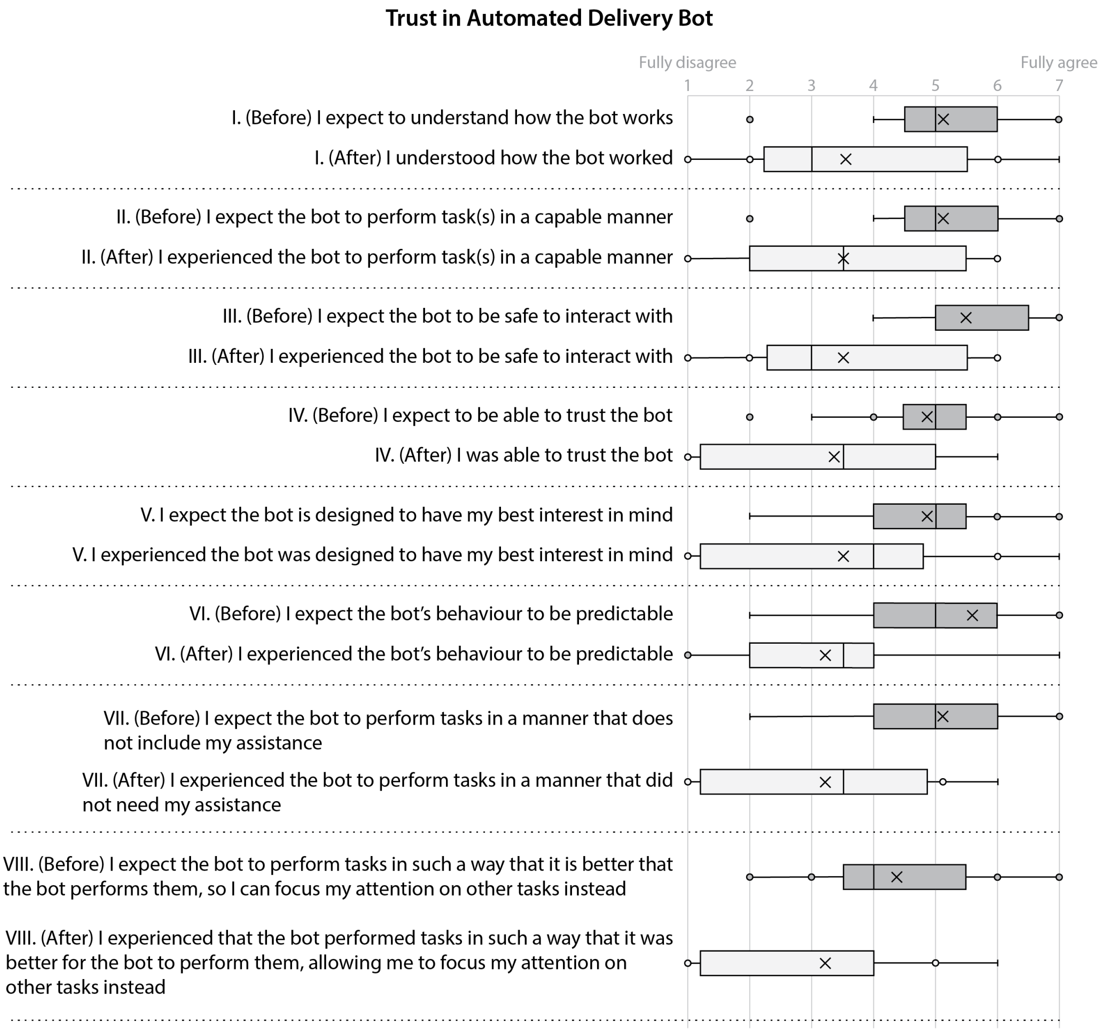

Figure 6 - pre-versus post-interaction rating of user trust in bot.

More specifically, in relation to how capable and reliable the bot was experienced, most of the participants experienced the bot to be less than optimal in terms of loading capacity, since it could only deliver smaller parcels. Secondly, and perhaps the biggest factor explaining why the participants did not fully trust the delivery bot, was that it was neither capable nor reliable in terms of its performance. This was because the bot malfunctioned, and the participants therefore perceived the bot as less than trustworthy.

The small capacity and unreliable performance of the bot had several negative effects on the logistics chain which in turn affected the participants' trust even more negatively, experiencing the introduction of the delivery bot as increasing risks, sub-optimizing the logistics chain and not resulting in any benefits. However, contextual aspects also directly affected the participants' trust in the bot.

Trust Affected by Contextual Aspects

The findings of Study III also show that user trust was affected not only by the delivery bot itself; trust in the bot was also affected by the effect the bot had on the logistics chain and by several contextual aspects. This included perceived risk – to oneself and others – task difficulty, workload, system complexity and benefits.

Perceived risk to oneself and others. There was a negative correlation between the participants' 'trust in the bot' and perceived risk ($r\ (10) = -.65$, $p = .042$). i.e. the participants' perception of the probability that the bot's tasks could lead to negative consequences for them and other people. This could partly be explained by the participants experiencing the bot as underperforming and not being reliable, since it malfunctioned a couple of times while it was part of the logistics system at the University. Thus, the participants felt that there is a risk of negative effects such as delays in the logistics system and consequently more work for them if the bot does not fulfil the purpose of delivering parcels as intended.

The participants' trust in the bot was also affected by perceived (traffic) risks for other road users. This fear seems to have been primarily triggered by the previously mentioned malfunctioning of the bot. However, the perceived risk to other road users also seems to have been affected by the size of the bot. The participants thought that the small-sized bot could be difficult to spot by other road users, which in turn could increase the risk of accidents. Finally, there was also a perceived risk of goods being stolen and/or damaged when transported by the bot, primarily since participants did not experience the bot as robust enough or sufficiently theft-proof.

Perceived task difficulty, workload and system complexity. There were negative correlations between 'trust in the bot' and perceived task difficulty, perceived workload, and perceived system complexity respectively (see Table 4). The results can be explained by the sub-optimal effects that the introduction of the bot had in the logistics chain at the University.

Table 4 – Correlation between trust in bot and contextual aspects (perceived task difficulty, perceived workload and perceived system complexity). See Appendix B for item description.

		Perceived task difficulty (Item IV)	**Perceived task difficulty (Item V)**	**Perceived Workload (Item VIII)**	**Perceived Workload (Item IX)**	**Perceived system complexity (Item XII)**
'Trust in bot'	R	-.80**	-.77*	-.76*	-.69*	-.96**
	Number of observations	10	9	8	9	9

*) Correlation is significant at $p \le 0.05$. **) Correlation is significant at $p \le 0.01$.

First, participants were concerned with how a delivery bot would be able to handle parcel delivery within the University's logistics system. This was a system the participants regarded as rather complicated, involving many different tasks and tasks that were considered difficult for the bot to handle.

Secondly, the participants experienced introduction of the bot into the logistics system as making their own work tasks harder rather than easier, increasing their own workload, in particular as an effect of making their logistics activities more complex and time-restrained due to more work with assisting the bot in completing tasks that the bot could not do by itself. One primary reason was that the bot was not dimensioned for the number of parcels to be delivered, which meant that the logistics personnel at TC had to deliver parcels in parallel with the bot. This was something that created both stress and frustration for the participants. A second reason was that logistics personnel at the departments also had to add new tasks to their schedule since they now had to wait for the bot so as to "assist" it with unlocking and opening doors, unloading, and reloading the bot, and then letting the bot out through the locked door. Therefore, by adding the responsibility of such tasks, the introduction of the bot led to sub-optimization for both user-groups, that is to say for personnel at TC as well as for personnel at the departments.

Perceived benefits. There was also a positive correlation between 'trust in the bot' and perceived benefits ($r(9) = .85$, $p = .004$) as both scored low. The perception of the bot as offering few benefits is primarily explained by the overall logistics chain sub-optimization. Instead of making the participants' work easier and freeing up time, introduction of the bot led to the opposite. Therefore, the participants did not see the usefulness of having a bot for LMD at the University. However, some of the participants believed, based on their experience with the bot, that the bot could be better suited in other ODs such as in hospitals or industrial plants. This since the participants believed that hospitals and industries had less complicated environments and therefore easier for the bot to operate. Thus, based on the interaction the participants created beliefs of where the bot would be more useful.

Organizational Effects on Trust

Finally, from the thematic analysis it was also possible to identify that (user) trust was not only affected by the bot itself, by the consequences the bot had on the logistics chain or by contextual aspects directly, but also by the organization by which the participants were employed as well as by the developers behind the bot. The bot had been implemented with little information given to logistics personnel and also without the logistics personnel being involved in the decision-making process, i.e. being able to explain their needs in order to get a solution for LMD that solved their actual needs for their daily work. Furthermore, when asked what was most important for them to be able to trust the bot, several mentioned the degree to which the bot was able to handle the tasks for which it had been implemented. However, when probed one participant mentioned that it is not about whether you trust the bot or not but rather your trust in the developer. Thus, negative attitudes towards their own organization and the developers for not being transparent or forthcoming seem to have also had a negative impact on the participants' attitude towards the bot.

Main Takeaways:

- *The experienced (and insufficient) capacity and reliability of the automated delivery bot decreased user trust.*
- *Due to the bot being perceived as unreliable, this increased the perception of risk to oneself and others (accidental users), which further decreased user trust.*
- *The introduction of the bot increased system complexity, task difficulty and increased the users' overall workload (helping the bot complete tasks). This in turn decreased user trust in the automated delivery bot.*
- *The perceived complex (logistical) system with complicated tasks affected user trust in the bot.*
- *Since the introduction of the automated delivery bot led to several negative consequences the users saw no benefits to using an automated delivery bot.*
- *Both the bot and the OD itself affected user trust. However, interaction between the two seemingly affected trust the most. Thus, the interdependence between AV and OD is fundamental for user trust.*
- *Based on the interaction, beliefs of where the bot would be more useful were created.*

4.3. STUDY III

4.3.1. Aim

The aim of Study III was to develop further knowledge of how professional bus drivers experience, accept and trust an Advanced Driver Assistance System (ADAS) used in an automated full-length public transport (PT) bus (see Figure 7). The system can assist bus drivers by automatically docking at bus stops (see Figure 8 – one of five bus stops). However, since the thesis only concerns trust in automation, any method and/or result concerning acceptance is left out and can therefore only be found in the appended paper (see Paper D).

4.3.2. Method

The study was conducted with ten professional bus drivers (nine male, one female) who got to use the automated bus that could automatically dock at bus stops. Their ages ranged from 32 to 71 (M = 52.3, SD = 11.9) and they had between 1 and 41 years of driving experience (M = 14.1, SD = 14.4). The study was conducted on a public route with five different bus stops in an industrial area. Objective and subjective data were collected during and after the test-drive. Subjective data was collected through interviews and two questionnaires to capture the participants' trust in the ADAS. The *Trust Questionnaire* and the *Questionnaire on Contextual Aspects*. Objective data was collected through video recordings, to document the driver's interaction with the user interface of the automated bus and identify any behavioural changes indicating user trust, such as hands on the steering wheel (or not).

Figure 7 - automated PT bus

Figure 8 - bus stop

4.3.3. Analysis

Conclusions from the study were drawn based on triangulation, that is to say a combination of multiple methods and measures used to cross-check findings and compensate for the weaknesses of one method with the strengths of another. The triangulation comprised a thematic analysis of the interview data, a compilation and analysis of the questionnaire data as well as an analysis of the video data. The data collected by the two questionnaires on user trust was done post test-drive. A non-parametric test, Spearman's rho, was then used to determine any association between trust in the automated bus and contextual aspects.

4.3.4. Findings

The findings show that a majority of the participants' trust in the ADAS formed while using it as they increased their understanding of the system's capabilities and limitations. Trust was primarily affected by the performance of the ADAS, with participants experiencing the system's driving behaviour as predictable and dependable.

Furthermore, trust was identified on two system levels (see Paper D), more specifically on the **operation level** which concerns users' interaction with the system's user interface for input and receiving feedback via in-vehicle interfaces and/or via the vehicle's driving behaviour such as acceleration and deceleration; and also on the **use level**, which is a more aggregated level in which use implies that someone is using something for a particular purpose (or task). For ADAS, this means drivers using the system to manage the driving of the bus.

On the **operation level** the findings show that the participants' trust in the ADAS system was quite high, evident from the Trust Questionnaire, with median scores for each item between 5.5 and 6.5 (a higher score is more positive) on a scale from 1 to 7 (see Figure 9).

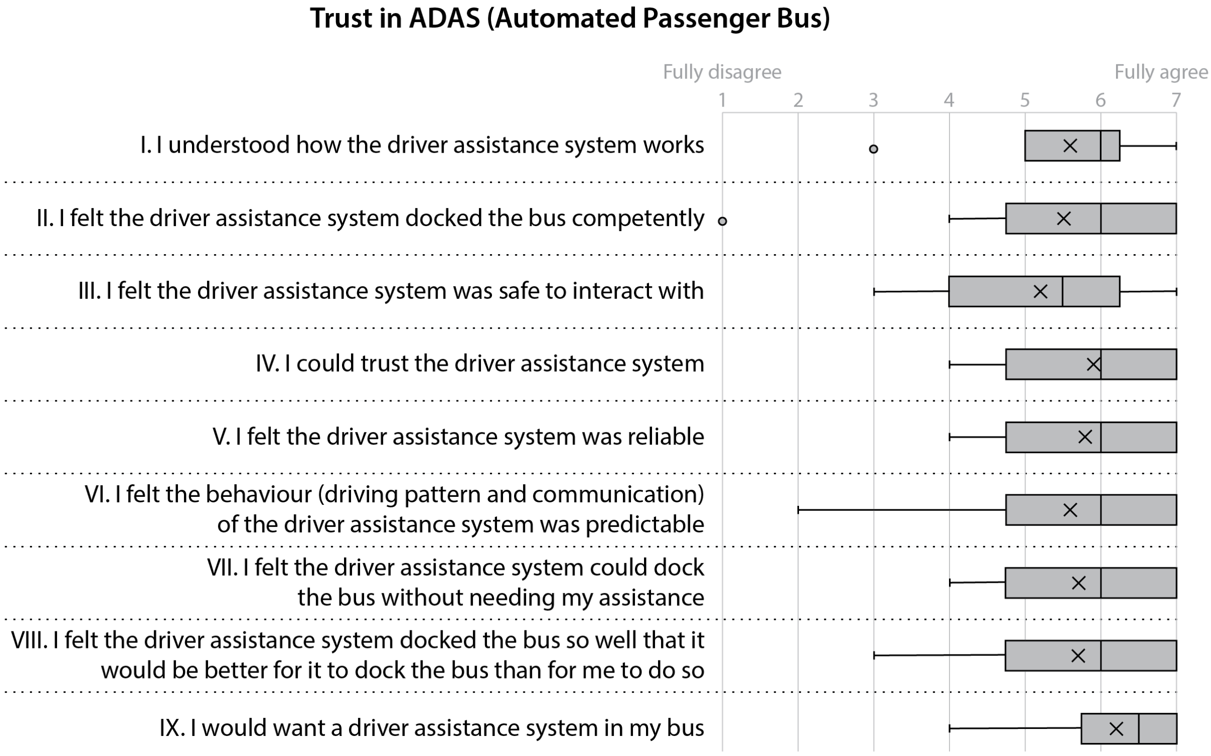

Figure 9 - Trust-questionnaire results

Predictability, dependability, safety, ease-of-use and ADAS complexity. According to the non-parametric test the participants' trust in the ADAS primarily correlated to how safe the system was perceived to be by the participants in their interaction with it (r (10) = .80, p=.006), how easy it was to understand how to use the ADAS (ADAS complexity) (r(10) = .73, p=.02) and also how complex the ADAS was to operate (r(10)= .86, p=.001). However, from the semi-structured interviews it was possible to see that some of the participants' trust was also primarily affected by how predictable and how dependable the system was (as mentioned in the opening paragraph of this section).

Moreover, although the participants generally had a high level of trust in the system's ability to carry out the manoeuvring task (with their trust increasing as they gained more experience with the system (see Paper D), some thought it performed unpredictably and unreliably during certain sequences. A common reference was the driver-to-ADAS control handover. In this situation, most of the participants did not fully trust the system as they found the handover unsettling. This was also reflected in the Trust Questionnaire, in which the item with the lowest ratings was whether the ADAS was safe to interact with (see Figure 9). Another sequence that some participants mentioned was when there were other road users present and the ADAS did not convey (via the driving behaviour) whether the ADAS had identified them or not, for instance by slowing down in close proximity to other road users. This in turn did not allow the participants to predict whether or not the ADAS would avoid the other road users. Thus, the participants took control of the bus.

On the **use level** participants stated that their level of trust was affected by the bus's ability to carry not only out the manoeuvring task but also the docking task within the OD (public transport, i.e. the regular traffic environment).

Perceived task difficulty, complexity and risk. Many participants said that they trusted the ADAS and that they could rely on the system to dock the bus. This was further supported by the correlation between the items relating to the degree to which the system could dock without assistance from the participants ($r(10) = .93$, $p=<.001$), how easy it was for the bus to dock ($r(10)=.76$, $p=.01$), the degree to which the ADAS assisted in decreasing the participants' difficulty in docking the bus ($r(10)= .67$, $p=.033$), and their trust in the ADAS. The participants felt they just needed to monitor the docking sequence.

Participants' trust in the ADAS was also highly affected by the traffic environment in which they experienced the ADAS, which led to several of the participants comparing the current traffic environment to rush-hour traffic in the city centre. The environment in which the tests were run involved light traffic. In other words, the number of cars, trucks, pedestrians and cyclists was quite limited. The participants felt that they could trust the system to conduct the docking sequence in this environment but that if the traffic had been denser, with higher levels of complexity, for instance within a working context carrying out the transportation of passengers on a bus line, they would not have trusted it to the same extent and would therefore not have relied on it as much due to the increased risk of incidents and delays (delays that may increase the participants' stress level too). Therefore, as the specific low-density traffic environment included little complexity, leading to lower perceived risks, the ADAS was experienced as trustworthy. However, if using the system for docking in a busy city centre, for instance, the participants seem to be less willing to trust the ability of the ADAS. Thus it seems that the interaction with the ADAS in the industrial area shaped the participants' beliefs of whether or not the ADAS was trustworthy in another context.

Main Takeaways:

- *Predictable and dependable driving behaviour increased user trust in the ADAS.*
- *ADAS was perceived as safe and easy to use (and understand) which in turn affected user trust, thus increasing user trust.*
- *Easy for the bus (ADAS) to handle the docking (perceived task difficulty), which increased user trust.*
- *Traffic environment (low degrees of complexity and risks) affected user trust, lowering the threshold for trusting the ADAS.*
- *User trust was identified and affected on two different levels: operation level (interaction with user interface/driving behaviour) and use level (trusting the ADAS to manage the bus, i.e. automatically dock).*
- *Belief in how trustworthy the automated bus was in other contexts (same OD), e.g. when used for everyday work, was shaped by the current user-AV interaction, i.e. where the automated bus was expected to be trustworthy or not to reach user goal(s).*

4.4. STUDY IV

4.4.1. Aim

The primary aim of Study IV was to investigate how frequent public transport (PT) users perceived, experienced and trusted travelling with a full-length, fully automated PT bus, and their expectations of how automated buses in the future might affect their commuting (see Paper E).

Figure 10 – Traffic situation #6.

4.4.2. Method

The study conducted included twenty-two participants (frequent PT users) recruited from two different cities in Sweden (Gothenburg and Borås). The participants were twelve male and ten female, aged between 21 and 70 (*M*=40.5, *SD*=16.5). The study was conducted over the course of two days on a test course with a route that included nine traffic situations that one would likely encounter in everyday traffic when travelling by PT, such as the bus slowing down in an intersection to allow a cyclist to pass (see Figure 10).

The automated bus used a L3 automation system but was intended to be perceived as an L5 since the bus was fully self-driving for the specific test route but could not drive by itself in any other environment. The system navigated via a comparison of continuous sensor data and a previously created map and predefined route. One could say that the bus followed a digital track that, via sensors, continuously compared the bus's position in relation to that track, in other words a route, similar to a train track but digital.

Subjective data on the participants' trust was collected post-study via the *Trust Questionnaire* to measure the participants' trust in the automated bus and also via the *Questionnaire on Contextual Aspects* to measure the participants' trust in the bus as affected by the context. The Trust Questionnaire included some new additions. Furthermore, semi-structured interviews were used to collect data on the participants' perception of the bus in its relation to driving behaviour and appearance, overall experience travelling with the bus such as how the bus was felt to perform in everyday traffic situations, and what effects the participants believed automated buses would have on their everyday commuting in future.

4.4.3. Analysis

The data collected in the two questionnaires analysed by a non-parametric test, Spearman's rho, to determine any association between trust in the automated bus and contextual aspects. All items in the Trust Questionnaire that showed a high positive correlation were combined into a new variable, *'Trust in automated bus'*. Correlation coefficients between 'Trust in automated bus' and each of the items in the *Questionnaire on Contextual Aspects* were then calculated to determine any relationship while a significance level of 0.05 was considered.

Finally, the interview data was analysed using an iterative thematic analysis to identify any themes and patterns in the interview transcripts. The analysis was conducted in two steps; in a first step the focus was on identifying broad themes related to the primary research questions of the study such as automated bus performance in relation to traffic situations, experience and trust as well as participants' expectation of the effects of introducing automated buses into the public transport system. In the second step these themes were further analysed to identify any patterns within them.

4.4.4. Findings

Trust in the automated bus

Most participants experienced the ride with the automated bus as very positive and felt that it was exciting and interesting. The participants also accepted travelling with the bus to a high degree. This was also reflected in the Trust Questionnaire with a median score for each item between 5 and 7 (a higher score is more positive) on a scale from 1 to 7 (see figure 11).

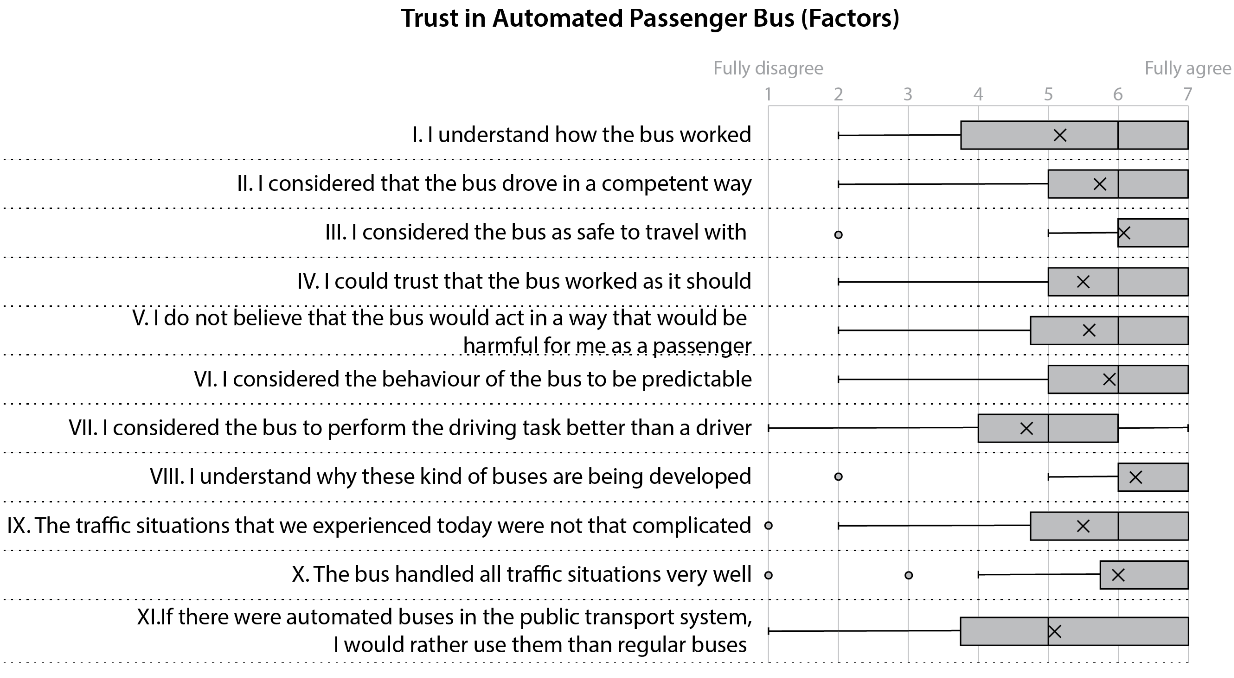

Figure 11 - Trust Questionnaire results. A higher score corresponds to a more positive score, on a scale from 1-7.

The general performance of the automated bus affected the participants' trust in the automated bus to a high degree ($r(22) = .65$, $p = .001$) but it was also noted that the bus handled the traffic situations for the most part in a capable way ($r(22) = .50$, $p = .018$) (see paper E for further description). This was further supported by the interview data with some explaining that the bus handled the traffic situations as well as the drivers did, so the bus was regarded as capable. The performance of the automated bus seems to have led the participants to feel safe, which in turn also seems to have further affected their trust in the automated bus (r(22) = .71, p = <.001). According to the participants, the automated bus drove in a "slow" and "calm" fashion and was therefore perceived as cautious, sometimes even too cautions (see paper E). The automated bus's driving behaviour was also experienced as predictable which seems to have affected the participants' trust in the bus (r(22) = .66, p = <.001). The cautious and predictable driving behaviour may also explain why the participants experienced the automated bus as generally feeling safe and not posing any risk which in turn seems to have positively affected the participants' trust in the automated bus (r(22) = .56, p = .007).

However, as Figure 12 shows, not only the bus per se affected the participants' trust (M_{Bus} =6) but also the environment, that is to say other road users and the traffic situations ($M_{Traffic\ Env.}$ =6), along with the developers of the bus, in other words the brand (M_{Brand} =6).

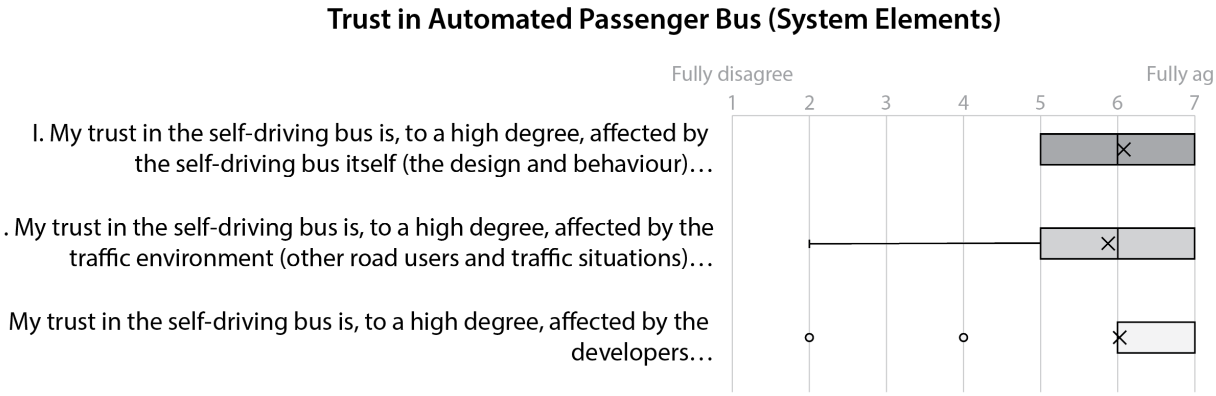

Figure 12 - Effect of contextual aspects on user trust.

Trust as affected by Traffic Environment & Developers

In terms of the **traffic environment** several participants described the environment as controlled where not too many unforeseen incidents could happen (due to study setup in an enclosed area). This seems to have affected the participants' trust as well where one participant even described the controlled environment as lowering the threshold for trusting the automated bus since nothing could really happen. This statement also coincides with the questionnaire data from the 'Questionnaire on Contextual Aspects' where the only item correlating to the new feature 'Trust in the Automated Bus' related to perceptions of decreased risk (r(22) = .5, p = .019) which could have been triggered by the experience of a rather risk-free environment which made the participants feel safe.

Furthermore, in relation to user trust being affected by the **developers** of the automated bus, the effect was twofold: (i) trust seems to have been affected by the mere presence of the representatives of the company. This is explained by creation of a perceived distance between the user and the automated bus which in turn gave the participants a feeling of not needing to rely on the bus itself to same extent as if they had not been there (they could instead could rely on the representatives of the company). Therefore, the involvement of the developers in the study affected user trust by decreasing the perception of risk. However, (ii) user trust was also affected by the brand of the company developing the automated bus. The automated bus was developed by a well-known company that has been around for years and not by an unknown company without any expertise. This increased the participants' expectations that the automated bus had to be reliable, that is to say they could trust the automated bus because otherwise the company would not have allowed the participants to travel on the bus.

Another interesting phenomenon was that the participants did in general trust the bus (to a high degree). However, they could additionally in the same interview session describe not trusting the bus due to specific tasks and actions carried out by the bus. One example of a task decreasing participants' trust was during negotiation of an intersection, more specifically a turn when the bus waited longer than it usually did to initiate the turn even though there was no traffic. The delay caused the participants to wonder whether this was on purpose. Not understanding why the bus did or did not do something also seems to have affected their trust in the bus. However, some actions also decreased the participants' trust. One of them was for instance during passenger boarding scenario (one of the extras waited at a makeshift bus stop for the bus) where some of the participants perceived the bus as braking too hard before stopping and picking up the waiting public transport user.

Thus, the findings indicate that there is user trust in different aspects of the automated bus, such as for the entirety of the automated bus, which also seems to be highly affected by the developers behind the technology and the overall brand, but also for specific tasks (negotiating an intersection) and actions (braking) in specific situations.

Main Takeaways:

- *The automated bus was experienced as capable, predictable & having a risk-mitigating driving behaviour (cautious) which in turn was perceived as safe. This in turn increased user trust.*
- *The traffic environment was experienced as controlled, predictable, including low perceived risks (perceived as safe); this lowered the threshold for trusting the automated bus.*
- *User trust was affected by the inclusion of safety measures (representatives of the developers oversaw the automated bus) as well as the by the brand itself, i.e. being viewed as a company that develops reliable technology.*
- *User trust was affected both by specific actions and tasks conducted by the bus and also by why actions and/or tasks were or were not conducted or not.*

5. FINDINGS

The fifth chapter of the thesis – Findings – describes the insights gathered from the four empirical studies. The chapter presents the findings from the within-study analysis answering RQ1 and RQ2.

5.1. FACTORS AFFECTING TRUST (DURING USE)

The first research question was formulated as *What trust factors affect user trust in AVs in different ODs during use?*

The most prevalent findings from the within-study analysis, that is to say findings that either were found to have (i) a significant correlation with user trust, (ii) frequently occurred in the interview material and/or (iii) were experienced by users to be important (or interpreted as important for the users) for user trust, are summarized in Table 5.

Table 5 - Identified Trust Factors

	USER: Regular Car Drivers AV: Automated Car OD: Traffic System	USER: Logistic Personnel AV: Automated Delivery Bot OD: Logistic Chain - Last Mile	USER: Bus Drivers AV: Automated PT Bus OD: Public Transport (PT)	USER: PT Passengers AV: Automated PT Bus OD: Public Transport (PT)
AV RELATED TRUST FACTORS	**STUDY I**	**STUDY II**	**STUDY III**	**STUDY IV**
CAPABILITY		✓		✓
RELIABILITY		✓		
PREDICTABILITY	✓		✓	✓
BENEVOLENT INTENTIONS	✓			
DEPENDABILITY			✓	
UNDERSTANDABILITY	✓			✓
EASE-OF-USE			✓	
CONFORMING TO EXPECTATIONS	✓	✓		

The following findings (section 5.1. and section 5.2.) will be presented as questions posed by a user of an AV. This, since it seems as if these questions are the ones the user of an AV wants answered to be able to assess the trustworthiness of an AV.

5.1.1. How Capable, Reliable & Predictable is the AV?

How well the AV was perceived to perform, that is to say how capable, reliable and predictable the AV was perceived to be by the user, consistently affected user trust throughout the four studies.

The perceived capability relates to how adequate the AV was judged to be based on its experienced ability to assist the user in reaching his or her goal (e.g. conducting a task). The perceived reliability of the AV is the degree to which the user experienced the AV as being consistently 'good' enough to assist him or her in reaching a goal. Both these affected user trust and perceived predictability, that is to say the degree to which a user can forecast a future behaviour (and a near future user-AV state) based on the current AV behaviour.

The effect of the AV's capability and reliability on user trust was only prevalent in Study II and Study IV. This can be explained by that in Study II the automated delivery bot malfunctioned unexpectedly a couple of times when driving: *"Suddenly it just breaks down. They [colleagues] saw when they [developers of the bot] were at it* [trying to troubleshoot what the problem was], *it* [the bot] *ran five metres then there were issues again…. No, there needs to be a machine you can trust, otherwise there will be trouble throughout the day"* (P3). Another explanation is that it was too small to handle the number of parcels that had to be delivered. Therefore, there were obvious signs that the automated delivery bot lacked capable and reliable performance in relation to the goal of transporting parcels twice a day, which in turn may explain why these factors affected user trust (dramatically decreasing user trust) more in Study II than in the other studies. When automation fails a task perceived as easy by the user, both user trust and reliance tend to dramatically decrease (Madhavan et al., 2006). Thus, when introducing an AV, the AV needs to handle the most fundamental task for which it has been designed, especially if that task is perceived as an easy task to perform by the user. Otherwise, user trust will be substantially and negatively affected.

In study IV, on the other hand, one of the passengers thought that the bus handled the traffic situations as well as the passenger him-/herself would have handled them and therefore the automated PT bus was experienced as capable: *"…it* [the automated bus] *solved a majority of the* [traffic] *situations as I would have solved them. Then the question becomes whether I am competent; personally I would argue that I am"* (P11). Similar findings have been made by Sun et al. (2020) who explain that if the driving behaviour of an AV mirrors the driving behaviour of the user it may increase willingness to trust the AV, for instance by being viewed as more capable. Thus, if the AV has (or is perceived to have) a similar driving behaviour as the user, this increases user trust.

The AV's perceived predictability, based on current AV behaviour, was communicated via the AV's driving behaviour in Study I, Study III and Study IV and seems to be the first factor the participants used as a way to assess whether the AV was trustworthy or not: *"I had a lot more trust than I expected. It didn't feel like it* [the automated car] *acted in an unpredictable way"* (P7). This shows similarities with interpersonal trust in terms of trust initially being more affected by how predictable specific behaviours are, as argued by Rempel et al. (1985) but also similarities in terms of human-automation trust, in other words that trust is first affected by the behaviour of the automation (Muir, 1987, 1994). The users' perception of the AV as being predictable increased user trust.

5.1.2. Can I Depend On the AV?

The perceived dependability of the AV, that is to say the degree to which the user perceives that the AV can be counted upon if needed, was found to affect user trust in Study III. This was primarily due to participants experiencing the AV as capable after they had experienced the AV throughout several docking sequences and started to hypothesize that the more you experience how capable it is the more you feel you can depend on the AV. This was expressed as: *"It was good in every way. If I only had done one lap I would perhaps have said* *no* [not understanding enough to dare to depend on the AV for the task of docking]. *But now I am starting to understand how it [the AV] works and it is really good."* (P5). Based on the 30 conducted dockings the user attributed the AV with qualities based on how well the cooperation went. Therefore, through experience with the AV the user built a level of dependency in the AV to handle the responsibility for docking.

These findings are consistent with earlier research where dependability is argued to be the second stage in the trust formation process both in interpersonal trust (Rempel et al., 1985) and in human-machine trust through experience with the automation system (Muir, 1994). A possible explanation of why dependability only became prevalent in Study III may be primarily due to the length of the experience (docking 30 times) but also due to the study taking place in a natural work context including professional users who fully understand what a tool needs to be able to do to be useful in assisting them in their everyday work. The perception of the automated bus as being dependable increased user trust in the bus, more specifically in the ADAS system.

5.1.3. Can I Understand What the AV is Doing?

Moreover, users' understanding of why an AV behaves as it does, also affected user trust. For instance, in Study I if the AV initiated an action without the user knowing or understanding why, this decreased user trust especially if the action was conducted in a (perceived) risk-filled situation. An example was a traffic situation where the automated car was overtaking another moving car: *"I didn't have clear sight of the road when he overtook [the other car], I experienced that as unpleasant."… "Why did it have to overtake [the other car], was it because of the slow speed?"* (P5).

In Study IV, the same scenario played out, but instead of reacting the AV did nothing even though the participants believed it should have done something: *"During the test itself. There was one of the turns where the bus seemed to have a small delay. Maybe because, I don't know if it was on purpose or not, but it stayed longer than it usually would since there was no traffic or anything"* (P24). Therefore, the motive behind the delay could not be understood, which in turn decreased user trust but not to the same extent as the situation referred to in Study I. Thus, not fully understanding certain AV behaviours seems to negatively affect user trust. This is consistent with earlier theory where understanding why automation behaves as it does is important for user trust (cf. understand underlying mechanisms suggested by Lee & See, 2004).

5.1.4. How Will the AV Act Toward others?

The frame of reference states that automation cannot be benevolent. Instead it refers to the designer behind the automation and his or her orientation towards the user and whether this intent is communicated to the user via the automation (Lee & See, 2004). The empirical findings from Study I may not necessarily contradict but provide added nuance to earlier findings, as they show that users attribute human-like qualities such as benevolence to machines as well: *"It* [the automated car] *kept its distance to the extent that the cyclist wouldn't feel that it* [the automated car] *was too close"* (P17). Therefore, if the automated car acted (via its driving behaviour) respectfully towards other road users (i.e. pedestrians and cyclists) and showed clear intentions (e.g. coming to a full-stop at a pedestrian crossing) user trust increased. This can be explained by people tending to anthropomorphise artefacts by projecting meaning and feelings onto them (Cupchik & Hilscher, 2008). Therefore, if the AV acts in a certain way people may exaggerate the actual capabilities of the AV.

However, the attribution of altruistic tendencies to the AV was only found in Study I. One explanation is the presence of the halo effect in Study I, a cognitive bias where someone assesses the entirety of an AV based merely on a small sample of information, such as one good characteristic of the system (Colley et al., 2021). In other words, Study I was conducted in 2017 during the height of AV (automated car) optimism and the automated car [Level 5, SAE] was generally perceived by the participants (regular car drivers without any expertise in the area of AVs) as being highly trustworthy and driving in a capable manner, which in turn may have caused them to attribute even higher levels of agency so as to regard the AVs as benevolent.

5.1.5. Is It Easy To Use the AV?

Ease of use in terms of the perceived effort needed to use a system (Davis, 1989) was only identified in Study III. Ease of use was defined in Study III as the effort needed for the participants to understand when the ADAS system (used for docking) could be used or not used. The bus drivers thought that it was easy to understand how to use the system which affected their trust in the ADAS system, more specifically it increased their trust. The results further support the arguments by Hoff and Bashir (2015) that ease of use may be an important trust factor not only in TiA but in the area of TiAV as well.

This trust-affecting factor was only found in Study III but can be explained by the fact that none of the other studies included this type of interaction as the users did not have to initiate or turn off any automated system. This suggests that when a user needs to interact with an automated system via an interface it is important that it is easy to understand when the system can and should be used.

5.1.6. Will the AV act in Accordance with My Expectations?

Conforming to expectations seems to affect user trust in both Study I and Study II. In Study I this was observed only during certain traffic situations due to the AV not conforming (according to the participants) to unwritten rules regarding deceleration. In Study II the AV did not conform to expectations of being able to handle the volume of parcels needed to be transported since the AV was not optimized for the task within the participants' work context (this in addition to malfunction incidents). This in turn severely decreased user trust. According to Zhang et al. (2020) user trust in AVs increases when perceived performance is higher than expected and that unexpected but positive notion led to the highest level of trust. One could argue that the opposite might be true as well. In other words, when perceived performance is lower than expected performance trust may decrease a substantial amount. Thus, it is important that users' expectations on the performance of the AV correlate with the actual performance of the AV to assist users in creating and maintaining appropriate user trust.

5.1.7. However, the most important Factors for Me as a user to Trust the AV is…

The trust factors identified and presented previously, which consistently affected user trust throughout the four conducted studies (Studies I-IV) primarily related to what an AV does during use to be perceived as capable, predictable and reliable. This relates to the information on performance, argued to be important for TiA, especially in the beginning of a relationship with automation (Lee & Moray, 1992; Lee & See, 2004).

However, the importance of performance information for appropriate user trust, or more specifically the importance of performance information being used as a way to design for creating and/or maintaining a level of appropriate user trust, has also been found to affect

TiAV, in particular regarding earlier research on trust in automated cars. For instance, Körber et al. (2018) found that performance information (cf. competence and reliability information) given prior to interaction with an automated driving system (ADS) not only influenced reported trust in automation but also reliance in the ADS. Other studies that found performance information important for user trust includ Beller et al. (2013). According to the driving simulator study it was found that providing uncertainty information, that is information related to how reliable the automated system is (Noah & Walker, 2017) during use, increased user trust and in turn led to better driver-automation cooperation. Thus, these findings suggest that user trust in an AV – whether it is in a simulator or an actual automated car, bus and/or delivery bot in realistic environments – seems to be affected by performance information and therefore empirically confirms earlier predictions.

Almost of equal importance was why the AV behaved as it did and whether the behaviour was perceived as appropriate or not. In other words, it is important for the user to understand why an AV did or did not conduct an action, and whether or not that action was appropriate in order to assist the user in reaching his or her goal. This refers to information on process. That is the degree to which the user can understand the underlying mechanisms of why the automation behaves as it does in the current situation, as well as if that behaviour is favourable for reaching the user's goals(s) (Lee & See, 2004). Process information has been identified empirically to also affect user trust, for instance by Koo et al. (2014) who found that providing information on why an action is conducted by the AV may decrease anxiety and increase user trust. Within the area of Human-Autonomy Teams (HATs) with robots it has also been found that user trust and the robot's trustworthiness declined due to unexpected behaviour by the robot and the most effective solution to mitigate the decrease in trust and (robot) trustworthiness was to give explanations for why an event occurred. This was achieved by displaying the robot's environmental awareness (Lyons et al., 2023). However, it is very important that the information given is tailored to the user, suited to his or her logic, so the accuracy of the automated system (cf. intelligent system) is estimated correctly and therefore becomes meaningful for the user (Nourani et al., 2019). It is equally important that neither too little nor too much information is communicated to the user (Ekman et al., 2016). However, it has also been found that merely giving an explanation for certain actions, for instance an explanation for a take-over request, had no notable effect on user trust which Korber et al. (2018) explain could have been an effect of the study setup, such as using a driving simulator that lacked a naturalistic setting. Thus, based on the findings it seems that information assisting the user in understanding why an AV did or did not perform an action and whether or not that was appropriate in order to assist the user in reaching his or her goal, is important for user trust. Having said that, this information needs to be carefully designed with the user in mind.

So, What Trust Factors Affect User Trust in AVs in different ODs? …

To conclude the answer to RQ1, the trust factor affecting user trust in different types of AVs in different ODs is primarily the performance of the AV. That is, how predictable, reliable and capable the AV is perceived to be when conducting e.g. a task, as well as how appropriate the behaviour of the AV is perceived to be for reaching a task and whether or not the user understands why the AV behaves as it does when conducting the task.

5.2. CONTEXTUALLY ORIGINATING USER TRUST ASPECTS (DURING USE)

The second research question concerns contextual aspects and was formulated as *What contextual aspects affect user trust in AVs in different ODs?*

The findings from the within-study analysis are summarized in Table 6.

Table 6 - Contextual trust aspects

CONTEXTUALLY RELATED TRUST ASPECTS	STUDY I (USER: Regular Car Drivers; AV: Automated Car; OD: Traffic System)	STUDY II (USER: Logistic Personnel; AV: Automated Delivery Bot; OD: Logistic Chain - Last Mile)	STUDY III (USER: Bus Drivers; AV: Automated PT Bus; OD: Public Transport (PT))	STUDY IV (USER: PT Passengers; AV: Automated PT Bus; OD: Public Transport (PT))
PERCEIVED WORKLOAD		✓		
PERCEIVED RISK(S)	✓	✓	✓	✓
PERCEIVED COMPLEXITY (WORK ENVIRONMENT)		✓	✓	
PERCEIVED TASK DIFFICULTY	✓	✓	✓	
ORGANIZATIONAL SETTING		✓	✓	
PERCEIVED BENEFITS		✓		

5.2.1. Will the AV Increase Workload?

Perceived workload as an effect of the introduction of an AV (automated delivery bot) was only identified in Study II. The AV was perceived as increasing the users' own workload, in particular as an effect of making their logistics activities more complex and time-restrained due to the added burden of assisting the bot in completing tasks that the bot could not do by itself: *"It would need to carry more load, we do not have any use for goods being delivered twice, no we don't. It leads to more work"… "then we have to deliver the goods [internally within each department] twice"* (P8). Therefore, user trust significantly decreased. It is generally argued that as the level of automation increases, user workload should decrease (in ordinary operation not accounting for unpredictable circumstances when workload can spike) and performance of the human-machine interaction should become more consistent (Balfe et al., 2015). However, here the opposite occurred: instead of mitigating high levels of workload the AV increased the workload, which in turn decreased user trust.

5.2.2. Are There any Risks for Me or Someone Else?

Perceived risk(s) was identified in all four studies and seems to be one of the most important factors affecting user trust and has also been identified many times in earlier studies (Lee & See, 2004; Li et al., 2019; Mayer et al., 1995). In the studies perceived risks were identified as twofold: perceived risk for oneself and for others. Perceived risks were often explained as an effect of the AV-OD (environment) interaction, either initiated by the users' experience of the

AV performance (as in Study II) –*"What happens if he [the bot] gets a blackout, goes down the stairs [there are several stairways on the University campus], hits somebody; we have a lot of visually impaired people [moving about on the campus], not too many disabled but there can still be a lot of injuries"* – (P4) **or** initiated by the complexity of the OD (as in Study III): *"Here [in an industrial area] it isn't a problem, but in rush-hour traffic, how will it react then?"* (P8). According to Li et al. (2019) risk is necessary to be able to study user trust in an adequate way. Since all studies presented in the thesis included some level of perceived risk, whether associated with negative effects on the whole OD due to the introduction of an AV leading for instance to a sub-optimal logistic system (Study II) **or** the perceived possibility of an accident between an AV and a pedestrian (Study I), a valid user trust measurement can be assumed. Therefore, even though it seems necessary to include risk when studying user trust, the level and type of risk (or uncertainty) may be less important. Thus, some level of perception of risk is needed for user trust and risk can be divided into two different types: risk to oneself and risk to others. Both seem to affect user trust.

5.2.3. Can the AV Be Used Here or There and What are the Consequences?

Perceived (work environment) complexity was identified in two of the studies, Study II and Study III. For example, in Study II the logistics system became more complex due to introduction of the AV (automated delivery bot) which was a major contributing aspect that led to an increased workload for the participants, which in turn decreased user trust. In Study IV, on the other hand, user trust in the AV (automated PT bus) was affected by the low level of complexity in the environment: *"Here [in an industrial area] it isn't a problem, but in rush-hour traffic, how will it react then?"* (P8). This explanation was in relation to the participants believing that there were far fewer road users, and far fewer different types of road users, in the context where the trial took place than in the city-centre during rush hours. Even though they trusted the AV (ADAS) for the task of docking in the industrial area, they expected that they might not trust the AV to the same extent (since perceived risk also increased in parallel with complexity) in the city-centre based on its current performance (in terms of braking, speed): *"Yes, the system* [ADAS system] *worked great, but it depends a lot on the roads, traffic, other road users, if there are cyclists and stuff like that. There are several question marks regarding that* [whether the participants would have trusted it or not]*"* (P5). Thus, perceived environmental complexity can increase due to the introduction of AVs which in turn may decrease user trust (as a negative effect) but can also affect how someone expects to trust an AV in another context (based on how the AV performed in one context).

5.2.4. How Will the AV Handle the Task and What's the Consequence For Me?

Perceived Task difficulty was also identified to affect user trust in two ways:

(i) if the AV was experienced in situations that, according to the user, were not that difficult for the AV to handle, such as travelling straight ahead in one lane while passing another vehicle in the other (Study I), this made it more difficult for the participants to judge whether the AV was trustworthy or not since the AV only showed the lowest level of capabilities necessary (due to the task being perceived as easy): *''It* [the AV] *handles itself well, but there was no manoeuvre needed to handle the situation"* (P8).

(ii) if the AV had an effect on tasks such as increasing perceived task difficulty for the user (Study II) or decreasing perceived task difficulty for the user (Study III): *"It makes things easier for bus drivers, much easier and simpler. Especially since the worst part is getting in and out of bus stops"* (P5). These results are in line with those of Monroe and Vangsness (2022) who found that task difficulty (and stress) had a significant effect on their participants' trust in automation. Therefore, it is important to consider both the user's perception of a task, that is whether it is perceived as easy and/or difficult, as well as the effect the AV itself might have on the task's difficulty, since both may affect user trust in terms of either increasing or decreasing the level of user trust.

5.2.5. AV Developers Should Support My Needs and Goals, but Can I Trust Them?

The organizational setting in which an AV is introduced may also affect user trust, if not directly then indirectly. Whether or not users are ultimately willing to use the AV seems to depend on the users' perception of and attitude towards the actors who introduce new technology into their work context.

This was observed in Study IV: *"It's* [the AV] *a machine, it does what it is told. Do I trust the person who has programmed it* [the AV]*? That is the discussion we should be having"* (P4). The participants' attitude was a consequence of not getting sufficient information regarding the purpose of the field study. At the same time the participants felt that they had bigger issues to address than introducing an AV (automated delivery bot) that would assist them in transporting parcels between different departments. It also seemed (based on their narrative) that there was a history of being steamrollered by their own management, not receiving the support needed for issues relating to IT systems and routines (therefore perhaps not fully trusting the management). In other words, they did not understand why the management had wanted to and allowed a study on the transport of parcels using an automated delivery bot, one of many activities they actually felt worked well. They also believed that doing this was part of a first step towards more automation, in other words less opportunity for work.

According to Mayer et al. (1995) organizational trust is, simply put, a party's willingness to be vulnerable towards the actions of another party. Since the participants in Study IV might already have had a low level of trust towards their own management, the idea of introducing an AV and willingly being part of the field study did not sit well with the participants. This was a field study that only seemed to further amplify the already negative attitude towards the whole organization behind the AV (including the developers of the AV) and this in turn seems to have affected their attitude towards the AV as well.

Study II and Study III were the only two studies in which the organizational setting affected the participants' trust, in Study II more so than in Study III. The explanation seems to be because Study II was the only one in which an AV was fully introduced into the users' daily work routines. In Study III the AV was used in the actual traffic environment, but the study lacked realistic routines and organizational goals, even though there were concerns about whether or not the AV would, when introduced into the public transport system, lead to more stress (due to slow docking) and delays. These concerns made some of the users believe that they perhaps would not trust the AV if introduced as part of their work activity in an actual public transport system (transporting passengers, keeping the timetable and so on).

Thus, from a trust perspective implementation is not 'plug and play' process – a mere matter of introducing AVs into a work context and users' daily routines. First there must exist a level of trust in the actors responsible for the introduction, second the responsible actors who are introducing the technology must explain the reasons why this is being done, which hopefully is for benevolent reasons, in other words to assist the users in reaching their goals. This relates in some way to information about purpose, that is to say the intended use of the AV and whether or not the organization behind the introduction has benevolent intentions (Lee & See, 2004).

5.2.6. What are The Overall Benefits For Me or Us? …

Perceived benefits in terms of lack of benefits was identified as affecting user trust only in Study II. The users' perception of the bot as offering few benefits was primarily explained by the overall logistics chain sub-optimization. Instead of making the participants' work easier and freeing up time, the introduction of the bot had the opposite effect. However, despite the sub-optimization and very few benefits, the participants still saw a possible future for AVs within the area of logistics, although in another OD – *"Unfortunately it doesn't do any work, but perhaps in the future, yes! But maybe not at the school [university] but in a shopping mall, at Volvo Trucks AB where everything is flat and indoors, or perhaps the hospital"* (P4).

The Study III participants, on the other hand, saw many possible benefits with the ADAS system for docking in the AV (automated PT bus), particularly in complex environments where it was difficult to dock and where the accuracy of the ADAS system could be useful according to the participants.

In Study IV the participants also expected certain benefits from the introduction of AVs (automated PT busses) especially as an enabler of new services that could increase the number of departures and new departure locations in geographically isolated areas such as the countryside. However, the benefits identified in Studies III and IV were not as prevalent or pronounced (not a significant correlation with user trust) and did not affect user trust to the same extent as the lack of benefits in Study II and were thus not included.

One possible explanation of why perceived benefits only affected user trust in Study II (correlated to a significant degree with user trust and noted in the interview material) is similar to the explanation of why the organizational setting affected user trust in Study II (and Study III): it can be attributed to the introduction of the AV into an naturalistic work context (Study II) or a realistic use context (Study III) where the participants' experienced at first hand the lack of benefits for them, thus negatively affecting user trust in the AV. Thus, if AVs are introduced into naturalistic contexts (or realistic ones), perceived benefits will probably be more important for user trust than one can identify in a simulated study. In other words, certain factors/aspects affecting user trust may be difficult to understand (and identify) outside a realistic and preferably naturalistic context. This, since simulator studies lack for instance uncertainties, routines and organizational goals and thus the realism. For instance, without any clear goals it is difficult for the user to understand what the AV needs to assist him or her with to reach a certain individual and/or organizational goal, thus making it harder to assess the AV's trustworthiness. Therefore, it is highly important to conduct studies on user trust in AVs in as realistic contexts as possible, preferably in naturalistic contexts in order to understand user trust to the fullest extent possible.

5.2.7. However, the Most Important Aspects for Me as a User to Trust the AV is…

The contextual aspects identified and presented previously, which consistently affected user trust throughout the four conducted studies (Studies I-IV) primarily related to perception of risk and task difficulty. Perception of risk affected user trust in all four studies, as perceived risk both to oneself and to others (which in turn affected the user's trust in the AV), and according to earlier research it is fundamental for user trust (Lee & See, 2004; Mayer et al., 1995). Otherwise the issue of trust may not even be relevant (Wicks et al., 1999) or provide valid results (Li et al., 2019). Based on the narratives it seemed that when perception of risk increased, users' willingness to trust the AV decreased, such as when bus drivers expected to not be as willing to trust the AV during rush-hour in the city centre. Earlier studies have reached contradictory conclusions on this matter; some suggest that automation is trusted less (and used less) in riskier situations (Perkins et al., 2010) while others have evidence of the opposite, suggesting that users rely more on automation in riskier situations (Lyons & Stokes, 2012). The findings here suggest the former.

In terms of perceived task difficulty, it seems as it may affect user trust in different ways. According to the findings of Monroe and Vangsness (2022) it seems that if a task is perceived as difficult, a user may not follow instructions from automation to the same extent, in other words not rely on the automation. However, the opposite has also been found – that when a task is perceived as difficult users may be more inclined to rely on the automation (Schwark et al., 2010). In the results from the studies presented here, task difficulty refers either to a task being perceived as (too) easy, so the user cannot judge the trustworthiness of the AV or the AV increasing the task difficulty for the user which in turn led to a decrease in trust in the AV. Thus, it seems that when task difficulty is low, other ways of communicating the capabilities may be important so as to understand the limitations of the AV, and that developers understand the effect AVs may have on users' tasks.

So, what Contextual Aspects Affect User Trust in AVs in different ODs? …

To conclude the answer to RQ2, the contextual aspects affecting user trust in different types of AVs in different ODs are primarily the perception of risk for oneself and others and task difficulty. Perception of risk is the probability of negative outcomes and degree of the seriousness of the outcome for both the user of the AV but also for others that possibly come in contact with the AV or are affected by the AVs behaviour e.g. other road users. Perception of task difficulty is a task being perceived as (too) easy, so the user cannot judge the trustworthiness of the AV, or the AV increasing the task difficulty for the user thus adding to negative outcomes.

5.3. IMPORTANT CONSIDERATIONS FROM A DESIGN PERSPECTIVE

RQ1 and RQ2 were formulated based on the literature on user trust summarized in Chapter 2. Frame of Reference.

With Research Question 3: *What are the important considerations from a design perspective in order to design for appropriate user trust?* the perspective changed from that of describing and explaining factors and aspects influencing user trust in AVs to prescribing a way to approach designing for appropriate user trust.

5.3.1. AV, Context & Interdependency affect User Trust

The answers to RQ1 and RQ2 provide AV-related trust factors and contextual aspects that are important to consider when designing for appropriate user trust. However, the cross-study analysis resulted in the identification of a third aspect, perhaps even more important to consider, that is the interdependency between AV and context. The importance of this interdependency for user trust became evident, for example in Study I when a participant responded to a question about what her trust was affected by and why: "I experienced it [the AV] as rolling forward when she crossed the street, *I believe it would have been nicer if it had stopped completely, for her sake"* (P9)

What the AV did – rolling slowly forward towards the pedestrian while she was crossing the street – negatively affected the user's trust in the AV. The user's trust was affected negatively because of the behaviour of the AV, as the AV was not perceived as 'benevolent' towards the pedestrian (see 5.1. Factors Affecting Trust). Thus, what the AV does and where – in a specific context such as a traffic situation where a pedestrian crosses the street – defines the interdependency that ultimately affects user trust.

Interdependency between automation and context has previously been identified as a factor that is important to consider when studying user trust. According to Muir (1987) and others there is a cause-and-effect connection between automation and environment. The more stable the environment is, the more predictable it is, and therefore user trust is inversely related to the amount of disturbance posed by the environment upon the machine and vice versa.

5.3.2. Information from AV, Context & Interdependency affect User Trust

However, the identified trust factors and contextual aspects affecting user trust are effects[11] of how users interpret what the AV does and where the AV does something, in other words, they describe users' interpretation of information originating from the AV, context and the interdependency between what the AV does and where.

In the cross-study analysis it was observed that the users always interpreted information from what the AV did in relation to a specific context, as well as from the context in relation to what the AV did and by which user trust was affected. This is illustrated by another example from Study I:*"Since I received early signals that the vehicle had understood that there was an obstacle or a situation ahead, to which it [the automated car] had to respond, this made me in some ways trust it more"* (P2).

[11] Effects - user perceptions which can be designed <u>for</u> but not designed per say.

The role played by information in affecting user trust is also acknowledged in earlier research on TiA. According to Lee and See (2004) providing information about what the automation does in terms of parameters such as capability and predictability (cf. performance information), information so the user understands the underlying mechanisms of the behaviour (cf. process) and intended use of the automation (cf. purpose) creates the basis for maintaining an appropriate level of user trust. In the area of TiAV, several researchers (e.g. Beggiato, Hartwich, et al., 2015; Hergeth, 2016; Kraus et al., 2019; Verberne et al., 2012) have focused both on designing information relating to one or several of the important information dimensions (cf. performance, purpose and process information) as for example described by Lee and See (2004), and on providing this information via in-car interfaces such as displays to study the effect on user trust in driving simulators.

However, even though information, intentionally designed to explain the behaviour of an AV and communicated via in-car displays is important in assisting the user of the AV in creating appropriate user trust, the empirical studies show that much of the information that the user interpreted originated from the AV per se, the context per se and the interdependency per se.

Hence, the user's experience and interpretation of these components and the interdependence between them were what actually affected user trust. It is the design of these fundamental and underlying components, that is the AV, its behaviour, in different contexts, that shape the information affecting user trust. Therefore, it is how to design these fundamental and underlying components that must be considered when designing for appropriate trust.

5.3.3. User Trust is affected by Information Originating from Underlying Components

From where the information originated in the empirical studies, that is the fundamental and underlying components, can be described in general terms as: **Who** (the AV) does something; **What** the AV does (on three separate levels of abstraction[12]); By **What means** the AV is doing something; **When** the AV does something, and **Where** (context) the AV does something (divided into three levels of decomposition[13]). In addition, users often questioned Why the AV behaved as it did, that is users wanted to understand underlying motives to AV behaviours which was found to be important for user trust.

This can be clarified with a generic example: the AV (**Who**) starts to decelerate (**What** i.e. deaccelerate) (**What means** i.e. driving behaviour) 100 metres (**When**) from an intersection (**Why** i.e. due to intersection ahead) in a city-centre during rush hour (**Where** i.e. context). This scenario is then interpreted by the user to be a highly predictable behaviour (**How** i.e. the effect of how this is experienced by the user) by the AV which in turn affects user trust.

All the presented components can be adjusted to assist users in creating appropriate levels of user trust and will hereafter be referred to as design variables from which the effects i.e. – **How** the user interprets the information – originate. Thus, 'how' is not a design variable but rather the dependent variable.

[12] Explained in section 5.4. The tentative Framework for Trust Analysis & Design – (2) What the Trustee Does
[13] Explained in section 5.4. The tentative Framework for Trust Analysis & Design – (6) Where

So, What are the Important Considerations from a Design Perspective in order to design for Appropriate User Trust? …

To conclude the answer to RQ3, trust factors and contextual aspects are important to consider as they describe (in part) the desired (or not) effects of a design process with the purpose to design for appropriate user trust in AVs. More fundamentally though the design process must consider the underlying components and therefore the following design variables and the interplay between them:

(I) The Who i.e. the AV, **(II) What** the AV does on three levels of abstraction (abstraction of activity), **(III)** by **What Means** the AV does something, **(IV) When** the AV does something, **(V) Why** the AV does something and **(VI) Where** the AV does something (on three levels of decomposition of context).

As design variables they can be varied to create the desired effect such as the AV being perceived as being predictable which in turn affect user trust. In other words, **(VII) How** the user interprets the AV based on information retrieved from the interplay between the design variables is what ultimately affect user trust.

Finally, even though there is an interplay between all the design variables there is an interdependent relationship between two of them: **(II) What** the AV does and **(VI) Where** this happens. The interdependency must be considered since user trust <u>is always</u> affected by what the AV does in relation to where this takes place. This interdependence lies on three levels of abstraction and decompositions alike.

5.4. TENTATIVE FRAMEWORK FOR TRUST ANALYSIS & DESIGN

Based on the six identified design variables, the interplay between them as well the interdependency between two of them ('what' the AV does and 'where') was further synthesized via the cross-study synthesis by trying to find the right pieces and combining them to generate a richer image of trust in AVs. The end-result is a tentative framework which also serves as a 'tool-for-thought' to support developers and researchers when designing for appropriate user trust.

(I) WHO refers to the trustee, in this case the AV.

(II) WHAT THE TRUSTEE DOES relates to information on performance as proposed by Lee and See (2004) and appears to be a fundamental underlying design variable.

However, in the frame of reference what the automation (in this case the AV) does is not clearly defined. According to the empirical findings it seems that what the AV does relates to three levels of abstraction. More specifically, the AV can conduct an action (lowest level of abstraction), a task (intermittent level) and an activity (highest level of abstraction).

> **Action:** An action can be viewed as single operation that the AV conducts. An action can for instance be an AV accelerating, braking, turning, and/or providing in-car information.
>
> **Task:** A task is created from several AV actions. A task may for instance be when an automated car negotiates a traffic situation, when an automated bus docks, and/or in the case of an automated delivery bot, it may for instance involve unloading parcels.
>
> **Activity:** An activity is created from several tasks (and actions). An activity is for instance an automated car, bus and/or delivery bot transporting people and/or parcels within their naturalistic OD. In other words, an activity may for example be commuting to and from work or transporting parcels within the logistics system.

These levels share some similarities with 'levels of detail' described by Lee and See (2004) as user trust being affected by different layers of the automation, for example affected by the main system, by functions of that system and by sub-functions of that system. Thus, the levels of detail are relevant since user trust seems to be directed towards and affected by different aspects of the AV. However, user trust in different 'levels of detail' refers, simply put, to an abstraction of the automated system, that is to say user trust in the system, user trust in functions of the system and user trust in sub-functions of the system. The *abstraction of activity* refers instead to what the AV does in terms of action, task and activity and the effect of that specific level of abstraction on how the user perceives and or experiences it, such as an action. Users appear not to focus on the technical aspects per se but rather on the effects of the technical aspects of the system. This is that which ultimately affects them and therefore it is more relevant to focus on abstraction of activities.

An example of an **action** that directly affected user trust: *"I experienced it* [the AV] *as rolling forward when she crossed the street, I believe it would have been nicer if it had stopped completely, for her sake"* (P9 – Study I). This was the participant's explanation to why the participant trust curve showed a negative tangency; in other words her trust in the AV (defensive driving behaviour) decreased because the AV kept slowly rolling towards the pedestrian, not completely stopping.

However, user trust was also affected by how the AV handled a **task** such as negotiating a traffic situation: *"At one point I was a little unsure; there was a semi-trailer rig, it was just* [showing about 1 decimetre with his hands] *this far from the mirror, he* [the semi-trailer] *took a short cut in the corner and it felt like the bus actually drew closer to him* [the semi-trailer] *so I had to do something and the brake was close so we came to a stop right there* [the bus driver stopped the bus]. *I was uncertain if this really works."*(P6 – Study III). Thus, based on the traffic situation with the semi-trailer, the user was not fully able to rely on the AV to handle the situation and therefore took over.

Finally, user trust was also affected by how the AV handled an **activity** such as transporting parcels within LMD: "*It would need to carry more load, we do not have any use for goods being delivered twice, no we don't. It leads to more work"… "then we have to deliver the goods [internally within each department] twice"* (P8 – Study II). Many of the users did not trust the AV for the activity of transporting parcels (LMD) due to the sub-optimal effects it had on the entire logistics system.

(III) WHAT MEANS refers to the communication channel used, for instance driving behaviour and/or in-car interfaces or anything relating to the AV from which the user receives information that he or she interprets. The behaviour and/or in-car interfaces may have been fully or partially actively designed with a purpose, but for various reasons some parts may not have been, which in turn can create unforeseen effects. In addition, the context may also alter driving behaviour, that is to say how the information from the behaviour is perceived.

(IV) WHEN refers to when in terms of distance relative to another object and/or situation as well as when in time the AV does something (or does not do something) – *"…it is especially important that it positions itself properly in the lane, and also with regard to when it brakes* [in relation to traffic situations]*"* (P1 – Study I) or when discussing the possible benefits of an AV (automated PT bus) – *"…it* [the AV] *might be more on time, you can trust it more"*(P4 – Study IV). Thus, user trust seems to be affected by when in relation to a traffic situation the AV performs an action (brakes) but also when in time relative to the user goals an AV arrives, that is to say how well the AV conducts an activity in comparison with a manually operated bus.

(V) WHY refers to why the AV acted as it did, for example why the AV performs an action, task and/or activity (or not) as well as in what way this is done: *"I didn't have clear sight of the road when he overtook [the other car], I experienced that as unpleasant."… "Why did it have to overtake [the other car], was it because of the slow speed?"* (P5 – Study II). The user clearly did not understand why the car acted as it did and seemed not to experience it to be the appropriate thing to do (i.e. overtake the other car). Similarly to information on process as suggested by Lee and See (2004) it is important that the user receives information on the underlying motives behind why, in this case why an AV acts as it does and that the behaviour is appropriate to the current circumstances. Otherwise, user trust may be affected. Thus, transparency is highly important to users in understanding both why an AV does something and the appropriateness of the behaviour in relation to the situation and the user preferences.

(VI) WHERE refers to the level of decomposition of context in which the user and AV are situated and/or to which the user refers. In other words, the number of elements such as user, AV, traffic situation and so on included in the level of analysis. In addition, the user does not only need to be situated in one level of decomposition but can also refer to one in which the

AV is trusted or not trusted (see section 5.4.3) ‘propagation of beliefs’ below for further explanation).

> **Micro level:** The Micro level is the lowest level. The micro level contains the user’s goal(s), the AV and the interaction between the two. A simple example is the user interacting with the AV. At the micro level the user may get information directly from one or more information channels related to the AV as well as information from the interaction itself (interaction between user and AV). A generic example could be user trust affected by the information provided by an action conducted by the AV and/or information from the user’s interaction with the AV, as well as how well the action coincides with reaching the user’s goal, which in turn results in/generates information by which user trust is affected. An example of the micro level can be found in Study II: *“It* [referring to the ‘Aggressive’ driving behaviour] *had twitchier steering behaviour, acceleration and braking which led to poorer comfort and also decreased trust.”* (P12). Thus, the context included the user, AV and the AV's actions.
>
> **Meso level:** The Meso level contains one or more micro contexts and encompasses, in addition to what is included in the micro level, also a task, a local environment, other systems such as technical systems, other users such as other road users and their goals. A generic example could be an AV conducting for example a task that affects user trust due to how the user perceives the AV to conduct the task and the effect of this on the user’s individual goal, task at hand, other users (and their goals) and/or the local environment. An example of trust at the meso level is illustrated by the following comment: *“Yes, from what I’ve experienced so far. I tested it* [docking with the automated bus] *six times and it worked. It was good in every way.”* (P5 – Study III). Thus, the context included the user, the AV, and the AV's task of docking at several bus stops.
>
> **Macro level:** The Macro level includes, in addition to one or several meso- (and micro) contexts, also organizational goals and routines, a global environment and other system ecologies. A generic example would be an AV that conducts an activity which affects user trust due to the effect of the AV's activity on the organizational goals, organizational routines, environment and/or other systems within the macro level: *“…the bot cannot be allowed to break down while in transit, because there are valuable items sent through this service. For instance, there are registered letters, I mean how would that work security-wise?”* (P7). The participant’s concern for the potential risk of theft if the automated delivery bot’s reliability did not improve affected several of the participants’ trust since the risk of the parcels not reaching their destination could affect the entire logistics system which includes the organization, other users, routines and more. Thus, trust existed on the macro level and was affected by how the AV conducted the activity of transporting parcels.

(VII) HOW refers to how the user interprets the interplay between the other design variables or rather the effect of the interplay of the other variables: *“Since I received early signals that the vehicle had understood that there was an obstacle or a situation ahead, to which it [the automated car] had to respond, this made me in some way trust it more”* (P2 – Study I). The participant expressed this based on the action conducted by the driving behaviour (starting to decelerate), which was interpreted by the user as the AV having understood that there was an obstacle, with the participant then assuming that the AV would react to the obstacle. In other words, this information helped the participant to predict what would happen next which in turn affected his trust, making this a subjective interpretation of the AV as identifying an object in the distance because it started to decelerate. It is here argued that How is the effect of the other design variables and the interplay between them. One could say that 'How' refers to how predictable, how dependable, how capable, how benevolent and so on the AV is perceived to be from the user's perspective. Thus, 'How' represents the trust factors and contextual aspects presented in section (5.1. Factors Affecting Trust and 5.2. Contextually originating User Trust Aspects) and is the effect of the interpretation of information created in the interplay and interdependence between the design variables (I-VI).

5.4.1. User Trust in Different Intersections

Since What the AV does and Where this is done (context) are interdependent, they always intersect with each other. In other words, user trust is based on information created from what the AV does, such as an action on a meso level. Put another way, user trust in an **action-meso intersection** (for example of activity-context intersections see figure 13). The analysis of the findings of the four conducted studies presented in the thesis (Studies I-IV) show that user trust is found in different activity-context intersections.

LEVEL OF ABSTRACTION - ACTIVITY (WHAT)	LEVEL OF DECOMPOSITION - CONTEXT (WHERE)		
	Micro-level	Meso-level	Macro-level
Activity	N/A	N/A	E.g. Trust in AV to handle the traffic system
Task	N/A	E.g. Trust in AV docking	E.g. Trust that the AV docks well enough in relation to the demands of the traffic system
Action	E.g. Trust in AV braking	E.g. Trust that the AV brakes well enough to cater for standing passengers	E.g. Trust that the AV brakes well enough to cater for standing passengers, while at the same time meeting the demands of the traffic system.

Figure 13 – An example of how user trust may exist in one intersection between a level of abstraction and a level of decomposition. In other words a activity-context intersection.

In **Study I** user trust was higher in the AV (automated car) when executing a 'Defensive' driving behaviour due to it being more predictable. In other words, 'Defensive' driving behaviour was generally more trusted. Thus one might say that the participants' trust was higher for the ‘Defensive’ driving behaviour in a **task-meso intersection**. This is because the study only included a series of simulated tasks – negotiating a series of traffic situations (tasks) including other road users (car drivers as well as pedestrians and cyclists) and therefore lacked a global environment and thus a macro-level context.

In **Study II** user trust in the AV (automated delivery bot) was in general low. Since the AV was introduced into the participants’ actual work context (macro-level) and not trusted for the activity of transporting parcels (activity-level) one can conclude that user trust was lacking in an **activity-macro intersection.** However, user trust was low in most intersections but since the AV was evaluated on a macro level for the activity of transporting parcels (aim) this is what needs to be emphasised. Thus, user trust was low in most intersections but since the purpose of the AV was to conduct the activity of transporting parcels within the participants’ work context, an appropriate level of user trust in the AV on an activity-macro level was key. Since the AV did not function as expected one could argue that user trust was appropriate since it would not be able to perform sufficient work for the participants.

Furthermore, in **Study III** the participants trusted the AV (automated PT bus – ADAS) to a high degree for docking the bus at bus stops. The AV was experienced in a naturalistic setting (actual traffic system), that is to say in a global environment. This despite the fact that the actual traffic system had no routines and no passengers and therefore the participants trusted the AV on a (meso-level) for the task of docking the bus at bus stops (task-level) which explains why the participants trusted the AV in a **task-meso intersection**.

Finally, in **Study IV** the participants also trusted the AV (Automated PT bus) to a high degree. Since the study was on a test-track as in Study I and only included a series of simulated tasks – negotiating a series of traffic situations (tasks) including other road users – the participants had a high level of user trust in the AV in a **task-meso intersection.**

5.4.2. Mismatches in User Trust

However, in addition to the design variables and activity-context intersections, two other characteristics of user trust were indicated in the within-study analysis and then identified during the cross-study analysis: (i) mismatches in user trust on different levels where abstraction of activity and decomposition intersect and (ii) a characteristic of user trust that relates to user trust formation during use which I term propagation of beliefs.

An example of mismatches in user trust between different activity-context intersections is when several of the participants experienced the AV executing a ‘Defensive’ driving behaviour in Study I as more trustworthy, in other words in a task-meso intersection due to being more predictable. However, for a specific traffic situation the participants trusted the ‘Aggressive’ driving behaviour more since the vehicle came to a full stop in response to a pedestrian waiting to cross the street. This was interpreted as the AV being benevolent and respectful towards the pedestrian which in turn increased user trust. The action of a complete stop for a pedestrian waiting to cross the street lies in an action-meso intersection. Thus, the ‘defensive’ AV was trusted more in a task-meso intersection but in one situation in an action-meso intersection, several of the participants had more trust in the ‘Aggressive’ driving behaviour (see figure 14).

LEVEL OF ABSTRACTION - ACTIVITY (WHAT)	LEVEL OF DECOMPOSITION - CONTEXT (WHERE)		
	Micro-level	Meso-level	Macro-level
Activity	N/A	N/A	?
Task	N/A	'Defensive' driving behaviour perceived as more trustworthy due to being more predictable (how). Since e.g. 'Defensive' braked (what) earlier (when).	?
Action	?	'Aggressive' driving behaviour perceived as more trustworthy due to acting benevolent (how) and respectful (how) due to completely stopping (what).	?

Figure 14 - User trust in AV driving behaviour in a task-meso intersection versus in an action-meso intersection.

5.4.3. Propagation of Beliefs

The second characteristic of user trust relating to trust formation that was found was 'propagation of beliefs'. Propagation of beliefs was identified during the revisit to the empirical material and is basically what the AV does (abstraction of activity – action, task and/or activity) and where this is done (decomposition of context – micro, meso and/or macro), based on which beliefs are created regarding whether or not the user would trust the AV for an action, task and/or activity at another intersection between a level of activity (action, task and/or activity) and a level of context, that is to say micro, meso and/or macro.

One example to illustrate the propagation of beliefs and mismatches in trust between intersections can be found in Study IV. In this case the participants trusted the AV for docking, i.e. a task-meso intersection, but some did not trust it in an activity-macro intersection due to the accumulative effects that several dockings were believed to have on the routines in the public transport system. The participants were concerned about not being able to work the public bus routes on time, due to slow AV docking: "*…I believe it would create a whole lot of stress for the drivers. We have a timetable to follow, and often we cannot make it on time. Especially during peak-hours, when the timetable is no way near enough* [for what it actually takes to drive the routes]" (P6). In other words, while docking in a task-meso intersection the participants created beliefs about the AV not being capable enough for implementation into their regular public transport routes due to negative effects on their current routines (activity-macro intersection). In addition, several participants were concerned about the potential risk of harm from using the AV in day-to-day working life: “*Here* [in an industrial area] *it isn't a problem* [the user trusts the automated bus], *but in rush-hour traffic, how will it react then* [and can I trust it then]?” (P8). They trusted it in a task-meso intersection but did not believe that they could trust it in an activity-macro intersection. Therefore, the participants experienced a mismatch in trust in different abstraction-decomposition intersections based on trust and beliefs (see figure 15).

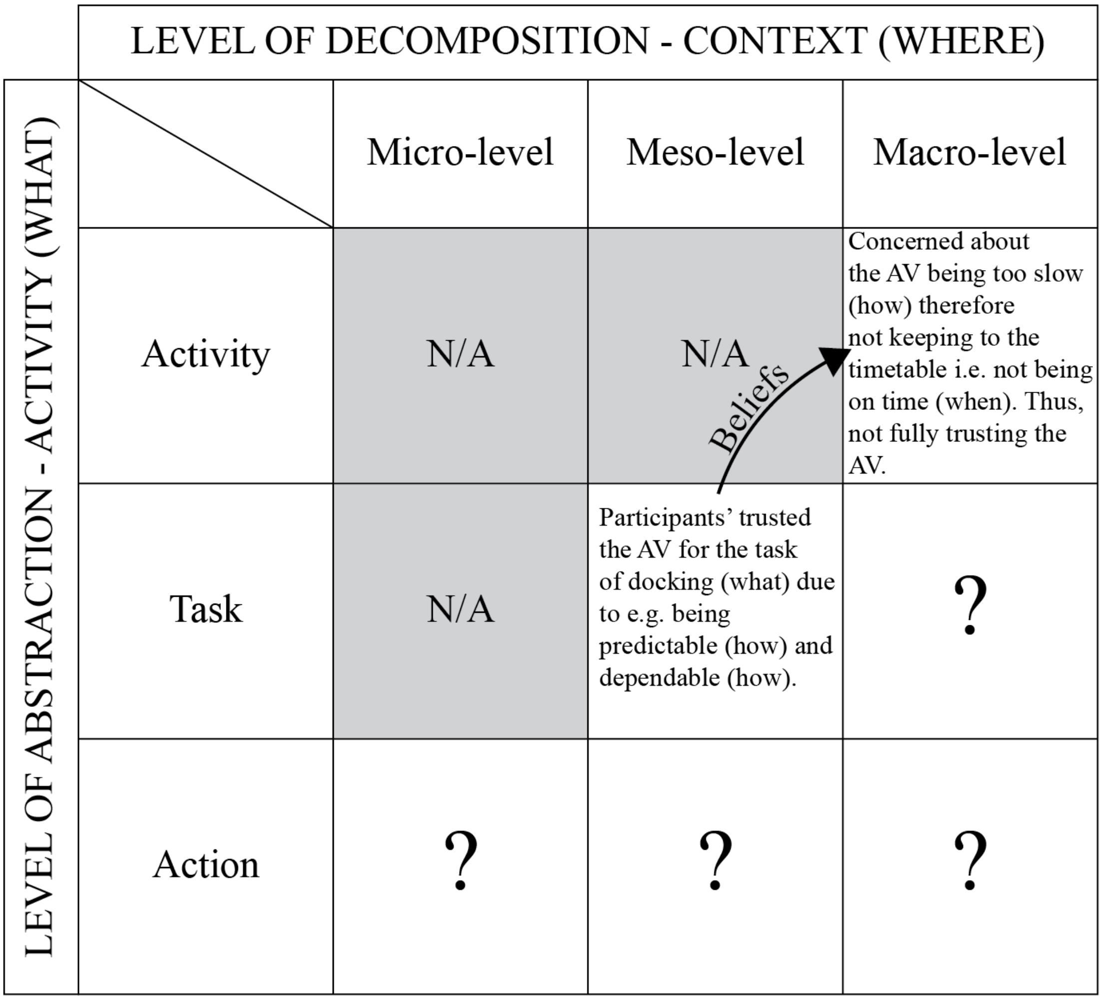

Figure 15 - user trust identified in the task-meso intersection and beliefs (of the trustworthiness of the AV) are propagated to an activity-macro intersection.

Another example of the phenomenon of propagation of beliefs was found in Study III and illustrates how trust was formed and then propagated as beliefs from an action-micro intersection to an action-meso intersection (the action in the bus line during work): *“As of now when he* [the AV - automated PT bus] *stops, it somewhat jerks and this needs to be remedied* [the harsh stop] *since there are a lot of old people* [passengers] *who stand in the bus often without holding on to something, and that is dangerous for them.”* (P5).

The first part of the interview excerpt describes information received from the AV braking, that is action-micro intersection, affecting user trust: *“As of now when he* [the automated bus] *stops, it somewhat jerks and this needs to be remedied…”*. The second part of the sentence – *“...since there are a lot of old people* [passengers] *who stand in the bus often without holding on to something, and that is dangerous for them”* relates to the participant’s concerns based on past experiences and thus knowledge about what is needed to safely transport passengers when docking, that is on a meso level. It is defined as a meso level since the participant only refers to allowing passengers to stand without referring to any demands and/or characteristics of the public transport system. In other words, if the participant had said that passengers often stand because no seats were available when travelling in the public transport system, it would have been on a macro level.

The participant did not fully trust the AV for the action of braking (action-micro intersection), not primarily due to the jerkiness and harsh stopping behaviour there and then, but rather due to the possible effects it could have on elderly passengers standing in the bus, that is in an action-meso intersection. Thus, the participant's beliefs about the effects of the action on a completely different context level (meso level) affected the participant's trust more than did the information given at the context level (micro level) where the participant and AV were situated (in which the participants also trusted the AV to a high degree).

Another example of a propagation of beliefs was indicated in Study IV when a participant was asked whether or not they could trust the AV (automated PT bus). The answer was: *"Yes, I believe so. Maybe you trust it more since there isn't much going on"* but when asked moments later whether the participant would trust the AV for daily commuting the response was: *"I am a bit hesitant about that since it needs to keep in its own lane all the time and I felt it was not capable of doing that"* (P2 – Study IV).

The first response –*"Yes, I believe so ...."* – describes user trust in a current task-meso intersection since the participants only experienced a series of tasks as the AV negotiated simulated traffic situations without any organizational goals and routines, and no global environment and/or other system ecologies as in the 'real world'. The second response – "*I am a bit hesitant about that ...."* – shows the user not fully trusting the AV in a task-macro intersection based on what the AV did (not completely keeping to its own lane while driving, i.e. the task) in the current context (meso level). It is defined as a task since the participant refers to the AV as not capable of driving within its own lane, and as a macro level since it is evident that the participant refers to a task in the real world (which is what the participant is asked about). Thus, the participant trusted the AV in a task-meso intersection but did not believe it was capable of performing adequately in a task-macro intersection.

Therefore, the examples not only show that user trust is based on an assessment of information or lack thereof, created in the interdependence between what the AV does and where (context) in relation to ones goal(s) as described and illustrated by the presented framework, but also, from the information retrieved, beliefs are created for what the AV can do in other contexts. It also nuances mismatches in user trust between different levels of intersections between abstraction of activity and decomposition of context, in other words, you can trust the AV 'here and now' but not necessarily 'there and then'.

6. GENERAL DISCUSSION

The sixth chapter of the thesis – Discussion – deals with design for user trust based on the findings in the thesis and methodological considerations.

6.1. DESIGN FOR APPROPRIATE USER TRUST

The aim of the thesis was to contribute to design of AVs by developing further knowledge on how to design for appropriate user trust, in different types of AVs within different ODs.

6.1.1. Approaching User Trust

I initiated the journey from the same place as several other researchers such as Hergeth (2016), Kraus (2020) and Walker (2021), examining the area of TiA or TiAV from the extensive work by Mayer et al. (1995), Lee and See (2004) and Hoff and Bashir (2015). However, based on my design perspective I have taken a slightly different approach. Instead of continuing to measure the effect on user trust of information deliberately designed to be conveyed via in-car displays such as graphical user interfaces (GUI) and conducting the studies via the use of driving simulators, I have focused on understanding and identifying what in the AV and context during use, in both realistic and naturalistic settings, creates the information, ultimately affecting user trust.

This approach assisted me in both adding to trust factors such as the importance of AV behaviour conforming to a user's expectations (see Section 5.1.), and nuancing previously presented contextual aspects such as perceived task difficulty, perceived environmental complexity and so on (see Section 5.2.) in the domain of AVs including automated cars, public transport busses and delivery bots.

Moreover, this approach helped me identify a set of design variables to be considered including: (I) Who, which refers to the automated artefact, i.e. the AV; (II) What, which refers to what the AV does; (III) By what means (the AV does something) which refers to the communication channel via which a user receives information from the AV; (IV) When, which refers to when the AV does something either in terms of distance to other objects or when in time. Furthermore, (V) Why refers to the user retrieving relevant information so as to help the user understand why the AV did/will do something, (VI) Where refers to the context in which the AV does something, and finally (VII) How, which is the dependent variable, that is to say the effect of the other design variables such as how the user interprets the information created.

Perhaps the most important finding presented in the thesis is the interdependency between (II) what the AV does in terms of an action, task and/or activity and (VI) where the AV does this in terms of a micro, meso and/or macro level. That is, the user's interpretation of the trustworthiness of the AV always depends on what the AV does and where. Thus, user trust changes not only based on what the AV does but also where it does it.

According to Muir (1987) and others there is a cause-and-effect connection between automation and the environment. The more stable the environment is, the more predictable it is and therefore user trust is inversely related to the amount of disturbance posed by the environment upon the machine and vice versa. This interdependency between the automated artefact and the context has been found in other areas as well. One such example is in the area of human-robot interaction (HRI), where robots were trusted differently partly depending on the robot's role, appearance and context (a care vs a production context) (Biermann et al., 2021).

However, to the author's knowledge little has been done to describe this interdependency in greater detail – or rather specify this interdependency – that influences user trust. What was identified during the dissertation project was a set of design variables. Two of these design variables relate to what the AV does and where the AV does something, which can be divided into three separate levels of abstraction and decomposition respectively. This is done to showcase in detail what the user actually trusts/does not trust in the AV, where the user does/does not trust the AV, and whether user trust is based on mere beliefs of the trustworthiness of the AV or actual user trust. Therefore, when designing for appropriate user trust in AVs, the AV-context interdependence needs to be considered not only due to the increased possibility of perception of risk, in turn achieving more valid results (Li et al., 2019), but more importantly because without a realistic context it is not possible to adequately assess trustworthiness:

"*…the meaning in a message depends on the context in which the message originated and the context in which it is interpreted.* " (Madden, 2000, p. 343).

Therefore, I would argue that the benefits of using simulators studying user trust are limited. The findings underscore the intricate interdependence between what the AV does, by what means, when, why and also where this takes place on several levels. In simple terms the interdependency between AV and context is what ultimately affects user trust. Therefore, one should be careful when using simulators and should be fully aware that the concept of an AV tested and/or evaluated in the best way possible represents the final AV and the context in which it will be used. Otherwise, the results obtained may be misleading which in turn could lead to badly informed design decisions which could in the end lead to misuse or disuse.

I would also argue that more focus should be put on how to design for user trust based on the interdependence between AV and context and their non-mutually exclusive effect on user trust. The findings presented here are a small step towards designing for appropriate user trust via the AV-context interdependence. However, more work needs to be done.

A suggestion would be to further develop methods from the area of human errors - Human Reliability Assessment (HRA) (Embrey, 2004), for example Cognitive Walkthrough (CW) or Predictive Human Error Analysis (PHEA) to also account for the user's understanding of the AV-context interdependency. The proposed framework for Trust Analysis & Design could be used to systematically assess the user's theoretical understanding of different actions, tasks or activities in different levels of contexts. This to identify e.g. actions on different levels of contexts where the user might interpret the information from the AV-context interdependency incorrectly. This could in turn give design suggestions for further development in the design process. Thus, the methods could be further developed to account for users understanding of AV-context interdependency to assist in designing for appropriate user trust. However, this needs to be further investigated.

6.1.2. Propagation of Beliefs

One phenomenon related to the previously described interdependency is 'propagation of beliefs'. In other words, based on the user's current interaction with the AV, for instance conducting a task on one level of context such as a bus docking (task) at a bus stop (meso-level) affected user trust. That is to say the user formed a level of trust in the AV for docking at the bus stop but the user also created beliefs for instance about how trustworthy the AV would be for another action, task, and/or activity on a completely different decomposition level (context level). Thus, showing that the user can have contradictory views about what the

AV should be trusted with doing such as an action, task and/or activity and where (micro, meso and/or macro context).

The propagation of beliefs could lead to both matches and mismatches in user trust between what the AV does in one context and what the user believes the AV can do in another context. Thus, propagation of beliefs for other not yet experienced contexts (and actions, tasks and activities) may lead to both disuse and in a worst case scenario misuse for instance in an aggregated context, that is to say a context that includes more elements and is therefore often more complex.

These results are supported by Walker (2021). According to Walker users having developed dynamic learned trust (trust formed over the course of many situations) showed more appropriate levels of situational trust (user trust in one specific situation) even for scenarios they had not yet experienced, which he calls ‘generalization’, and this in turn contributes to dynamic learned trust. Generalization can help users calibrate user trust in the AV without needing to experience all traffic situations. However, he further points out that user trust from limited experience may also cause inappropriate user trust.

Thus, I would argue that it is highly important to understand users’ discrepancy between trust in what the AV does in one context and belief in what the AV can do in another context. In the worst case scenario it may lead to misuse, causing accidents.

Furthermore, based on propagation of beliefs and matches and mismatches between user trust and beliefs for different abstraction of activities and decomposition of context levels, it also seems that the framework presented here includes consequences for how interviews and questionnaires should be conducted. For example sometimes single item questionnaire are used to measure user trust (e.g. Hergeth et al., 2016; Korber et al., 2018) how can we know what we are measuring – user beliefs of how trustworthy the AV is for something or user trust in something specific? Or are we instead measuring user trust in the AV conducting an action or task in a micro context or macro context? The tentative Framework of Trust Analysis and Design assists in specifying in what abstraction of activity or decomposition of context we are measuring user trust. Thus, instruments should be designed to be able to account for not only the design variables but also these two dimensions.

6.2. METHODOLOGICAL CONSIDERATIONS

First and foremost, the approach to studying user trust in different AVs in different but realistic and naturalistic contexts alike can be regarded as both a strength and a weakness. A strength since I can identify similarities and differences between TiA and TiAV, for instance, and similarities between what affects user trust in different AVs and ODs. However, the drawback is that Studies I-IV were all different so I have been forced to adapt research designs and instruments accordingly. I will here present the means by which I have tried to ensure the quality of the dissertation project.

Due to my ontological background as a critical realist – acknowledging that an objective world exists with or without our presence although that world is shaped and affected by our interpretation of the same (Guba & Lincoln, 1994) – and the fact that I conduct primarily qualitative research, the quality-assuring concepts of reliability, validity and generalisability are not as important as some other criteria. However, what are more important are quality-assuring criteria such as credibility, dependability, confirmability, transferability and reflexivity (Stenfors et al., 2020) and the degree to which the findings are ecologically valid (Osborne-Crowley, 2020).

The credibility of research findings presented in this thesis is supported by the fact that the methods chosen are both explained and justified for the specific aim of collecting data on user trust (Stenfors et al., 2020) This is done both by using methods that are regarded by the research community as valid for collecting data on user trust, and by using mixed-method research – combining and using different valid methods to collect data on user trust in each of the four conducted studies. Using a combination of methods is important since it allows data sets to be compared and/or related against one another to either confirm or disconfirm the data (Creswell, 2014) and makes it possible to ensure valid and reliable measurements (Kohn et al., 2021). Furthermore, since many of the findings are coherent with earlier research, with similar factors and contextual aspects being identified with the exception of some differences, this approach seems credible.

Furthermore, in terms of dependability, that is to say the extent to which the research conducted and presented in this thesis is replicable in similar conditions (Stenfors et al., 2020), there has been considerable focus on presenting each step as well as the rationale behind these steps as thoroughly as possible. However, in terms of the thematic analyses and whether or not the similar results could be replicated by another researcher, this is difficult to say. In the case of another design researcher the answer is highly likely, but from a researcher who primarily works with quantitative data the answer is not equally clear. This is because conducting a thematic analysis in an adequate and valid manner demands experience and many iterations to ensure depth. However, to compensate for the fact that interview material can be interpreted differently all apart from one study were analysed by two researchers, that is to say Intercoder agreement (cf. Creswell, 2014) which involves cross-checking to determine consistency between the two researchers. Unfortunately, due to contextual circumstances while the data collected in Study II was analysed there was a shortage of colleagues (during the COVID pandemic) who could assist in conducting the analysis, as well as time restrictions, so I did the analysis myself.

Moreover, confirmability is also important for rigorous qualitative research, underscoring that there exists a connection between data and findings and that this connection is clear. To show this connection in a clear way, I have tried to explain and describe the data extracted from the interview material in as clear and rich a way as possible, for instance by including quotes that provide further support to, and give a more nuanced explanation of, a specific finding presented in the thesis. Similar to confirmability, transferability is also important for the context in which the studies are explained and in that the researcher explains how the context may have shaped the findings. This is important so the findings can be transferred to another context and/or group (Stenfors et al., 2020). Since I have tried to understand user trust in different contexts, this is explained via my findings, for instance that a user may not trust an AV in the same way in one context as in another. Further, I have tried to be as transparent as possible in terms of shortcomings, detailing for instance how in Study II the AV (automated delivery bot) did not function properly, including a couple of malfunctions directly in front of the participants, and the developers behind the automated delivery bot had to step in. This in turn affected user trust. However, from my perspective nevertheless this translated into data on how this malfunction affected the participants' trust, increasing the overall richness and depth of the data collected throughout the four studies. Since richness and depth are important as long as the data is appropriate for answering the research question, all are markers of high-quality research (Stenfors et al., 2020).

I would like to add a few words about reflexivity, that is to say the degree to which the researcher conducting the research reflects on his or her role in shaping the aim of the study, for instance through the researcher's relationship with the participants (Stenfors et al., 2020). First of all, all participants were recruited either via advertisements (Studies I and IV), by a third party (Study III) or had no earlier connection with the participants. Secondly, as a researcher I have tried to build a rapport with the participants to get them to feel relaxed, letting them know that there is no right or wrong answers, while at the same time focusing on staying neutral. This was important in order to allow the participants to feel they say can whatever they want (even though it may be something highly negative towards a technology, for instance) while at the same time doing my utmost to understand their narrative. To do so I have in every study employed probing techniques (Studies I-IV), asking follow-up questions such as 'can you tell me more about that?' or 'why do you think this is so?' or 'how do you feel about that?' According to Kelly et al. (2010) probing techniques are a means of eliciting rich and clear answers from participants. However, I would also argue that probing increases the chance of identifying phenomena that would not be captured by the predefined interview questions in the interview protocol. Thus, probing increases richness and clarity of data while at the same time increasing the chance of identifying new phenomena.

Finally, I created two instruments (i) *Trust Questionnaire* – which had its theoretical basis in the questionnaire created by Jian et al. (2000) and incorporated antecedents to performance, purpose and process information as described by Lee and See (2004) (see Appendix A) and (ii) *The Questionnaire on Contextual Aspects* – which was based on the aspects identified by Hoff and Bashir (2015) and identified aspects related to the context that may affect user trust in automation (see Appendix B).

The choice of creating my own questionnaires stems from the issue of primarily focusing on a high level of ecological validity (that is also why I did not conduct any driving simulator studies), i.e. that the study incorporates stimuli that resemble the stimuli encountered in the real world (Osborne-Crowley, 2020). In my case this involves stimuli from the AV such as driving behaviour and context, for example traffic situations, physical environment and work routines. I could have chosen a validated instrument, which is important to be able to compare results and understand what you are actually measuring in terms of user trust (Kohn et al., 2021). However, the issue would be the same since the instrument needs to be adapted to the specific study, including user group, AV type as well as OD. Therefore, the instrument would no longer be a validated instrument. Thus, I felt it is better to create my own instruments which can be adapted to the specific criteria and limitations of each study (Study I-Study IV) but still be based on trust theories and therefore more useful. However, they need to be further validated and further developed to account for the findings in the thesis. In other words there is a need for instruments such as questionnaires that measure user trust for different things that the AV does such as an action, task and/or activity, Why the AV does something, When the AV does something, in different contexts (Where – micro, meso and/or macro context) as well as accounting for users' beliefs created during use (cf. dynamically learned trust and situational trust Hoff & Bashir, 2015). All this in order for us as researchers to really know what we are measuring and for developers to know what to focus on when designing for appropriate user trust. Thus, the questionnaires need to be validated and further developed.

7. CONCLUSION

The aim of the thesis was to contribute to AV design by developing further knowledge of how to design for appropriate user trust, in different types of AVs within different operative domains (ODs). The aim was operationalised into two research questions: ***What Trust Factors Affect User Trust in AVs in different ODs?*** and ***What Contextual Aspects Affect User Trust in AVs in different ODs?*** To answer the research questions, four user studies were conducted of user trust in automated cars, public transport buses and delivery bots for transportation of parcels.

Addressing the first research question ***What Trust Factors Affect User Trust in AVs in different ODs?***, the user studies showed that user trust was primarily affected by what an AV did during use and to what degree this behaviour was perceived as capable, predictable and reliable. Trust was also affected by information that helped the user in understanding why an AV did or did not conduct an action and whether or not that action was appropriate in order to assist the user in reaching his or her goal such as completing a task in a sufficient manner.

The second research question posed was ***What Contextual Aspects Affect User Trust in AVs in different ODs?*** The user studies showed that there were primarily two contextual aspects affecting user trust in different types of AVs in different ODs; the perception of risk for oneself and others. That is, the probability of negative outcomes and degree of the seriousness of the outcome for both the user of the AV but also for others that possibly comes in contact with the AV or are affected by the AVs behaviour e.g. other road users. Another aspect was perception of task difficulty in terms of a task being perceived as (too) easy, so the user cannot judge the trustworthiness of the AV, or the AV increasing the task difficulty for the user and hereby adding to negative outcomes.

These findings show that regardless of AV type and OD it is important to consider how well the AV performs in the use context, how much information the user gets to understand the behaviour of AV and, further, if that behaviour helps the user to reach his or her goal. Thus, the findings further support earlier presented theories on Trust in Automation (TiA) and in addition nuances contextual aspects by further explaining how they affect user trust.

Furthermore, the third and last research question posed in response to the aim was ***What are the Important Considerations from a Design Perspective in order to design for Appropriate User Trust?***. The conclusion is that the trust factors and contextual aspects are effects of information received from the AV, the context and the interdependency between them, that is, the factors and contextual aspects are the user's interpretations of the interplay between a set of design variables. In other words, a set of variables that can be adjusted and designed to create the effect sought after e.g. the user interprets the AV as predictable and thus ultimately affecting user trust accordingly. This since user perceptions can be designed for but not designed per say.

These design variables are: **(I) Who** – the automated artefact, i.e. the AV, **(II) What** – what the AV does, (III) By **what means** (the AV does something) – the communication channel via which a user receives information from the AV, for example driving behaviour and/or in-car interfaces, **(IV)** When – when the AV does something either in terms of distance to other objects or when in time, e.g. early or late, **(V) Why** – the user must have access to relevant information so they can understand why the AV does/will do something and **(VI) Where** – the context in which the AV does something. Finally, **(VII) How** – is the dependent variable which is the effect of the other design variables such as how the user interprets the information created, for example the AV being perceived as capable since it handles all traffic situations with ease and is therefore regarded as trustworthy. In other words, **(VII) How** represents the trust factors and contextual aspects since they are the user's interpretation of the interplay between the underlying and fundamental design variables.

However, another important finding was the interdependence between AV and context, their non-mutually exclusive relationship and how that relationship affected user trust. The interdependence was empirically derived to exist between two of the design variables; **(II) What** the AV does and **(VI) Where**, i.e. in which context the AV does something.

The results suggest further that what the AV does can be differentiated into three different levels of abstraction (or abstraction of activity). That is the user can trust (or not) the AV to conduct either an action, task and/or an activity. Similarly, the user can also trust the AV to conduct either an action, task and/or an activity in different decompositions of context such as on a micro, meso and macro level. What the AV does and where can, as mentioned earlier, never be separated and therefore always intersect somewhere in an activity-context intersection.

The interdependence between the variables **(II) What** the AV does and **(VI) Where** the AV does this is presented and illustrated in the tentative Framework for Trust Analysis & Design. The Framework shows how one user (or group of users) may trust the AV in one context but not believe the AV can be trusted in another due to a so-called 'propagation of beliefs' showing a mismatch in user trust between one activity-context intersection and another. Thus, the user may trust the AV here and now but not necessarily there and then.

REFERENCES

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

Archer, B., Baynes, K., & Roberts, P. (2005). A framework for Design and Design Education. A Reader Containing Papers From 1970s and 80s.

Balfe, N., Sharples, S., & Wilson, J. R. (2015). Impact of automation: Measurement of performance, workload and behaviour in a complex control environment. Applied ergonomics, 47, 52-64. https://doi.org/https://doi.org/10.1016/j.apergo.2014.08.002

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

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I., & Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation. 7. Tagung Fahrerassistenzsysteme,

Beggiato, M., Pereira, M., Petzoldt, T., & Krems, J. (2015). Learning and development of trust, acceptance and the mental model of ACC. A longitudinal on-road study. Transportation research part F: traffic psychology and behaviour, 35, 75-84. https://doi.org/https://doi.org/10.1016/j.trf.2015.10.005

Behringer, R., Sundareswaran, s., Gregory, B., Elsley, R., Addison, B., Guthmiller, W., Daily, R., & Bevly, D. (2004). The DARPA grand challenge - development of an autonomous vehicle. https://doi.org/10.1109/IVS.2004.1336386

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

Biermann, H., Brauner, P., & Ziefle, M. (2021). How context and design shape human-robot trust and attributions. Paladyn, Journal of Behavioral Robotics, 12(1), 74-86.

Biros, D. P., Daly, M., & Gunsch, G. (2004). The Influence of Task Load and Automation Trust on Deception Detection. Group Decision and Negotiation, 13(2), 173-189. https://doi.org/10.1023/B:GRUP.0000021840.85686.57

Box, G. E., & Draper, N. R. (2007). Response surfaces, mixtures, and ridge analyses. John Wiley & Sons.

Boysen, N., Fedtke, S., & Schwerdfeger, S. (2021). Last-mile delivery concepts: a survey from an operational research perspective. OR Spectrum, 43(1), 1-58. https://doi.org/10.1007/s00291-020-00607-8

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. 3(2), 77-101.

Chang, M., Büchel, D., Reinecke, K., Lehmann, T., & Baumeister, J. (2022). Ecological validity in exercise neuroscience research: A systematic investigation. European Journal of Neuroscience, 55(2), 487-509. https://doi.org/https://doi.org/10.1111/ejn.15595

Charters, E. J. B. E. J. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. 12(2).

Colley, M., Belz, J. H., & Rukzio, E. (2021). Investigating the effects of feedback communication of autonomous vehicles. 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,

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.

Cummings, M., Sasangohar, F., Thornburg, K., Xing, J., & D'Agostino, A. (2010). Human-system interface complexity and opacity part i: literature review. Massachusettes Institute of Technology, Cambridge, MA.
Cupchik, G. C., & Hilscher, M. C. (2008). Holistic perspectives on the design of experience. In Product experience (pp. 241-255). Elsevier.
Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008
Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. Cognition, Technology & Work, 4(4), 240-244.
Dixon, L. (2020). Autonowashing: The Greenwashing of Vehicle Automation. Transportation Research Interdisciplinary Perspectives, 5, 100113. https://doi.org/https://doi.org/10.1016/j.trip.2020.100113
Dorst, K. (2011). The core of 'design thinking' and its application. Design Studies, 32(6), 521-532. https://doi.org/https://doi.org/10.1016/j.destud.2011.07.006
Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. International Journal of Human-Computer Studies, 58(6), 697-718. https://doi.org/10.1016/s1071-5819(03)00038-7
Edelmann, A., Stümper, S., & Petzoldt, T. (2019). Specific Feedback Matters-The Role of Specific Feedback in the Development of Trust in Automated Driving Systems. 2019 IEEE Intelligent Vehicles Symposium (IV),
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,
Embrey, D. (2004). Qualitative and quantitative evaluation of human error in risk assessment. Human Factors for Engineers; Sandom, C., Harvey, RS, Eds, 151-202.
Endsley, M. R. (2017). Autonomous Driving Systems: A Preliminary Naturalistic Study of the Tesla Model S. Journal of Cognitive Engineering and Decision Making, 11(3), 225-238. https://doi.org/10.1177/1555343417695197
Forlizzi, J., Zimmerman, J., & Stolterman, E. (2009). From design research to theory: Evidence of a maturing field. International Assoc. of Societies of design research conference,
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. 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. https://doi.org/10.1007/s10111-011-0194-3
Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. Handbook of qualitative research, 2(163-194), 105.
Habibovic, A., Andersson, J., Nilsson, M., Lundgren, V. M., & Nilsson, J. (2016, 19-22 June 2016). Evaluating interactions with non-existing automated vehicles: three Wizard of Oz approaches. 2016 IEEE Intelligent Vehicles Symposium (IV),
Hagenzieker, M., Boersma, R., Velasco, J., Ozturker, M., Zubin, I., & Heikop, D. (2020). Automated Buses in Europe. An Inventory of Pilots ", Version 0.5. TU Delft.
Hergeth, S. (2016). Automation trust in conditional automated driving systems: approaches to operationalization and design.
Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. Human

factors, 58(3), 509-519. https://journals.sagepub.com/doi/10.1177/0018720815625744?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed
Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). Applied Statistics For The Bhevaioural Sciences (Second Edition ed.). Hougthon Mifflin Company.
Hoff, K. A., & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors, 57(3), 407-434. https://doi.org/10.1177/0018720814547570
Itoh, M. (2012). Toward overtrust-free advanced driver assistance systems. Cognition Technology & Work, 14(1), 51-60. https://doi.org/10.1007/s10111-011-0195-2
Janssen, C., Donker, S., Brumby, D., & Kun, A. (2019). History and Future of Human-Automation Interaction. International Journal of Human-Computer Studies. https://doi.org/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.
Karlsson, M. (1996). User requirements elicitation-a framework for the study of the relation between user and artefact. Chalmers University of Technology.
Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. ACM Transactions on Information Systems (TOIS), 2(1), 26-41.
Kelly, S. E., Bourgeault, I., & Dingwall, R. (2010). Qualitative interviewing techniques and styles. The SAGE handbook of qualitative methods in health research, 307-326.
Kohn, S. C., de Visser, E. J., Wiese, E., Lee, Y.-C., & Shaw, T. H. (2021). Measurement of Trust in Automation: A Narrative Review and Reference Guide [Review]. Frontiers in Psychology, 12(4138). https://doi.org/10.3389/fpsyg.2021.604977
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. https://doi.org/10.1007/s12008-014-0227-2
Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. Applied ergonomics, 66, 18-31.
Korber, M., Prasch, L., & Bengler, K. (2018). Why Do I Have to Drive Now? Post Hoc Explanations of Takeover Requests. Human factors, 60(3), 305-323. https://doi.org/10.1177/0018720817747730
Kraus, J. M. (2020). Psychological processes in the formation and calibration of trust in automation Universität Ulm].
Kraus, J. M., Forster, Y., Hergeth, S., & Baumann, M. (2019). Two routes to trust calibration: effects of reliability and brand information on trust in automation. International Journal of Mobile Human Computer Interaction (IJMHCI), 11(3), 1-17.
Kujala, S., Roto, V., Vaananen-Vainio-Mattila, K., Karapanos, E., & Sinnela, A. (2011). UX Curve: A method for evaluating long-term user experience. Interacting with Computers, 23(5), 473-483. https://doi.org/10.1016/j.intcom.2011.06.005
Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics, 35(10), 1243-1270. https://doi.org/10.1080/00140139208967392
Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80. https://doi.org/DOI 10.1518/hfes.46.1.50.30392
Li, M., Holthausen, B. E., Stuck, R. E., & Walker, B. N. (2019). No Risk No Trust: Investigating Perceived Risk in Highly Automated Driving. Proceedings of the 11th

International Conference on Automotive User Interfaces and Interactive Vehicular Applications,
Lundgren, V. M., Andersson, J., Enerbäck, O., & Dolins, S. (2020). User acceptance of mixed-traffic autonomous shuttles in Gothenburg, Sweden. IOP Conference Series: Earth and Environmental Science,
Lyons, J. B., Hamdan, I. a., & Vo, T. Q. (2023). Explanations and trust: What happens to trust when a robot partner does something unexpected? Computers in Human Behavior, 138, 107473. https://doi.org/https://doi.org/10.1016/j.chb.2022.107473
Lyons, J. B., & Stokes, C. K. (2012). Human–human reliance in the context of automation. Human factors, 54(1), 112-121.
Madden, A. (2000). A definition of information. Aslib Proceedings,
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.
Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation Failures on Tasks Easily Performed by Operators Undermine Trust in Automated Aids. 48(2), 241-256. https://doi.org/10.1518/001872006777724408
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. https://doi.org/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.
Monroe, S., & Vangsness, L. (2022). The Effects of Task Difficulty and Stress on Trust in an Automated Navigation Aid. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 66(1), 1080-1084. https://doi.org/10.1177/1071181322661406
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.
Muir, B. M. (1987). Trust between Humans and Machines, and the Design of Decision Aids. International Journal of Man-Machine Studies, 27(5-6), 527-539. https://doi.org/Doi 10.1016/S0020-7373(87)80013-5
Muir, B. M. (1994). Trust in Automation .1. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems. Ergonomics, 37(11), 1905-1922. https://doi.org/Doi 10.1080/00140139408964957
Müller, A. I., Weinbeer, V., & Bengler, K. (2019). Using the wizard of Oz paradigm to prototype automated vehicles: methodological challenges. Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications: Adjunct proceedings,
Noah, B. E., & Walker, B. N. (2017). Trust Calibration through Reliability Displays in Automated Vehicles Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria. https://doi.org/10.1145/3029798.3034802
Nourani, M., Kabir, S., Mohseni, S., & Ragan, E. D. (2019). The effects of meaningful and meaningless explanations on trust and perceived system accuracy in intelligent systems. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing,

Numan, J. H. (1998). Knowledge-based systems as companions: Trust, human computer interaction and complex systems University of Groningen].
Osborne-Crowley, K. (2020). Social cognition in the real world: reconnecting the study of social cognition with social reality. Review of General Psychology, 24(2), 144-158.
Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, misuse, disuse, abuse. Human factors, 39(2), 230-253. https://doi.org/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. https://doi.org/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. https://doi.org/10.1518/155534308x284417
Park, J., Amendah, E., Lee, Y., & Hyun, H. (2019). M-payment service: Interplay of perceived risk, benefit, and trust in service adoption. Human Factors and Ergonomics in Manufacturing & Service Industries, 29(1), 31-43.
Perkins, L., Miller, J. E., Hashemi, A., & Burns, G. (2010). Designing for human-centered systems: Situational risk as a factor of trust in automation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
Popkin, S. M. (1999). An examination and comparison of workload and subjective measures collected from railroad dispatchers. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
Raats, K., Fors, V., & Pink, S. (2019). Understanding Trust in Automated Vehicles Proceedings of the 31st Australian Conference on Human-Computer-Interaction, Fremantle, WA, Australia. https://doi.org/10.1145/3369457.3369493
Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in Close Relationships. Journal of Personality and Social Psychology, 49(1), 95-112. https://doi.org/Doi 10.1037/0022-3514.49.1.95
SAE. (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles-J3016-202104. https://doi.org/10.4271/J3016_202104
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. https://doi.org/10.1177/1071181312561483
Schwark, J., Dolgov, I., Graves, W., & Hor, D. (2010). The influence of perceived task difficulty and importance on automation use. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,
Sheridan, T. B. (1992). Telerobotics, automation, and human supervisory control. MIT press.
Shladover, S. E. (2018). Connected and automated vehicle systems: Introduction and overview. Journal of Intelligent Transportation Systems, 22(3), 190-200. https://doi.org/10.1080/15472450.2017.1336053
Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences. (2nd editon).
Sindi, S., & Woodman, R. (2020, 20-23 Sept. 2020). Autonomous Goods Vehicles for Last-mile Delivery: Evaluation of Impact and Barriers. 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC),
Sonneberg, M.-O., Leyerer, M., Kleinschmidt, A., Knigge, F., & Breitner, M. H. (2019). Autonomous unmanned ground vehicles for urban logistics: optimization of last mile

delivery operations. Proceedings of the 52nd Hawaii International Conference on System Sciences,
Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. Theoretical Issues in Ergonomics Science, 1(4), 315-331. https://doi.org/10.1080/14639220052399131
Stenfors, T., Kajamaa, A., & Bennett, D. (2020). How to … assess the quality of qualitative research. The Clinical Teacher, 17(6), 596-599. https://doi.org/https://doi.org/10.1111/tct.13242
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. https://doi.org/10.1016/j.promfg.2015.07.797
Sun, X., Li, J., Tang, P., Zhou, S., Peng, X., Li, H. N., & Wang, Q. (2020). Exploring Personalised Autonomous Vehicles to Influence User Trust. Cognitive Computation, 12(6), 1170-1186. https://doi.org/10.1007/s12559-020-09757-x
Thill, S., Hemeren, P. E., & Nilsson, M. (2014). The apparent intelligence of a system as a factor in situation awareness. 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. https://doi.org/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. https://doi.org/10.1177/0018720812443825
Walker, F. (2021). To trust or not to trust?: Assessment and calibration of driver trust in automated vehicles University of Twente].
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. https://doi.org/10.1016/j.jesp.2014.01.005
Wicks, A. C., Berman, S. L., & Jones, T. M. (1999). The structure of optimal trust: Moral and strategic implications. Academy of Management Review, 24(1), 99-116.
Yu, B., Bao, S., Zhang, Y., Sullivan, J., & Flannagan, M. (2021). Measurement and prediction of driver trust in automated vehicle technologies: an application of hand position transition probability matrix. Transportation Research Part C: Emerging Technologies, 124, 102957.
Zhang, Q., Yang, X. J., & Robert, L. P. (2020). Expectations and trust in automated vehicles. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems,

APPENDIX A

Trust Questionnaire

Trust Questionnaire. 7-point Likert Scale ranging from 1= Strongly Disagree, 7= Strongly Agree.

Table - Trust questionnaire as used in Study II. The items were the same in all studies except for the artefact designation.

	Pre-Interaction Phase	Post-Interaction Phase
Item I	I expect the bot to perform tasks in a capable manner	I experienced the bot as able to perform task(s) in a capable manner
Item II	I expect the bot to be safe to interact with	I experienced the bot to be safe to interact with
Item III	I expect to be able to trust the bot	I was able to trust the bot
Item IV	I expect that the bot is designed to have my best interests in mind	I experienced the bot as being designed to have my best interests in mind
Item V	I expect the bot's behaviour to be predictable	I experienced the bot's behaviour as predictable
Item VI	I expect the bot to perform tasks in a manner that does not include my assistance	I experienced the bot as performing tasks in a manner that did not need my assistance
Item VII	I expect the bot to perform tasks in such a manner that it is better that the bot performs them, and I can instead turn my attention to other tasks	I experienced the bot as performing tasks in such a manner that it is better that the bot performs them, allowing me to instead turn my attention to other tasks

APPENDIX B

The Questionnaire on Contextual Aspects

APPENDIX B: Questionnaire on Contextual Aspects. 7-point scale.

Table – Questionnaire on Contextual aspects as used in Study II. The items were the same in all studies except for the artefact designation.

PERCEIVED RISK	Item I	According to you, what is the probability that the bot's tasks could lead to negative consequences? (e.g., delays, costs, injuries)	1= Very Low, 7= Very High
	Item II	According to you, how serious could these potential consequences be?	1= Very Trivial, 7= Very Serious
PERCEIVED TASK DIFFICULTY	Item III	According to you, how challenging are your everyday task(s) to perform?	1= Very Easy, 7= Very Hard
PERCEIVED TASK DIFFICULTY	Item IV	According to you, how complicated will it be for the bot to carry out tasks?	1= Very Easy, 7= Very Hard
PERCEIVED TASK DIFFICULTY	Item V	According to you, how did your activities change due to the bot?	1= Much Easier, 7= Much Harder
CONFORMING TO EXPECTATIONS	Item VI	According to you, how well did your expectations on the bot performing the task(s) match the experience?	1= Did not match expectations at all, 7= Fully Matched Expectations
WORKLOAD	Item VII	According to you, to what degree do you experience your work as being Hard after the bot was implemented?	1= Too a very low degree, 7= Too a very high degree
WORKLOAD	Item VIII	According to you, to what degree do you experience your work as being Complex after the bot was implemented?	1= Too a very low degree, 7= Too a very high degree
WORKLOAD	Item IX	According to you, to what degree do you experience your work as being Time-pressured after the bot was implemented?	1= Too a very low degree, 7= Too a very high degree
WORKLOAD	Item X	According to you, how big a part of your total work capacity did you use to meet the demands put on you in your work, after the bot was implemented?	1= A very small part, 7= A very big part

PERCEIVED BENEFITS	Item XI	According to you, how do you feel the introduction of the bot has affected your work?	1= Only Negatively, 7= Only Positively
SYSTEM COMPLEXITY	Item XII	According to you, are there many tasks in your work environment that the bot needs to handle to be able to perform the task successfully?	1= Strongly Disagree,7= Strongly Agree
SYSTEM COMPLEXITY	Item XIII	According to you, do the tasks that the bot needs to handle differ a lot from each other?	1= Strongly Disagree, 7= Strongly Agree
SYSTEM COMPLEXITY	Item XIV	According to you, do these tasks affect each other and if so, to what degree?	1= Strongly Disagree, 7= Strongly Agree